A BFGS-IP Algorithm for Solving Strongly Convex Optimization Problems with Feasibility Enforced by an Exact Penalty Approach

Jean Charles Gilbert, Sophie Jan, Paul Armand

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A BFGS-IP algorithm for solving strongly convex optimization problems with feasibility enforced by an exact penalty approach

Paul ARMAND — Jean Charles GILBERT — Sophie JAN

N° 4087

28 décembre 2000

THÈME 4

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--4087--FR+ENG

A BFGS-IP algorithm for solving strongly convex optimization problems with feasibility enforced by an exact penalty approach

Paul ARMAND*, Jean Charles GILBERT†, Sophie JAN‡

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet Estime

Rapport de recherche n° 4087 — 28 décembre 2000 — 33 pages

Abstract: This paper introduces and analyses a new algorithm for minimizing a convex function subject to a finite number of convex inequality constraints. It is assumed that the Lagrangian of the problem is strongly convex. The algorithm combines interior point methods for dealing with the inequality constraints and quasi-Newton techniques for accelerating the convergence. Feasibility of the iterates is progressively enforced thanks to shift variables and an exact penalty approach. Global and q-superlinear convergence is obtained for a fixed penalty parameter; global convergence to the analytic center of the optimal set is ensured when the barrier parameter tends to zero, provided strict complementarity holds.

Key-words: analytic center, BFGS quasi-Newton approximations, constrained optimization, convex programming, infeasible iterates, interior point algorithm, line-search, primal-dual method, shift and slack variables, superlinear convergence.

* LACO - Université de Limoges, Faculté des Sciences, 123, avenue Albert Thomas, 87060 Limoges (France) ; e-mail : `armand@unilim.fr`.

† INRIA Rocquencourt, projet Estime, BP 105, 78153 Le Chesnay Cedex (France) ; e-mail : `Jean-Charles.Gilbert@inria.fr`.

‡ MIP - UFR MIG - Université Paul Sabatier - 118, route de Narbonne - 31062 Toulouse Cedex 4 (France) ; e-mail : `jan@mip.ups-tlse.fr`.

Un algorithme de points intérieurs et de BFGS pour résoudre un problème d'optimisation fortement convexe avec admissibilité obtenue par pénalisation exacte

Résumé : Cet article présente et analyse un nouvel algorithme pour la minimisation d'une fonction convexe en présence d'un nombre fini de contraintes d'inégalité convexes. Le lagrangien du problème est supposé fortement convexe. L'algorithme allie une approche par points intérieurs pour prendre en compte efficacement les contraintes d'inégalité et une technique quasi-newtonienne pour accélérer la convergence. L'admissibilité des itérés est obtenue grâce à l'utilisation de variables de décalage qui sont progressivement ramenées à zéro par pénalisation exacte. La convergence globale et q-superlinéaire de l'algorithme est démontrée pour des paramètres de pénalisation fixés. Lorsque le paramètre de pénalisation logarithmique tend vers zéro et que l'on a complémentarité stricte, les itérés externes convergent vers le centre analytique de l'ensemble des solutions.

Mots-clés : algorithme de points intérieurs, approximation quasi-newtonienne de BFGS, centre analytique, convergence superlinéaire, itéré non admissible, méthode primale-duale, optimisation convexe, optimisation sous contraintes, recherche linéaire, variables de décalage et d'écart.

1 Introduction

This paper introduces and analyzes a new algorithm for solving a convex minimization problem of the form

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

where $f : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}^m$ are continuously differentiable functions on the whole space $\mathbb{R}^n$. We assume that f is convex and that each component $c_{(i)}$ of c, for $1 \leq i \leq m$, is *concave*. The feasible set of Problem (1.1) is then convex. The algorithm combines interior point (IP) ideas for dealing with the inequality constraints and quasi-Newton techniques for approximating second derivatives and providing fast convergence. The motivation for introducing such an algorithm has been detailed in [2].

The main contributions of this paper are twofold. First, we improve the capabilities of the primal-dual IP algorithm introduced in [2], by allowing the iterates to be infeasible. This property is useful when it is difficult to find a strictly feasible starting point. In the proposed algorithm, feasibility and optimality are obtained simultaneously. The iterates remain inside a region obtained by shifting the boundary of the feasible set and their asymptotic feasibility is enforced by means of an exact penalty approach. This one shifts back monotonically that boundary to its original position. By our second contribution, we enlarge the class of problems that this algorithm can solve. The strong convexity hypothesis that is necessary to settle the algorithm has been weakened. Instead of assuming the strong convexity of one of the functions f, $-c_{(1)}, \ldots, -c_{(m)}$, as in [2], our analysis shows that it is sufficient to assume the strong convexity of the Lagrangian of Problem (1.1). We believe that these contributions improve significantly the applicability of the algorithm.

In our approach, Problem (1.1) is transformed, using *shift* variables $s \in \mathbb{R}^m$, in an equivalent form (see [9]):

$$\begin{cases} \min f(x), \\ c(x) + s \geq 0, \\ s = 0. \end{cases} \tag{1.2}$$

The interest of this modification is that it is now easy to find an initial pair (x_1, s_1) satisfying $c(x_1) + s_1 > 0$. Of course Problem (1.2) is as difficult to solve as (1.1), but it is now possible to control the feasibility of the *inequality* constraints. In the chosen approach, the inequality $c(x) + s > 0$ is maintained throughout the iterations thanks to the logarithmic barrier function, while the equality $s = 0$ is relaxed and asymptotically enforced by exact penalization. Another key feature of this transformation is that the convexity of the original problem is preserved in (1.2). This would not have been the case if instead we had introduced *slack* variables $\tilde{s} \in \mathbb{R}^m$, as in the problem

$$\begin{cases} \min f(x), \\ c(x) = \tilde{s}, \\ \tilde{s} \geq 0. \end{cases} \tag{1.3}$$

With such a transformation, the positivity of the slacks would be maintained in the algorithm and the constraint $c(x) = \tilde{s}$ would be progressively enforced (see [5] for example). The drawback of (1.3) in the present context is that the nonlinear constraint cannot be viewed as a convex constraint, since the set that it defines may be nonconvex. This is a source of difficulties, preventing the extension of the analysis carried out in [2].

Provided the constraints satisfy some qualification assumptions, the Karush-Kuhn-Tucker (KKT) optimality conditions of Problem (1.1) can be written as follows (see [7, 14] for example): there exists a vector of multipliers $\lambda \in \mathbb{R}^m$ such that

$$\begin{cases} \nabla f(x) - A(x)^\top \lambda = 0, \\ C(x)\lambda = 0, \\ (c(x), \lambda) \geq 0, \end{cases} \tag{1.4}$$

where $\nabla f(x)$ is the gradient of f at x (for the Euclidean scalar product), $A(x)$ is the $m \times n$ Jacobian matrix of c, and $C(x) = \operatorname{diag}(c_{(1)}(x), \dots, c_{(m)}(x))$.

We consider a relaxed barrier problem associated with (1.2):

$$\begin{cases} \min \Big(\varphi_\mu(x,s) := f(x) - \mu \sum_{i=1}^m \log \big(c_{(i)}(x) + s_{(i)} \big) \Big), \\ s = r_\mu, \end{cases} \tag{1.5}$$

where φ_μ is the barrier function parameterized by $\mu > 0$, the arguments of the logarithm are implicitly assumed to be positive, and $r_\mu \in \mathbb{R}^m$ is a vector relaxing the equality constraint of (1.2). There is nothing in r_μ that is fundamental for the convergence of the algorithm, and one could set $r_\mu = 0$ (this is what is done in Sections 3 and 4, actually). For its efficiency, however, it may be more appropriate to force the feasibility progressively as μ goes to zero (of course it is required to have $r_\mu \to 0$ when $\mu \to 0$). This topics is further discussed in the introduction of Section 5.

Let us go back to the barrier problem (1.5). Its optimality conditions can be written

$$\begin{cases} \nabla f(x) - A(x)^\top \lambda = 0, \\ (C(x) + S)\lambda = \mu e, \\ s = r_\mu, \\ (c(x) + s, \lambda) > 0, \end{cases} \tag{1.6}$$

where $S = \operatorname{diag}(s_{(1)}, \dots, s_{(m)})$ and $e = (1 \cdots 1)^\top$ is the vector of all ones. Note that by eliminating s in (1.6) and by setting $r_\mu = 0$, one recovers the system (1.4), in which the complementarity conditions $C(x)\lambda = 0$ are perturbed in $C(x)\lambda = \mu e$, a frequently used technique in primal-dual IP methods. We prefer keeping s in the system (1.6), in particular in its second equation, since in the algorithm the iterate s needs to be nonzero when the iterate x is infeasible, in order to ensure the positivity of $c(x) + s$.

Our primal-dual IP algorithm computes approximate solutions of (1.6) for a sequence of parameters $\mu > 0$ decreasing to zero. For a fixed value of μ, it uses a sequence of quasi-Newton iterations for solving the first three equations of (1.6), using the BFGS update formula. These iterations are called *inner iterations*, while an *outer iteration* is the collection of inner iterations corresponding to the same value of μ.

The global convergence of the quasi-Newton iterates is ensured by a backtracking line-search on some merit function. A classical merit function associated with a constraint problem like (1.5) is the following exact penalty function:

$$\Theta_{\mu,\sigma}(x,s) := \varphi_\mu(x,s) + \sigma\|s - r_\mu\|_P,$$

where $\sigma > 0$ is the penalty parameter and $\|\cdot\|_P$ is an arbitrary norm. Let $\|\cdot\|_D$ be the dual norm associated with $\|\cdot\|_P$:

$$\|v\|_D := \sup_{\|u\|_P \le 1} v^\top u.$$

It is well known (see [4, Chapter 12] for example) that, for convex problems, the penalty function $\Theta_{\mu,\sigma}$ is exact (i.e., the solutions of (1.5) minimize $\Theta_{\mu,\sigma}$), if

$$\sigma \ge \|\lambda_\mu\|_D.$$

A property that plays an important role in our analysis is the convexity of $\Theta_{\mu,\sigma}$. Starting with Problem (1.3) instead of Problem (1.2) would have led to the merit function $f(x) - \mu\sum_{i=1}^m \log \tilde{s}_i + \sigma\|c(x) - \tilde{s}\|_P$, which is not necessarily convex. This is another way of motivating the choice of transforming the original problem (1.1) by using shift variables instead of slack variables.

Since our algorithm generates primal-dual iterates, we have chosen, as in [2], to use a primal-dual merit function by adding to $\Theta_{\mu,\sigma}$ a centralization term $\mathcal{V}_\mu$:

$$\psi_{\mu,\sigma}(x,s,\lambda) = \Theta_{\mu,\sigma}(x,s) + \tau\mathcal{V}_\mu(x,s,\lambda),$$

where τ is some positive constant and

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

Since $t \mapsto t - \mu\log t$ is minimized at $t = \mu$, function $\mathcal{V}_\mu$ has its minimal value at points satisfying the second equation of (1.6).

The strategy that consists in forcing the decrease of $\psi_{\mu,\sigma}$ from (x,s,λ) along a direction $d = (d^x, d^s, d^\lambda)$ can work well, provided d is a descent direction of $\psi_{\mu,\sigma}$. We shall show that this is actually the case if d is a (quasi-)Newton direction on the system (1.6) and if σ is large enough: $\sigma \ge \|\lambda + d^\lambda\|_D$ must hold. Satisfying this inequality does not raise any difficulty, since it is sufficient to increase σ whenever necessary. If σ is modified continually, however, the merit function changes from iteration to iteration and it is difficult to prove convergence. In order to avoid the instability of the penalty parameter, there are rules ensuring that either the sequence of generated σ's is unbounded or σ takes a fixed value after finitely many changes. Of course, only the latter situation is desirable. We have not succeeded, however, in proving that this situation actually occurs with our algorithm, despite the convexity of the problem and the assumed qualification of the constraints (Slater's condition). At this point, we quote that Pshenichnyj [16, Theorem 2.4] has proven an interesting result on the stabilization of the penalty parameter, but with an algorithm that may require a restart at the initial point when σ is updated. We did not want to go along this line, which is not attractive in practice, and have preferred to assume the boundedness of the sequence of σ's. With this assumption, we have been able to show that, for a fixed μ, the whole sequence of inner iterates converges to the solution to the barrier problem (1.5). This favorable situation can occur *only if* the Slater condition holds.

The paper is organized as follows. Section 2 provides notation and tools from convex analysis that are used throughout the paper. The quasi-Newton-IP algorithm for solving the barrier problem is presented in Section 3, while Section 4 focuses on the proof of its superlinear convergence. The last section describes the overall algorithm and provides conditions ensuring the convergence of the outer iterates towards the analytic center of the optimal set.

2 Background

In this paper, we always assume that $\mathbb{R}^n$ is equipped with the Euclidean scalar product and denote by $\|\cdot\|$ the associated ℓ_2 norm. Extending the algorithm to take into account an arbitrary scalar product, which is important in practice, should not present any difficulties.

A function is of class C^1 if it is continuously differentiable and of class $C^{1,1}$ if in addition its derivative is Lipschitz continuous.

A function $\xi : \mathbb{R}^n \to \mathbb{R}$ is *strongly convex* with modulus $\kappa > 0$, if the function $\xi(\cdot) - \frac{\kappa}{2}\|\cdot\|^2$ is convex. When ξ is differentiable, an equivalent property is the strong monotonicity of its gradient, that is: for all $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$ one has $(\nabla\xi(x) - \nabla\xi(y))^\top (x - y) \geq \kappa\|x - y\|^2$ (for other equivalent definitions, see for example [10, Chapter IV]).

Consider now a convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ that can take the value $+\infty$. The *domain* of f is defined by $\operatorname{dom} f := \{x \in \mathbb{R}^n : f(x) < \infty\}$ and its epigraph is $\operatorname{epi} f := \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} : f(x) \leq \alpha\}$. The set of such convex functions that are proper ($\operatorname{dom} f \neq \emptyset$) and closed ($\operatorname{epi} f$ is closed) is denoted by $\overline{\operatorname{Conv}}(\mathbb{R}^n)$. The *asymptotic derivative* of a function $f \in \overline{\operatorname{Conv}}(\mathbb{R}^n)$ is the function $f'_\infty \in \overline{\operatorname{Conv}}(\mathbb{R}^n)$ defined for $d \in \mathbb{R}^n$ by

$$f'_\infty(d) := \lim_{t \to +\infty} \frac{f(x_0 + t\,d) - f(x_0)}{t} = \lim_{t \to +\infty} \frac{f(x_0 + t\,d)}{t},$$

where x_0 is an arbitrary point in $\operatorname{dom} f$ (see for example [10, Section IV.3.2]). It follows immediately from this definition that the asymptotic derivative is positively homogeneous and that $f'_\infty(0) = 0$. The concept of asymptotic derivative is useful since it allows us to verify a topological property of compactness by a simple calculation: the level sets of a convex function f are compact if and only if $f'_\infty(d) > 0$ for all nonzero $d \in \mathbb{R}^n$ (see [10, Proposition IV.3.2.5]). A variant of this result is given in the following lemma (see [17, Corollary 27.3.3]).

Lemma 2.1 *If Problem* (1.1) *is feasible, then its solution set is nonempty and compact if and only if there is no nonzero vector* $d \in \mathbb{R}^n$ *such that* $f'_\infty(d) \leq 0$ *and* $(-c_{(i)})'_\infty(d) \leq 0$*, for all* $i = 1, \ldots, m$.

The following chain rule for asymptotic derivatives is proven in [3, Proposition 2.1]. Let $\eta \in \overline{\text{Conv}}(\mathbb{R})$ be nondecreasing and such that $\eta'_\infty(1) > 0$, and let $f \in \overline{\text{Conv}}(\mathbb{R}^n)$ be such that $(\operatorname{dom} \eta) \cap f(\mathbb{R}^n) \neq \emptyset$. Consider the composite function

$$g(x) = \begin{cases} \eta\Big(f(x)\Big) & \text{if } x \in \operatorname{dom} f, \\ +\infty & \text{otherwise.} \end{cases}$$

Then $g \in \overline{\text{Conv}}(\mathbb{R}^n)$ and

$$g'_\infty(d) = \begin{cases} \eta'_\infty\Big(f'_\infty(d)\Big) & \text{if } d \in \operatorname{dom} f'_\infty, \\ +\infty & \text{otherwise.} \end{cases} \tag{2.1}$$

The *subdifferential* $\partial f(x)$ of a convex function $f : \mathbb{R}^n \to \mathbb{R}$ at $x \in \mathbb{R}^n$ is the set of vectors $g \in \mathbb{R}^n$, called *subgradients*, such that: $f'(x;h) \geq g^\top h$, for all $h \in \mathbb{R}^n$. Clearly, x minimizes f if and only if $0 \in \partial f(x)$. If $\|\cdot\|_P$ is a norm on $\mathbb{R}^m$:

$$v \in \partial(\|\cdot\|_P)(u) \qquad \Longleftrightarrow \qquad \|v\|_D \leq 1 \ \text{ and } \ u^\top v = \|u\|_P, \tag{2.2}$$

where the dual norm $\|\cdot\|_D$ was defined in the introduction.

Despite this paper essentially deals with convex issues, occasionally we shall have to consider nonconvex functions, say $\psi : \mathbb{R}^n \to \mathbb{R}$, having however directional derivatives. If a point x minimizes ψ, there holds $\psi'(x;h) \geq 0$ for all $h \in \mathbb{R}^n$. If, in addition, ψ is of the form $\psi = \phi + f$, where ϕ is differentiable at x and f is convex, then the latter property can equivalently be written $-\nabla\phi(x) \in \partial f(x)$.

3 Solving the barrier problem

This section presents step by step the ingredients composing the algorithm for solving the barrier problem (1.5) for fixed μ: Algorithm A_μ. In Section 3.1, we introduce the basic assumptions for the well-posedness of the algorithm and draw some consequences from them, including existence and uniqueness of the solution to the barrier problem. Section 3.2 defines the direction along which the next iterate is searched. The next two sections analyze the primal merit function $\Theta_{\mu,\sigma}$, obtained by exact penalization of the constraint of the barrier problem (Section 3.3), and the primal-dual merit function $\psi_{\mu,\sigma}$, obtained by adding a centralization term to $\Theta_{\mu,\sigma}$ (Section 3.4). It is this latter function that is used in the algorithm to ensure its robustness. Algorithm A_μ is finally presented in Section 3.5.

In this section and in Section 4, we set the relaxation vector r_μ of the barrier problem (1.5) to zero:

$$\begin{cases} \min \varphi_\mu(x,s), \\ s = 0. \end{cases} \tag{3.1}$$

We shall see in Section 5, that there is no limitation in doing so, because a simple change of variables will allow us to to recover the results of Sections 3 and 4 for the case when r_μ is nonzero. The optimality conditions of the barrier problem becomes

$$\begin{cases} \nabla f(x) - A(x)^\top \lambda = 0, \\ (C(x)+S)\lambda = \mu e, \\ s = 0, \\ (c(x)+s, \lambda) > 0. \end{cases} \tag{3.2}$$

The Lagrangian associated with Problem (1.1) is the real-valued function ℓ defined for $(x,\lambda) \in \mathbb{R}^n \times \mathbb{R}^m$ by

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

When f and c are twice differentiable, the gradient and Hessian of ℓ with respect to x are given by

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

The following formulæ will be often useful in the sequel (from now on, we drop most of the dependencies in x and λ):

$$\nabla \varphi_\mu(x,s) = \begin{pmatrix} \nabla f(x) - \mu A^\top (C+S)^{-1} e \\ -\mu (C+S)^{-1} e \end{pmatrix} \tag{3.3}$$

$$\nabla^2 \varphi_\mu(x,s) = \begin{pmatrix} \nabla^2_{xx}\varphi_\mu(x,s) & \mu A^\top (C+S)^{-2} \\ \mu (C+S)^{-2} A & \mu (C+S)^{-2} \end{pmatrix}, \tag{3.4}$$

where

$$\nabla^2_{xx}\varphi_\mu(x,s) = \nabla^2_{xx}\ell(x, \mu(C+S)^{-1}e) + \mu A^\top (C+S)^{-2} A.$$

3.1 The barrier problem

Our minimal assumptions refer to the convexity and smoothness of Problem (1.1).

Assumptions 3.1 The functions f and $-c_{(i)}$ $(1 \le i \le m)$ are convex and differentiable from $\mathbb{R}^n$ to $\mathbb{R}$; and there exists $\check{\lambda} \in \mathbb{R}^m$, such that the Lagrangian $\ell(\cdot, \check{\lambda})$ is strongly convex with modulus $\check{\kappa} > 0$.

The second part of these assumptions is weaker than Assumption 2.1-(i) in [2], which requires the strong convexity of at least one of the functions f, $-c_{(1)}, \ldots, -c_{(m)}$. For example the problem of two variables $\min\{x_1^2 : 1 - x_2^2 \ge 0\}$ satisfies Assumptions 3.1, but not Assumption 2.1-(i) in [2].

We now derive three consequences of Assumptions 3.1. Lemma 3.2 shows that for any positive multiplier λ, $\ell(\cdot,\lambda)$ is strongly convex. In turn, with some other mild assumptions, this implies that Problem (1.1) has a compact set of solutions (Proposition 3.3) and that the barrier problem (3.1) has a unique primal-dual solution (Proposition 3.4). For $t \in \mathbb{R}$, we define $t^+ := \max(0,t)$.

Lemma 3.2 *Suppose that Assumptions* 3.1 *hold. Then, for any* $\lambda > 0$*, the Lagrangian* $\ell(\cdot, \lambda)$ *is strongly convex with modulus* $\kappa = \check{\kappa} \min(1, \lambda_{(1)}/\check{\lambda}^+_{(1)}, \ldots, \lambda_{(m)}/\check{\lambda}^+_{(m)}) > 0$*.*

Proof. It suffices to show that $\ell(\cdot, \lambda) - \frac{1}{2}\kappa\|\cdot\|^2$ is convex. One has

$$\ell(x, \lambda) - \frac{1}{2}\kappa\|x\|^2 = \left(1 - \frac{\kappa}{\check{\kappa}}\right) f(x) - \left(\lambda - \frac{\kappa}{\check{\kappa}}\check{\lambda}\right)^\top c(x) + \frac{\kappa}{\check{\kappa}} \left(\ell(x, \check{\lambda}) - \frac{1}{2}\check{\kappa}\|x\|^2\right).$$

The result then follows from the convexity of the functions f, $-c_{(1)}$, ... , $-c_{(m)}$ and $\ell(\cdot, \check{\lambda}) - \frac{1}{2}\check{\kappa}\|\cdot\|^2$, and from the inequalities $1 \geq \frac{\kappa}{\check{\kappa}} \geq 0$ and $\lambda - \frac{\kappa}{\check{\kappa}}\check{\lambda} \geq \lambda - \frac{\kappa}{\check{\kappa}}\check{\lambda}^+ \geq 0$. □

Proposition 3.3 *Suppose that Assumptions* 3.1 *hold. Then, there is no nonzero vector* $d \in \mathbb{R}^n$ *such that* $f'_\infty(d) < \infty$ *and* $(-c_{(i)})'_\infty(d) < \infty$*, for all* $i = 1, \ldots, m$*. In particular, if Problem* (1.1) *is feasible, then its solution set is nonempty and compact.*

Proof. Lemma 3.2 implies that $\ell(\cdot, e)$ is strongly convex. In particular, for all $d \neq 0$, $(\ell(\cdot, e))'_\infty(d) = f'_\infty(d) + \sum_{i=1}^m (-c_{(i)})'_\infty(d) = +\infty$. The first part of the proposition follows. The second part is then a consequence of Lemma 2.1. □

Proposition 3.4 *Suppose that Assumptions* 3.1 *hold. Then the barrier function* φ_μ *is strictly convex on the set* $\{(x, s) \in \mathbb{R}^n \times \mathbb{R}^m : c(x)+s > 0\}$*. Moreover if Problem* (1.1) *is strictly feasible, then the barrier problem* (3.1) *has a unique primal-dual solution. This one is denoted by* $\hat{z}_\mu := (\hat{x}_\mu, \hat{s}_\mu, \hat{\lambda}_\mu)$ *and we have* $\hat{s}_\mu = 0$ *and* $\hat{\lambda}_\mu = \mu C(\hat{x}_\mu)^{-1}e$*.*

Proof. Assumptions 3.1 imply that for all $x \neq x'$ and $\alpha \in (0, 1)$, $f(\alpha x + (1 - \alpha)x') \leq \alpha f(x) + (1 - \alpha)f(x')$ and $c_{(i)}(\alpha x + (1 - \alpha)x') \geq \alpha c_{(i)}(x) + (1 - \alpha)c_{(i)}(x')$, for all $i = 1, \ldots, m$; and at least one inequality is strictly satisfied (otherwise we would have $\ell(\alpha x + (1 - \alpha)x', e) = \alpha\ell(x, e) + (1 - \alpha)\ell(x', e)$, contradicting the strong convexity of $l(\cdot, e)$). Now consider two pairs $(x, s) \neq (x', s')$. If $x \neq x'$, then the strict convexity of φ_μ follows from the previous remark and the properties of the log function (strict monotonicity and concavity). If $x = x'$, then $s \neq s'$, $c(x) + s \neq c(x') + s'$, and the result follows from the monotonicity and strict concavity of the logarithm.

To prove the second part of the proposition, note that the pair $(\hat{x}_\mu, \hat{s}_\mu)$ is a solution to the barrier problem (3.1) if and only if $\hat{s}_\mu = 0$ and $\hat{x}_\mu$ is a solution to the unconstrained problem $\min\{\phi_\mu(x) : x \in \mathbb{R}^n\}$, where $\phi_\mu(\cdot) := \varphi_\mu(\cdot, 0)$. To prove that this problem has a solution, let us show that $(\phi_\mu)'_\infty(d) > 0$ for any nonzero $d \in \mathbb{R}^n$ (see Section 2). Let us introduce the increasing function $\eta \in \overline{\mathrm{Conv}}(\mathbb{R})$ defined by

$$\eta(t) = \begin{cases} -\log(-t) & \text{if } t < 0, \\ +\infty & \text{otherwise.} \end{cases} \tag{3.5}$$

By using the chain rule (2.1) we obtain

$$(\phi_\mu)'_\infty(d) = f'_\infty(d) + \mu \sum_{i=1}^m \eta'_\infty\left((-c_{(i)})'_\infty(d)\right),$$

with the convention that $\eta'_\infty(+\infty) = +\infty$. Since $\eta'_\infty(t) = +\infty$ if $t > 0$ and is zero otherwise, and since $f'_\infty(d^x) > -\infty$, $(\phi_\mu)'_\infty(d) \leq 0$ only if $f'_\infty(d) \leq 0$ and $(-c_{(i)})'_\infty(d) \leq 0$ for all $i = 1, \dots, m$. This is not possible by Proposition 3.3. The positivity of $(\phi_\mu)'_\infty$ implies the compactness of the level sets of ϕ_μ. Now, the fact that ϕ_μ is finite for some point, implies the existence of a minimizer of that function.

The uniqueness of the solution $(\hat{x}_\mu, \hat{s}_\mu)$ follows from the strict convexity of φ_μ. Existence of the dual solution $\hat{\lambda}_\mu$ is a consequence of the linearity of the constraint in (3.1) and its value is given by the second equation in (3.2). □

3.2 The Newton step

The Newton step $(d^x, d^s, d^\lambda) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ on (3.2) is a solution to

$$\begin{pmatrix} M & 0 & -A^\top \\ \Lambda A & \Lambda & C+S \\ 0 & I & 0 \end{pmatrix} \begin{pmatrix} d^x \\ d^s \\ d^\lambda \end{pmatrix} = \begin{pmatrix} -\nabla_x \ell \\ \mu e - (C+S)\lambda \\ -s \end{pmatrix}, \tag{3.6}$$

in which M is the Hessian of the Lagrangian $\nabla^2_{xx}\ell(x,\lambda)$ and $\Lambda = \operatorname{diag}(\lambda_{(1)}, \dots, \lambda_{(m)})$. In the quasi-Newton algorithm that we consider, M is a positive definite approximation to $\nabla^2_{xx}\ell(x,\lambda)$, obtained by BFGS updates.

The third equation in (3.6) determines d^s uniquely:

$$d^s = -s. \tag{3.7}$$

This makes it possible to eliminate d^s from the second equation:

$$\begin{pmatrix} M & -A^\top \\ A & \Lambda^{-1}(C+S) \end{pmatrix} \begin{pmatrix} d^x \\ d^\lambda \end{pmatrix} = \begin{pmatrix} -\nabla_x \ell \\ \mu\Lambda^{-1}e - c \end{pmatrix}. \tag{3.8}$$

Proposition 3.5 *Suppose that M is positive definite and that $(c(x)+s, \lambda) > 0$, then the system (3.6) has a unique solution.*

Proof. Writing

$$Q := M + A^\top \Lambda (C+S)^{-1} A,$$

and eliminating d^λ from (3.8) give

$$Q d^x = -\nabla_x \ell + A^\top (C+S)^{-1}(\mu e - C\lambda). \tag{3.9}$$

Since Q is positive definite, this equation determines d^x, while d^λ is given by the second equation in (3.8) and d^s by (3.7). □

3.3 A primal merit function

An exact penalty function associated with (3.1) is the function $\Theta_{\mu,\sigma}$ defined by

$$\Theta_{\mu,\sigma}(x,s) := \varphi_\mu(x,s) + \sigma \|s\|_P, \tag{3.10}$$

where $\sigma > 0$ is a penalty parameter and $\|\cdot\|_P$ is an arbitrary norm. The following proposition focuses on the connections between the minimizer of this merit function and the solution to the barrier problem (3.1).

Proposition 3.6 *Suppose that Assumptions* 3.1 *hold. Then* $\Theta_{\mu,\sigma}$ *is strictly convex on the set* $\{(x,s)\in\mathbb{R}^n\times\mathbb{R}^m : c(x){+}s>0\}$, *its level sets are compact and it has a unique minimizer, denoted by* $(\hat{x}_{\mu,\sigma},\hat{s}_{\mu,\sigma})$. *This one is characterized by the existence of* $\hat{\lambda}_{\mu,\sigma}\in\mathbb{R}^m$ *such that:*

$$\left\{\begin{array}{l} \nabla_x\ell(\hat{x}_{\mu,\sigma},\hat{\lambda}_{\mu,\sigma})=0,\\ (C(\hat{x}_{\mu,\sigma}){+}\hat{S}_{\mu,\sigma})\hat{\lambda}_{\mu,\sigma}=\mu e,\\ \hat{\lambda}_{\mu,\sigma}\in\sigma\partial(\|\cdot\|_P)(\hat{s}_{\mu,\sigma}),\\ (c(\hat{x}_{\mu,\sigma})+\hat{s}_{\mu,\sigma},\hat{\lambda}_{\mu,\sigma})>0. \end{array}\right. \tag{3.11}$$

The vector $\hat{\lambda}_{\mu,\sigma}$ *is uniquely determined and we note* $\hat{z}_{\mu,\sigma}:=(\hat{x}_{\mu,\sigma},\hat{s}_{\mu,\sigma},\hat{\lambda}_{\mu,\sigma})$. *Furthermore, if* $\sigma>\|\hat{\lambda}_{\mu,\sigma}\|_D$, *then* $\hat{z}_{\mu,\sigma}=\hat{z}_\mu$ *or equivalently* $\sigma\geq\|\hat{\lambda}_\mu\|_D$*; in particular Problem* (1.1) *has a strictly feasible point.*

Proof. The strict convexity of $\Theta_{\mu,\sigma}$ follows from that of φ_μ (Proposition 3.4). As in the proof of Proposition 3.4, we show that the level sets of $\Theta_{\mu,\sigma}$ are compact by proving that its asymptotic derivatives $(\Theta_{\mu,\sigma})'_\infty(d^x,d^s)$ are positive for any nonzero $(d^x,d^s)\in\mathbb{R}^n\times\mathbb{R}^m$. Using the function η defined in (3.5), one has (see (2.1)):

$$(\Theta_{\mu,\sigma})'_\infty(d^x,d^s)=f'_\infty(d^x)+\mu\sum_{i=1}^m\eta'_\infty\left((-c_{(i)})'_\infty(d^x)-d^s_{(i)}\right)+\sigma\|d^s\|_P,$$

with the convention that $\eta'_\infty(+\infty)=+\infty$. Since $\eta'_\infty(t)=+\infty$ if $t>0$ and is zero otherwise, and since $f'_\infty(d^x)>-\infty$, the asymptotic derivative $(\Theta_{\mu,\sigma})'_\infty(d^x,d^s)$ is nonpositive only if $f'_\infty(d^x)\leq-\sigma\|d^s\|_P$ and $(-c_{(i)})'_\infty(d^x)\leq d^s_{(i)}$, for all $i=1,\ldots,m$. According to Proposition 3.3, this is not possible when $d^x\neq 0$. This is not possible when $d^x=0$ either, since these inequalities would imply that $(d^x,d^s)=0$.

The compactness of the level sets of $\Theta_{\mu,\sigma}$ and the fact that $\Theta_{\mu,\sigma}$ is finite for some (x,s) satisfying $c(x)+s>0$ imply the existence of a minimizer of that function. Uniqueness follows from the strict convexity of $\Theta_{\mu,\sigma}$.

The solution pair $(\hat{x}_{\mu,\sigma},\hat{s}_{\mu,\sigma})$ is characterized by the optimality condition $0\in\partial\Theta_{\mu,\sigma}(\hat{x}_{\mu,\sigma},\hat{s}_{\mu,\sigma})$, which can also be written:

$$\nabla_x\varphi_\mu(\hat{x}_{\mu,\sigma},\hat{s}_{\mu,\sigma})=0\quad\text{and}\quad-\nabla_s\varphi_\mu(\hat{x}_{\mu,\sigma},\hat{s}_{\mu,\sigma})\in\sigma\partial(\|\cdot\|_P)(\hat{s}_{\mu,\sigma}).$$

Using (3.3), we obtain (3.11). The vector $\hat{\lambda}_{\mu,\sigma}$ is uniquely determined by the second condition in (3.11).

Suppose now that $\sigma>\|\hat{\lambda}_{\mu,\sigma}\|_D$. According to (2.2), the third condition in (3.11) can also be written:

$$\|\hat{\lambda}_{\mu,\sigma}\|_D\leq\sigma\quad\text{and}\quad\hat{\lambda}_{\mu,\sigma}^\top\hat{s}_{\mu,\sigma}=\sigma\|\hat{s}_{\mu,\sigma}\|_P. \tag{3.12}$$

The generalized Cauchy-Schwarz inequality then provides $\sigma\|\hat{s}_{\mu,\sigma}\|_P\leq\|\hat{\lambda}_{\mu,\sigma}\|_D\|\hat{s}_{\mu,\sigma}\|_P$, so that $\hat{s}_{\mu,\sigma}=0$. Hence Problem (1.1) has a strictly feasible point $\hat{x}_{\mu,\sigma}$ and $\hat{z}_{\mu,\sigma}$ satisfies (3.2). Since the latter system characterizes the unique solution to the barrier problem (3.1), $\hat{z}_{\mu,\sigma} = \hat{z}_\mu$. Note that, when $\hat{z}_{\mu,\sigma} = \hat{z}_\mu$, $\hat{\lambda}_{\mu,\sigma} = \hat{\lambda}_\mu$ and the first inequality in (3.12) implies that $\sigma \geq \|\hat{\lambda}_\mu\|_D$; conversely, if $\sigma \geq \|\hat{\lambda}_\mu\|_D$, by definition of $\hat{z}_\mu$ (Proposition 3.4), one has $\nabla_x \ell(\hat{x}_\mu, \hat{\lambda}_\mu) = 0$, $(C(\hat{x}_\mu)+\hat{S}_\mu)\hat{\lambda}_\mu = \mu e$, $\|\hat{\lambda}_\mu\|_D \leq \sigma$, and $\hat{\lambda}_\mu^\top \hat{s}_\mu = \sigma \|\hat{s}_\mu\|_P$ $(= 0)$, so that $\hat{z}_\mu$ satisfies (3.11) and $\hat{z}_{\mu,\sigma} = \hat{z}_\mu$. □

In the following proposition we give a condition on σ such that (d^x, d^s) is a descent direction of the merit function $\Theta_{\mu,\sigma}$.

Proposition 3.7 *Suppose that (x, s) satisfies $c(x) + s > 0$ and that M is symmetric positive definite. Let (d^x, d^s, d^λ) be the unique solution of* (3.6). *Then*

$$\Theta'_{\mu,\sigma}(x, s; d^x, d^s) = -(d^x)^\top M d^x - \left\| \Lambda^{1/2}(C+S)^{-1/2}(Ad^x + d^s) \right\|^2 + (\lambda+d^\lambda)^\top s - \sigma \|s\|_P. \quad (3.13)$$

Moreover, if

$$\sigma \geq \|\lambda+d^\lambda\|_D, \quad (3.14)$$

then (d^x, d^s) is a descent direction of $\Theta_{\mu,\sigma}$ at a point $(x, s) \neq (\hat{x}_\mu, \hat{s}_\mu)$, meaning that $\Theta'_{\mu,\sigma}(x, s; d^x, d^s) < 0$.

Proof. Using (3.3) and (3.7), the directional derivative of $\Theta_{\mu,\sigma}$ can be written

$$\Theta'_{\mu,\sigma}(x, s; d^x, d^s) = \left(\nabla f - \mu A^\top (C+S)^{-1} e \right)^\top d^x - \left(\mu (C+S)^{-1} e \right)^\top d^s - \sigma \|s\|_P. \quad (3.15)$$

From the first equation in (3.6):

$$\nabla f = -Md^x + A^\top (\lambda+d^\lambda). \quad (3.16)$$

On the other hand, by multiplying the second equation in (3.6) by $(C+S)^{-1}$, we obtain

$$\mu(C+S)^{-1} e = (\lambda+d^\lambda) + \Lambda (C+S)^{-1}(Ad^x + d^s). \quad (3.17)$$

With (3.16), (3.17), and $d^s = -s$, (3.15) becomes

$$\begin{aligned} \Theta'_{\mu,\sigma}(x, s; d^x, d^s) &= -\left(Md^x + A^\top \Lambda (C+S)^{-1}(Ad^x + d^s) \right)^\top d^x \\ &\quad - \left((\lambda+d^\lambda) + \Lambda(C+S)^{-1}(Ad^x + d^s) \right)^\top d^s - \sigma \|s\|_P \\ &= -(d^x)^\top M d^x + (\lambda+d^\lambda)^\top s - \left\| \Lambda^{1/2}(C+S)^{-1/2}(Ad^x + d^s) \right\|^2 - \sigma \|s\|_P \\ &\leq -(d^x)^\top M d^x - \left\| \Lambda^{1/2}(C+S)^{-1/2}(Ad^x + d^s) \right\|^2 + (\|\lambda+d^\lambda\|_D - \sigma)\|s\|_P. \end{aligned}$$

Formula (3.13) follows from this calculation.

When M is positive definite, the last inequality and (3.14) imply that the directional derivative of $\Theta_{\mu,\sigma}$ is nonpositive. If $\Theta'_{\mu,\sigma}(x,s;d^x,d^s)$ vanishes, then $d^x = 0$ (by the positive definiteness of M) and $d^s = 0$ (since $(c+s,\lambda) > 0$). Since (d^x, d^s, d^λ) is the solution to the system (3.6), we deduce that $(x, s, \lambda + d^\lambda)$ is a solution of (3.2). Now it follows from Proposition 3.4 that $(x,s) = (\hat{x}_\mu, \hat{s}_\mu)$. □

3.4 A primal-dual merit function

The merit function actually used in the algorithm is the function $\psi_{\mu,\sigma}$ obtained by adding to $\Theta_{\mu,\sigma}$ a centralization term:

$$\psi_{\mu,\sigma}(x,s,\lambda) = \Theta_{\mu,\sigma}(x,s) + \tau\mathcal{V}_\mu(x,s,\lambda), \tag{3.18}$$

where $\tau > 0$ is some positive constant and

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

This term was already considered in [11, 12, 1, 8, 2]. Its role here is to scale the displacement in λ.

To simplify the notation, we denote by z the triple (x,s,λ) and by $\mathcal{Z}$ the domain of $\psi_{\mu,\sigma}$:

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

We shall use the following derivatives:

$$\nabla\mathcal{V}_\mu(z) = \begin{pmatrix} A^\top\left(\lambda - \mu(C+S)^{-1}e\right) \\ \lambda - \mu(C+S)^{-1}e \\ c+s-\mu\Lambda^{-1}e \end{pmatrix} \tag{3.19}$$

and

$$\nabla^2\mathcal{V}_\mu(z) = \begin{pmatrix} \nabla^2_{xx}\mathcal{V}_\mu(z) & \mu A^\top(C+S)^{-2} & A^\top \\ \mu(C+S)^{-2}A & \mu(C+S)^{-2} & I \\ A & I & \mu\Lambda^{-2} \end{pmatrix}, \tag{3.20}$$

where

$$\nabla^2_{xx}\mathcal{V}_\mu(z) = \sum_{i=1}^m \left(\lambda_{(i)} - \frac{\mu}{c_{(i)}+s_{(i)}}\right)\nabla^2 c_{(i)} + \mu A^\top(C+S)^{-2}A.$$

The directional derivative of $\mathcal{V}_\mu$ along a direction $d = (d^x, d^s, d^\lambda)$ satisfying the second equation of (3.6) (i.e., the linearized perturbed complementarity conditions) shows that such a d is a descent direction of $\mathcal{V}_\mu$:

$$\begin{aligned}
&\nabla \mathcal{V}_\mu(z)^\top d \\
&= \left(\lambda - \mu(C{+}S)^{-1}e\right)^\top A d^x + \left(\lambda - \mu(C{+}S)^{-1}e\right)^\top d^s + \left(c + s - \mu\Lambda^{-1}e\right)^\top d^\lambda \\
&= \left(\lambda - \mu(C{+}S)^{-1}e\right)^\top \left(A d^x + d^s + \Lambda^{-1}(C{+}S)d^\lambda\right) \\
&= \left(\lambda - \mu(C{+}S)^{-1}e\right)^\top \Lambda^{-1}\left(\mu e - (C{+}S)\lambda\right) \\
&= -\left(\mu e - (C{+}S)\lambda\right)^\top \Lambda^{-1}(C{+}S)^{-1}\left(\mu e - (C{+}S)\lambda\right) \\
&= -\left\|\Lambda^{-1/2}(C{+}S)^{-1/2}\left(\mu e - (C{+}S)\lambda\right)\right\|^2 . \qquad (3.21)
\end{aligned}$$

The merit function $\psi_{\mu,\sigma}$ is not necessarily convex (see [2] for an example), but this will not raise any difficulty, since it has a unique minimizer.

Proposition 3.8 *Suppose that Assumptions* 3.1 *hold. Then, $\psi_{\mu,\sigma}$ has for unique minimizer the triple $\hat{z}_{\mu,\sigma} = (\hat{x}_{\mu,\sigma}, \hat{s}_{\mu,\sigma}, \hat{\lambda}_{\mu,\sigma})$ given by Proposition* 3.6.

Proof. Since $(\hat{x}_{\mu,\sigma}, \hat{s}_{\mu,\sigma})$ is a minimizer of $\Theta_{\mu,\sigma}$:

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

On the other hand, since $t \mapsto t{-}\mu \log t$ is minimized at $t = \mu$ and since (by the perturbed complementarity) $\mu = \left(c_{(i)}(\hat{x}_{\mu,\sigma}) + (\hat{s}_{\mu,\sigma})_{(i)}\right)(\hat{\lambda}_{\mu,\sigma})_{(i)}$ for all index i, we have

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

Adding up the preceding two inequalities gives $\psi_{\mu,\sigma}(\hat{z}_{\mu,\sigma}) \le \psi_{\mu,\sigma}(z)$ for all $z \in \mathcal{Z}$. Hence $\hat{z}_{\mu,\sigma}$ minimizes $\psi_{\mu,\sigma}$.

We still have to show that $\hat{z}_{\mu,\sigma}$ is the unique minimizer of $\psi_{\mu,\sigma}$. If $z = (x, s, \lambda)$ minimizes the nondifferentiable function $\psi_{\mu,\sigma}$, $\psi'_{\mu,\sigma}(z; d) \ge 0$ for all d or, equivalently, $\nabla_{(x,\lambda)}\psi_{\mu,\sigma}(z) = 0$ and $-\left(\nabla_s\varphi_\mu(x, s) + \tau\nabla_s\mathcal{V}_\mu(z)\right) \in \sigma\partial(\|\cdot\|_P)(s)$. By (3.3) and (3.19), this can be written:

$$\begin{cases}
\nabla f(x) - (1{+}\tau)\mu A(x)^\top (C(x){+}S)^{-1}e + \tau A(x)^\top \lambda = 0 \\
-(1{+}\tau)\mu(C(x){+}S)^{-1}e + \tau\lambda + \nu = 0 \\
c(x) + s - \mu\Lambda^{-1}e = 0,
\end{cases}$$

for some $\nu \in \sigma\partial(\|\cdot\|_P)(s)$. By the third equation above, $\lambda = \mu(C(x){+}S)^{-1}e$, so that $\lambda = \nu$ by the second equation. Then the previous system becomes

$$\begin{cases}
\nabla_x \ell(x, \lambda) = 0 \\
(C(x){+}S)\lambda = \mu e \\
\lambda \in \sigma\partial(\|\cdot\|_P)(s).
\end{cases}$$

By Proposition 3.6, $z = \hat{z}_{\mu,\sigma}$. □

We have seen with Proposition 3.7, that the quasi-Newton direction d solving (3.6) is a descent direction of $\Theta_{\mu,\sigma}$. According to the calculation (3.21), it is not surprising that d is also a descent direction of $\psi_{\mu,\sigma}$.

Proposition 3.9 *Suppose that $z \in \mathcal{Z}$ and that M is symmetric positive definite. Let $d = (d^x, d^s, d^\lambda)$ be the unique solution of* (3.6). *Then*

$$\begin{aligned}\psi'_{\mu,\sigma}(z;d) &= -(d^x)^\top M d^x - \left\|\Lambda^{1/2}(C{+}S)^{-1/2}(Ad^x + d^s)\right\|^2 + (\lambda{+}d^\lambda)^\top s - \sigma\|s\|_P \\ &\quad -\tau \left\|\Lambda^{-1/2}(C{+}S)^{-1/2}\left(\mu e - (C{+}S)\lambda\right)\right\|^2,\end{aligned}$$

so that, if $\sigma \geq \|\lambda{+}d^\lambda\|_D$ and $z \neq \hat{z}_\mu$, d is a descent direction of $\psi_{\mu,\sigma}$, meaning that $\psi'_{\mu,\sigma}(z;d) < 0$.

Proof. We have $\psi'_{\mu,\sigma}(z;d) = \Theta'_{\mu,\sigma}(x,s;d^x,d^s) + \tau\nabla\mathcal{V}_\mu(z)^\top d$. The formula for $\psi'_{\mu,\sigma}(z;d)$ thus follows from (3.13) and (3.21). Suppose now that $\sigma \geq \|\lambda{+}d^\lambda\|_D$, then from the generalized Cauchy-Schwarz inequality

$$\begin{aligned}\psi'_{\mu,\sigma}(z;d) &\leq -(d^x)^\top M d^x - \left\|\Lambda^{1/2}(C{+}S)^{-1/2}(Ad^x + d^s)\right\|^2 \\ &\quad -\tau \left\|\Lambda^{-1/2}(C{+}S)^{-1/2}\left(\mu e - (C{+}S)\lambda\right)\right\|^2,\end{aligned}$$

which is nonpositive. If $\psi'_{\mu,\sigma}(z;d)$ vanishes, one deduces that $d^x = 0$ (since M is positive definite) and next that $d^s = 0$ and $(C{+}S)\lambda - \mu e = 0$ (since $(c(x) + s, \lambda) > 0$). Now, using the second equation of (3.6), we obtain that $d = 0$. Since then, the right hand side of (3.6) vanishes, (3.2) holds and $z = \hat{z}_\mu$ by Proposition 3.4. We have thus proven that $\psi'_{\mu,\sigma}(z;d) < 0$ if in addition $z \neq \hat{z}_\mu$. □

3.5 Algorithm A_μ

We can now state one iteration of the algorithm used to solve the perturbed KKT system (3.2), with fixed $\mu > 0$. The constants $\omega \in\,]0, \frac{1}{2}[$ (Armijo's slope), $0 < \xi \leq \xi' < 1$ (backtracking reduction coefficients), $\tau > 0$ (centralization factor), and $\bar{\sigma} > 0$ (penalty factor threshold) are given independently of the iteration index. At the beginning of the iteration, the current iterate $z = (x, s, \lambda) \in \mathcal{Z}$ is supposed available, as well as a positive scalar σ_{old} (the penalty factor used in the preceding iteration) and a positive definite matrix M approximating the Hessian of the Lagrangian $\nabla^2_{xx}\ell(x,\lambda)$.

ALGORITHM A_μ for solving (3.2) (one iteration, from (z, M) to (z_+, M_+))

1. Compute $d := (d^x, d^s, d^\lambda)$, the unique solution to the linear system (3.6). If $d = 0$, stop (z solves the system (3.2)).
2. Update σ using the following rule : if $\sigma_{\text{old}} \geq \|\lambda{+}d^\lambda\|_D + \bar{\sigma}$, then $\sigma := \sigma_{\text{old}}$, else $\sigma := \max(1.1\,\sigma_{\text{old}}, \|\lambda{+}d^\lambda\|_D + \bar{\sigma})$.
3. Compute a stepsize $\alpha > 0$ by backtracking:

   3.0. Set $\alpha = 1$.

   3.1. While $z + \alpha d \notin \mathcal{Z}$, choose a new stepsize α in $[\xi\alpha, \xi'\alpha]$.

3.2. While the *sufficient decrease condition*

$$\psi_{\mu,\sigma}(z+\alpha d) \leq \psi_{\mu,\sigma}(z) + \omega\alpha\, \psi'_{\mu,\sigma}(z;d) \tag{3.22}$$

is not satisfied, choose a new stepsize α in $[\xi\alpha, \xi'\alpha]$.

3.3. Set $z_+ := z + \alpha d$.

4. Update M by the BFGS formula

$$M_+ := M - \frac{M\delta\delta^\top M}{\delta^\top M\delta} + \frac{\gamma\gamma^\top}{\gamma^\top\delta}, \tag{3.23}$$

where γ and δ are given by

$$\delta := x_+ - x \quad \text{and} \quad \gamma := \nabla_x \ell(x_+, \lambda_+) - \nabla_x \ell(x, \lambda_+). \tag{3.24}$$

At this stage of the presentation, the meaning of the steps forming Algorithm A_μ should be quite clear. In Step 2, the penalty parameter σ_{old} is updated into σ in order to ensure that the direction d computed in Step 1 be a descent direction of $\psi_{\mu,\sigma}$ (use Proposition 3.9 and observe that at this stage, $z \neq \hat{z}_\mu$). Then, the backtracking line-search in Step 3 is guaranteed to find a stepsize $\alpha > 0$ ensuring $z + \alpha d \in \mathcal{Z}$ and satisfying the Armijo condition (3.22). In Step 4, the matrix M is updated into M_+ by the BFGS formula to be a better approximation of the Hessian of the Lagrangian. As usual, $\gamma \in \mathbb{R}^n$ is the change in the gradient of the Lagrangian, with a multiplier fixed at its new value $\lambda_+ > 0$. The matrix M_+ is positive definite, since $\gamma^\top\delta > 0$ by the strong convexity of the Lagrangian (Lemma 3.2).

4 Analysis of Algorithm A_μ

In this section we prove that if μ is fixed and if the sequence of penalty parameters remains bounded, then the sequence of iterates converges q-superlinearly to a point on the central path. Hence, in all this section, we assume:

Assumption 4.1 The sequence $\{\sigma_k\}$ generated by Algorithm A_μ is bounded.

By Step 2 of Algorithm A_μ, each time σ_k is updated, it is at least multiplied by a factor not smaller than 1. Therefore, Assumption 4.1 implies that the sequence $\{\sigma_k\}$ is stationary for k large enough:

$$\sigma_k = \sigma \text{ for } k \text{ large.}$$

With Algorithm A_μ, the boundedness of $\{\sigma_k\}$ is equivalent to that of $\{\lambda_k + d_k^\lambda\}$. A limited number of experiments with Algorithm A_μ has shown that the latter sequence is actually bounded when Problem (1.1) has a strictly feasible point (Slater's condition). Of course, if Algorithm A_μ converges to the solution to the barrier problem (3.1),

$c(\hat{x}_\mu) > 0$ and Slater's condition holds, but we do not know whether this is a sufficient condition for stabilizing the penalty factors.

The proof of convergence is organized in three stages. First, we show the global convergence of the sequence $\{z_k\}$ to $\hat{z}_\mu$, the unique primal-dual solution to the barrier problem (3.1) (Section 4.1). With the update rule of σ, $\sigma > \|\hat{\lambda}_\mu\|_D$ and $\hat{z}_{\mu,\sigma} = \hat{z}_\mu$. It is then possible to use the behavior of $\psi_{\mu,\sigma}$ around $\hat{z}_{\mu,\sigma} = \hat{z}_\mu$ to show the r-linear convergence of $\{z_k\}$ (Section 4.2). This result is then used to show that the updated matrix M_k provides a good value $M_k d_k^x$ along the search direction d_k^x. This allows us to prove that the unit stepsize is accepted by the line-search after finitely many iterations. The q-superlinear convergence of $\{z_k\}$ then follows easily (Section 4.3).

4.1 Convergence

We denote by $z_1 = (x_1, s_1, \lambda_1) \in \mathcal{Z}$ the first iterate obtained with a penalty parameter set to σ and by

$$\mathcal{L}^{\mathrm{PD}}_{1,\sigma} := \{z \in \mathcal{Z} : \psi_{\mu,\sigma}(z) \le \psi_{\mu,\sigma}(z_1)\}$$

the level set of $\psi_{\mu,\sigma}$ determined by z_1. We denote by

$$\widetilde{\psi}_\mu(z) = \varphi_\mu(x,s) + \tau \mathcal{V}_\mu(z) \tag{4.1}$$

the differentiable part of $\psi_{\mu,\sigma}$.

The following Lemma gives the contribution of the line-search to the convergence of the sequence generated by Algorithm A_μ. Such a result is standard when the objective function is differentiable and finite-valued. It dates back at least to Zoutendijk [18] (for a proof, see [6]). Since $\psi_{\mu,\sigma}$ is nondifferentiable and takes infinite values, we prefer giving a specific proof, which in fact is very close to the original one.

Lemma 4.2 *Suppose that $\widetilde{\psi}_\mu$ is $C^{1,1}$ on an open convex neighborhood $\mathcal{N}$ of the level set $\mathcal{L}^{\mathrm{PD}}_{1,\sigma}$, and denote by L the Lipschitz modulus of $\widetilde{\psi}'_\mu$. Then for any descent direction d of $\psi_{\mu,\sigma}$ satisfying $d^s = -s$ and for α determined by the line-search in Step 3 of algorithm A_μ, one of the following inequalities holds:*

$$\psi_{\mu,\sigma}(z+\alpha d) \le \psi_{\mu,\sigma}(z) - K_0 |\psi'_{\mu,\sigma}(z;d)|,$$

$$\psi_{\mu,\sigma}(z+\alpha d) \le \psi_{\mu,\sigma}(z) - K_0 \frac{|\psi'_{\mu,\sigma}(z;d)|^2}{\|d\|^2},$$

where K_0 is the constant $\min\left(\omega, \frac{2\xi\omega(1-\omega)}{L}\right)$.

Proof. If the line-search is satisfied with $\alpha = 1$, the first inequality holds with $K_0 = \omega$. Suppose now that $\alpha < 1$, which means that some stepsize $\bar{\alpha}$ satisfying $\xi\bar{\alpha} \le \alpha \le \xi'\bar{\alpha}$ is not accepted by the line-search. This rejection of $\bar{\alpha}$ may have two reasons. Either $z + \bar{\alpha} d \notin \mathcal{Z}$ or $z + \bar{\alpha} d \in \mathcal{Z}$ but $\psi_{\mu,\sigma}(z+\bar{\alpha} d) > \psi_{\mu,\sigma}(z) + \omega\bar{\alpha}\,\psi'_{\mu,\sigma}(z;d)$. In the first case, there exists $\tilde{\alpha} \in \,]\alpha, \bar{\alpha}[$ such that $z + \tilde{\alpha} d \in \mathcal{N} \backslash \mathcal{L}^{\mathrm{PD}}_{1,\sigma}$ ($\mathcal{N}$ must be included in $\mathcal{Z}$). Then

$$\psi_{\mu,\sigma}(z+\tilde{\alpha} d) > \psi_{\mu,\sigma}(z_1) \ge \psi_{\mu,\sigma}(z) > \psi_{\mu,\sigma}(z) + \omega\tilde{\alpha}\,\psi'_{\mu,\sigma}(z;d).$$

If we set $\tilde{\alpha} := \bar{\alpha}$ in the second case, in either case, we have $\alpha \geq \xi\bar{\alpha} \geq \xi\tilde{\alpha}$ and

$$\psi_{\mu,\sigma}(z+\tilde{\alpha}d) > \psi_{\mu,\sigma}(z) + \omega\tilde{\alpha}\,\psi'_{\mu,\sigma}(z;d). \tag{4.2}$$

Using a Taylor expansion of $\widetilde{\psi}_\mu$, the Cauchy-Schwarz inequality, the Lipschitz continuity of $\nabla\widetilde{\psi}_\mu$ and $\tilde{\alpha} \leq 1$:

$$\begin{aligned}
&\psi_{\mu,\sigma}(z+\tilde{\alpha}d) - \psi_{\mu,\sigma}(z)\\
&\quad= \widetilde{\psi}_\mu(z+\tilde{\alpha}d) - \widetilde{\psi}_\mu(z) + \sigma(\|s+\tilde{\alpha}\,d^s\|_P - \|s\|_P)\\
&\quad= \tilde{\alpha}\nabla\widetilde{\psi}_\mu(z)^\top d + \int_0^1 \Big(\nabla\widetilde{\psi}_\mu(z+t\tilde{\alpha}d) - \nabla\widetilde{\psi}_\mu(z)\Big)^\top (\tilde{\alpha}d)\,dt + \sigma(\|s-\tilde{\alpha}s\|_P - \|s\|_P)\\
&\quad\leq \tilde{\alpha}\nabla\widetilde{\psi}_\mu(z)^\top d + \int_0^1 Lt\tilde{\alpha}^2\|d\|^2 dt - \sigma\tilde{\alpha}\|s\|_P\\
&\quad= \tilde{\alpha}\psi'_{\mu,\sigma}(z;d) + \frac{1}{2}L\tilde{\alpha}^2\|d\|^2.
\end{aligned}$$

Then (4.2) yields a lower bound on $\tilde{\alpha}$:

$$\tilde{\alpha} > \frac{2(1-\omega)}{L}\frac{|\psi'_{\mu,\sigma}(z;d)|}{\|d\|^2}.$$

Now, the sufficient decrease condition (3.22) is satisfied with α:

$$\begin{aligned}
\psi_{\mu,\sigma}(z+\alpha d) &\leq \psi_{\mu,\sigma}(z) - \omega\alpha\,|\psi'_{\mu,\sigma}(z;d)|\\
&\leq \psi_{\mu,\sigma}(z) - \omega\xi\tilde{\alpha}\,|\psi'_{\mu,\sigma}(z;d)|\\
&\leq \psi_{\mu,\sigma}(z) - \frac{2\xi\omega(1-\omega)}{L}\frac{|\psi'_{\mu,\sigma}(z;d)|^2}{\|d\|^2},
\end{aligned}$$

so that the second inequality holds with $K_0 = \frac{2\xi\omega(1-\omega)}{L}$. □

A consequence of the following lemma is that, because the iterates (x,s,λ) remain in the level set $\mathcal{L}^{\mathrm{PD}}_{1,\sigma}$, the sequence $\{(c(x)+s,\lambda)\}$ is bounded and bounded away from zero. This property plays an important role to control the contribution of the IP aspect of the algorithm.

Lemma 4.3 *Suppose that Assumptions* 3.1 *and* 4.1 *hold. Then, the level set $\mathcal{L}^{\mathrm{PD}}_{1,\sigma}$ is compact and there exist positive constants K_1 and K_2 such that*

$$K_1 \leq (c(x)+s,\lambda) \leq K_2, \quad \textit{for all } z \in \mathcal{L}^{\mathrm{PD}}_{1,\sigma}.$$

Proof. Since $\mathcal{V}_\mu(z)$ is bounded below by $m\mu(1-\log\mu)$, there is a constant $K'_1 > 0$ such that $\Theta_{\mu,\sigma}(x,s) \leq K'_1$ for all $z \in \mathcal{L}^{\mathrm{PD}}_{1,\sigma}$. By Assumptions 3.1 and Proposition 3.6, the level set $\mathcal{L}' := \{(x,s) : c(x)+s > 0,\ \Theta_{\mu,\sigma}(x,s) \leq K'_1\}$ is compact. By continuity of $(x,s) \mapsto c(x)+s$, $\{c(x)+s : (x,s) \in \mathcal{L}'\}$ is compact, so that $c(x)+s$ is bounded for $(x,s) \in \mathcal{L}'$, hence for $z \in \mathcal{L}^{\mathrm{PD}}_{1,\sigma}$. It is now also clear that $c(x)+s$ is bounded away from zero for $z \in \mathcal{L}^{\mathrm{PD}}_{1,\sigma}$, because $\varphi_\mu(x,s) \leq K'_1$ and $f(x)$ is bounded below.

By the compactness of $\mathcal{L}'$, $\Theta_{\mu,\sigma}$ is bounded below on $\mathcal{L}'$, hence $\mathcal{V}_\mu$ is bounded above on $\mathcal{L}^{\rm PD}_{1,\sigma}$. Now, from the form of the function $t \mapsto t - \mu \log t$, one deduces that, for some positive constants K'_2 and K'_3: $K'_2 \leq \lambda_{(i)}(c_{(i)}(x) + s_{(i)}) \leq K'_3$, for all $z \in \mathcal{L}^{\rm PD}_{1,\sigma}$ and all index i. Therefore, the λ-components of the z's in $\mathcal{L}^{\rm PD}_{1,\sigma}$ are bounded and bounded away from zero.

We have shown that $\mathcal{L}^{\rm PD}_{1,\sigma}$ is included in a bounded set. Hence, it is compact by continuity of $\psi_{\mu,\sigma}$. □

The search direction d_k of Algorithm A_μ is determined by the system (3.6). This one highlights the two aspects of the method: the IP approach is represented by the matrices Λ_k and $C(x_k)+S_k$, while the quasi-Newton technique manifests itself through the matrix M_k. One can view Lemma 4.3 as a way of controlling the contribution of the IP approach; while the next lemma allows us to master what is supplied by the BFGS updates. Lemma 4.4 claims indeed that, at least for a proportion of the iterations, there are bounds on various effects of the matrices M_k on the displacement δ_k (see Byrd and Nocedal [6] for a proof). 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 ceiling operator: $\lceil x \rceil = i$, when $i-1 < x \leq i$ and $i \in \mathbb{N}$.

Lemma 4.4 *Let $\{M_k\}$ be positive definite matrices generated by the BFGS formula using pairs of vectors $\{(\gamma_k, \delta_k)\}_{k\geq 1}$, satisfying for all $k \geq 1$*

$$\gamma_k^\top \delta_k \geq a_1\|\delta_k\|^2 \quad \textit{and} \quad \gamma_k^\top\delta_k \geq a_2\|\gamma_k\|^2, \tag{4.3}$$

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 $k \geq 1$,

$$b_1 \leq \cos\theta_j \quad \textit{and} \quad b_2 \leq \frac{\|M_j\delta_j\|}{\|\delta_j\|} \leq b_3, \tag{4.4}$$

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

The next lemma shows that the assumptions (4.3) made on γ_k and δ_k are satisfied in our context.

Lemma 4.5 *Suppose that Assumptions* 3.1 *and* 4.1 *hold, and that f and c are of class $C^{1,1}$. There exist constants $a_1 > 0$ and $a_2 > 0$ such that for all $k \geq 1$*

$$\gamma_k^\top \delta_k \geq a_1\|\delta_k\|^2 \quad \textit{and} \quad \gamma_k^\top\delta_k \geq a_2\|\gamma_k\|^2.$$

Proof. Let us first observe that, because $\psi_{\mu,\sigma}$ decreases at each iteration, the iterates generated by Algorithm A_μ stay in the level set $\mathcal{L}^{\rm PD}_{1,\sigma}$ and Lemma 4.3 can be applied. According to Lemma 3.2 and the fact that λ is bounded away from zero (Lemma 4.3), the Lagrangian is strongly convex on $\mathcal{L}^{\rm PD}_{1,\sigma}$, with a modulus $\kappa > 0$. Therefore

$$\gamma_k^\top\delta_k = (\nabla_x \ell(x_{k+1}, \lambda_{k+1}) - \nabla_x \ell(x_k, \lambda_{k+1}))^\top (x_{k+1} - x_k) \geq \kappa\|\delta_k\|^2$$

and the first inequality holds with $a_1 = \kappa$. The second one can be deduced from first inequality and $\|\gamma_k\| \leq K'\|\delta_k\|$, which follows from the Lipschitz continuity of ∇f and ∇c, and the boundedness of λ given by Lemma 4.3. □

We are now in position to prove that the sequence $\{z_k\}$ converges to $\hat{z}_\mu$. Since $\hat{z}_\mu$ is strictly feasible, necessarily, this event can occur only if Problem (1.1) has a strictly feasible point.

Theorem 4.6 *Suppose that Assumptions* 3.1 *and* 4.1 *hold, that f and c are of class $C^{1,1}$, and that Algorithm A_μ does not stop in Step* 1. *Then, Algorithm A_μ generates a sequence $\{z_k\}$ converging to $\hat{z}_\mu$ and we have $\sigma \geq \|\hat{\lambda}_\mu\|_D$.*

Proof. We denote by K'_1, K'_2, ... positive constants (independent of the iteration index). Given an iteration index j, we use the notation

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

The bounds on $(c(x)+s, \lambda)$ given by Lemma 4.3 and the fact that f and c are of class $C^{1,1}$ imply that $\widetilde{\psi}_\mu$ given by (4.1) is of class $C^{1,1}$ on some open convex neighborhood of the level set $\mathcal{L}^{\mathrm{PD}}_{1,\sigma}$. For example, one can take the neighborhood

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

where $g : (x,s) \mapsto c(x) + s$ and $\mathcal{O}$ is an open bounded convex set containing $\mathcal{L}^{\mathrm{PD}}_{1,\sigma}$ (this set $\mathcal{O}$ is used to have c' bounded on the given neighborhood).

Therefore, by the line-search and Lemma 4.2, there is a positive constant K'_1 such that either

$$\psi_{\mu,\sigma}(z_{k+1}) \leq \psi_{\mu,\sigma}(z_k) - K'_1 |\psi'_{\mu,\sigma}(z_k; d_k)| \tag{4.5}$$

or

$$\psi_{\mu,\sigma}(z_{k+1}) \leq \psi_{\mu,\sigma}(z_k) - K'_1 \frac{|\psi'_{\mu,\sigma}(z_k; d_k)|^2}{\|d_k\|^2}. \tag{4.6}$$

Let us now apply Lemma 4.4: fix $r \in]0,1[$ and denote by J the set of indices j for which (4.4) holds. Since Algorithm A_μ ensures $\sigma \geq \|\lambda_j + d^\lambda_j\|_D + \bar{\sigma}$, using Proposition 3.9 and the bounds from Lemma 4.3, one has for $j \in J$:

$$\begin{aligned}
|\psi'_{\mu,\sigma}(z_j; d_j)| \ &\geq \ (d^x_j)^\top M_j d^x_j + \|\Lambda_j^{1/2}(C_j+S_j)^{-1/2}(A_j d^x_j + d^s_j)\|^2 \\
&\quad + \tau\|\Lambda_j^{-1/2}(C_j+S_j)^{-1/2}((C_j+S_j)\lambda_j - \mu e)\|^2 + \bar{\sigma}\|s_j\|_P \\
&\geq \ \frac{b_1}{b_3}\|M_j d^x_j\|^2 + K_1 K_2^{-1}\|A_j d^x_j + d^s_j\|^2 \\
&\quad + \tau K_2^{-2}\|(C_j+S_j)\lambda_j - \mu e\|^2 + \bar{\sigma}\|s_j\|_P \\
&\geq \ K'_2\left(\|M_j d^x_j\|^2 + \|A_j d^x_j + d^s_j\|^2 + \|(C_j+S_j)\lambda_j - \mu e\|^2 + \|s_j\|_P\right).
\end{aligned}$$

On the other hand, by (3.6) and the fact that $\{s_j\}$ is bounded:

$$\begin{aligned}
\|d_j\|^2 &= \|d_j^x\|^2 + \|d_j^s\|^2 + \|d_j^\lambda\|^2 \\
&= \|d_j^x\|^2 + \|s_j\|^2 + \|(C_j{+}S_j)^{-1}\left(\mu e - (C_j{+}S_j)\lambda_j - \Lambda_j(A_j d_j^x + d_j^s)\right)\|^2 \\
&\leq \frac{1}{b_2^2}\|M_j d_j^x\|^2 + \|s_j\|^2 + 2K_1^{-2}\|(C_j{+}S_j)\lambda_j - \mu e\|^2 + 2K_1^{-2}K_2^2\|A_j d_j^x + d_j^s\|^2 \\
&\leq K_3'\left(\|M_j d_j^x\|^2 + \|(C_j{+}S_j)\lambda_j - \mu e\|^2 + \|A_j d_j^x + d_j^s\|^2 + \|s_j\|_P\right).
\end{aligned}$$

Combining these last two estimates with (4.5) or (4.6) gives for some positive constant K_4' and for any $j \in J$:

$$\begin{aligned}
&\psi_{\mu,\sigma}(z_{j+1}) - \psi_{\mu,\sigma}(z_j) \\
&\quad\leq -K_4'\left(\|M_j d_j^x\|^2 + \|(C_j{+}S_j)\lambda_j - \mu e\|^2 + \|A_j d_j^x + d_j^s\|^2 + \|s_j\|_P\right). \qquad (4.7)
\end{aligned}$$

Now, since the sequence $\{\psi_{\mu,\sigma}(z_k)\}$ is decreasing (by the line-search) and bounded below (by $\psi_{\mu,\sigma}(\hat{z}_{\mu,\sigma})$), it converges and we deduce from the preceding inequality that $M_j d_j^x$, $(C_j{+}S_j)\lambda_j - \mu e$, $A_j d_j^x + d_j^s$ and s_j tend to zero when $j \to \infty$ in J. According to the linear system (3.6), this implies that $\nabla_x \ell(x_j, \lambda_j + d_j^\lambda)$ and $(C_j{+}S_j)d_j^\lambda$ tend to zero. Therefore, $d_j^\lambda \to 0$ ($C_j{+}S_j$ is bounded away from zero) and $A(x_j)^\top d_j^\lambda \to 0$ (x_j is in the compact $\mathcal{L}_{1,\sigma}^{\rm PD}$). Finally $\nabla_x \ell(x_j, \lambda_j)$, $(C_j{+}S_j)\lambda_j - \mu e$, and s_j tend to zero. This means that any limit point z of $\{z_j\}_{j\in J}$ satisfies (3.2), i.e., $z = \hat{z}_\mu$. Since $\{z_j\}$ remains in the compact $\mathcal{L}_{1,\sigma}^{\rm PD}$ (Lemma 4.3), the whole sequence $\{z_j\}_{j\in J}$ converges to $\hat{z}_\mu$.

Using the update rule of σ in Algorithm A_μ, we have $\sigma \geq \|\hat{\lambda}_\mu\|_D$ and, thanks to Proposition 3.6, $\hat{z}_\mu = \hat{z}_{\mu,\sigma}$. Therefore $\{\psi_{\mu,\sigma}(z_k)\}$ converges to $\psi_{\mu,\sigma}(\hat{z}_{\mu,\sigma})$. In addition, $\{z_k\}$ remains in the compact $\mathcal{L}_{1,\sigma}^{\rm PD}$ and $\psi_{\mu,\sigma}$ has a unique minimizer $\hat{z}_{\mu,\sigma}$ (Proposition 3.8). As a result, the whole sequence $\{z_k\}$ converges to $\hat{z}_\mu$. □

4.2 R-linear convergence

Knowing that the sequence $\{z_k\}$ converges to the unique solution $\hat{z}_\mu$ of the barrier problem (3.1) and that $\hat{z}_\mu = \hat{z}_{\mu,\sigma}$, we can now study its speed of convergence. The analysis of the q-superlinear convergence in Section 4.3 requires that we first show

$$\sum_{k\geq 1} \|z_k - \hat{z}_\mu\| < \infty.$$

The convergence of this series results directly from the r-linear convergence of $\{z_k\}$:

$$\limsup_{k\to\infty} \|z_k - \hat{z}_\mu\|^{1/k} < 1,$$

which is proven in this section (Theorem 4.10). This one results from the strong convexity of $\psi_{\mu,\sigma}$ near $\hat{z}_{\mu,\sigma}$ (see Lemmas 4.7 and 4.9; it is therefore important to have $\hat{z}_\mu = \hat{z}_{\mu,\sigma}$ to take advantage of this convexity property) and from the contribution of the BFGS and IP techniques summarized in Lemmas 4.3 and 4.4.

Lemma 4.7 *Suppose that Assumptions* 3.1 *hold. Then, the functions* $\widetilde{\psi}_\mu$ *and* $\psi_{\mu,\sigma}$ *are strongly convex in the neighborhood of any point* $z = (x, s, \lambda) \in \mathcal{Z}$ *satisfying the centrality condition* $(C(x) + S)\lambda = \mu e$.

Proof. Let $z \in \mathcal{Z}$ be a point satisfying $(C(x) + S)\lambda = \mu e$. Using (3.4), (3.20), and the fact that $(C(x) + S)\lambda = \mu e$, the Hessian of $\widetilde{\psi}_\mu$ at z can be written

$$\nabla^2\widetilde{\psi}_\mu(x, s, \lambda) = \begin{pmatrix} \nabla^2_{xx}\ell(x,\lambda) + (1{+}\tau)\mu A^\top(C{+}S)^{-2}A & (1{+}\tau)\mu A^\top(C{+}S)^{-2} & \tau A^\top \\ (1{+}\tau)\mu(C{+}S)^{-2}A & (1{+}\tau)\mu(C{+}S)^{-2} & \tau I \\ \tau A & \tau I & \tau\mu^{-1}(C{+}S)^2 \end{pmatrix}.$$

To establish that $\widetilde{\psi}_\mu$ is strongly convex in the neighborhood of z, it is enough to show that the matrix above is positive definite. Multiplying this matrix on both sides by a vector $(u, v, w) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ gives

$$u^\top\nabla^2_{xx}\ell(x,\lambda)u + \mu\|(C{+}S)^{-1}(Au + v)\|^2 + \tau\|\mu^{1/2}(C{+}S)^{-1}(Au + v) + \mu^{-1/2}(C{+}S)w\|^2,$$

which is nonnegative. If it vanishes, one deduces that $u = 0$ (since $\nabla^2_{xx}\ell(x, \lambda)$ is positive definite for fixed $\lambda > 0$, a consequence of Assumptions 3.1 and Lemma 3.2), and next that $v = w = 0$ (since $c(x) + s > 0$). Hence $\nabla^2\widetilde{\psi}_\mu(z)$ is positive definite.

To extend this property to $\psi_{\mu,\sigma}$, it suffices to observe that $\psi_{\mu,\sigma}$ is the sum of $\widetilde{\psi}_\mu$ and of the convex function $(x, s, \lambda) \mapsto \sigma\|s\|_P$. □

Lemma 4.8 *Let* a, α, *and* β *be nonnegative numbers, such that* $a \le \alpha\, a^{1/2} + \beta$. *Then* $a \le \alpha^2 + 2\beta$.

Proof. If $a \le \beta$, the result clearly holds. Otherwise, $a > \beta$ and squaring both sides of $a - \beta \le \alpha\, a^{1/2}$ yields $a^2 \le \alpha^2 a + 2a\beta - \beta^2$. Since $a > 0$, the result follows. □

Lemma 4.9 *Suppose that Assumptions* 3.1 *and* 4.1 *hold. Then there exist a constant* $a > 0$ *and an open neighborhood* $\mathcal{N} \subset \mathcal{Z}$ *of* $\hat{z}_{\mu,\sigma}$, *such that for all* $z \in \mathcal{N}$

$$\begin{aligned} a\|z - \hat{z}_{\mu,\sigma}\|^2 &\le \psi_{\mu,\sigma}(z) - \psi_{\mu,\sigma}(\hat{z}_{\mu,\sigma}) \\ &\le \frac{1}{a}\left(\|\nabla_x\ell(x,\lambda)\|^2 + \|(C{+}S)\lambda - \mu e\|^2 + \|s - \hat{s}_{\mu,\sigma}\|_P\right). \end{aligned} \tag{4.8}$$

Proof. For the inequality on the left in (4.8), we first use the strong convexity of $\psi_{\mu,\sigma}$ in the neighborhood of $\hat{z}_{\mu,\sigma}$ (Lemma 4.7): for some neighborhood $\mathcal{N} \subset \mathcal{Z}$ of $\hat{z}_{\mu,\sigma}$, there exists a positive constant a' such that, for all $z \in \mathcal{N}$, $\psi_{\mu,\sigma}(z) \ge \psi_{\mu,\sigma}(\hat{z}_{\mu,\sigma}) + \psi'_{\mu,\sigma}(\hat{z}_{\mu,\sigma}; z - \hat{z}_{\mu,\sigma}) + a'\|z - \hat{z}_{\mu,\sigma}\|^2$. Since $\psi_{\mu,\sigma}$ is minimized at $\hat{z}_{\mu,\sigma}$, one gets $\psi'_{\mu,\sigma}(\hat{z}_{\mu,\sigma}; z - \hat{z}_{\mu,\sigma}) \ge 0$. Thus the inequality on the left in (4.8) holds on $\mathcal{N}$ with $a = a'$.

For the inequality on the right, we first use the convexity of $\widetilde{\psi}_\mu$ (see (4.1)) near $\hat{z}_{\mu,\sigma}$ (Lemma 4.7), (3.3) and (3.19) to write

$$\begin{aligned} \psi_{\mu,\sigma}(z) - \psi_{\mu,\sigma}(\hat{z}_{\mu,\sigma}) &= \widetilde{\psi}_\mu(z) - \widetilde{\psi}_\mu(\hat{z}_{\mu,\sigma}) + \sigma(\|s\|_P - \|\hat{s}_{\mu,\sigma}\|_P) \\ &\le \nabla\widetilde{\psi}_\mu(z)^\top(z - \hat{z}_{\mu,\sigma}) + \sigma\|s - \hat{s}_{\mu,\sigma}\|_P \\ &= g^\top(z - \hat{z}_{\mu,\sigma}) - \lambda^\top(s - \hat{s}_{\mu,\sigma}) + \sigma\|s - \hat{s}_{\mu,\sigma}\|_P, \end{aligned}$$

where

$$g = \begin{pmatrix} \nabla_x \ell(x,\lambda) + (1+\tau)A^\top (C+S)^{-1}\left((C+S)\lambda - \mu e\right) \\ (1+\tau)(C+S)^{-1}\left((C+S)\lambda - \mu e\right) \\ \tau\Lambda^{-1}\left((C+S)\lambda - \mu e\right) \end{pmatrix}.$$

For λ in a neighborhood of $\hat{\lambda}_{\mu,\sigma}$, there exists $a'' > 0$ such that

$$\psi_{\mu,\sigma}(z) - \psi_{\mu,\sigma}(\hat{z}_{\mu,\sigma}) \le g^\top (z - \hat{z}_{\mu,\sigma}) + a'' \|s - \hat{s}_{\mu,\sigma}\|_P.$$

With the Cauchy-Schwarz inequality and the inequality on the left of (4.8):

$$\psi_{\mu,\sigma}(z) - \psi_{\mu,\sigma}(\hat{z}_{\mu,\sigma}) \le \|g\| \left(\frac{\psi_{\mu,\sigma}(z) - \psi_{\mu,\sigma}(\hat{z}_{\mu,\sigma})}{a'} \right)^{1/2} + a'' \|s - \hat{s}_{\mu,\sigma}\|_P.$$

Now Lemma 4.8 yields

$$\psi_{\mu,\sigma}(z) - \psi_{\mu,\sigma}(\hat{z}_{\mu,\sigma}) \quad \le \quad \frac{1}{a'}\|g\|^2 + 2a''\|s - \hat{s}_{\mu,\sigma}\|_P.$$

The inequality on the right in (4.8) now follows by using the bounds obtained in Lemma 4.3. □

Theorem 4.10 *Suppose that Assumptions* 3.1 *and* 4.1 *hold, that f and c are $C^{1,1}$ functions, and that Algorithm A_μ does not stop in Step* 1. *Then, Algorithm A_μ generates a sequence $\{z_k\}$ converging r-linearly to $\hat{z}_\mu$, 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 know from Theorem 4.6, that Algorithm A_μ generates a sequence $\{z_k\}$ converging to $\hat{z}_\mu$. In addition $\sigma \ge \|\hat{\lambda}_\mu\|_D$, so that $\hat{z}_\mu = \hat{z}_{\mu,\sigma}$ (Proposition 3.6).

Now, let us fix $r \in]0,1[$ and denote by J the set of indices j for which (4.4) holds. Since d solves the linear system (3.6), one has for $j \in J$

$$\begin{aligned} &\|\nabla_x \ell(x_j,\lambda_j)\|^2 \\ &\quad = \quad \|M_j d_j^x - A_j^\top d_j^\lambda\|^2 \\ &\quad \le \quad 2\|M_j d_j^x\|^2 + 2\left\| A_j^\top (C_j+S_j)^{-1}\Big(\mu e - (C_j+S_j)\lambda_j - \Lambda_j(A_j d_j^x + d_j^s)\Big)\right\|^2, \end{aligned}$$

and, with the bounds from Lemma 4.3, we have for some positive constant K_1':

$$\begin{aligned} &\|\nabla_x \ell(x_j,\lambda_j)\|^2 + \|(C_j+S_j)\lambda_j - \mu e\|^2 + \|s_j\|_P \\ &\quad \le \quad K_1'\left(\|M_j d_j^x\|^2 + \|(C_j+S_j)\lambda_j - \mu e\|^2 + \|A_j d_j^x + d_j^s\|^2 + \|s_j\|_P\right). \end{aligned}$$

We have shown during the proof of Theorem 4.6, see (4.7), that there exists a positive constant K_2' such that for any $j \in J$:

$$\begin{aligned} &\psi_{\mu,\sigma}(z_{j+1}) - \psi_{\mu,\sigma}(z_j) \\ &\quad \le \quad -K_2'\left(\|M_j d_j^x\|^2 + \|(C_j+S_j)\lambda_j - \mu e\|^2 + \|A_j d_j^x + d_j^s\|^2 + \|s_j\|_P\right). \end{aligned}$$

Combining these inequalities gives for the constant $K_3' = K_2'/K_1'$ and for any $j \in J$:

$$\psi_{\mu,\sigma}(z_{j+1}) \;\leq\; \psi_{\mu,\sigma}(z_j) - K_3' \left(\|\nabla_x \ell(x_j,\lambda_j)\|^2 + \|(C_j{+}S_j)\lambda_j - \mu e\|^2 + \|s_j\|_P\right).$$

From the convergence of the sequence $\{z_j\}$ to $\hat{z}_\mu$ and Lemma 4.9 (note that here $\hat{s}_{\mu,\sigma} = 0$ since $\hat{z}_{\mu,\sigma} = \hat{z}_\mu$), there exists an index $j_0 \in J$, such that for $j \in J$ and $j \geq j_0$, z_j is in the neighborhood $\mathcal{N}$ given by Lemma 4.9 and

$$\begin{aligned}
&\psi_{\mu,\sigma}(z_{j+1}) - \psi_{\mu,\sigma}(\hat{z}_\mu)\\
&\quad\leq\; \psi_{\mu,\sigma}(z_j) - \psi_{\mu,\sigma}(\hat{z}_\mu) - K_3' \left(\|\nabla_x \ell(x_j,\lambda_j)\|^2 + \|(C_j{+}S_j)\lambda_j - \mu e\|^2 + \|s_j\|_P\right)\\
&\quad\leq\; \tau^{\frac{1}{r}}(\psi_{\mu,\sigma}(z_j) - \psi_{\mu,\sigma}(\hat{z}_\mu)),
\end{aligned}$$

where the constant $\tau := (1 - K_3' a)^r$ is necessarily in $[0,1[$. On the other hand, by the line-search, $\psi_{\mu,\sigma}(z_{k+1}) - \psi_{\mu,\sigma}(\hat{z}_\mu) \leq \psi_{\mu,\sigma}(z_k) - \psi_{\mu,\sigma}(\hat{z}_\mu)$, for all $k \geq 1$. According to Lemma 4.4, $|[1,k] \cap J| \geq \lceil rk \rceil$, so that for $k \geq j_0$: $|[j_0,k] \cap J| \geq \lceil rk \rceil - j_0 + 1 \geq rk - j_0 + 1$. Let $k_0 := \lceil j_0/r \rceil$, so that $|[j_0,k] \cap J| > 0$ for all $k \geq k_0$. By the last inequality, one has for $k \geq k_0$:

$$\psi_{\mu,\sigma}(z_{k+1}) - \psi_{\mu,\sigma}(\hat{z}_\mu) \leq K_4' \tau^k,$$

where K_4' is the positive constant $(\psi_{\mu,\sigma}(z_{j_0}) - \psi_{\mu,\sigma}(\hat{z}_\mu))/\tau^{(j_0-1)/r}$. Now, using the inequality on the left in (4.8), one has for all $k \geq k_0$:

$$\|z_{k+1} - \hat{z}_\mu\| \leq \frac{1}{\sqrt{a}} (\psi_{\mu,\sigma}(z_{k+1}) - \psi_{\mu,\sigma}(\hat{z}_\mu))^{\frac{1}{2}} \leq \left(\frac{K_4'}{a}\right)^{\frac{1}{2}} \tau^{\frac{k}{2}},$$

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

4.3 Q-superlinear convergence

Using shift variables s has a worth noting consequence. These ones are updated by the formula $s_{k+1} = s_k + \alpha_k d_k^s = (1 - \alpha_k) s_k$. Therefore, if the unit stepsize $\alpha_k = 1$ is ever accepted by the line-search, the shift variables are set to zero, and this value is maintained at all the subsequent iterations. If this event occurs, the algorithm becomes identical to the feasible algorithm in [2], which has been proven to be q-superlinear convergent (see Theorem 4.4 in [2]). As a result, to prove the q-superlinear convergence of algorithm A_μ, it is sufficient to show that $\alpha_k = 1$ for some index k.

In Proposition 4.11 below we show that the unit stepsize is indeed accepted by the Armijo condition (3.22), if the matrices M_k satisfy the estimate

$$(d_k^x)^\top \left(M_k - \hat{M}_\mu\right) d_k^x \geq o(\|d_k^x\|^2), \tag{4.9}$$

where we used the notation $\hat{M}_\mu := \nabla_{xx}^2 \ell(\hat{x}_\mu, \hat{\lambda}_\mu)$. This one is itself a consequence of the stronger estimate

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

Although Assumptions 3.1 are weaker than Assumptions 2.1 in [2], since $\gamma_k^\top \delta_k > 0$ still holds, we can show that (4.10) holds using the same arguments that those of the proof of Theorem 4.4 in [2].

Proposition 4.11 *Suppose that Assumptions* 3.1 *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* (4.9) *when $k \to \infty$. Then the sufficient decrease condition* (3.22) *is satisfied with $\alpha_k = 1$ for k sufficiently large.*

Proof. Observe first that the positive definiteness of $\hat{M}_\mu$ and (4.9) imply that

$$(d_k^x)^\top M_k d_k^x \geq K' \|d_k^x\|^2, \tag{4.11}$$

for some positive constant K' and sufficiently large k. Observe also that $d_k \to 0$ (for $d_k^x \to 0$, use (3.9), (4.11), $\nabla_x \ell(x_k, \lambda_k) \to 0$, and $C(x_k)\lambda_k \to \mu e$). Therefore, for k large enough, z_k and $z_k + d_k$ are near $\hat{z}_\mu$ and one can expand $\psi_{\mu,\sigma}(z_k + d_k)$ about z_k.

Decomposing $\psi_{\mu,\sigma}(z_k) = \widetilde{\psi}_\mu(z_k) + \sigma\|s_k\|_P$, we have for k large enough:

$$\begin{aligned}
&\psi_{\mu,\sigma}(z_k + d_k) - \psi_{\mu,\sigma}(z_k) - \omega\psi'_{\mu,\sigma}(z_k; d_k) \\
&\quad= \widetilde{\psi}_\mu(z_k + d_k) - \widetilde{\psi}_\mu(z_k) - \omega\nabla\widetilde{\psi}_\mu(z_k)^\top d_k - (1-\omega)\sigma\|s_k\|_P \\
&\quad= (1-\omega)\nabla\widetilde{\psi}_\mu(z_k)^\top d_k + \frac{1}{2} d_k^\top \nabla^2\widetilde{\psi}_\mu(z_k) d_k - (1-\omega)\sigma\|s_k\|_P + o(\|d_k\|^2) \\
&\quad= \left(\frac{1}{2} - \omega\right)\nabla\widetilde{\psi}_\mu(z_k)^\top d_k + \frac{1}{2}\left(\nabla\widetilde{\psi}_\mu(z_k)^\top d_k + d_k^\top\nabla^2\widetilde{\psi}_\mu(z_k) d_k\right) \\
&\qquad - (1-\omega)\sigma\|s_k\|_P + o(\|d_k\|^2).
\end{aligned} \tag{4.12}$$

We want to show that the latter expression is nonpositive when k is large.

For this, we start by evaluating the terms $\nabla\widetilde{\psi}_\mu(z_k)^\top d_k$ and $d_k^\top\nabla^2\widetilde{\psi}_\mu(z_k)d_k$. From (3.13), (3.21), and (3.6):

$$\begin{aligned}
&\nabla\widetilde{\psi}_\mu(z_k)^\top d_k \\
&\quad= -(d_k^x)^\top M_k d_k^x - \left\|\Lambda_k^{1/2}(C_k+S_k)^{-1/2}(A_k d_k^x + d_k^s)\right\|^2 + (\lambda_k+d_k^\lambda)^\top s_k \\
&\qquad - \tau\left\|\Lambda_k^{1/2}(C_k+S_k)^{-1/2}\left(A_k d_k^x + d_k^s + \Lambda_k^{-1}(C_k+S_k)d_k^\lambda\right)\right\|^2 \\
&\quad= D_k + (\lambda_k+d_k^\lambda)^\top s_k,
\end{aligned} \tag{4.13}$$

where

$$\begin{aligned}
D_k \;:=\; & -(d_k^x)^\top M_k d_k^x - (1+\tau)\left\|\Lambda_k^{1/2}(C_k+S_k)^{-1/2}(A_k d_k^x + d_k^s)\right\|^2 \\
& - \tau\left\|\Lambda_k^{-1/2}(C_k+S_k)^{1/2} d_k^\lambda\right\|^2 - 2\tau(A_k d_k^x + d_k^s)^\top d_k^\lambda.
\end{aligned}$$

On the other hand, from (3.4) and (3.20) and a calculation very similar to the one done in Lemma 4.7, one has

$$\begin{aligned}
&d_k^\top\nabla^2\widetilde{\psi}_\mu(z_k)d_k \\
&\quad= (d_k^x)^\top\nabla^2_{xx}\ell(x_k,\tilde{\lambda}_k)d_k^x + (1+\tau)\mu\|(C_k+S_k)^{-1}(A_k d_k^x + d_k^s)\|^2 \\
&\qquad + 2\tau(A_k d_k^x + d_k^s)^\top d_k^\lambda + \tau\mu\|\Lambda_k^{-1}d_k^\lambda\|^2,
\end{aligned} \tag{4.14}$$

where $\tilde{\lambda}_k = (1{+}\tau)\mu(C_k{+}S_k)^{-1}e - \tau\lambda_k$. Injecting (4.13) and (4.14) in (4.12), and using (4.9), $z_k \to \hat{z}_\mu$, and $\sigma \geq \|\lambda_k + d_k^\lambda\|_D$, give

$$\begin{aligned}
&\psi_{\mu,\sigma}(z_k + d_k) - \psi_{\mu,\sigma}(z_k) - \omega\psi'_{\mu,\sigma}(z_k; d_k)\\
&= \left(\frac{1}{2} - \omega\right) D_k - \frac{1}{2}(d_k^x)^\top \left(M_k - \nabla^2_{xx}\ell(x_k, \tilde{\lambda}_k)\right) d_k^x\\
&\quad + \frac{1}{2}(1{+}\tau)(A_k d_k^x + d_k^s)^\top \left(\mu(C_k{+}S_k)^{-2} - (C_k{+}S_k)^{-1}\Lambda_k\right)(A_k d_k^x + d_k^s)\\
&\quad + \frac{\tau}{2}(d_k^\lambda)^\top \left(\mu\Lambda_k^{-2} - (C_k{+}S_k)\Lambda_k^{-1}\right) d_k^\lambda\\
&\quad + (1-\omega)(\lambda_k + d_k^\lambda)^\top s_k - (1-\omega)\sigma\|s_k\|_P + o(\|d_k\|^2)\\
&\leq \left(\frac{1}{2} - \omega\right) D_k + (1-\omega)(\|\lambda_k + d_k^\lambda\|_D - \sigma)\|s_k\|_P + o(\|d_k\|^2)\\
&\leq \left(\frac{1}{2} - \omega\right) D_k + o(\|d_k\|^2).
\end{aligned}$$

To conclude, we still have to show that the negative terms in D_k can absorb the term in $o(\|d_k\|^2)$. For this, we use the Cauchy-Schwarz inequality on the last term in D_k to obtain

$$\begin{aligned}
&2\tau\left|(A_k d_k^x + d_k^s)^\top d_k^\lambda\right|\\
&\leq 2\tau\left\|(C_k{+}S_k)^{-1/2}\Lambda_k^{1/2}(A_k d_k^x + d_k^s)\right\|\,\left\|(C_k{+}S_k)^{1/2}\Lambda_k^{-1/2}d_k^\lambda\right\|\\
&\leq \frac{1{+}2\tau}{2}\left\|(C_k{+}S_k)^{-1/2}\Lambda_k^{1/2}(A_k d_k^x + d_k^s)\right\|^2 + \frac{2\tau^2}{1{+}2\tau}\left\|(C_k{+}S_k)^{1/2}\Lambda_k^{-1/2}d_k^\lambda\right\|^2.
\end{aligned}$$

Then, for some positive constants K'_1, K'_2, and K'_3, one has

$$\begin{aligned}
D_k &\leq -(d_k^x)^\top M_k d_k^x - \frac{1}{2}\left\|\Lambda_k^{1/2}(C_k{+}S_k)^{-1/2}(A_k d_k^x + d_k^s)\right\|^2\\
&\quad - \frac{\tau}{1{+}2\tau}\left\|\Lambda_k^{-1/2}(C_k{+}S_k)^{1/2}d_k^\lambda\right\|^2\\
&\leq -K'_1\|d_k^x\|^2 - K'_2\|A_k d_k^x + d_k^s\|^2 - K'_3\|d_k^\lambda\|^2.
\end{aligned}$$

For any $\epsilon > 0$:

$$\begin{aligned}
\|A_k d_k^x + d_k^s\|^2 &= \|A_k d_k^x\|^2 + 2(A_k d_k^x)^\top d_k^s + \|d_k^s\|^2\\
&\geq \|A_k d_k^x\|^2 - (1{+}\epsilon)\|A_k d_k^x\|^2 - \frac{1}{1{+}\epsilon}\|d_k^s\|^2 + \|d_k^s\|^2\\
&\geq -\epsilon\|A_k\|^2\|d_k^x\|^2 + \frac{\epsilon}{1{+}\epsilon}\|d_k^s\|^2.
\end{aligned}$$

Set now $\epsilon := K'_1/(2K'_2\|A_k\|^2)$ to conclude that

$$D_k \leq -\frac{K'_1}{2}\|d_k^x\|^2 - \frac{\epsilon K'_2}{1{+}\epsilon}\|d_k^s\|^2 - K'_3\|d_k^\lambda\|^2.$$

This negative upper bound of D_k can absorb the term $o(\|d_k\|^2)$. □

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 4.12 *Suppose that Assumptions* 3.1 *and* 4.1 *hold, that f and c are $C^{1,1}$ functions, twice continuously differentiable near $\hat{x}_\mu$ with locally radially Lipschitzian Hessians at $\hat{x}_\mu$, and that Algorithm A_μ does not stop in Step* 1*. Then the sequence $\{z_k\} = \{(x_k, s_k, \lambda_k)\}$ generated by this algorithm converges to $\hat{z}_\mu = (\hat{x}_\mu, \hat{s}_\mu, \hat{\lambda}_\mu)$ with a q-superlinear speed of convergence and, for k sufficiently large, the unit stepsize $\alpha_k = 1$ is accepted by the line-search.*

Proof. According to the observation made at the beginning of this section and the previous proposition, we only have to show that the estimate (4.10) holds to guarantee the q-superlinear convergence of $z_k \to \hat{z}_\mu$. This can be done exactly as in the proof of Theorem 4.4 in [2], using a standard result from the BFGS theory (see [15, Theorem 3] and [6]). □

A consequence of this result is that the q-superlinear convergence of $\{z_k\}$ does not depend on the value of the positive factor τ multiplying the centralization term $\mathcal{V}_\mu$ in the merit function.

5 The overall primal-dual algorithm

We have already mentioned at the beginning of Section 4.3 that, as soon as the unit stepsize $\alpha = 1$ is accepted by the line-search, all the iterates become strictly feasible. This is because the shift variables are updated by the rule: $s_+ = s + \alpha d^s = (1 - \alpha)s$. This is not necessarily an advantage. For some problems, it is difficult to find a point satisfying the constraints, so that the property above becomes a drawback: for many iterations the unit stepsize is not accepted. This may well slow down the speed of convergence of Algorithm A_μ.

To prevent this effect from occurring, one can relax the constraint $s = 0$ of Problem (1.2), substituting it into $s = r_\mu$, as in the barrier problem (1.5). The function $r : \mu \in [0, +\infty[\mapsto r_\mu \in \mathbb{R}^m$ is supposed to be continuous at $\mu = 0$ and to satisfy $r_0 = 0$. To overcome the difficulty mentioned in the previous paragraph, it is natural to take $r_\mu \geq 0$, although this is not required by the analysis below. An example of relaxation vector is $r_\mu = (\mu/\mu^1)s^1$. In this approach, feasibility in Problem 1.1 is only obtained asymptotically.

Let us mention that the results of Sections 3 and 4 are still valid when $r_\mu \neq 0$, since μ is fixed in these sections and the constraint $s = 0$ can be recovered in Problem (1.5) thanks to the substitutions

$$c(x) \to c(x) - r_\mu \qquad \text{and} \qquad s \to s + r_\mu.$$

We can now describe the overall algorithm. We index the outer iterations with superscripts $j \in \mathbb{N}\backslash\{0\}$ and note $r^j = r_{\mu^j}$. At the beginning of the jth outer iteration an approximation $z_1^j := (x_1^j, s_1^j, \lambda_1^j) \in \mathcal{Z}$ of the solution $\hat{z}$ of (1.4) is supposed available, as well as a positive definite matrix M_1^j approximating the Hessian of the Lagrangian. Values for the penalty parameters $\mu^j > 0$, for the relaxation vector $r^j \in \mathbb{R}^m$, and for a precision threshold $\epsilon^j := (\epsilon_l^j, \epsilon_c^j, \epsilon_s^j) > 0$ are also known.

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

1. Starting from z_1^j, use Algorithm A$_\mu$ until $z^j := (x^j, s^j, \lambda^j)$ satisfies

$$\begin{cases} \|\nabla f(x^j) - A(x^j)^\top \lambda^j\| \leq \epsilon_l^j \\ \|(C(x^j)+S^j)\lambda^j - \mu^j e\| \leq \epsilon_c^j \\ \|s^j - r^j\|_P \leq \epsilon_s^j. \end{cases} \tag{5.1}$$

2. Set the new penalty parameters $\mu^{j+1} > 0$ and $\sigma^{j+1} > 0$, the precision thresholds $\epsilon^{j+1} := (\epsilon_l^{j+1}, \epsilon_c^{j+1}, \epsilon_s^{j+1}) > 0$, and the new relaxation vector r^{j+1}, such that $\{\mu^j\}$, $\{\epsilon^j\}$, and $\{r^j\}$ converge to zero when $j \to \infty$. 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}.

The following lemma gives an over-estimate of the function value at an outer iteration.

Lemma 5.1 *Suppose that Assumptions* 3.1 *hold. If* (x^j, s^j, λ^j) *satisfies* (5.1)*, then for any* $x \in \mathbb{R}^n$*, one has*

$$f(x^j) \leq f(x) - (\lambda^j)^\top \left(c(x) + s^j\right) + m^{\frac{1}{2}} \epsilon_c^j + m\mu^j + \epsilon_l^j \|x^j - x\|. \tag{5.2}$$

Proof. Using the convexity of the Lagrangian: $\ell(x^j, \lambda^j) + \nabla_x \ell(x^j, \lambda^j)^\top (x - x^j) \leq \ell(x, \lambda^j)$, for any $x \in \mathbb{R}^n$. Therefore, the first criterion in (5.1) yields

$$\begin{aligned} f(x^j) &\leq f(x) - (\lambda^j)^\top c(x) + (\lambda^j)^\top c(x^j) + \epsilon_l^j \|x^j - x\| \\ &\leq f(x) - (\lambda^j)^\top \left(c(x) + s^j\right) + (\lambda^j)^\top \left(c(x^j) + s^j\right) + \epsilon_l^j \|x^j - x\|. \end{aligned}$$

Now with the second criterion in (5.1):

$$(\lambda^j)^\top (c(x^j) + s^j) = e^\top \left(\left(C(x^j) + S^j\right) \lambda^j - \mu^j e\right) + m\mu^j \leq m^{\frac{1}{2}} \epsilon_c^j + m\mu^j.$$

The result follows from these last two inequalities. □

Theorem 5.2 *Suppose that Assumptions* 3.1 *hold and that f and c are $C^{1,1}$ functions. Suppose also that Algorithm A generates a sequence $\{z^j\}$, which means that the stopping criteria* (5.1) *can be satisfied at every outer iteration. Then, depending on the problem data, the following situations occur:*

(*i*) *The limit points $\bar{z} := (\bar{x}, \bar{s}, \bar{\lambda})$ of $\{z^j\}$, if any, are such that $\bar{s} = 0$ (in fact all the sequence $\{s^j\} \to 0$) and $(\bar{x}, \bar{\lambda})$ is a primal-dual solution to Problem* (1.1).

(*ii*) *If Problem* (1.1) *has no feasible point, $\|x^j\| \to \infty$ and $\|\lambda^j\| \to \infty$.*

(*iii*) *If Problem* (1.1) *has a feasible point and Algorithm A takes $r^j \geq 0$ and $s_1^j \geq 0$, the sequence $\{x^j\}$ is bounded and its limit points are feasible.*

(*iv*) *If the Slater condition holds, the sequence $\{z^j\}$ is bounded, so that $\{z^j\}$ has indeed a limit point that satisfies the properties given in point* (*i*).

Proof. Let $\bar{z}$ be a limit point of $\{z^j\}$. Taking the limit in (5.1) shows that $\bar{s} = 0$ and that $(\bar{x}, \bar{\lambda})$ satisfies the KKT conditions (1.4). Therefore, $(\bar{x}, \bar{\lambda})$ is a primal-dual solution to the convex problem (1.1). Point (*i*) is proven.

Before proceeding with the other points, let us show the following intermediate result, which will be used three times. Suppose that, for some subsequence of indices J, $\|x^j\| \to \infty$ when $j \to \infty$ in J. Then $\|\lambda^j\| \to \infty$ when $j \to \infty$ in J and, for any point $\check{x} \in \mathbb{R}^n$:

$$(\exists\, i \in \{1, \ldots, m\})\, (\exists \text{ subsequence } J' \subset J)\, \big(\forall\, j \in J'\big) \quad c_{(i)}(\check{x}) + s^j_{(i)} < 0. \tag{5.3}$$

Indeed, define $t^j := \|x^j - \check{x}\|$, which is nonzero for j large in J. One has $t^j \to \infty$ when $j \to \infty$ in J, and $d^j := (x^j - \check{x})/t^j \to d \neq 0$ for some subsequence $J_0 \subset J$. Since $-c(x^j) < s^j$ and $\{s^j\}$ tends to zero, we have $(-c_{(i)})'_\infty(d) \leq 0$, for all $i = 1, \ldots, m$. It follows from Proposition 3.3 that $f'_\infty(d) = +\infty$. On the other hand, dividing both sides of inequality (5.2) by t^j provides

$$\frac{f(x^j) - f(\check{x})}{t^j} \leq -\frac{(\lambda^j)^\top (c(\check{x}) + s^j)}{t^j} + \frac{m^{\frac{1}{2}} \epsilon_c^j + m\mu^j}{t^j} + \epsilon_l^j \frac{\|x^j - \check{x}\|}{t^j}.$$

The left hand side tends to $f'_\infty(d) = +\infty$ when $j \to \infty$ in J_0, while the last two terms in the right hand side tends to zero. Therefore $\|\lambda^j\| \to \infty$ when $j \to \infty$ in J_0 (even more rapidly than t^j) and, since $\lambda^j > 0$, (5.3) must hold. By applying the result just proven to any subsequent of J, we see that $\{\lambda^j\}_{j \in J}$ cannot have a limit point, so that $\|\lambda^j\| \to \infty$ for $j \to \infty$ in all J.

Consider now point (*ii*), assuming that Problem (1.1) is not feasible. Then, the sequence $\{x^j\}$ cannot have a limit point $\bar{x}$, otherwise the limit in $c(x^j) + s^j > 0$ for an appropriate subsequence would imply that $\bar{x}$ is feasible for Problem (1.1). Then $\|x^j\| \to \infty$. Also $\|\lambda^j\| \to \infty$ by the intermediate result proven above.

Consider now point (*iii*) and let $\check{x}$ be feasible for Problem (1.1). The nonnegativity of r^j and s_1^j imply that $s^j \geq 0$. Indeed, during the jth outer iteration, the shift variables are updated by $s_+ := s - \alpha(s - r^j) = (1-\alpha)s + \alpha r^j$; hence s_+ is nonnegative by induction and so is s^j. Then (5.3) cannot hold (because $c_{(i)}(\check{x}) \geq 0$ and $s^j_{(i)} \geq 0$). Hence $\{x^j\}$ is bounded by the intermediate result. On the other hand $c(x^j) + s^j > 0$ and $s^j \to 0$, so that the limit point of $\{x^j\}$ are feasible.

Consider finally point (iv), assuming that Problem (1.1) has a strictly feasible point $\check{x}$. Since $c(\check{x})+s^j > 0$ for large j, (5.3) cannot hold, which implies the boundedness of $\{x^j\}$. We still have to show that $\{\lambda^j\}$ is bounded. We proceed by contradiction, assuming that $\|\lambda^j\| \to \infty$ for $j \to \infty$ in some subsequence J. Then, for some subsequence $J' \subset J$, the bounded sequence $\{(x^j, \lambda^j/\|\lambda^j\|)\}_{j\in J'}$ converges to $(\bar{x}, \bar{\lambda})$, say. Dividing the first two inequalities in (5.1) by $\|\lambda^j\|$ and taking limits when $j \to \infty$, $j \in J'$, we deduce that $\bar{\lambda} \geq 0$, $A(\bar{x})^\top \bar{\lambda} = 0$ and $(\bar{\lambda})^\top c(\bar{x}) = 0$. Using the concavity of the components of c and the strict feasibility of $\check{x}$, one has

$$c(\bar{x}) + A(\bar{x})^\top (\check{x} - \bar{x}) \geq c(\check{x}) > 0.$$

Multiplying by $\bar{\lambda}$, we deduce that $(\bar{\lambda})^\top c(\check{x}) = 0$, and thus $\bar{\lambda} = 0$, which is in contradiction with $\|\bar{\lambda}\| = 1$. □

We now exhibit conditions ensuring that the whole sequence of outer iterates $\{z^j\}$ converges to the analytic center of the primal-dual optimal set. To get that property, it is necessary to assume that the Slater condition is satisfied.

Assumption 5.3 *There exists $x \in \mathbb{R}^n$ such that $c(x) > 0$.*

Let us first recall the definition of analytic center of the optimal sets which, under Assumptions 3.1 and 5.3, is uniquely defined. We denote by $\mathrm{opt}(P)$ and $\mathrm{opt}(D)$ the sets of primal and dual solutions to 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 the latter case, the following index set

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

is nonempty (otherwise, for any $\lambda > 0$, the Lagrangian $\ell(\cdot, \lambda)$ would be constant on a segment of optimal points, not reduced to a single point, which is in contradiction with Lemma 3.2). By concavity of c, $\{\hat{x} \in \mathrm{opt}(P) : c_B(\hat{x}) > 0\}$ is nonempty either. The analytic center of $\mathrm{opt}(P)$ is then defined as the unique solution to the following problem:

$$\max_{\substack{\hat{x}\in\mathrm{opt}(P)\\ c_B(\hat{x})>0}} \left(\sum_{i\in B} \log c_{(i)}(\hat{x}) \right). \tag{5.4}$$

The fact that this problem is well defined and has a unique solution is highlighted in Lemma 5.4 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

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

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 to the following problem:

$$\max_{\substack{\hat{\lambda}\in\mathrm{opt}(D)\\ \hat{\lambda}_N>0}} \left(\sum_{i\in N} \log \hat{\lambda}_{(i)} \right). \tag{5.5}$$

Lemma 5.4 *Suppose that Assumptions* 3.1 *and* 5.3 *hold. If* $\mathrm{opt}(P)$ (*resp.* $\mathrm{opt}(D)$) *is not reduced to a singleton, then Problem* (5.4) (*resp.* (5.5)) *has a unique solution.*

Proof. The proof of this lemma is very similar to the one of Lemma 5.2 in [2], although Assumptions 3.1 and 5.3 are weaker than Assumptions 2.1 in [2].

Consider first Problem (5.4) and suppose that $\mathrm{opt}(P)$ is not a singleton. We have seen that the feasible set in (5.4) is nonempty. Let $\hat{x}_0$ be a point satisfying the constraints in (5.4). 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 (Proposition 3.3) and closed. Therefore, Problem (5.4) has a solution. To prove its uniqueness, suppose that $\hat{x}_1$ and $\hat{x}_2$ are two distinct solutions to Problem (5.4). Then, any point in the nontrivial segment $[\hat{x}_1, \hat{x}_2]$ is also optimal for this problem, so that, by the strict concavity of the log, c_B has a constant value over the segment. On the other hand, f is also constant on the segment (which is contained in $\mathrm{opt}(P)$), as well as $c_{(i)}$ for $i \notin B$ (which vanishes on the segment). It follows that the Lagrangian has a constant value over a nontrivial segment, a contradiction with its assumed strong convexity.

Using similar arguments (including the fact that Assumption 5.3 implies the boundedness of $\mathrm{opt}(D)$ and the fact that the objective function in (5.5) is strictly concave), one can show that Problem (5.5) 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 $B \cup N = \{1, \ldots, m\}$.

Theorem 5.5 *Suppose that Assumptions* 3.1 *and* 5.3 *hold and that f and c are $C^{1,1}$ functions. Suppose also that Problem* (1.1) *has the strict complementarity property and that the sequences $\{r^j\}$ and $\{\epsilon^j\}$ in Algorithm A satisfies the estimate $r^j = o(\mu^j)$ and $\epsilon^j = o(\mu^j)$. Then the sequence $\{z^j\}$ generated by Algorithm A converges to the point $\hat{z}_0 := (\hat{x}_0, 0, \hat{\lambda}_0)$, 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

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

and with the upper bound of $f(x^j)$ given by inequality (5.2), we obtain

$$\begin{aligned} 0 &\leq -\hat{\lambda}^\top c(x^j) - (\lambda^j)^\top (c(\hat{x}) + s^j) + m^{\frac{1}{2}} \epsilon_c^j + m\mu^j + \epsilon_l^j \|x^j - \hat{x}\| \\ &\leq -\hat{\lambda}^\top w^j - (\lambda^j)^\top c(\hat{x}) + (\hat{\lambda} - \lambda^j)^\top s^j + m^{\frac{1}{2}} \epsilon_c^j + m\mu^j + \epsilon_l^j \|x^j - \hat{x}\|, \end{aligned}$$

where $w^j := c(x^j) + s^j$. According to Theorem 5.2, $\{x^j\}$ and $\{\lambda^j\}$ are bounded, and by definition of B and N: $c_{(i)}(\hat{x}) = 0$ for $i \notin B$, and $\hat{\lambda}_{(i)} = 0$ for $i \notin N$. Hence

$$\hat{\lambda}_N^\top w_N^j + (\lambda_B^j)^\top c_B(\hat{x}) \leq m\mu^j + O(\|\epsilon^j\|) + O(\|s^j\|).$$

Now, using $s^j = O(\|\epsilon^j\|) + O(\|r^j\|)$ from the third criterion in (5.1), and the assumptions $r^j = o(\mu^j)$ and $\epsilon^j = o(\mu^j)$, we obtain finally

$$\hat{\lambda}_N^\top w_N^j + (\lambda_B^j)^\top c_B(\hat{x}) \leq m\mu^j + o(\mu^j). \tag{5.6}$$

We pursue by adapting an idea used by McLinden [13] to give properties of the limit points of the path $\mu \mapsto (\hat{x}_\mu, \hat{\lambda}_\mu)$. Let us define $\Gamma^j := \Lambda^j w^j - \mu^j e$. One has for all indices i:

$$w_{(i)}^j = \frac{\mu^j + \Gamma_{(i)}^j}{\lambda_{(i)}^j} \quad \text{and} \quad \lambda_{(i)}^j = \frac{\mu^j + \Gamma_{(i)}^j}{w_{(i)}^j}.$$

Substituting this in (5.6) and dividing by μ^j give

$$\sum_{i\in N} \frac{\hat{\lambda}_{(i)}}{\lambda_{(i)}^j} \frac{\mu^j + \Gamma_{(i)}^j}{\mu^j} + \sum_{i\in B} \frac{c_{(i)}(\hat{x})}{w_{(i)}^j} \frac{\mu^j + \Gamma_{(i)}^j}{\mu^j} \leq m + \frac{o(\mu^j)}{\mu^j}.$$

By assumptions, $\epsilon^j = o(\mu^j)$, so that the second inequality in (5.1) implies that $\Gamma_{(i)}^j = o(\mu^j)$. Let $(\hat{x}_0, \hat{\lambda}_0)$ be a limit point of $\{(x^j, \lambda^j)\}$. Taking the limit in the preceding estimate yields

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

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

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

or

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

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

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

This shows that $\hat{x}_0$ is a solution of (5.4) and that $\hat{\lambda}_0$ is a solution of (5.5). Since the problems in (5.4) and (5.5) have a unique solution, all the sequence $\{x^j\}$ converges to $\hat{x}_0$ and all the sequence $\{\lambda^j\}$ converges to $\hat{\lambda}_0$. □

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] P. Armand, J.Ch. Gilbert, S. Jan-Jégou (2000). A feasible BFGS interior point algorithm for solving strongly convex minimization problems. *SIAM Journal on Optimization*, 11, 199–222.

[3] 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.

[4] J.F. Bonnans, J.Ch. Gilbert, C. Lemaréchal, C. Sagastizábal (2001). *Numerical Optimization – Theoretical and Practical Aspects*. Springer Verlag, Berlin.

[5] R.H. Byrd, J.Ch. Gilbert, J. Nocedal (2000). A trust region method based on interior point techniques for nonlinear programming. *Mathematical Programming*, 89, 149–185.

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

[8] A. Forsgren, P.E. Gill (1998). Primal-dual interior methods for nonconvex nonlinear programming. *SIAM Journal on Optimization*, 8, 1132–1152.

[9] D.M. Gay, M.L. Overton, M.H. Wright (1998). A primal-dual interior method for nonconvex nonlinear programming. In Y.-X. Yuan (editor), *Advances in Nonlinear Programming*. Kluwer Academic Publishers.

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

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

[12] M. Kojima, S. Mizuno, T. Noma (1990). Limiting behavior of trajectories generated by a continuation method for monotone complementarity problems. *Mathematics of Operations Research*, 15, 662–675.

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

[14] J. Nocedal, S.J. Wright (1999). *Numerical Optimization*. Springer Series in Operations Research. Springer, New York.

[15] 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.

[16] B.N. Pshenichnyj (1994). *The Linearization Method for Constrained Optimization*. Computational Mathematics 22. Springer-Verlag.

[17] R.T. Rockafellar (1970). *Convex Analysis*. Princeton Mathematics Ser. 28. Princeton University Press, Princeton, New Jersey.

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

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rhône-Alpes : 655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

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