Laboratoire d'Arithmétique, de Calcul formel et d'Optimisation
ESA - CNRS 6090

A quasi-Newton penalty barrier method for convex minimization problems

Paul Armand

Rapport de recherche n° 2002-01

Université de Limoges, 123 avenue Albert Thomas, 87060 Limoges Cedex
Tél. 05 55 45 73 23 - Fax. 05 55 45 73 22 - laco@unilim.fr

http://www.unilim.fr/laco/

A quasi-Newton penalty barrier method for convex minimization problems

Paul ARMAND[†]

October 14, 2002

Abstract. We describe an infeasible interior point algorithm for convex minimization problems. The method uses quasi-Newton techniques for approximating the second derivatives and providing superlinear convergence. We propose a new feasibility control of the iterates by introducing shift variables and by penalizing them in the barrier problem. We prove global convergence under standard conditions on the problem data, without any assumption on the behavior of the algorithm.

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

AMS subject classification. 65K05, 90Cxx, 90C25, 90C30, 90C51, 90C53.

1 Introduction

We consider the smooth convex minimization problem

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

where $f : \mathbb{R}^n \to \mathbb{R}$, $c : \mathbb{R}^n \to \mathbb{R}^m$, and where f and the components c_i are convex and differentiable functions. We solve (1.1) by a quasi-Newton interior point method based on the algorithm introduced by Armand, Gilbert and Jan-Jégou [1] and further developed in [2]. Our new method can start from an infeasible initial point, but it differs from the algorithm proposed in [2] about the feasibility control of the iterates. As a result of our new approach, the algorithm is simpler to analyze and it has stronger convergence properties.

To handle infeasible iterates, we adopt a strategy similar to the one developed in [2]. The boundary of the feasible set is shifted, so that the iterates remain inside a feasible region and an exact penalization is used to shift back the boundary to its original position. But a difficult, and unsolved, issue in [2] was the control of the penalty parameter. In that paper, the authors have not succeeded in proving that the sequence of penalty parameters generated by their algorithm remains bounded, and therefore the whole convergence analysis has been done under a boundedness hypothesis of this sequence. In the present paper, shift and penalization are introduced in a different way and an explicit control of the penalty parameters is proposed. Our main result shows that the sequence of penalty parameters generated by the algorithm remains bounded.

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

The whole convergence analysis is done by using the same assumptions on the problem data as those in [2], but without any assumption on the behavior of the algorithm.

We associate to problem (1.1) the following penalized problem

$$\left\{\begin{array}{l} \min\ f(x) + \sigma^\top s, \\ c(x) \le s, \\ s \ge 0, \end{array}\right. \tag{1.2}$$

where the objective function is parametrized with a penalty vector $\sigma \in \mathbb{R}^m$. The motivation for introducing this problem is the following. The penalization is exact, in the sense that if (x, λ) is a primal-dual solution of (1.1), then (x, s) with $s = 0$, is solution of (1.2) whenever $\sigma \ge \lambda$ (the inequality is understood componentwise). On the other hand, given any initial point x_1, not necessarily feasible for (1.1), it is easy to find s_1 such that the pair (x_1, s_1) is strictly feasible for (1.2). Therefore, any interior point algorithm, that maintains strict feasibility of the iterates, can be applied to solve problem (1.2). During the iterations for solving (1.2), it then suffices to update conveniently the components of the penalty vector σ in order to drive s to zero, and in this way to obtain a convergent sequence to a solution of the original problem (1.1).

The primal-dual interior point algorithm to solve (1.2) is as follows. We associate to (1.2) the penalized barrier problem

$$\left\{\begin{array}{l} \min\ \varphi_{\sigma,\mu}(x,s) := f(x) + \sigma^\top s - \mu \sum_{i=1}^m \log\big((s_i - c_i(x))s_i\big), \\ (s - c(x), s) > 0, \end{array}\right. \tag{1.3}$$

where $\varphi_{\sigma,\mu}$ is a barrier function parametrized by $\mu > 0$ and $\sigma \in \mathbb{R}^m$. This problem is unconstrained, the positivity constraint being implicit, and, as is said above, a strictly feasible starting point can be easily computed. For fixed parameters μ and σ, problem (1.3) is approximately solved by applying a sequence of primal-dual quasi-Newton (BFGS) iterations to the perturbed optimality conditions of the penalized problem (1.2), see [1]. The convergence of the iterates is guaranteed by means of a line search procedure on some merit function. These iterations are called inner iterations. Once an approximate solution of (1.3) is found, the parameters μ and σ are updated. The barrier parameter μ is reduced to zero, to force convergence to a solution of (1.2). Using the estimate of the optimal multipliers available at the end of the inner iterations, the components of the penalty parameter σ are possibly increased, to force the convergence of s to zero. The stabilization of σ's values is obtained by standard update rules ensuring that the sequence of penalty parameters is either stationary or unbounded. Then, a new sequence of inner iterations is applied, until convergence to a solution of (1.1). The collection of inner iterations corresponding to the same value of μ and σ is called an outer iteration.

Let us mention the main differences with the strategy proposed in [2]. In that paper, problem (1.1) is transformed into the equivalent form:

$$\left\{\begin{array}{l} \min\ f(x), \\ c(x) \le s, \\ s = 0, \end{array}\right. \tag{1.4}$$

where $s \in \mathbb{R}^n$ is a vector of shift variables. The advantage of the transformation was to preserve the convexity of the original problem, that would not have been the case with the use of slack variables, that is with the transformation $c(x)+s=0$, $s \geq 0$ (see [3], for example). The barrier problem associated to (1.4) is the following equality constraint problem:

$$\begin{cases} \min \ \phi_\mu(x) := f(x) - \mu \sum_{i=1}^m \log\big(s_i - c_i(x)\big), \\ s = 0. \end{cases}$$

In this approach, the inequality $s - c(x) > 0$ is maintained thanks to the logarithmic barrier function, while the equality $s = 0$ is relaxed and asymptotically enforced by exact penalization. The merit function, used to force the global convergence of the inner iterates, contains the term $\phi_\mu(x) + \rho\|s\|$, where ρ is a penalty parameter. Since the algorithm is a line search method, the quasi-Newton direction computed at each inner iteration must be a descent direction of the merit function. This property is guaranteed by a possible increase of the penalty parameter value before performing the line search, hence the need of a stability property of the sequence of penalty parameters to prove the convergence of the inner iterates. In our new approach this problem is avoided, because the penalty parameter is fixed during an inner iteration, the update of its value is done at the outer iteration only, which simplifies both the algorithm and its convergence analysis. Let us mention also that in [2] the shift variables are updated according to $s_+ = (1-\alpha)s$, where α is the step length. This formula follows from the linearization of the equation $s = 0$ in (1.4). For the efficiency of the algorithm (to prevent a non-acceptance of the unit step length by the line search), the constraint $s = 0$ is relaxed into $s = r_\mu$, where r_μ is an additional parameter that converges to 0 with μ. Our new algorithm does not need such a relaxation, the shift variables are progressively driven to 0 thanks to the combination of the logarithmic barrier function and the penalization term.

We briefly mention a link between our approach and previous works in nonlinear programming. In [9], Mayne and Polak proposed a scheme to incorporate equality constraints in methods that solve inequality constraints problems by generating feasible iterates. This scheme has been used in some feasible directions algorithms, see [6, 8] and the recent work of Tits, Urban, Bakhtiari and Lawrence [15]. It consists of keeping the iterates on one side of each equality constraint and penalizing the iterates to force them to go more and more near the boundary of these fictive inequality constraints. This is exactly what is done when the equivalent formulation (1.4) of problem (1.1) is transformed into (1.2).

The paper is organized as follows. In the next section we introduce the notation and state the assumption used throughout the paper. The algorithm for solving the penalized barrier problem is presented in Section 3 and its convergence properties are stated. In Section 4 the overall algorithm is presented and the convergence of the outer iterates is analyzed. Part of the results of the paper, especially those concerning the inner iterations, are proved by using similar techniques to those used in [1]. To simplify the presentation and to highlight the main contribution of the paper, some proofs are relegated in appendix at the end of the paper.

2 Notation and assumption

Vector inequalities are understood componentwise: $u \geq 0$ means that each component satisfies $u_i \geq 0$ and $u \geq v$ means that $u - v \geq 0$, with similar meanings when $\geq$ is replaced by $>$. Given two vectors x, $y \in \mathbb{R}^n$, $x^\top y$ denotes the Euclidean scalar product and $\|x\|$ denotes the associated ℓ_2 norm. A function is of class $C^{1,1}$ if it is continuously differentiable and if its derivative is Lipschitz continuous.

Let us recall some definitions of convex analysis (see for example [7]). A function $\phi : \mathbb{R}^n \to \mathbb{R}$ is strongly convex with modulus $\kappa > 0$, if the function $\phi(\cdot) - \frac{\kappa}{2}\|\cdot\|^2$ is convex. For a differentiable function, the strong convexity is equivalent to the strong monotonicity of its gradient, that is: for all $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$, $(\nabla\phi(x) - \nabla\phi(y))^\top (x - y) \geq \kappa\|x - y\|^2$. Consider now a closed proper convex function $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$. Its asymptotic derivative is the closed proper convex function defined for $d \in \mathbb{R}^n$ by

$$f^\infty(d) := \lim_{t\to+\infty} \frac{f(x + t\,d) - f(x)}{t},$$

where x is an arbitrary point in the domain of f. The level sets of f are compact if and only if $f^\infty(d) > 0$ for all nonzero $d \in \mathbb{R}^n$ (see [7, Proposition IV.3.2.5]). As corollary (see [14, Corollary 27.3.3]), if problem (1.1) is feasible, then its solution set is nonempty and compact if and only if the following property holds:

$$(\forall\, d \neq 0) \quad c_i^\infty(d) \leq 0 \quad \forall\, i = 1, \ldots, m \quad \Longrightarrow \quad f^\infty(d) > 0. \tag{2.1}$$

The Lagrangian associated with problem (1.1) is the function $\ell : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ defined 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

$$\nabla_x \ell(x, \lambda) = \nabla f(x) + \nabla c(x)\lambda \quad \text{and} \quad \nabla^2_{xx}\ell(x, \lambda) = \nabla^2 f(x) + \sum_{i=1}^m \lambda_i \nabla^2 c_i(x),$$

where $\nabla f(x)$ is the gradient of f at x (for the Euclidean scalar product), $\nabla c(x)$ is the matrix whose columns are the gradients $\nabla c_i(x)$. Provided the constraints satisfy some qualification assumption, the Karush-Kuhn-Tucker (KKT) optimality conditions of problem (1.1) can be written as follows (see [12] for example): if x is a solution of (1.1) there exists a vector of multipliers $\lambda \in \mathbb{R}^m$ such that

$$\begin{cases} \nabla f(x) + \nabla c(x)\lambda = 0, \\ c(x) \leq 0, \\ C(x)\lambda = 0, \\ \lambda \geq 0, \end{cases} \tag{2.2}$$

where $C(x)$ is the diagonal matrix $\operatorname{diag}(c_1(x), \ldots, c_m(x))$.

Our minimal assumption refers to the convexity and smoothness of problem (1.1).

Assumption 2.1 The functions f and c_i $(1 \leq i \leq m)$ are convex and differentiable from $\mathbb{R}^n$ to $\mathbb{R}$. There exists $\lambda \in \mathbb{R}^m$, such that the Lagrangian $\ell(\cdot, \lambda)$ is strongly convex.

A first consequence of this assumption is the strong convexity of the Lagrangian $\ell(\cdot, \lambda)$ for *any* multiplier $\lambda > 0$, with a modulus depending continuously on λ, see [2, Lemma 3.2]. This property plays a key role in the convergence analysis of the minimization algorithm.

A second consequence of Assumption 2.1 is the compactness of the solution set of problem (1.1).

Proposition 2.2 [2] *Suppose that Assumption* 2.1 *holds. 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.*

3 Solving the penalized barrier problem

This section presents the algorithm for solving the penalized barrier problem (1.3) for fixed parameters μ and σ. We first prove some elementary properties of problems (1.2) and (1.3), that provide the basis for the minimization method. Next, the quasi-Newton step and the merit function are defined, then the line search minimization algorithm, is described. At last, the convergence results are stated.

The optimality conditions of the penalized problem (1.2) can be written (from now on, to simplify the notation we drop most of the dependencies in x)

$$\left\{\begin{array}{l} \nabla f + \nabla c\, \lambda = 0, \\ (S - C)\, \lambda = 0, \\ S(\sigma - \lambda) = 0, \\ (s - c, s, \lambda, \sigma - \lambda) \geq 0, \end{array}\right. \tag{3.1}$$

where $S = \mathrm{diag}(s_1, \ldots, s_m)$. Note that $\sigma - \lambda$ corresponds to the vector of multipliers associated with the nonnegativity constraint of (1.2). Assumption 2.1 implies that the solution set of the penalized problem (1.2) is nonempty and compact, even if problem (1.1) is infeasible.

Proposition 3.1 *Suppose that Assumption 2.1 holds. Then, the penalized problem* (1.2) *is strictly feasible and for any penalty vector* $\sigma > 0$ *its solution set is nonempty and compact. If* $(\hat{x}, \hat{\lambda})$ *is a primal-dual solution of* (1.1)*, then for any* $\sigma \geq \hat{\lambda}$*, the vector triple* $(\hat{x}, \hat{s}, \hat{\lambda})$*, with* $\hat{s} = 0$*, is a solution of* (3.1)*, which means that* $(\hat{x}, \hat{s}, \hat{\lambda}, \sigma - \hat{\lambda})$ *is a primal-dual solution of the penalized problem* (1.2)*. Conversely, let* $(\hat{x}_\sigma, \hat{s}_\sigma, \hat{\lambda}_\sigma)$ *be a solution of* (3.1)*, if* $\sigma > \hat{\lambda}_\sigma$ *then* $\hat{s}_\sigma = 0$ *and* $(\hat{x}_\sigma, \hat{\lambda}_\sigma)$ *is a primal-dual solution of* (1.1).

Proof. It is clear that problem (1.2) is strictly feasible. Let us prove that its solution set is nonempty and compact by using the characterization (2.1). Let $d := (d^x, d^s) \in \mathbb{R}^n \times \mathbb{R}^m$, $d \neq 0$, such that $c_i^\infty(d^x) \leq d_i^s$ for all $i = 1, \ldots, m$ and $d^s \geq 0$. By Proposition 2.2, it follows that either $f^\infty(d^x) = \infty$ or $d^x = 0$. In both cases, since $\sigma > 0$, one has $f^\infty(d^x) + \sigma^\top d^s > 0$.

Let $(\hat{x}, \hat{\lambda})$ be a solution of (2.2). If $\sigma - \hat{\lambda} \geq 0$, then by setting $\hat{s} = 0$, $(\hat{x}, \hat{s}, \hat{\lambda})$ solves (3.1). Conversely, suppose that $(\hat{x}_\sigma, \hat{s}_\sigma, \hat{\lambda}_\sigma)$ solves (3.1). Since $\sigma - \hat{\lambda}_\sigma > 0$, the second complementarity condition in (3.1) implies $\hat{s}_\sigma = 0$, and thus $(\hat{x}_\sigma, \hat{\lambda}_\sigma)$ solves (2.2). $\square$

A usual way to introduce interior point strategy for solving an inequality constraint problem, is to perturb the complementarity conditions by a parameter $\mu > 0$ (see for example [16]). Let us denote by $e = (1 \cdots 1)^\top$, the vector of all ones. Conditions (3.1) are transformed into

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

The substitution $\mu(S - C)^{-1}e$ for λ in the first and third equalities of (3.2) gives the optimality conditions of the penalized barrier problem (1.3).

Proposition 3.2 *Suppose that Assumption 2.1 holds. Then, for any penalty vector $\sigma > 0$ and any barrier parameter $\mu > 0$, the barrier function $\varphi_{\sigma,\mu}$ is strictly convex on its domain and its level sets are compact. In particular, the penalized barrier problem* (1.3) *has a unique solution, denoted by $(\hat{x}_{\sigma,\mu}, \hat{s}_{\sigma,\mu})$. This one is characterized by the existence of $\hat{\lambda}_{\sigma,\mu} \in \mathbb{R}^m$ such that $(\hat{x}_{\sigma,\mu}, \hat{s}_{\sigma,\mu}, \hat{\lambda}_{\sigma,\mu})$ is a solution of* (3.2).

Proof. Assumption 2.1 implies 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 $c_i(\alpha x + (1 - \alpha)x') \leq \alpha c_i(x) + (1 - \alpha)c_i(x')$, 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 $\ell(\cdot, e)$). Now consider two pairs $(x, s) \neq (x', s')$. If $s = s'$, then $x \neq x'$ and the strict convexity of $\varphi_{\sigma,\mu}$ follows from the previous remark and the properties of the log function (strict monotonicity and concavity). If $s \neq s'$, the result follows from the strict convexity of the function $s \to \sigma^\top s - \mu \sum_i \log s_i$, the monotonicity and the concavity of the log.

The compactness of the level sets of $\varphi_{\sigma,\mu}$ is a consequence of Proposition 3.1 and of the well known compactness property of the level sets of the log barrier function associated to an inequality convex problem, see [5, Lemma 12]. It is clear that the domain of $\varphi_{\sigma,\mu}$ is nonempty, therefore its minimum exits and is unique. It satisfies

$$\begin{cases} \nabla f + \mu \nabla c(S - C)^{-1}e &= 0, \\ \sigma - \mu(S - C)^{-1}e - \mu S^{-1}e &= 0. \end{cases}$$

Defining $\hat{\lambda}_{\sigma,\mu}$ by $(\hat{\lambda}_{\sigma,\mu})_i := \mu/(\hat{s}_{\sigma,\mu} - c(\hat{x}_{\sigma,\mu}))_i$, the result is proved. □

Equations (3.2) are approximately solved by applying a sequence of Newton iterations. The Newton step $(d^x, d^s, d^\lambda) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ at a given iterate (x, s, λ) is a solution of the following linear system

$$\begin{pmatrix} M & 0 & \nabla c \\ -\Lambda \nabla c^\top & \Lambda & S - C \\ 0 & \Sigma - \Lambda & -S \end{pmatrix} \begin{pmatrix} d^x \\ d^s \\ d^\lambda \end{pmatrix} = - \begin{pmatrix} \nabla_x \ell \\ (S - C)\lambda - \mu e \\ S(\sigma - \lambda) - \mu e \end{pmatrix}, \tag{3.3}$$

in which $M = \nabla^2_{xx}\ell(x, \lambda)$, Λ and Σ are the diagonal matrices $\mathrm{diag}(\lambda_1, \ldots, \lambda_m)$ and $\mathrm{diag}(\sigma_1, \ldots, \sigma_m)$. In the quasi-Newton algorithm that we consider, M is a positive definite approximation to the Hessian of the Lagrangian, updated by the BFGS formula.

Proposition 3.3 *Suppose that M is positive definite and that $(s-c, s, \lambda, \sigma-\lambda) > 0$, then the linear system* (3.3) *has a unique solution.*

Proof. By permuting the last two lines of the square matrix in (3.3), one has the following block LU decomposition:

$$\begin{pmatrix} M & 0 & \nabla c \\ 0 & \Sigma-\Lambda & -S \\ -\Lambda\nabla c^\top & \Lambda & S-C \end{pmatrix} = \begin{pmatrix} I & 0 & 0 \\ 0 & I & 0 \\ -\Lambda\nabla c^\top M^{-1} & \Lambda(\Sigma-\Lambda)^{-1} & I \end{pmatrix} \begin{pmatrix} M & 0 & \nabla c \\ 0 & \Sigma-\Lambda & -S \\ 0 & 0 & K \end{pmatrix},$$

where $K = S - C + \Lambda\left(\nabla c^\top M^{-1}\nabla c + (\Sigma-\Lambda)^{-1}S\right)$. The lower triangular factor is nonsingular because it has only ones on its diagonal. The assumptions made on the data imply that the three blocks M, $\Sigma-\Lambda$ and K are positive definite. Therefore, the block upper triangular factor is nonsingular. □

To simplify the notation, we denote by z the vector triple (x, s, λ) and define the following domain:

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

To force convergence of the quasi-Newton iterates, we use the following primal-dual merit function (see [1]):

$$\psi_{\sigma,\mu}(z) = \varphi_{\sigma,\mu}(x, s) + \tau\mathcal{V}_{\sigma,\mu}(z), \tag{3.4}$$

where $\tau > 0$ is a scaling parameter and

$$\begin{aligned} \mathcal{V}_{\sigma,\mu}(z) \quad &= \quad \lambda^\top(s - c(x)) - \mu\sum_{i=1}^m \log\Big(\lambda_i\,(s_i - c_i(x))\Big) \\ &\qquad + s^\top(\sigma-\lambda) - \mu\sum_{i=1}^m \log\Big(s_i\,(\sigma_i - \lambda_i)\Big), \end{aligned}$$

is a centralization term. Its purpose is to control the displacement in the dual space. Indeed, its minimum value (with respect to z) is achieved if and only if the perturbed complementarity conditions (second and third equations in (3.2)) are satisfied. The following proposition shows that $\psi_{\sigma,\mu}$ can act as merit function.

Proposition 3.4 *Suppose that Assumption* 2.1 *holds. Then, $\psi_{\sigma,\mu}$ has for unique minimizer the vector triple $\hat{z}_{\sigma,\mu} := (\hat{x}_{\sigma,\mu}, \hat{s}_{\sigma,\mu}, \hat{\lambda}_{\sigma,\mu})$ given by Proposition* 3.2 *and it has no other stationary point. Furthermore, suppose that M is symmetric positive definite. Let $z \in \mathcal{Z}_\sigma$ and let d be the unique solution of* (3.3)*. If $z \neq \hat{z}_{\sigma,\mu}$, then d is a descent direction of $\psi_{\sigma,\mu}$ at z, meaning that $\nabla\psi_{\sigma,\mu}(z)^\top d < 0$.*

The proof is given in Appendix A.1.

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

ALGORITHM $\mathrm{A}_{\sigma,\mu}$ for solving (3.2) (one inner iteration).

1. Compute $d := (d^x, d^s, d^\lambda)$, the unique solution to the linear system (3.3). If $d = 0$, stop (z solves the system (3.2)).
2. Compute a step length $\alpha > 0$ by backtracking:
   - 2.1. Set $\alpha = 1$.
   - 2.2. While $z + \alpha d \notin \mathcal{Z}_\sigma$, choose a new step length α in $[\xi\alpha, \xi'\alpha]$.
   - 2.3. While the sufficient decrease condition (or Armijo condition)
   $$\psi_{\sigma,\mu}(z + \alpha d) \le \psi_{\sigma,\mu}(z) + \omega\alpha\, \nabla\psi_{\sigma,\mu}(z)^\top d \tag{3.5}$$
   is not satisfied, choose a new step length α in $[\xi\alpha, \xi'\alpha]$.
   - 2.4. Set $z_+ := z + \alpha d$.
3. Update M by the BFGS formula
   $$M_+ := M - \frac{M\delta\delta^\top M}{\delta^\top M\delta} + \frac{\gamma\gamma^\top}{\gamma^\top\delta}, \tag{3.6}$$
   where γ and δ are given by
   $$\delta := x_+ - x \quad \text{and} \quad \gamma := \nabla_x\ell(x_+, \lambda_+) - \nabla_x\ell(x, \lambda_+). \tag{3.7}$$

Under Assumption 2.1, every step of the algorithm is well defined. In Step 1, the direction d exists because $z \in \mathcal{Z}_\sigma$ and M is positive definite (Proposition 3.3). The algorithm does not loop in Step 2 because $z \ne \hat{z}_{\sigma,\mu}$ (Step 1), so that d is a descent direction of $\psi_{\sigma,\mu}$ (Proposition 3.4) and the backtracking line search guarantees a sufficient step length reduction so that $z + \alpha d \in \mathcal{Z}_\sigma$ and (3.5) can be satisfied. In Step 3, formula (3.6) is well defined because $\gamma^\top\delta > 0$, due to the strong convexity of the Lagrangian (see the remark following Assumption 2.1).

The convergence analysis of this kind of algorithm has been fully detailed in [1, 2], but it is worth noting that Algorithm $\mathrm{A}_{\sigma,\mu}$ presented here is not Algorithm A_μ of [1] applied to the perturbed optimality conditions of the penalized problem (1.2). To be convinced, suppose that we apply directly the algorithm of [1]. The perturbed KKT conditions to solve would then be

$$\begin{cases} \nabla f + \nabla c\,\lambda = 0, \\ \sigma - \lambda - \xi = 0, \\ (S - C)\,\lambda = \mu e, \\ S\xi = \mu e, \\ (s - c, s, \lambda, \xi) > 0. \end{cases}$$

Though these conditions are equivalent to the formulation (3.2), applying the Newton method to this system does not produce the same iterates as Algorithm $A_{\sigma,\mu}$, unless $\xi = \sigma - \lambda$. Moreover, the required strong convexity assumption on the Lagrangian associated to problem (1.2), does not hold here because this Lagrangian function is only linear is s. Nevertheless, the r-linear and q-superlinear convergence properties also apply to Algorithm $A_{\sigma,\mu}$ and they are stated in the next two results.

Theorem 3.5 *Suppose that Assumption 2.1 holds and that f and c are of class $C^{1,1}$ in a neighborhood of the level set $\{z \in \mathcal{Z}_\sigma : \psi_{\sigma,\mu}(z) \le \psi_{\sigma,\mu}(z_1)\}$, where $z_1 \in \mathcal{Z}_\sigma$ is a starting point. Suppose that Algorithm $A_{\sigma,\mu}$ does not stop in Step 1. Then, the generated sequence $\{z_k\}$ converges to $\hat{z}_{\sigma,\mu}$, the unique solution of (3.2). The rate of convergence is r-linear, which means that $\limsup_{k\to\infty} \|z_k - \hat{z}_{\sigma,\mu}\|^{1/k} < 1$. In particular, the series $\sum_{k\ge1} \|z_k - \hat{z}_{\sigma,\mu}\|$ is convergent.*

The proof is given in Appendix B.1.

A stronger convergence result can be proved by using an additional assumption on the second derivatives. 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

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

Theorem 3.6 *Suppose that all assumptions of Theorem 3.5 hold and in addition that f and c are twice continuously differentiable near $\hat{x}_{\sigma,\mu}$ with locally radially Lipschitzian Hessians at $\hat{x}_{\sigma,\mu}$. Then, the sequence $\{z_k\}$ generated by this algorithm converges to $\hat{z}_{\sigma,\mu}$ q-superlinearly, which means that $\|z_{k+1} - \hat{z}_{\sigma,\mu}\| = o(\|z_k - \hat{z}_{\sigma,\mu}\|)$. Moreover, for k sufficiently large, the unit step length $\alpha_k = 1$ is accepted by the line search.*

The proof is given in Appendix B.2.

4 Overall algorithm

This section presents and analyses the overall algorithm for solving problem (1.1). Part of the analysis requires the existence of a strictly feasible point.

Assumption 4.1 *There exists $x \in \mathbb{R}^n$ such that $c(x) < 0$.*

This property is usually called the Slater condition. It is equivalent to the boundedness of the set of dual solutions of problem (1.1).

The perturbed KKT conditions (3.2) are approximately solved for a sequence of barrier parameters μ decreasing to zero. The control of the penalty parameter σ is based on Proposition 3.1. The latter suggests that σ should be increased whenever the difference $\sigma - \lambda$ is to close to zero. A stable control is obtained with standard update rules ensuring boundedness of the sequence of penalty parameters if and only if their values change finitely often.

We state now the overall algorithm for solving problem (1.1). A constant vector $\underline{\sigma} \in \mathbb{R}^m$, $\underline{\sigma} > 0$, is given independently of the iteration index. At the beginning of the jth outer iteration, the penalty parameter $\sigma^j \geq \underline{\sigma}$ and an approximation $z_1^j := (x_1^j, s_1^j, \lambda_1^j) \in \mathcal{Z}_{\sigma^j}$ of the solution of (3.2) are supposed available, as well as a positive definite matrix M_1^j approximating the Hessian of the Lagrangian. The barrier parameter $\mu^j > 0$ and the 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 $\mathrm{A}_{\sigma,\mu}$ until $z^j := (x^j, s^j, \lambda^j)$ satisfies $z^j \in \mathcal{Z}_{\sigma^j}$ and
$$\left\{ \begin{array}{l} \|\nabla f(x^j) + \nabla c(x^j)\lambda^j\| \leq \epsilon_l^j, \\ \|(S^j - C(x^j))\lambda^j - \mu^j e\| \leq \epsilon_c^j, \\ \|S^j(\sigma^j - \lambda^j) - \mu^j e\| \leq \epsilon_s^j. \end{array} \right. \tag{4.1}$$
2. Update the penalty vector with the following rule: For all $i \in \{1 \dots m\}$, if $\sigma_i^j \geq \lambda_i^j + \underline{\sigma}_i$, then $\sigma_i^{j+1} := \sigma_i^j$, else $\sigma_i^{j+1} := \max(1.1\sigma_i^j, \lambda_i^j + \underline{\sigma}_i)$.
3. Set the new barrier parameter $\mu^{j+1} > 0$, the precision thresholds $\epsilon^{j+1} := (\epsilon_l^{j+1}, \epsilon_c^{j+1}, \epsilon_s^{j+1}) > 0$, such that $\{\mu^j\}$ and $\{\epsilon^j\}$ converge to zero when $j \to \infty$. Choose a new starting iterate $z_1^{j+1} \in \mathcal{Z}_{\sigma^{j+1}}$ for the next outer iteration and a new positive definite matrix M_1^{j+1}.

If the functions f and c are $C^{1,1}$ and if Assumption 2.1 holds, then Algorithm A is well defined. Indeed, Theorem 3.5 implies that the stopping criterion (4.1) is satisfied after a finite number of iterations of Algorithm $\mathrm{A}_{\sigma,\mu}$. In Step 2, the update rule of the penalty parameters implies that for all index i, each sequence $\{\sigma_i^j\}$ is nondecreasing and is either unbounded or stationary. In Step 3, a possible choice is to set $z_1^{j+1} = z^j$ (note that $\mathcal{Z}_{\sigma^j}$ is included in $\mathcal{Z}_{\sigma^{j+1}}$) and $M_1^{j+1} = M^j$.

Lemma 4.2 *For all $i = 1, \dots, m$, the nondecreasing sequence of penalty parameters $\{\sigma_i^j\}$ is bounded if and only if the sequence of multipliers $\{\lambda_i^j\}$ is bounded. In that case, there exist an index j_i and $\sigma_i > 0$ such that for all $j \geq j_i$,*
$$\sigma_i^j = \sigma_i,$$
and in addition the sequence $\{s_i^j\}$ tends to 0 when $j \to \infty$.

Proof. The first part follows directly from the update rule in Step 2 of Algorithm A. If $\{\lambda_i^j\}$ is bounded, then $\sigma_i^j - \lambda_i^j = \sigma_i - \lambda_i^j \geq \underline{\sigma}_i > 0$ for all $j \geq j_i$. The third inequality of the stopping criterion (4.1) and the convergence of the sequences $\{\mu^j\}$ and $\{\epsilon_c^j\}$ to zero imply that the products $s_i^j(\sigma_i^j - \lambda_i^j)$ tend to zero, and thus $s_i^j \to 0$ when $j \to \infty$. □

The following lemma is a consequence of the convexity of the Lagrangian, it gives an estimate of the penalty function value at an outer iteration.

Lemma 4.3 *Suppose that Assumption 2.1 holds. If* (x^j, s^j, λ^j) *satisfies* (4.1)*, then for any* $x \in \mathbb{R}^n$ *one has*

$$f(x^j) + (\sigma^j)^\top s^j \le \ell(x, \lambda^j) + m^{\frac{1}{2}}(\epsilon_c^j + \epsilon_s^j) + 2m\mu^j + \|x^j - x\|\epsilon_l^j. \tag{4.2}$$

Proof. Let x be any vector of $\mathbb{R}^n$. The convexity of the Lagrangian implies

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

Using the Cauchy-Schwarz inequality and the first inequality of (4.1) we obtain

$$f(x^j) + (\sigma^j)^\top s^j \le \ell(x, \lambda^j) + (\sigma^j)^\top s^j - (\lambda^j)^\top c(x^j) + \|x^j - x\|\epsilon_l^j.$$

The last two inequalities of (4.1) imply

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

and

$$(\sigma^j - \lambda^j)^\top s^j = e^\top \big(S^j(\sigma^j - \lambda^j) - \mu^j e\big) + m\mu^j \le m^{\frac{1}{2}}\epsilon_s^j + m\mu^j.$$

Writing $(\sigma^j)^\top s^j - (\lambda^j)^\top c(x^j) = (\sigma^j - \lambda^j)^\top s^j + (\lambda^j)^\top (s^j - c(x^j))$ and combining the three preceding inequalities, we obtain (4.2). □

Proposition 4.4 *Suppose that Assumption 2.1 holds and that* f *and* c *are* $C^{1,1}$ *functions. Then, Algorithm A generates a sequence* $\{z^j\}$ *satisfying* (4.1) *and one of the following three situations occurs.*

- (i) *The sequence* $\{x^j\}$ *is unbounded. In this case, problem* (1.1) *is infeasible and* $\{\lambda^j\}$ *is unbounded.*
- (ii) *The sequence* $\{x^j\}$ *is bounded, but the sequence* $\{\lambda^j\}$ *is unbounded. In this case, the sequence* $\{s^j\}$ *is bounded and for any of its limit point* $\bar{s}$*, the set* $\{x : c(x) \le \bar{s}\}$ *is nonempty but does not satisfy the Slater condition.*
- (iii) *Both sequences* $\{x^j\}$ *and* $\{\lambda^j\}$ *are bounded. In this situation,* $\{s^j\}$ *tends to zero and any limit point of* $\{(x^j, \lambda^j)\}$ *is a primal-dual solution of problem* (1.1).

Proof. Assumption 2.1 and Theorem 3.5 imply that the stopping criterion (4.1) can be satisfied after a finite number of iterations of Algorithm $\mathrm{A}_{\sigma,\mu}$. The sequence $\{z^j\}$ is then well defined.

To prove (i), suppose that there exists a subset J of indices such that $\|x^j\| \to \infty$ when $j \to \infty$ in J. Let x be any point in $\mathbb{R}^n$. Let us show that x is not feasible for (1.1) and that $\{\lambda^j\}_{j\in J}$ is unbounded. Define $t^j := \|x^j - x\|$ and $d^j := (x^j - x)/t^j$. One has $t^j \to \infty$ when $j \to \infty$ in J and $d^j \to d \ne 0$ for some subsequence $J' \subset J$. Using (4.2) and next $\underline{\sigma} > 0$, $\sigma^j > \underline{\sigma}$ and $(s^j, s^j - c(x^j)) > 0$, for any $x \in \mathbb{R}^n$ one has

$$\begin{aligned} f(x^j) - f(x) + \underline{\sigma}^\top \big(c(x^j) - c(x)\big) &\le (\lambda^j - \underline{\sigma})^\top c(x) + \underline{\sigma}^\top \big(c(x^j) - s^j\big) + (\underline{\sigma} - \sigma^j)^\top s^j \\ &\quad + m^{\frac{1}{2}}(\epsilon_c^j + \epsilon_s^j) + 2m\mu^j + \|x^j - x\|\epsilon_l^j \\ &\le (\lambda^j - \underline{\sigma})^\top c(x) + (1 + t^j)\xi^j, \end{aligned} \tag{4.3}$$

where $\{\xi^j\}$ is a sequence that converges to zero. Dividing both sides of (4.3) by t^j we obtain

$$\frac{f(x^j)-f(x)}{t^j}+\sum_{i=1}^{m}\underline{\sigma}_i\frac{c_i(x^j)-c_i(x)}{t^j}\le\frac{(\lambda^j-\underline{\sigma})^\top c(x)}{t^j}+(\frac{1}{t^j}+1)\xi^j.$$

Since $\underline{\sigma}>0$, we deduce from Proposition 2.2 that the left hand side tends to

$$f^\infty(d)+\sum_{i=1}^{m}\underline{\sigma}_i\, c_i^\infty(d)=+\infty$$

when $j\to\infty$ in J'. It follows that $(\lambda^j)^\top c(x)\to\infty$, therefore x is infeasible and $\|\lambda^j\|\to\infty$ when $j\to\infty$ in J'.

To prove the first part of outcome (ii), we proceed by contradiction. Assuming $\{x^j\}$ bounded and $\{\lambda^j\}$ unbounded, suppose that there exists an index i such that $s_i^j\to\infty$ when $j\to\infty$ for a subsequence J. When $j\to\infty$ in J, one has $\sigma_i^j-\lambda_i^j\to 0$ by the third inequality in (4.1), $\lambda_i^j\to\infty$ by the update rule of σ_i^j in Step 2 of Algorithm A and next $s_i^j-c_i(x^j)\to 0$ by the second inequality in (4.1), a contradiction with the boundedness of $\{x^j\}$.

To prove the second part of (ii), suppose that $\{x^j\}$ is bounded and that $\|\lambda^j\|\to\infty$ when $j\to\infty$ for a subsequence J. There exists $J'\subset J$ such that the subsequence $\{(x^j,s^j,\lambda^j/\|\lambda^j\|)\}_{j\in J'}$ converges to $(\bar{x},\bar{s},\bar{\lambda})$. Dividing the first two inequalities in (4.1) by $\|\lambda^j\|$ and taking limits when $j\to\infty$ in J', we deduce that

$$\nabla c(\bar{x})\bar{\lambda}=0\quad\text{and}\quad\bar{\lambda}^\top(c(\bar{x})-\bar{s})=0.$$

Using the convexity of the components of c, for all $x\in\mathbb{R}^n$ one has

$$c(\bar{x})-\bar{s}+\nabla c(\bar{x})^\top(x-\bar{x})\le c(x)-\bar{s}.$$

Multiplying by $\bar{\lambda}$ we obtain

$$0\le\bar{\lambda}^\top(c(x)-\bar{s}).\tag{4.4}$$

Since $\bar{\lambda}\ge 0$ and $\|\bar{\lambda}\|=1$, we deduce that the set $\{x:c(x)\le\bar{s}\}$ has no strictly feasible point. Suppose now that $\tilde{s}$ is a limit point of $\{s^j\}$ and let us show that the Slater condition is not satisfied for the set $\{x:c(x)\le\tilde{s}\}$. Let J'' be a subsequence such that $(x^j,s^j)\to(\tilde{x},\tilde{s})$ when $j\to\infty$ in J''. Since $s^j-c(x^j)>0$, it is clear that $c(\tilde{x})\le\tilde{s}$. If $\{\lambda^j\}_{j\in J''}$ is unbounded, the previous reasoning with $J=J''$ applies, so that the set $\{x:c(x)\le\tilde{s}\}$ has no strictly feasible point. On the other hand, the sequence $\{\lambda^j\}_{j\in J''}$ is bounded. For all index i, each sequence $\{\sigma_i^j\}$ is either stationary or tends to infinity. In both cases, Lemma 4.2 or the third inequality in (4.1) implies that $\tilde{s}_i=0$ for all index i, and thus $\tilde{s}=0$. Since the set $\{x:c(x)\le 0\}$ is included in $\{x:c(x)\le\bar{s}\}$, the former does not satisfy the Slater condition either.

To prove the third outcome, suppose that $\{x^j\}$ and $\{\lambda^j\}$ are bounded. Lemma 4.2 implies that σ^j has constant value for j large enough, say σ, and that $s^j\to 0$. Let

$(\bar{x}, \bar{\lambda})$ be any limit point of the sequence $\{(x^j, \lambda^j)\}$. Since $z^j \in \mathcal{Z}_\sigma$ one has $c(\bar{x}) \le 0$, $\bar{\lambda} \ge 0$ and by taking limits in (4.1) $\nabla f(\bar{x}) + \nabla c(\bar{x})\bar{\lambda} = 0$ and $C(\bar{x})\bar{\lambda} = 0$. This shows that $(\bar{x}, \bar{\lambda})$ satisfies (2.2), therefore $(\bar{x}, \bar{\lambda})$ is a primal-dual solution of problem (1.1). □

The following result summarizes the behavior of the algorithm with respect to the feasibility of problem (1.1).

Theorem 4.5 *Suppose that Assumption 2.1 holds and that f and c are $C^{1,1}$ functions. Then, Algorithm A generates a sequence $\{z^j\}$ satisfying* (4.1) *and the following properties hold:*

- (*i*) *If problem* (1.1) *is infeasible, then the sequence of multipliers $\{\lambda^j\}$ is unbounded and for at least one index i the sequence $\{\sigma_i^j\}$ tends to infinity.*
- (*ii*) *If problem* (1.1) *is feasible, then the sequence $\{(x^j, s^j)\}$ is bounded.*
- (*iii*) *If problem* (1.1) *is strictly feasible, then the sequence of penalty parameters $\{\sigma^j\}$ is stationary, the sequence $\{s^j\}$ tends to zero, the sequence $\{(x^j, \lambda^j)\}$ is bounded and any of its limit point is a primal-dual solution of problem* (1.1).

Proof. If the sequence $\{\lambda^j\}$ is bounded, then assertion (*i*) of Proposition 4.4 implies that the sequence $\{x^j\}$ is bounded and assertion (*iii*) implies that problem (1.1) is feasible. This proves that whenever (1.1) is infeasible, the sequence of multipliers is unbounded and, by Lemma 4.2, there exists i such that the nondecreasing sequence $\{\sigma_i^j\}$ is unbounded.

To prove (*ii*), suppose that problem (1.1) is feasible. Proposition 4.4 implies that $\{x^j\}$ is bounded and whether the sequence of multipliers is bounded or not, the sequence $\{s^j\}$ is bounded.

To prove the third outcome, suppose that problem (1.1) is strictly feasible. For any limit point $(\bar{x}, \bar{s})$ of $\{(x^j, s^j)\}$, since $\bar{s} \ge 0$, the set $\{x : c(x) \le \bar{s}\}$ satisfies the Slater condition. It follows from (*ii*) of Proposition 4.4 that $\{\lambda^j\}$ is bounded, so that conclusion (*iii*) of Proposition 4.4 applies. □

When problem (1.1) is feasible but not strictly feasible, it is not guaranteed that the limit points of $\{x^j\}$ are feasible. To obtain this property, it suffices to modify the update rule of the penalty parameters, in order to force the convergence of the whole sequence $\{s^j\}$ to zero. Let $\underline{\rho} > 0$ be a constant, set the value of the penalty parameter with

$$\sigma^j := \rho^j e, \tag{4.5}$$

where $\{\rho^j\}$ is a nondecreasing sequence of positive numbers updated according to the following rule:

$$\text{If } \rho^j \ge \|\lambda^j\|_\infty + \underline{\rho}, \text{ then } \rho^{j+1} := \rho^j, \text{ else } \rho^{j+1} := \max(1.1\rho^j, \|\lambda^j\|_\infty + \underline{\rho}). \tag{4.6}$$

It follows that the sequence $\{\rho^j\}$ is nondecreasing and either tends to infinity, or is bounded and so is stationary.

Lemma 4.6 *The nondecreasing sequence* $\{\rho^j\}$ *is bounded if and only if the sequence of multipliers* $\{\lambda^j\}$ *is bounded. Moreover, for any index* i*, if the sequence* $\{\lambda_i^j\}$ *is bounded then the sequence* $\{s_i^j\}$ *converges to zero.*

Proof. The first part is a straightforward consequence of (4.6). Suppose that $\{\lambda_i^j\}$ is bounded for some index i. The third inequality in (4.1) implies that the product $s_i^j(\rho^j - \lambda_i^j)$ tends to zero when $j \to \infty$. If $\{\rho^j\}$ is bounded, then by (4.6) $\rho^j - \lambda_i^j \geq \underline{\rho} > 0$ for sufficiently large j. On the other hand, if $\{\rho^j\}$ is unbounded, then $(\rho^j - \lambda_i^j) \to \infty$ when $j \to \infty$. In both cases, we deduce that the variables s_i^j tend to zero when $j \to \infty$. □

It is easy to see that the modifications of the computation of σ^j in Algorithm A do not alter the conclusions of Proposition 4.4. Taking into account the new setting of σ^j, we can restate Theorem 4.5 in a stronger form.

Theorem 4.7 *Suppose that Assumption 2.1 holds and that* f *and* c *are* $C^{1,1}$ *functions. Suppose that Algorithm A is modified according to* (4.5) *for setting the penalty parameter and to* (4.6) *for the update rule in Step 2. Then the modified algorithm generates a sequence* $\{z^j\}$ *satisfying* (4.1) *and the following properties hold:*

- (*i*) *If problem* (1.1) *is infeasible, then the sequence of multipliers* $\{\lambda^j\}$ *is unbounded and the penalty parameters* ρ^j *tend to infinity.*
- (*ii*) *If problem* (1.1) *is feasible, then the sequence* $\{x^j\}$ *is bounded and* $\{s^j\}$ *tends to zero, in particular any limit point of* $\{x^j\}$ *is feasible.*
- (*iii*) *If problem* (1.1) *is strictly feasible, then the sequence of penalty parameters* $\{\rho^j\}$ *is stationary,* $\{s^j\}$ *tends to zero,* $\{(x^j, \lambda^j)\}$ *is bounded and any of its limit point is a primal-dual solution of problem* (1.1).

Proof. Points (*i*) and (*iii*) follow straightforwardly from Theorem 4.5 and Lemma 4.6.

To prove the second outcome, we proceed by contradiction. Suppose that (1.1) is feasible and that for some index i the variables s_i^j do not tend to zero. The feasibility assumption and Proposition (4.4) imply that the sequence $\{(x^j, s^j)\}$ is bounded. Since $\{s_i^j\}$ does not tend to zero, Lemma 4.6 implies that $\{\lambda_i^j\}$ is unbounded and so $\{\rho^j\}$ tends to infinity. There exist $\bar{s}$, with $\bar{s}_i > 0$ and a subsequence J such that $s^j \to \bar{s}$ when $j \to \infty$ in J. Using the third inequality in (4.1) one has $(\rho^j - \lambda_i^j) \to 0$ when $j \to \infty$ in J, and thus $\|\lambda^j\|_\infty \to \infty$ when $j \to \infty$ in J. Algorithm A guarantees that $z^j \in \mathcal{Z}_{\sigma^j}$ for all j, therefore $\|\lambda^j\|_\infty < \rho^j$. For all j one has $0 \leq \|\lambda^j\|_\infty - \lambda_i^j < \rho^j - \lambda_i^j$. Dividing both sides by $\|\lambda^j\|_\infty$ and taking the limit $j \to \infty$ in J, we obtain $\lambda_i^j / \|\lambda^j\|_\infty \to 1$. Following the proof of outcome (*ii*) of Proposition 4.4, there exists a subsequence $J' \subset J$, such that $\left\{(x^j, s^j, \lambda^j/\|\lambda^j\|_\infty)\right\}$ tends to $(\bar{x}, \bar{s}, \bar{\lambda})$ when $j \to \infty$ in J' and such that inequality (4.4) holds for some feasible point x. Since $(\bar{s}, \bar{\lambda}) \geq 0$ and $c(x) \leq 0$, we obtain $0 \leq \bar{\lambda}^\top \bar{s} \leq \bar{\lambda}^\top c(x) \leq 0$, and thus $\bar{\lambda}^\top \bar{s} = 0$, a contradiction with $\bar{\lambda}^\top \bar{s} \geq \bar{\lambda}_i \bar{s}_i = \bar{s}_i > 0$. □

In [1] it is proved that if, in addition to the Slater condition, problem (1.1) has the strict complementarity property, then the whole sequence of outer iterates converges to the analytic center of the primal-dual optimal set. A similar result is proved in [2], on condition that the sequence of penalty parameters remains bounded. In the present paper, we prove also convergence to a particular point of the primal-dual optimal set. Since the barrier problem is associated with a penalized problem, this limit point depends on the value reached by the penalty vector. This point is the analytic center of the primal-dual optimal set of the penalized problem (1.2). In the primal space, this center corresponds to the analytic center of the primal optimal set associated to (1.1), because the limit value of the variables s^j is zero. On the other hand, in the dual space the vector of multipliers associated to the constraint $s \geq 0$ is $\sigma - \lambda$ (see (3.1)) and therefore the limit point of the multipliers is not the analytic center of the dual optimal set. We call this point the σ-center of the dual optimal set and define it below.

Let us first recall the definition of the analytic center of an optimal set (see [11] for related results). Let us denote by $\mathrm{opt}(P)$ and by $\mathrm{opt}(D)$ the sets of primal and dual solutions of problem (1.1). If $\mathrm{opt}(P)$ is reduced to a single point, the analytic center is that point. Otherwise, $\mathrm{opt}(P)$ is a convex set with more than one point and the following index set

$$B := \{i : \exists\, x \in \mathrm{opt}(P) \text{ such that } c_i(x) < 0\}$$

is nonempty (otherwise, for any $\lambda > 0$, the Lagrangian $\ell(\cdot, \lambda)$ would be constant on a nontrivial segment of optimal points, a contradiction with the strong convexity assumption). By convexity of the component of c, $\{x \in \mathrm{opt}(P) : c_B(x) < 0\}$ is nonempty either. The analytic center of $\mathrm{opt}(P)$ is then defined as the unique solution to the following problem:

$$\begin{cases} \max \sum_{i \in B} \log\big(-c_i(x)\big), \\ x \in \mathrm{opt}(P), \\ c_B(x) < 0. \end{cases} \tag{4.7}$$

Lemma 4.8 [2] *Suppose that Assumption 2.1 and 4.1 hold. Then, problem* (4.7) *has a unique solution.*

Similarly, let us define the σ-center of the set of dual solutions of problem (1.1). Let $\sigma \in \mathbb{R}^m$ be such that $\sigma > \lambda$ for at least one $\lambda \in \mathrm{opt}(D)$. If $\mathrm{opt}(D)$ is reduced to a single point, the σ-center is that point. In case of multiple dual solutions, the index set

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

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

$$\begin{cases} \max \sum_{i \in N} \log \lambda_i + \sum_{i=1}^{m} \log(\sigma_i - \lambda_i), \\ \lambda \in \mathrm{opt}(D), \\ \lambda_N > 0, \\ \sigma - \lambda > 0. \end{cases} \tag{4.8}$$

Lemma 4.9 *Suppose that Assumption 2.1 and 4.1 hold. Then, for any $\sigma \in \mathbb{R}^m$ such that $\sigma > \lambda$ for at least one multiplier $\lambda \in \mathrm{opt}(D)$, problem* (4.8) *has a unique solution.*

Proof. The feasibility of problem (4.8) follows from the convexity of opt(D). Let β the objective value at some feasible solution. Assumption 4.1 is equivalent to the compactness of opt(D)(see for example [7]). Then, the set

$$\left\{\lambda : \lambda \in \mathrm{opt}(D), \lambda_N > 0, \sigma - \lambda > 0 \text{ and } \sum_{i\in N} \log \lambda_i + \sum_{i=1}^{m} \log(\sigma_i - \lambda_i) \geq \beta\right\}$$

is nonempty and compact. Therefore, problem (4.8) has a solution and since the objective function is strictly concave, the solution is unique. □

By complementarity (i.e., $C(x)\lambda = 0$) 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 $B \cup N = \{1, \dots, m\}$.

Theorem 4.10 *Suppose that Assumptions* 2.1 *and* 4.1 *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 stopping tolerance $\epsilon^j := (\epsilon_l^j, \epsilon_c^j, \epsilon_s^j)$ used in Algorithm A satisfies the estimate $\epsilon^j = o(\mu^j)$. Then, Algorithm A generates a sequence of penalty parameters $\{\sigma^j\}$ and a sequence of iterates $\{z^j\}$ such that $\sigma^j = \sigma$ for j large enough, and $\{z^j\}$ converges to the point $z^* := (x^*, s^*, \lambda^*)$, where x^* is the analytic center of the primal optimal set, $s^* = 0$ and λ^* is the σ-center of the dual optimal set.*

Proof. Since the Slater condition is satisfied, we know from Theorem 4.5 that the sequence $\{\sigma^j\}$ is stationary, $\{(x^j, \lambda^j)\}$ is bounded, $\{s^j\}$ tends to zero and any limit point of $\{(x^j, \lambda^j)\}$ is a primal-dual solution of (1.1). Let σ be the value reached by σ^j. Step 2 of Algorithm A implies that $\sigma \geq \lambda^j + \underline{\sigma}$ for j large enough. It follows that any limit point λ of the sequence $\{\lambda^j\}$ satisfies $\sigma - \lambda > 0$, therefore problem (4.8) is feasible and the σ-center of the dual optimal set is well defined.

Let $(\bar{x}, \bar{\lambda})$ be a primal-dual solution of problem (1.1) such that $\sigma - \bar{\lambda} > 0$. The Lagrangian $\ell(\cdot, \bar{\lambda})$ is minimized at $\bar{x}$ and $\bar{\lambda}^\top c(\bar{x}) = 0$, so that

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

Using the upper bound on $f(x^j) - f(\bar{x})$ given by (4.2) at $x = \bar{x}$ and using the fact that $\sigma^j = \sigma$ for large j, one has

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

for sufficiently large j. Since $\{x^j\}$ is bounded and $\epsilon^j = o(\mu^j)$, we deduce

$$\bar{\lambda}^\top w^j - (\lambda^j)^\top c(\bar{x}) + (\sigma - \bar{\lambda})^\top s^j \leq 2m\mu^j + o(\mu^j),$$

where $w^j := s^j - c(x^j)$. By definition of the sets B and N, we obtain finally

$$\bar{\lambda}_N^\top w_N^j - (\lambda_B^j)^\top c_B(\bar{x}) + (\sigma - \bar{\lambda})^\top s^j \leq 2m\mu^j + o(\mu^j). \tag{4.9}$$

The remainder of the proof uses an argument due to McLinden [10]. Let us define $\Gamma^j := \Lambda^j w^j - \mu^j e$ and $\Delta^j := S^j(\sigma - \lambda^j) - \mu^j e$. One has for all indices i:

$$w_i^j = \frac{\mu^j + \Gamma_i^j}{\lambda_i^j}, \quad \lambda_i^j = \frac{\mu^j + \Gamma_i^j}{w_i^j} \quad \text{and} \quad s_i^j = \frac{\mu^j + \Delta_i^j}{\sigma_i - \lambda_i^j}.$$

Substituting this in (4.9) gives

$$\sum_{i \in N} \frac{\bar{\lambda}_i}{\lambda_i^j} \frac{\mu^j + \Gamma_i^j}{\mu^j} + \sum_{i \in B} \frac{-c_i(\bar{x})}{w_i^j} \frac{\mu^j + \Gamma_i^j}{\mu^j} + \sum_{i=1}^{m} \frac{\sigma_i - \bar{\lambda}_i}{\sigma_i - \lambda_i^j} \frac{\mu^j + \Delta_i^j}{\mu^j} \leq 2m + o(1).$$

By assumption $\epsilon^j = o(\mu^j)$, so that the last two inequalities in (4.1) imply that $\Gamma_i^j = o(\mu^j)$ and $\Delta_i^j = o(\mu^j)$. Let (x^*, λ^*) be a limit point of $\{(x^j, \lambda^j)\}$. Taking the limits in the preceding estimate yields (recall that $s^j \to 0$, so that $w^j \to -c(x^*)$)

$$\frac{1}{2m} \left(\sum_{i \in N} \frac{\bar{\lambda}_i}{\lambda_i^*} + \sum_{i \in B} \frac{c_i(\bar{x})}{c_i(x^*)} + \sum_{i=1}^{m} \frac{\sigma_i - \bar{\lambda}_i}{\sigma_i - \lambda_i^*} \right) \leq 1.$$

Necessarily $\lambda_N^* > 0$, $c_B(x^*) < 0$ and $\sigma - \lambda^* > 0$. By strict complementarity, there are exactly $2m$ positive fractions in the left-hand side of the last inequality. Therefore, by monotonicity and concavity of the log function, one has

$$\sum_{i \in N} \log \frac{\bar{\lambda}_i}{\lambda_i^*} + \sum_{i \in B} \log \frac{c_i(\bar{x})}{c_i(x^*)} + \sum_{i=1}^{m} \log \frac{\sigma_i - \bar{\lambda}_i}{\sigma_i - \lambda_i^*} \leq 0.$$

Taking successively $\bar{\lambda} = \lambda^*$ and $\bar{x} = x^*$ in this inequality, we obtain

$$\sum_{i \in B} \log\big(-c_i(\bar{x})\big) \leq \sum_{i \in B} \log\big(-c_i(x^*)\big)$$

and

$$\sum_{i \in N} \log \bar{\lambda}_i + \sum_{i=1}^{m} \log(\sigma_i - \bar{\lambda}_i) \leq \sum_{i \in N} \log \lambda_i^* + \sum_{i=1}^{m} \log(\sigma_i - \lambda_i^*).$$

We deduce that x^* is a solution of (4.7) and λ^* is a solution of (4.8). Since these problems have a unique solution, the conclusion follows. □

5 Concluding remarks

In this paper we have extended the BFGS interior point method described in [1] to an infeasible algorithm. We have studied a different scheme than those proposed in [2] and have proved that it leads to a simpler algorithm with stronger convergence properties. It is worth noting that the strategy described in Section 4 can be used in other algorithmic frameworks than line search quasi-Newton methods.

Adding linear equality constraints which are not satisfied at each iteration, should not present any difficulty. Suppose that the problem to solve has the following form:

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

where $A = (a_i^\top)_{i=1}^p$ is a $p \times n$ matrix. By possibly changing the sign of some lines of (A, b) (depending on the sign of $a_i^\top x^1 - b_i$ at the starting point x^1), the penalized problem becomes

$$\begin{cases} \min f(x) + \rho^\top (Ax - b) + \sigma^\top s, \\ c(x) \leq s, \\ Ax \geq b, \\ s \geq 0, \end{cases}$$

where ρ is an additional vector of penalty parameters. The associated KKT conditions can be written

$$\begin{cases} \nabla f(x) + \nabla c(x)\lambda + A^\top \xi = 0, \\ (S - C(x))\,\lambda = 0, \\ L(x)(\rho - \xi) = 0, \\ S(\sigma - \lambda) = 0, \\ (s - c(x), Ax - b, s, \lambda, \rho - \xi, \sigma - \lambda) \geq 0, \end{cases}$$

where $L(x) = \operatorname{diag}(a_1^\top x - b_1, \ldots, a_p^\top x - b_p)$ and $\xi \in \mathbb{R}^p$. Note that $\rho - \xi$ is the vector of multipliers associated with the constraint $Ax \geq b$. The update rule of ρ is a straight extension of the one of σ and one can expect that the sequences of penalty parameters are bounded whenever the Slater condition holds. Note that, since the additional terms are only linear in x, the convergence properties of the inner iterates also apply.

A Merit function properties

In this appendix we prove Proposition 3.4, then we give additional properties of $\psi_{\sigma,\mu}$ that will be used by the convergence proofs of Appendix B.

A.1 Descent direction

Lemma A.1 *Let $z \in \mathcal{Z}_\sigma$ and let $d := (d^x, d^s, d^\lambda)$ be a solution of* (3.3)*, then*

$$\begin{aligned} \nabla \psi_{\sigma,\mu}(z)^\top d \;=\; & -(d^x)^\top M d^x \\ & - \|\Lambda^{1/2}(S - C)^{-1/2}(\nabla c^\top d^x - d^s)\|^2 - \|(\Sigma - \Lambda)^{1/2} S^{-1/2} d^s\|^2 \\ & - \tau \|(S - C)^{-1/2}\Lambda^{-1/2}(\mu e - (S - C)\lambda)\|^2 \\ & - \tau \|S^{-1/2}(\Sigma - \Lambda)^{-1/2}(\mu e - S(\sigma - \lambda))\|^2. \end{aligned} \tag{A.1}$$

Proof. From the definition (3.4) of $\psi_{\sigma,\mu}$, one has

$$\nabla \psi_{\sigma,\mu}(z)^\top d = \nabla \varphi_{\sigma,\mu}(x, s)^\top \begin{pmatrix} d^x \\ d^s \end{pmatrix} + \tau \nabla \mathcal{V}_{\sigma,\mu}(z)^\top d. \tag{A.2}$$

The gradient of $\varphi_{\sigma,\mu}$ is given by

$$\nabla\varphi_{\sigma,\mu}(x,s) = \begin{pmatrix} \nabla f + \mu\nabla c(S-C)^{-1}e \\ \sigma - \mu((S-C)^{-1} + S^{-1})e \end{pmatrix}. \tag{A.3}$$

From the first and third equations of (3.3), one has

$$\begin{pmatrix} M & 0 \\ 0 & \Sigma - \Lambda \end{pmatrix} \begin{pmatrix} d^x \\ d^s \end{pmatrix} + \begin{pmatrix} \nabla c \\ -S \end{pmatrix} (\lambda + d^\lambda) = -\begin{pmatrix} \nabla f \\ S\sigma - \mu e \end{pmatrix}, \tag{A.4}$$

while the second equation gives

$$\lambda + d^\lambda = (S-C)^{-1} \begin{pmatrix} \Lambda\nabla c^\top & -\Lambda \end{pmatrix} \begin{pmatrix} d^x \\ d^s \end{pmatrix} + \mu(S-C)^{-1}e.$$

Eliminating $\lambda + d^\lambda$ in $(A.4)$ and using $(A.3)$, we obtain

$$\begin{aligned} &\nabla\varphi_{\sigma,\mu}(x,s) \qquad\qquad (A.5) \\ &= -\left(\begin{pmatrix} M & 0 \\ 0 & S^{-1}(\Sigma-\Lambda) \end{pmatrix} + \begin{pmatrix} \nabla c(S-C)^{-1}\Lambda\nabla c^\top & -\nabla c(S-C)^{-1}\Lambda \\ -(S-C)^{-1}\Lambda\nabla c^\top & (S-C)^{-1}\Lambda \end{pmatrix} \right) \begin{pmatrix} d^x \\ d^s \end{pmatrix} \end{aligned}$$

From this last formula, we deduce

$$\begin{aligned} \nabla\varphi_{\sigma,\mu}(x,s)^\top \begin{pmatrix} d^x \\ d^s \end{pmatrix} &= -(d^x)^\top M d^x - \|\Lambda^{1/2}(S-C)^{-1/2}(\nabla c^\top d^x - d^s)\|^2 \\ &\quad -\|(\Sigma-\Lambda)^{1/2}S^{-1/2}d^s\|^2. \end{aligned} \tag{A.6}$$

It remains to compute the second directional derivative in the right hand side of $(A.2)$. The gradient of $\mathcal{V}_{\sigma,\mu}$ is given by

$$\nabla\mathcal{V}_{\sigma,\mu}(z) = \begin{pmatrix} -\nabla c\left(\lambda - \mu(S-C)^{-1}e\right) \\ \lambda - \mu(S-C)^{-1}e \\ s - c - \mu\Lambda^{-1}e \end{pmatrix} + \begin{pmatrix} 0 \\ \sigma - \lambda - \mu S^{-1}e \\ -s + \mu(\Sigma-\Lambda)^{-1}e \end{pmatrix}. \tag{A.7}$$

By using the last two equations in (3.3), one has

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

□

Proof of Proposition 3.4. Since $(\hat{x}_{\sigma,\mu}, \hat{s}_{\sigma,\mu})$ is the unique minimizer of $\varphi_{\sigma,\mu}$ (Proposition 3.2), one has

$$\varphi_{\sigma,\mu}(\hat{x}_{\sigma,\mu}, \hat{s}_{\sigma,\mu}) \leq \varphi_{\sigma,\mu}(x,s) \quad \text{for all } (x,s) \text{ such that } (s-c(x), s) > 0.$$

On the other hand, since $t \to t - \mu \log t$ is minimized at $t = \mu$ and since $\hat{z}_{\sigma,\mu}$ satisfies the last two equalities of (3.2), one has

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

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

It remains to show that $\psi_{\sigma,\mu}$ has a unique stationary point. Using the definition (3.4) of $\psi_{\sigma,\mu}$, formula $(A.3)$ and $(A.7)$, and the fact that $\tau > 0$, any stationary point must satisfy the following three equations:

$$\nabla f + \mu \nabla c (S-C)^{-1} e + \tau \nabla c (\mu (S-C)^{-1} e - \lambda) = 0, \tag{A.8}$$

$$\sigma - \mu ((S-C)^{-1} + S^{-1}) e = 0, \tag{A.9}$$

$$(s - c - \mu \Lambda^{-1} e) - (s - \mu (\Sigma - \Lambda)^{-1} e) = 0. \tag{A.10}$$

After some factorizations, equation $(A.9)$ can be written

$$\Lambda (S-C)^{-1} (s - c - \mu \Lambda^{-1} e) + S^{-1} (\Sigma - \Lambda)(s - \mu (\Sigma - \Lambda)^{-1} e) = 0.$$

It follows that $(A.9)$ and $(A.10)$ are of the form

$$\begin{cases} Au + Bv = 0, \\ u - v = 0, \end{cases}$$

with positive definite matrices A and B. We deduce that $S(\sigma - \lambda) = \mu e$ and $(S-C)\lambda = \mu e$. Substituting λ for $\mu (S-C)^{-1} e$ in $(A.8)$ gives $\nabla f + \nabla c \, \lambda = 0$. According to Proposition 3.2, one has $(x, s, \lambda) = (\hat{x}_{\sigma,\mu}, \hat{s}_{\sigma,\mu}, \hat{\lambda}_{\sigma,\mu})$. The first part of the proposition is proved.

Lemma A.1 and the positive definiteness of M imply that $\nabla \psi_{\sigma,\mu}(z)^\top d \leq 0$. If this directional derivative vanishes, then $d^x = 0$, $d^s = 0$, $(S-C)\lambda = \mu e$ and $S(\sigma - \lambda) = \mu e$. Since (d^x, d^s, d^λ) is the solution of (3.3), we deduce that $d^\lambda = 0$ and that (x, s, λ) is solution of (3.2). It follows from Proposition 3.2 that $(x, s, \lambda) = (\hat{x}_{\sigma,\mu}, \hat{s}_{\sigma,\mu}, \hat{\lambda}_{\sigma,\mu})$. □

A.2 Boundedness and local strong convexity

The following derivatives will be useful in the sequel:

$$\nabla^2 \varphi_{\sigma,\mu}(x,s) = \begin{pmatrix} \nabla^2_{xx} \varphi_{\sigma,\mu}(x,s) & -\mu \nabla c (S-C)^{-2} \\ -\mu (S-C)^{-2} \nabla c^\top & \mu ((S-C)^{-2} + S^{-2}) \end{pmatrix}, \tag{A.11}$$

where

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

and

$$\nabla^2\mathcal{V}_{\sigma,\mu}(z) = \begin{pmatrix} \nabla^2_{xx}\mathcal{V}_{\sigma,\mu}(z) & -\mu\nabla c(S-C)^{-2} & -\nabla c \\ -\mu(S-C)^{-2}\nabla c^\top & \mu(S-C)^{-2} & I \\ -\nabla c^\top & I & \mu\Lambda^{-2} \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & \mu S^{-2} & -I \\ 0 & -I & \mu(\Sigma-\Lambda)^{-2} \end{pmatrix}, \tag{A.12}$$

where $\nabla^2_{xx}\mathcal{V}_{\sigma,\mu}(z) = \sum_{i=1}^m (\frac{\mu}{s_i-c_i} - \lambda_i)\nabla^2_{xx}c_i + \mu\nabla c(S-C)^{-2}\nabla c^\top$. Note that

$$\begin{aligned} &\begin{pmatrix} u \\ v \end{pmatrix}^\top \nabla^2\varphi_{\sigma,\mu}(x,s)\begin{pmatrix} u \\ v \end{pmatrix} \qquad (A.13)\\ &= u^\top\nabla^2_{xx}\ell(x,\mu(S-C)^{-1}e)u + \mu\|(S-C)^{-1}(\nabla c^\top u - v)\|^2 + \mu\|S^{-1}v\|^2. \end{aligned}$$

The next result shows that the level sets of the merit function are compact and that the vector quadruple $(s-c(x), s, \lambda, \sigma-\lambda)$ is bounded and bounded away from zero over these sets.

Lemma A.2 *Suppose that Assumption 2.1 holds. Then, for any $r \in \mathbb{R}$, the level set $\mathcal{L}_r := \{z \in \mathcal{Z}_\sigma : \psi_{\sigma,\mu}(z) \le r\}$ is compact and there exist positive numbers K_1 and K_2 such that*

$$K_1 \le (s-c(x), s, \lambda, \sigma-\lambda) \le K_2 \quad \textit{for all } z \in \mathcal{L}_r.$$

Proof. The function $\mathcal{V}_{\sigma,\mu}$ is bounded below by $2m\mu(1-\log\mu)$ on its domain. Therefore, there exists K_1' such that $\varphi_{\sigma,\mu}(x,s) \le K_1'$ for all $z \in \mathcal{L}_r$. Assumption 2.1 and Proposition 3.2 imply that the level set $\mathcal{L}' := \{(x,s) : \varphi_{\sigma,\mu}(x,s) \le K_1'\}$ is compact. The set $\{(s-c(x), s) : (x,s) \in \mathcal{L}'\}$ is the image of a compact set by a continuous function, therefore it is compact. It follows that $(s-c(x), s)$ is bounded for all $(x,s) \in \mathcal{L}'$, and hence for $z \in \mathcal{L}_r$. It is also bounded away from zero, because $\varphi_{\sigma,\mu}(x,s) \le K_1'$ and $f(x) + \sigma^\top s$ is bounded below for all $z \in \mathcal{L}_r$.

The function $\varphi_{\sigma,\mu}$ is bounded below on the compact set $\mathcal{L}'$, then $\mathcal{V}_{\sigma,\mu}$ is bounded above on the level set $\mathcal{L}_r$. It follows that there exist positive numbers K_2' and K_3' such that $K_2' \le (\lambda_i(s_i - c_i(x)), s_i(\sigma_i - \lambda_i)) \le K_3'$ for all $z \in \mathcal{L}_r$ and all index i. We then deduce that $(\lambda, \sigma-\lambda)$ is bounded and bounded away from zero for all $z \in \mathcal{L}_r$. Since $\mathcal{L}_r$ is contained in a bounded set and $\psi_{\sigma,\mu}$ is continuous, then it is compact. □

Though the merit function $\psi_{\sigma,\mu}$ is not convex it is strongly convex in a neighborhood of its minimizer $\hat{z}_{\sigma,\mu}$ and more generally, in the neighborhood of any point satisfying the centrality conditions.

Lemma A.3 *Suppose that Assumption 2.1 holds. If $z \in \mathcal{Z}_\sigma$ satisfies the centrality conditions $(S-C)\lambda = \mu e$ and $S(\sigma-\lambda) = \mu e$, then the merit function $\psi_{\sigma,\mu}$ is locally strongly convex at z.*

Proof. Let $z \in \mathcal{Z}_\sigma$ satisfying the centrality conditions. Let us show that the Hessian matrix of $\psi_{\sigma,\mu}$ is positive definite at z. Using $(A.11)$, $(A.12)$ and the fact that z satisfies $(S-C)\lambda = \mu e$ and $S(\sigma - \lambda) = \mu e$, the Hessian of $\psi_{\sigma,\mu}$ can be written

$$
\begin{aligned}
&\nabla^2 \psi_{\sigma,\mu}(z) \\
&= \begin{pmatrix} \nabla^2 \varphi_{\sigma,\mu}(x,s) & 0 \\ & 0 \\ 0 \;\; 0 & 0 \end{pmatrix} + \tau \begin{pmatrix} \mu \nabla c(S-C)^{-2} \nabla c^\top & -\mu \nabla c (S-C)^{-2} & -\nabla c \\ -\mu (S-C)^{-2} \nabla c^\top & \mu (S-C)^{-2} & I \\ -\nabla c^\top & I & \mu^{-1}(S-C)^2 \end{pmatrix} \\
&\quad + \tau \begin{pmatrix} 0 & 0 & 0 \\ 0 & \mu S^{-2} & -I \\ 0 & -I & \mu^{-1} S^2 \end{pmatrix}.
\end{aligned}
$$

Multiplying on both sides by $(u, v, w) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ and using $(A.13)$ give

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

which is nonnegative. If it is equal to zero, the positive definiteness of the Hessian of the Lagrangian implies that $u = 0$, and next $v = w = 0$. □

The following result shows that, on the level set of $\psi_{\sigma,\mu}$, the distance of a point to the minimizer is bounded by the value of the merit function itself or by the norm of its gradient. This result is a consequence of the local strong convexity of $\psi_{\sigma,\mu}$ at $\hat{z}_{\sigma,\mu}$ (Lemma A.3) and of the compactness of the level set $\mathcal{L}_r$ (Lemma A.2).

Lemma A.4 *Suppose that Assumption 2.1 holds. Then, for any $r \in \mathbb{R}$ there exists a number $a > 0$ such that for any $z \in \mathcal{L}_r$*

$$
a \|z - \hat{z}_{\sigma,\mu}\|^2 \le \psi_{\sigma,\mu}(z) - \psi_{\sigma,\mu}(\hat{z}_{\sigma,\mu}) \le \frac{1}{a} \|\nabla \psi_{\sigma,\mu}(z)\|^2. \tag{A.14}
$$

The proof is omitted, see [1, Proposition 3.2].

B Convergence of Algorithm $\mathrm{A}_{\sigma,\mu}$

In this appendix, we give the convergence proofs of Algorithm $\mathrm{A}_{\sigma,\mu}$.

B.1 R-linear convergence

Let us denote by $z_1 := (x_1, s_1, \lambda_1)$ the starting iterate of Algorithm $\mathrm{A}_{\sigma,\mu}$. We define the level set

$$
\mathcal{L}_1 := \{ z \in \mathcal{Z}_\sigma : \psi_{\sigma,\mu}(z) \le \psi_{\sigma,\mu}(z_1) \}.
$$

Since the merit function decreases at each iteration, all the iterates stay in $\mathcal{L}_1$.

The convergence analysis rests on two lemmas. The first one is a consequence of the sufficient decrease condition (3.5). The second one is a property of the BFGS update formula (3.6).

Lemma B.1 [4] *Suppose that $\psi_{\sigma,\mu}$ is $C^{1,1}$ on an open convex neighborhood $\mathcal{N}$ of the level set $\mathcal{L}_1$. There exists a constant $K_3 > 0$ such that for any $z \in \mathcal{L}_1$ and for any descent direction d of $\psi_{\sigma,\mu}$, if the step length α is determined by the line search in Step 2 of Algorithm $A_{\sigma,\mu}$, then one of the following inequalities holds:*

$$\psi_{\sigma,\mu}(z + \alpha d) \leq \psi_{\sigma,\mu}(z) - K_3 |\nabla \psi_{\sigma,\mu}(z)^\top d|, \tag{B.1}$$

$$\psi_{\sigma,\mu}(z + \alpha d) \leq \psi_{\sigma,\mu}(z) - K_3 \frac{|\nabla \psi_{\sigma,\mu}(z)^\top d|^2}{\|d\|^2}. \tag{B.2}$$

Lemma B.2 [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 $k \geq 1$*

$$\gamma_k^\top \delta_k \geq a_1 \|\delta_k\|^2 \quad \text{and} \quad \gamma_k^\top \delta_k \geq a_2 \|\gamma_k\|^2, \tag{B.3}$$

where $a_1 > 0$ and $a_2 > 0$ are independent of k. Then, there exist positive constants b_1, b_2, and b_3, such that

$$\cos \theta_k := \frac{\delta_k^\top M_k \delta_k}{\|M_k \delta_k\| \|\delta_k\|} \geq b_1 \quad \text{and} \quad b_2 \leq \frac{\|M_k \delta_k\|}{\|\delta_k\|} \leq b_3, \tag{B.4}$$

for at least half of the iterations.

Lemma B.3 *Suppose that Assumption 2.1 holds and that f and c are of class $C^{1,1}$ in a neighborhood of the level set $\mathcal{L}_1$. Suppose that the step length α_k is determined according to the backtracking line search of Algorithm $A_{\sigma,\mu}$ (Step 2). Then, there exists a constant $K > 0$ such that for any iteration k satisfying (B.4),*

$$\psi_{\sigma,\mu}(z_k + \alpha_k d_k) \leq \psi_{\sigma,\mu}(z_k) - K \|\nabla \psi_{\sigma,\mu}(z_k)\|^2. \tag{B.5}$$

Proof. To simplify the notation we do not use the iteration index and denote by z the current iterate. We denote by K_1', K_2', ... some positive constants (independent of the iteration index).

The bounds on $(s - c, s, \lambda, \sigma - \lambda)$ given by Lemma A.2 and the fact that f and c are of class $C^{1,1}$ imply that $\psi_{\sigma,\mu}$ is of class $C^{1,1}$ on some open convex neighborhood of the level set $\mathcal{L}_1$. Therefore, by the line search and Lemma B.1, either (B.1) or (B.2) is satisfied.

Suppose now that the current iterate satisfies the bounds of Lemma B.2. Using (A.1), the bounds of Lemma A.2 and (B.4), one has

$$\begin{aligned}
&|\nabla \psi_{\sigma,\mu}(z)^\top d| \\
&\quad = (d^x)^\top M d^x + \|\Lambda^{1/2}(S - C)^{-1/2}(\nabla c^\top d^x - d^s)\|^2 + \|(\Sigma - \Lambda)^{1/2} S^{-1/2} d^s\|^2 \\
&\qquad + \tau \|(S - C)^{-1/2} \Lambda^{-1/2}(\mu e - (S - C)\lambda)\|^2 \\
&\qquad + \tau \|S^{-1/2}(\Sigma - \Lambda)^{-1/2}(\mu e - S(\sigma - \lambda))\|^2 \\
&\quad \geq \frac{b_1}{b_3} \|M d^x\|^2 + K_1 K_2^{-1}(\|\nabla c^\top d^x - d^s\|^2 + \|d^s\|^2) \\
&\qquad + \tau K_2^{-2}(\|\mu e - (S - C)\lambda\|^2 + \|\mu e - S(\sigma - \lambda)\|^2) \\
&\quad \geq K_1'(\|M d^x\|^2 + \|\nabla c^\top d^x - d^s\|^2 + \|d^s\|^2 + \|\mu e - (S - C)\lambda\|^2 + \|\mu e - S(\sigma - \lambda)\|^2).
\end{aligned}$$

Let us denote by K_2' an upper bound of $\|\nabla c\|$ on $\mathcal{L}_1$. Using $(A.5)$, $(A.7)$ and next the inequality $\|\sum_{i=1}^p \xi_i\|^2 \le p\sum_{i=1}^p \|\xi_i\|^2$, we obtain

$$
\begin{aligned}
&\|\nabla\psi_{\sigma,\mu}(z)\|^2 \\
&\quad= \|\nabla_x\psi_{\sigma,\mu}(z)\|^2 + \|\nabla_s\psi_{\sigma,\mu}(z)\|^2 + \|\nabla_\lambda\psi_{\sigma,\mu}(z)\|^2 \\
&\quad= \|Md^x + \nabla c(S-C)^{-1}\Lambda(\nabla c^\top d^x - d^s) + \tau\nabla c(S-C)^{-1}((S-C)\lambda - \mu e)\|^2 \\
&\qquad + \|S^{-1}(\Sigma-\Lambda)d^s - (S-C)^{-1}\Lambda(\nabla c^\top d^x - d^s) + \tau(\mu(S-C)^{-1}e - \sigma + \mu S^{-1}e)\|^2 \\
&\qquad + \tau^2\|c + \mu\Lambda^{-1}e - \mu(\Sigma-\Lambda)^{-1}e\|^2 \\
&\quad\le 3\|Md^x\|^2 + 3(K_2'K_1^{-1}K_2)^2\|\nabla c^\top d^x - d^s\|^2 + 3(\tau K_2'K_1^{-1})^2\|\mu e - (S-C)\lambda\|^2 \\
&\qquad + 4(K_1^{-1}K_2)^2(\|d^s\|^2 + \|\nabla c^\top d^x - d^s\|^2) \\
&\qquad + 6(\tau K_1^{-1})^2(\|\mu e - (S-C)\lambda\|^2 + \|\mu e - S(\sigma-\lambda)\|^2) \\
&\quad\le K_3'(\|Md^x\|^2 + \|\nabla c^\top d^x - d^s\|^2 + \|d^s\|^2 + \|\mu e - (S-C)\lambda\|^2 + \|\mu e - S(\sigma-\lambda)\|^2).
\end{aligned}
$$

Form the Newton system (3.3), one has

$$
\begin{aligned}
\|d\|^2 &= \|d^x\|^2 + \|d^s\|^2 + \|d^\lambda\|^2 \\
&= \|d^x\|^2 + \|d^s\|^2 + \|S^{-1}(\Sigma-\Lambda)d^s + \sigma - \lambda - \mu S^{-1}e\|^2 \\
&\le b_2^{-2}\|Md^x\|^2 + (1 + 2(K_1^{-1}K_2)^2)\|d^s\|^2 + 2K_1^{-2}\|\mu e - S(\sigma-\lambda)\|^2 \\
&\le K_4'(\|Md^x\|^2 + \|d^s\|^2 + \|\mu e - S(\sigma-\lambda)\|^2).
\end{aligned}
$$

Finally, combining these three inequalities with $(B.1)$ or $(B.2)$ and taking we obtain $(B.5)$ with the constant $K = K_3K_1'/K_3'\min(1, K_1'/K_4')$. □

Proof of Theorem 3.5. By virtue of the strong convexity assumption (Assumption 2.1), there exist positive constants a_1 and a_2 such that the inequalities $(B.3)$ are satisfied (see [2, Lemma 4.5]). From Lemma B.2, Lemma B.3 and Lemma A.4 there exist constants $K > 0$ and $a > 0$ such that

$$\psi_{\sigma,\mu}(z_{k+1}) - \psi_{\sigma,\mu}(\hat{z}_{\sigma,\mu}) \le (1 - aK)(\psi_{\sigma,\mu}(z_k) - \psi_{\sigma,\mu}(\hat{z}_{\sigma,\mu})),$$

is satisfied for at least half of the iterations. On the other hand, by (3.5) one has

$$\psi_{\sigma,\mu}(z_{k+1}) - \psi_{\sigma,\mu}(\hat{z}_{\sigma,\mu}) \le \psi_{\sigma,\mu}(z_k) - \psi_{\sigma,\mu}(\hat{z}_{\sigma,\mu}),$$

for any iteration index k. It follows that

$$\psi_{\sigma,\mu}(z_{k+1}) - \psi_{\sigma,\mu}(\hat{z}_{\sigma,\mu}) \le (1 - aK)^{k/2}(\psi_{\sigma,\mu}(z_1) - \psi_{\sigma,\mu}(\hat{z}_{\sigma,\mu})),$$

for all $k \ge 1$. Finally, using the left inequality in $(A.14)$, we conclude that for $k \ge 1$

$$\|z_{k+1} - \hat{z}_{\sigma,\mu}\| \le K'(1 - aK)^{k/4},$$

where $K' = a^{-1/2}(\psi_{\sigma,\mu}(z_1) - \psi_{\sigma,\mu}(\hat{z}_{\sigma,\mu}))^{1/2}$. □

B.2 Q-superlinear convergence

Recall that $\{z_k\}$ is q-superlinear convergent if $z_{k+1} - \hat{z}_{\sigma,\mu} = o(\|z_k - \hat{z}_{\sigma,\mu}\|)$, meaning that $\|z_{k+1} - \hat{z}_{\sigma,\mu}\|/\|z_k - \hat{z}_{\sigma,\mu}\| \to 0$ (assuming $z_k \neq \hat{z}_{\sigma,\mu}$).

Lemma B.4 *Suppose that Assumptions* 2.1 *holds and that f and c are twice continuously differentiable near $\hat{x}_{\sigma,\mu}$. Suppose also that the sequence $\{z_k\}$ generated by Algorithm $A_{\sigma,\mu}$ converges to $\hat{z}_{\sigma,\mu}$ and that the positive definite matrices M_k satisfy the estimate*

$$(d_k^x)^\top (M_k - \hat{M}_{\sigma,\mu}) d_k^x \geq o(\|d_k^x\|^2), \tag{B.6}$$

where

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

Then, for k sufficiently large the unit step length is accepted by the line search condition (3.5)*, that is*

$$\psi_{\sigma,\mu}(z_k + d_k) \leq \psi_{\sigma,\mu}(z_k) + \omega\, \nabla\psi_{\sigma,\mu}(z_k)^\top d_k.$$

Proof. To simplify the notation we do not use the iteration index k and denote by z the current iterate. Since we prove an asymptotic property, it is implicitly assumed that k is sufficiently large so that the various mentioned properties are satisfied and that all the limits are taken when $k \to \infty$. We denote by K_1', K_2', ... positive constants (independent of the iteration index).

Observe first that the positive definiteness of $\hat{M}_{\sigma,\mu}$ and $(B.6)$ imply that

$$(d^x)^\top M d^x \geq K_1' \|d^x\|^2. \tag{B.7}$$

Observe also that $d \to 0$ (for $(d^x, d^s) \to 0$, use $(A.6)$, the bounds of Lemma A.2, $(B.7)$ and $\nabla\varphi_{\sigma,\mu}(x,s) \to 0$). Therefore, z and $z+d$ are near $\hat{z}_{\sigma,\mu}$ and one can expand $\psi_{\sigma,\mu}(z+d)$ about z. Using a Taylor series expansion to the second order, one has

$$\begin{aligned} &\psi_{\sigma,\mu}(z+d) - \psi_{\sigma,\mu}(z) - \omega\nabla\psi_{\sigma,\mu}(z)^\top d \\ &= \big(\frac{1}{2} - \omega\big)\nabla\psi_{\sigma,\mu}(z)^\top d + \frac{1}{2}\big(\nabla\psi_{\sigma,\mu}(z)^\top d + d^\top \nabla^2\psi_{\sigma,\mu}(z) d\big) + o(\|d\|^2). \end{aligned} \tag{B.8}$$

We begin by showing that the second term in parenthesis in the right-hand side is smaller than a term of order $o(\|d\|^2)$. Using $(A.1)$ and (3.3) one has

$$\begin{aligned} \nabla\psi_{\sigma,\mu}(z)^\top d &= -(d^x)^\top M d^x \\ &\quad - \|\Lambda^{1/2}(S-C)^{-1/2}(\nabla c^\top d^x - d^s)\|^2 - \|(\Sigma - \Lambda)^{1/2} S^{-1/2} d^s\|^2 \\ &\quad - \tau\|(S-C)^{-1/2}\Lambda^{1/2}(\nabla c^\top d^x - d^s) - (S-C)^{1/2}\Lambda^{-1/2} d^\lambda\|^2 \\ &\quad - \tau\|S^{-1/2}(\Sigma-\Lambda)^{1/2} d^s - S^{1/2}(\Sigma - \Lambda)^{-1/2} d^\lambda\|^2 \\ &= -(d^x)^\top M d^x - (1+\tau)\|\Lambda^{1/2}(S-C)^{-1/2}(\nabla c^\top d^x - d^s)\|^2 \\ &\quad - (1+\tau)\|(\Sigma-\Lambda)^{1/2} S^{-1/2} d^s\|^2 - \tau\|(S-C)^{1/2}\Lambda^{-1/2} d^\lambda\|^2 \\ &\quad - \tau\|S^{1/2}(\Sigma-\Lambda)^{-1/2} d^\lambda\|^2 + 2\tau(d^\lambda)^\top(\nabla c^\top d^x - d^s) \\ &\quad + 2\tau(d^\lambda)^\top d^s. \end{aligned} \tag{B.9}$$

On the other hand, using $(A.11)$ and $(A.12)$ we obtain

$$\begin{aligned}
& d^\top \nabla^2 \psi_{\sigma,\mu}(z) d \\
&= (d^x)^\top \nabla^2_{xx} \ell(x, \tilde{\lambda}) d^x + (1+\tau)\mu \|(S-C)^{-1}(\nabla c^\top d^x - d^s)\|^2 \\
&\quad - 2\tau (d^\lambda)^\top (\nabla c^\top d^x - d^s) + \tau\mu \|\Lambda^{-1} d^\lambda\|^2 + (1+\tau)\mu \|S^{-1} d^s\|^2 - 2\tau (d^\lambda)^\top d^s \\
&\quad + \tau\mu \|(\Sigma - \Lambda)^{-1} d^\lambda\|^2, \qquad \text{(B.10)}
\end{aligned}$$

where $\tilde{\lambda} = (1+\tau)\mu(S-C)^{-1}e - \tau\lambda$. By using $(B.6)$, $(B.9)$, $(B.10)$ and the fact that $z \to \hat{z}_{\sigma,\mu}$ we deduce

$$\begin{aligned}
& \nabla \psi_{\sigma,\mu}(z)^\top d + d^\top \nabla^2 \psi_{\sigma,\mu}(z) d \\
&= (d^x)^\top (\nabla^2_{xx} \ell(x, \tilde{\lambda}) - M) d^x \\
&\quad + (1+\tau)(\nabla c^\top d^x - d^s)^\top (\mu(S-C)^{-2} - \Lambda(S-C)^{-1})(\nabla c^\top d^x - d^s) \\
&\quad + (1+\tau)(d^s)^\top (\mu S^{-2} - (\Sigma - \Lambda) S^{-1}) d^s + \tau (d^\lambda)^\top (\mu \Lambda^{-2} - (S-C)\Lambda^{-1}) d^\lambda \\
&\quad + \tau (d^\lambda)^\top (\mu(\Sigma - \Lambda)^{-2} - S(\Sigma - \Lambda)^{-1}) d^\lambda \\
&\leq o(\|d\|^2).
\end{aligned}$$

Combining this estimate with $(B.8)$ we obtain

$$\psi_{\sigma,\mu}(z+d) - \psi_{\sigma,\mu}(z) - \omega \nabla \psi_{\sigma,\mu}(z)^\top d \leq (\frac{1}{2} - \omega) \nabla \psi_{\sigma,\mu}(z)^\top d + o(\|d\|^2).$$

Since $\omega < \frac{1}{2}$, the proof will be completed if we show that $\nabla \psi_{\sigma,\mu}(z)^\top d \leq -K' \|d\|^2$ for some positive constant K'. Using the Cauchy-Schwarz inequality one has

$$\begin{aligned}
2\tau (d^\lambda)^\top (\nabla c^\top d^x - d^s) &\leq \frac{1+2\tau}{2} \|\Lambda^{1/2}(S-C)^{-1/2}(\nabla c^\top d^x - d^s)\|^2 \\
&\quad + \frac{2\tau^2}{1+2\tau} \|(S-C)^{1/2} \Lambda^{-1/2} d^\lambda\|^2
\end{aligned}$$

and

$$2\tau (d^\lambda)^\top d^s \leq \frac{1+2\tau}{2} \|S^{-1/2}(\Sigma - \Lambda)^{1/2} d^s\|^2 + \frac{2\tau^2}{1+2\tau} \|S^{1/2}(\Sigma - \Lambda)^{-1/2} d^\lambda\|^2.$$

Combining these two inequalities with $(B.9)$, and next using $(B.7)$ and Lemma A.2 we obtain

$$\begin{aligned}
\nabla \psi_{\sigma,\mu}(z)^\top d &= -(d^x)^\top M d^x \\
&\quad - \frac{1}{2} \|\Lambda^{1/2}(S-C)^{-1/2}(\nabla c^\top d^x - d^s)\|^2 - \frac{1}{2} \|(\Sigma - \Lambda)^{1/2} S^{-1/2} d^s\|^2 \\
&\quad - \frac{\tau}{1+2\tau} \|(S-C)^{1/2} \Lambda^{-1/2} d^\lambda\|^2 - \frac{\tau}{1+2\tau} \|S^{1/2}(\Sigma - \Lambda)^{-1/2} d^\lambda\|^2 \\
&\leq -K_1' \|d^x\|^2 - K_2' \|\nabla c^\top d^x - d^s\|^2 - K_3' \|d^s\|^2 - K_4' \|d^\lambda\|^2.
\end{aligned}$$

For any $\epsilon > 0$:

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

Set now $\epsilon := K_1'/(2K_2'\|\nabla c^\top\|^2)$ to conclude that

$$\nabla\psi_{\sigma,\mu}(z)^\top d \leq -\frac{K_1'}{2}\|d^x\|^2 - (K_3' + \frac{\epsilon K_2'}{1+\epsilon})\|d^s\|^2 - K_4'\|d^\lambda\|^2 \leq -K'\|d\|^2.$$

□

The following lemma gives a characterization of the q-superlinear convergence of Algorithm $\mathrm{A}_{\sigma,\mu}$. The proof is analogous to the one of Proposition 4.2 in [1] and is omitted.

Lemma B.5 *Suppose that Assumption* 2.1 *holds and that f and c are twice differentiable at $\hat{x}_{\sigma,\mu}$. Suppose that the sequence $\{z_k\}$ generated by Algorithm $A_{\sigma,\mu}$ converges to $\hat{z}_{\sigma,\mu}$ and that, for k sufficiently large, the unit step length $\alpha_k = 1$ is accepted by the line search. Then $\{z_k\}$ converges q-superlinearly towards $\hat{z}_{\sigma,\mu}$ if and only if*

$$(M_k - \hat{M}_{\sigma,\mu})d_k^x = o(\|d_k\|). \tag{B.11}$$

Proof of Theorem 3.6. According to the proof of Theorem 4.4 in [1], the convergence of the series $\sum_{k\geq 1}\|z_k - \hat{z}_{\sigma,\mu}\|$ (Theorem 3.5), the local radial Lipschitz continuity of the Hessians of f and c and a standard result from the BFGS theory (see [13, Theorem 3] and [4]), imply

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

This estimate implies that (B.6) holds, therefore the unit step length is accepted (Lemma B.4). It implies also that (B.11) holds, and thus q-superlinear convergence follows (Lemma B.5). □

References

[1] Paul Armand, Jean Charles Gilbert, and Sophie Jan-Jégou. A feasible BFGS interior point algorithm for solving convex minimization problems. *SIAM Journal on Optimization*, 11(1):199–222, 2000.

[2] Paul Armand, Jean Charles Gilbert, and Sophie Jan-Jégou. A BFGS-IP algorithm for solving strongly convex optimization problems with feasibility enforced by an exact penalty approach. *Math. Program.*, 92(3, Ser. B):393–424, 2002. ISMP 2000, Part 2 (Atlanta, GA).

[3] Richard H. Byrd, Jean Charles Gilbert, and Jorge Nocedal. A trust region method based on interior point techniques for nonlinear programming. *Mathematical Programming*, 89(1, Ser. A):149–185, 2000.

[4] Richard H. Byrd and Jorge Nocedal. A tool for the analysis of quasi-Newton methods with application to unconstrained minimization. *SIAM Journal on Numerical Analysis*, 26:727–739, 1989.

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

[6] José Herskovits. A two-stage feasible directions algorithm for nonlinear constrained optimization. *Mathematical Programming*, 36(1):19–38, 1986.

[7] Jean-Baptiste Hiriart-Urruty and Claude Lemaréchal. *Convex analysis and minimization algorithms. I*. Springer-Verlag, Berlin, 1993. Fundamentals.

[8] Craig T. Lawrence and André L. Tits. Nonlinear equality constraints in feasible sequential quadratic programming. *Optimization Methods and Software*, 6:265–282, 1996.

[9] D. Q. Mayne and E. Polak. Feasible directions algorithms for optimization problems with equality and inequality constraints. *Mathematical Programming*, 11(1):67–80, 1976.

[10] L. McLinden. An analogue of Moreau's proximation theorem, with application to the nonlinear complementarity problem. *Pacific J. Math.*, 88(1):101–161, 1980.

[11] Renato D. C. Monteiro and Fangjun Zhou. On the existence and convergence of the central path for convex programming and some duality results. *Comput. Optim. Appl.*, 10(1):51–77, 1998.

[12] Jorge Nocedal and Stephen J. Wright. *Numerical optimization*. Springer-Verlag, New York, 1999.

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

[14] R. Tyrrell Rockafellar. *Convex analysis*. Princeton University Press, Princeton, N.J., 1970.

[15] André L. Tits, Thomas J. Urban, Sasan Bakhtiari, and Craig T. Lawrence. A primal-dual interior-point method for nonlinear programming with strong global and local convergence properties. Technical Report TR 2001-3, Institute for Systems Research, University of Maryland, College Park, July 2001.

[16] Stephen J. Wright. *Primal-dual interior-point methods*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.