Math. Program., Ser. A (2013) 137:199–223
DOI 10.1007/s10107-011-0493-8

FULL LENGTH PAPER

A quasi-Newton strategy for the sSQP method for variational inequality and optimization problems

Damián Fernández

Received: 25 April 2009 / Accepted: 2 September 2011 / Published online: 8 October 2011

Abstract The quasi-Newton strategy presented in this paper preserves one of the most important features of the stabilized Sequential Quadratic Programming method, the local convergence without constraint qualifications assumptions. It is known that the primal-dual sequence converges quadratically assuming only the second-order sufficient condition. In this work, we show that if the matrices are updated by performing a minimization of a Bregman distance (which includes the classic updates), the quasi-Newton version of the method converges superlinearly without introducing further assumptions. Also, we show that even for an unbounded Lagrange multipliers set, the generated matrices satisfies a bounded deterioration property and the Dennis-Moré condition.

Keywords Stabilized sequential quadratic programming · Karush–Kuhn–Tucker system · Variational inequality · Quasi-Newton methods · Superlinear convergence

Mathematics Subject Classification (2000) 65K05 · 90C30 · 90C53

1 Introduction

Given smooth mappings $F : \mathbb{R}^n \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^m$, we consider the following variational problem:

$$\text{find } x \in D \quad \text{such that} \quad \langle F(x), y - x\rangle \geq 0 \quad \forall\, y \in x + \mathcal{T}_D(x), \tag{1}$$

D. Fernández (✉)
FaMAF-UNC, Faculty of Mathematics, Astronomy and Physics,
National University of Córdoba, Medina Allende s/n,
Ciudad Universitária, Córdoba, Argentina
e-mail: dfernandez@famaf.unc.edu.ar

where

$$D = \left\{x \in \mathbb{R}^n \mid g_i(x) \leq 0,\ i = 1, \ldots, m\right\},$$

and $\mathcal{T}_D(x)$ is the (standard) tangent cone to the set D at the point $x \in D$, i.e., $d \in \mathcal{T}_D(x)$ if there is a sequence of vectors $d^k \to d$ along with a sequence of scalars $t_k \to 0^+$ such that $x + t_k d^k \in D$. Throughout the paper, we assume that at least F is once and g is twice continuously differentiable at a solution $\bar{x}$ of (1).

To simplify formulas, we consider only inequality constraints. For equality constraints or mixed constraints the extension is the obvious.

When for some smooth function $f : \mathbb{R}^n \to \mathbb{R}$ it holds that

$$F(x) = f'(x), \quad x \in \mathbb{R}^n, \tag{2}$$

then (1) describes (primal) first-order necessary optimality conditions for the optimization problem

$$\min\ \ f(x) \quad \text{subject to} \quad x \in D. \tag{3}$$

When the feasible set D is convex, the variational problem (1) is equivalent to the classical variational inequality:

$$\text{find } x \in D \quad \text{such that} \quad \langle F(x), y - x\rangle \geq 0 \quad \forall\, y \in D.$$

See [7] for more details.

The main topic of this work is the study of a quasi-Newton version of the method introduced in [8] to solve problem (1). This method is an adaptation of the *stabilized Sequential Quadratic Programming* (sSQP) method, introduced by Wright in [22].

The sSQP method was created to obtain fast convergence in presence of degenerate constraints in optimization problems. The local analysis of this method, and its good behavior, has been studied in [14,19,22–25]. In [9], the method was studied as a particular case of a more general iterative framework. In [8], the sSQP method was adapted to solve (1), and it was shown that without any constraint qualification assumption, the method generates a primal-dual sequence that converges superlinearly/quadratically for initial points near a solution satisfying the second-order sufficient condition.

In order to complete the local analysis of the sSQP method, we study a quasi-Newton strategy for this method, showing that under perturbations the method still guarantee a good local behavior without introducing additional hypotheses. In many aspects of this paper we take as a guide the work of Bonnans [3], devoted to the study of the Josephy-Newton method [17,18]. In [3], Bonnans improved the classical result of [2] for optimization problems, showing consistency and convergence of this kind of methods under the weakest set of assumptions in the literature.

One of the most important differences with our approach, is the fact that the stabilization procedure allows us to work with degenerate solutions, and hence, with a possible unbounded Lagrange multipliers set. The existence of an isolated primal-dual solution and the existence of solutions under small perturbations (called "semistability" and "hemistability" in [3]), have been replaced by a calmness condition (introduced in [21]) of the solution set. This calmness condition is automatically fulfilled under the usual second-order sufficient condition (see [15]).

The quasi-Newton update proposed in this work must be a matrix satisfying the secant equation and with a minimal change respect to the previous one, measured using a Bregman distance. This kind of update contains as a particular case the BFGS, PSB and Broyden update, among others. We show that with this kind of update, the calmness of the solution set is sufficient to show that the method is well-defined and that a bounded deterioration condition holds. Also, we show that a Dennis–Moré type condition is satisfied and that the generated primal-dual sequence converges superlinealy.

The rest of the paper is organized as follows. In Sect. 2, we introduce the quasi-Newton algorithm and define concepts that will be used. We prove in Sect. 3 that the second-order sufficient condition is sufficient to guarantee solvability of the subproblems under small perturbations of the data. In Sect. 4 we describe how to perform the matrix update, we show the consistency of the algorithm and we state the superlinear convergence of the generated primal-dual sequence. In Sect. 5, we comment some facts about the primal convergence of the method. We give a numerical example in Sect. 6 and some conclusions and lines for future research are stated in Sect. 7.

Some words about our notation. We use $\langle \cdot, \cdot \rangle$ to denote the Euclidean inner product, $\| \cdot \|$ the associated norm, and B the closed unit ball (the space is always clear from the context). We use $\langle\!\langle \cdot, \cdot \rangle\!\rangle$ to denote an inner product in the space of matrices and $\| \cdot \|$ to denote the associated norm. We use I to denote the identity matrix (the dimension is always clear from the context). For any matrix M, $M_{\mathcal{I}}$ denotes the submatrix of M with rows indexed by the set $\mathcal{I}$. When in matrix notation, vectors are considered columns, and for a vector x we denote by $x_{\mathcal{I}}$ the subvector of x with coordinates indexed by $\mathcal{I}$. For simplicity, we use (x, μ) for the column vector $(x^\top, \mu^\top)^\top$. We use the notation $\xi(t) = o(t)$ for any function $\xi : \mathbb{R}_+ \to \mathbb{R}^q$ such that $\lim_{t \to 0} t^{-1} \xi(t) = 0$. For a function $\Psi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^q$, we denote by $\Psi'(\bar{x}, \bar{\mu})$ the full derivative of Ψ at the point $(\bar{x}, \bar{\mu})$, and by $\Psi'_x(\bar{x}, \bar{\mu})$ the partial derivative of Ψ with respect to x at $(\bar{x}, \bar{\mu})$. For a set $S \subset \mathbb{R}^l$ and a point $z \in \mathbb{R}^l$, the distance from z to S is defined as $\mathrm{dist}(z, S) = \inf_{s \in S} \|z - s\|$. Then $\Pi_S(z) = \{s \in S \mid \mathrm{dist}(z, S) = \|z - s\|\}$ is the set of all points in S that have minimal distance to z.

2 Quasi-Newton strategy

In order to solve (1) we will use the method introduced in [8], replacing the higher order derivatives by a matrix that can be computed easily.

To this end let $c \in [1, \infty)$ be an arbitrary but fixed constant and consider a sequence $\{(x^k, \mu^k)\}$ generated by the following process:

Given $(x^k, \mu^k) \in \mathbb{R}^n \times \mathbb{R}^m$, $M_k \in \mathbb{R}^{n \times n}$ and $\sigma_k > 0$, compute $(x^{k+1}, \mu^{k+1}) \in \mathbb{R}^n \times \mathbb{R}^m$ solution of the *affine* variational inequality

$$\text{find } (y, \nu) \in \Delta_k \quad \text{s.t.} \quad \langle \Phi_k(y, \nu), (z, \lambda) - (y, \nu) \rangle \geq 0 \quad \forall\, (z, \lambda) \in \Delta_k, \tag{4}$$

satisfying

$$\left\| \begin{bmatrix} y - x^k \\ \nu - \mu^k \end{bmatrix} \right\| \leq c\, \sigma_k, \tag{5}$$

where

$$\Phi_k : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m, \quad \Phi_k(y, \nu) = \begin{bmatrix} F(x^k) + M_k(y - x^k) \\ \sigma_k \nu \end{bmatrix},$$

and

$$\Delta_k = \left\{ (y, \nu) \in \mathbb{R}^n \times \mathbb{R}^m \mid g(x^k) + g'(x^k)(y - x^k) - \sigma_k(\nu - \mu^k) \leq 0 \right\}.$$

As can be seen, $\left(x^k, \max\left\{0, \mu^k + \frac{1}{\sigma_k} g(x^k)\right\}\right) \in \Delta_k$. Thus, subproblems (4) are always feasible independent of any constraint qualification assumption. In Theorem 1, we show the existence of a solution of (4) satisfying (5) for a suitable parameter σ_k. Hence, for optimization problems (3), if the matrix M_k is symmetric and positive definite, then (x^{k+1}, μ^{k+1}) will be the unique solution of the quadratic problem

$$\begin{array}{cl} \min\limits_{(y,\nu)\in\mathbb{R}^n\times\mathbb{R}^m} & \left\langle f'(x^k), y - x^k \right\rangle + \frac{1}{2} \left\langle M_k(y - x^k), y - x^k \right\rangle + \frac{\sigma_k}{2} \|\nu\|^2 \\ \text{s.t.} & g(x^k) + g'(x^k)(y - x^k) - \sigma_k(\nu - \mu^k) \leq 0. \end{array} \tag{6}$$

By the affine structure of the subproblem, we know that (y, ν) is a solution of (4) if and only if there exists $\lambda \in \mathbb{R}^m$ such that

$$\begin{aligned} 0 &= F(x^k) + M_k(y - x^k) + g'(x^k)^\top \lambda, \\ 0 &= \sigma_k \nu - \sigma_k \lambda, \\ 0 &\leq \lambda \perp \left[g(x^k) + g'(x^k)(y - x^k) - \sigma_k(\nu - \mu^k) \right] \leq 0, \end{aligned}$$

where $\nu \perp u$ means that $\langle \nu, u \rangle = 0$. Hence, since $\sigma_k > 0$, we conclude that (y, ν) is a solution of (4) if and only if (y, ν) satisfies

$$\begin{aligned} 0 &= F(x^k) + M_k(y - x^k) + g'(x^k)^\top \nu, \\ 0 &\leq \nu \perp \left[g(x^k) + g'(x^k)(y - x^k) - \sigma_k(\nu - \mu^k) \right] \leq 0. \end{aligned} \tag{7}$$

The matrix $M_k \in \mathbb{R}^{n\times n}$ will be updated in the classical form, as in most quasi-Newton algorithm (see details in Sect. 4).

The parameter $\sigma_k > 0$ is some computable quantity measuring the error produced by (x^k, μ^k) with respect to the solution set of the Karush–Kuhn–Tucker (KKT) system for (1), which is

$$\begin{aligned} 0 &= \Psi(x, \mu), \\ 0 &\leq \mu \perp g(x) \leq 0, \end{aligned} \tag{8}$$

where

$$\Psi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n, \quad \Psi(x, \mu) = F(x) + g'(x)^\top \mu.$$

To be more specific, let Σ be a set-valued map from $\mathbb{R}^n \times \mathbb{R}^m$ to $\mathbb{R}^n \times \mathbb{R}^m$ such that

$$\Sigma(p) = \left\{ w \in \mathbb{R}^n \times \mathbb{R}^m \mid 0 \in G(w) + \mathcal{N}_{\mathbb{R}^n \times \mathbb{R}^m_+}(w) + p \right\}, \tag{9}$$

where for $w = (x, \mu)$,

$$G(x, \mu) = \begin{bmatrix} \Psi(x, \mu) \\ -g(x) \end{bmatrix}, \quad \mathcal{N}_{\mathbb{R}^n \times \mathbb{R}^m_+}(x, \mu) = \begin{bmatrix} 0 \\ \mathcal{N}_{\mathbb{R}^m_+}(\mu) \end{bmatrix}, \tag{10}$$

and

$$\mathcal{N}_{\mathbb{R}^m_+}(\mu) = \begin{cases} \{\nu \in \mathbb{R}^m \mid \nu \leq 0, \langle \nu, \mu \rangle = 0\} & \text{if } \mu \geq 0, \\ \emptyset, \ \text{otherwise}, \end{cases}$$

is the normal cone to $\mathbb{R}^m_+$ at $\mu \in \mathbb{R}^m$. Hence, $\Sigma(0)$ is the solution set of the KKT system (8) and we have that the Lagrange multipliers set associated to $\bar{x}$ is defined as

$$\mathcal{M}(\bar{x}) = \{\mu \mid (\bar{x}, \mu) \in \Sigma(0)\}.$$

Also, we will say that a function $\sigma : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m$ provides a local error bound for the solution set of (8) at a point $\bar{w} = (\bar{x}, \bar{\mu})$ if there exist a neighborhood $\mathcal{V}$ of $\bar{w}$ and constants $\beta_2 \geq \beta_1 > 0$ such that

$$\beta_1 \operatorname{dist}(w, \Sigma(0)) \leq \sigma(w) \leq \beta_2 \operatorname{dist}(w, \Sigma(0)), \quad \forall w \in \mathcal{V}. \tag{11}$$

An example of a computable parameter σ_k satisfying (11) is given by the natural residual of (8) at (x^k, μ^k), that is,

$$\sigma_k = \sigma(x^k, \mu^k), \ \text{with} \ \sigma(x, \mu) = \left\| \begin{bmatrix} \Psi(x, \mu) \\ \min\{-g(x), \mu\} \end{bmatrix} \right\|.$$

By [9, Theorem 2], we know that the natural residual provides an error bound for the solution set of (8) at a point $\bar{w} = (\bar{x}, \bar{\mu})$ if $\bar{w}$ satisfy the following calmness condition: there exist $\bar{\varepsilon}, \gamma, \tau > 0$ such that

$$\Sigma(p) \cap (\bar{w} + \bar{\varepsilon} B) \subseteq \Sigma(0) + \tau \|p\| B \quad \forall p \in \gamma B. \tag{12}$$

As can be seen in the literature (see [15, Lemma 2] or [8, Lemma 2]), the calmness condition (12) holds at a point $\bar{w} = (\bar{x}, \bar{\mu})$ if $\bar{w}$ satisfies the second-order condition (13). Furthermore, the property (12) is also equivalent to the assumption that the multiplier $\bar{\mu}$ in $\bar{w}$ is noncritical [16]. We shall not introduce the latter notion here, as for the purposes of this paper it is enough to mention that $\bar{\mu}$ is noncritical if the following second-order condition holds at $\bar{w}$.

We say that $(\bar{x}, \bar{\mu})$, with $\bar{\mu} \in \mathcal{M}(\bar{x})$, satisfies the second-order condition (SOC) for the KKT system (8) if

$$\left\langle \Psi'_x(\bar{x}, \bar{\mu})u, u \right\rangle > 0 \quad \forall\, u \in \mathcal{C}(\bar{x}; D, F)\backslash\{0\}, \tag{13}$$

where

$$\begin{aligned} \mathcal{C}(\bar{x}; D, F) &= \{u \in \mathbb{R}^n \mid \langle F(\bar{x}), u\rangle = 0,\ \langle g'_i(\bar{x}), u\rangle \leq 0\, \forall\, i \in \mathcal{I}(\bar{x})\} \\ &= \left\{ u \in \mathbb{R}^n \,\middle|\, \begin{array}{ll} \langle g'_i(\bar{x}), u\rangle = 0 & \forall\, i \in \mathcal{I}_+(\bar{x}, \mu) \\ \langle g'_i(\bar{x}), u\rangle \leq 0 & \forall\, i \in \mathcal{I}_0(\bar{x}, \mu) \end{array} \right\}, \mu \in \mathcal{M}(\bar{x}), \end{aligned} \tag{14}$$

with

$$\mathcal{I} = \mathcal{I}(\bar{x}) = \{i = 1, \ldots, m \mid g_i(\bar{x}) = 0\}$$

being the set of constraints active at $\bar{x}$, and

$$\mathcal{I}_+(\bar{x}, \mu) = \{i \in \mathcal{I}(\bar{x}) \mid \mu_i > 0\}, \quad \mathcal{I}_0(\bar{x}, \mu) = \mathcal{I}(\bar{x})\backslash\mathcal{I}_+(\bar{x}, \mu),$$

being the set of strongly and weakly active constraints, respectively. As is well known, the second expression for $\mathcal{C}(\bar{x}; D, F)$ does not depend on the choice of $\mu \in \mathcal{M}(\bar{x})$. In the case of the optimization problem (3), $\mathcal{C}(\bar{x}; D, F)$ is the standard critical cone at $\bar{x}$, and (13) is the standard second-order condition, which is sufficient for optimality of the point $\bar{x}$.

Also, since SOC (13) implies that the primal part $\bar{x}$ of the solution is locally unique, we can guarantee that there exists a neighborhood $\mathcal{U}$ of $(\bar{x}, \bar{\mu})$ such that

$$\Sigma(0) \cap \mathcal{U} = (\{\bar{x}\} \times \mathcal{M}(\bar{x})) \cap \mathcal{U}.$$

Hence, we have that for (x, μ) near enough to $(\bar{x}, \bar{\mu})$, (11) can be written as

$$\beta_1\Big(\|x - \bar{x}\| + \operatorname{dist}(\mu, \mathcal{M}(\bar{x}))\Big) \leq \sigma(x, \mu) \leq \beta_2\Big(\|x - \bar{x}\| + \operatorname{dist}(\mu, \mathcal{M}(\bar{x}))\Big).$$

3 Solvability of subproblems

In this section we show that locally, the SOC (13) at $(\bar{x}, \bar{\mu})$ is sufficient to guarantee the solvability of the subproblems (4) when M_k is taken in a neighborhood of $\Psi'_x(\bar{x}, \bar{\mu})$. The proof of the existence of solutions follows extending the results showed in [8] for exact derivatives to the quasi-Newton scheme. We include those results for completeness.

First, let us show that under small perturbations, the SOC (13) implies some copositivity property in the primal-dual space.

Proposition 1 *Suppose that SOC* (13) *holds at* $(\bar{x}, \bar{\mu})$. *Then there exist constants* $\gamma_1, \varepsilon_1, \delta_1, \zeta_1 > 0$ *such that if* $\|x - \bar{x}\| < \varepsilon_1$, $\|M - \Psi'_x(\bar{x}, \bar{\mu})\| < \delta_1$ *and* $0 < \sigma < \zeta_1$, *it holds that*

$$\langle Mu, u\rangle + \sigma\|v\|^2 \geq \gamma_1\left(\|u\|^2 + \sigma\|v\|^2\right) \quad \forall (u, v) \in K(x, \sigma), \tag{15}$$

where

$$K(x, \sigma) = \left\{(u, v) \in \mathbb{R}^n \times \mathbb{R}^{|\mathcal{I}|} \;\middle|\; \begin{array}{ll} \langle g'_i(x), u\rangle = \sigma v_i, & i \in \mathcal{I}_+(\bar{x}, \bar{\mu}) \\ \langle g'_i(x), u\rangle \leq \sigma v_i, & i \in \mathcal{I}_0(\bar{x}, \bar{\mu}) \end{array}\right\}. \tag{16}$$

Proof Suppose the contrary, i.e., that there exist

$$\left\{\left(x^k, M_k, \sigma_k\right)\right\} \to \left(\bar{x}, \Psi'_x(\bar{x}, \bar{\mu}), 0\right)$$

and $(u^k, v^k) \in K(x^k, \sigma_k)$ such that

$$\left\langle M_k u^k, u^k\right\rangle + \sigma_k\left\|v^k\right\|^2 < \frac{1}{k}\left(\left\|u^k\right\|^2 + \sigma_k\left\|v^k\right\|^2\right). \tag{17}$$

Evidently, (17) subsumes that $(u^k, v^k) \neq 0$. Let $\eta_k = \left\|\left(u^k, \sqrt{\sigma_k}v^k\right)\right\| > 0$. Passing onto a subsequence, if necessary, we can assume that

$$\frac{1}{\eta_k}\begin{bmatrix} u^k \\ \sqrt{\sigma_k}v^k \end{bmatrix} \to \begin{bmatrix} \bar{u} \\ \bar{w} \end{bmatrix} \neq 0. \tag{18}$$

Observe that since $\sigma_k \to 0$ and $\sqrt{\sigma_k}v^k/\eta_k$ is bounded, it holds that

$$\sigma_k\frac{v^k}{\eta_k} = \sqrt{\sigma_k}\frac{\sqrt{\sigma_k}v^k}{\eta_k} \to 0. \tag{19}$$

Since $(u^k, v^k) \in K(x^k, \sigma_k)$, dividing now relations in (16) by η_k, passing onto the limit and taking into account (19) we obtain that

$$\left\langle g'_i(\bar{x}), \bar{u}\right\rangle = 0 \;\forall\, i \in \mathcal{I}_+(\bar{x}, \bar{\mu}), \quad \left\langle g'_i(\bar{x}), \bar{u}\right\rangle \leq 0 \;\forall\, i \in \mathcal{I}_0(\bar{x}, \bar{\mu}),$$

i.e., $\bar{u} \in \mathcal{C}(\bar{x}; D, F)$.

On the other hand, dividing (17) by η_k^2 and taking limits, we have that

$$\left\langle \Psi'_x(\bar{x}, \bar{\mu})\bar{u}, \bar{u}\right\rangle + \|\bar{w}\|^2 \leq 0. \tag{20}$$

This shows that $\left\langle \Psi'_x(\bar{x}, \bar{\mu})\bar{u}, \bar{u}\right\rangle \leq 0$ for $\bar{u} \in \mathcal{C}(\bar{x}; D, F)$. Hence, $\bar{u} = 0$. Now from (20) we have that $\bar{w} = 0$ also, in contradiction with (18). □

Corollary 1 *Suppose that SOC* (13) *holds at* $(\bar{x}, \bar{\mu})$. *Then there exist constants* $\varepsilon_1, \delta_1, \zeta_1 > 0$ *such that if* $\|x - \bar{x}\| < \varepsilon_1$, $\|M - \Psi'_x(\bar{x}, \bar{\mu})\| < \delta_1$ *and* $0 < \sigma < \zeta_1$, *the matrix*

$$\begin{bmatrix} M & g'_{\mathcal{I}}(x)^\top \\ -g'_{\mathcal{I}}(x) & \sigma I \end{bmatrix} \tag{21}$$

is nonsingular.

Proof By Proposition 1, there exist constants $\gamma_1, \varepsilon_1, \delta_1, \zeta_1 > 0$ such that (15) holds. Let $\|x - \bar{x}\| < \varepsilon_1$, $\|M - \Psi'_x(\bar{x}, \bar{\mu})\| < \delta_1$, $0 < \sigma < \zeta_1$ and suppose that (u, v) is a vector in the kernel of the matrix given in (21), i.e.,

$$0 = Mu + g'_{\mathcal{I}}(x)^\top v, \tag{22}$$
$$0 = -g'_{\mathcal{I}}(x)u + \sigma v. \tag{23}$$

By (23) we have that $\langle g'_i(x), u\rangle = \sigma v_i$ for all $i \in \mathcal{I}$. This shows that $(u, v) \in K(x, \sigma)$ defined in (16). Also, multiplying (23) by $v^\top$ we have

$$\left\langle g'_{\mathcal{I}}(x)u, v\right\rangle = \sigma \|v\|^2.$$

Multiplying by $u^\top$ both sides in (22), we then obtain that

$$0 = \langle Mu, u\rangle + \left\langle g'_{\mathcal{I}}(x)^\top v, u\right\rangle = \langle Mu, u\rangle + \sigma\|v\|^2.$$

Then, by (15), we have that $0 \geq \gamma_1 \left(\|u\|^2 + \sigma\|v\|^2\right)$. Hence, $u = 0$ and $v = 0$, implying that the matrix in (21) is nonsingular. □

The previous partial results, allow us to show the existence of solutions for a problem involving only the active constraints in (7) and satisfying (5).

Proposition 2 *Suppose that SOC* (13) *holds at* $(\bar{x}, \bar{\mu})$. *Then there exist constants* $\varepsilon_1, \delta_1, \zeta_1 > 0$ *such that if* $\|x - \bar{x}\| < \varepsilon_1$, $\|M - \Psi'_x(\bar{x}, \bar{\mu})\| < \delta_1$ *and* $0 < \sigma < \zeta_1$, *the mixed complementarity problem of finding* $(y, v_{\mathcal{I}}) \in \mathbb{R}^n \times \mathbb{R}^{|\mathcal{I}|}$ *such that*

$$\begin{aligned} &0 = F(x) + M(y - x) + g'_{\mathcal{I}}(x)^\top v_{\mathcal{I}}, \\ &0 = g_i(x) + \langle g'_i(x), y - x\rangle - \sigma(v_i - \mu_i),\ i \in \mathcal{I}_+(\bar{x}, \bar{\mu}), \\ &0 \leq \lambda_i \perp g_i(x) + \langle g'_i(x), y - x\rangle - \sigma(v_i - \mu_i) \leq 0,\ i \in \mathcal{I}_0(\bar{x}, \bar{\mu}), \end{aligned} \tag{24}$$

has a nonempty compact solution set.

Proof The same proof of [8, Proposition 2], changing $\Psi'_x(x, \mu)$ by M, $\sigma(x, \mu)$ by σ and $K(x, \mu)$ by $K(x, \sigma)$. □

Proposition 3 *Suppose that SOC (13) holds at $(\bar{x}, \bar{\mu})$ and σ is a function satisfying (11). Then there exist constants $\gamma_2, \varepsilon_1, \delta_1 > 0$ such that if $\|M - \Psi_x'(\bar{x}, \bar{\mu})\| < \delta_1$ and $(x, \mu) \in ((\bar{x}, \bar{\mu}) + \varepsilon_1 B) \cap (\mathbb{R}^n \times \mathbb{R}^m_+)$ with $\sigma(x, \mu) > 0$, it holds that*

$$\left\| \begin{bmatrix} y - x \\ \nu_{\mathcal{I}} - \mu_{\mathcal{I}} \end{bmatrix} \right\| \leq \gamma_2 \sigma(x, \mu),$$

where $(y, \nu_{\mathcal{I}})$ is any solution of (24).

Proof The same proof of [8, Proposition 3], changing $\Psi_x'(x^k, \mu^k)$ by M_k with $M_k \to \Psi_x'(\bar{x}, \bar{\mu})$. □

Now, the existence of solutions under small perturbations, can be written as follows.

Theorem 1 *Suppose that SOC (13) holds at $(\bar{x}, \bar{\mu})$ and σ is a function satisfying (11). Then there exist constants $c, \varepsilon_1, \delta_1 > 0$ such that if $\|M - \Psi_x'(\bar{x}, \bar{\mu})\| < \delta_1$ and $(x, \mu) \in ((\bar{x}, \bar{\mu}) + \varepsilon_1 B) \cap (\mathbb{R}^n \times \mathbb{R}^m_+)$ with $\sigma(x, \mu) > 0$, there exists $(\bar{y}, \bar{\nu})$, a solution of the mixed complementarity problem of finding $(y, \nu) \in \mathbb{R}^n \times \mathbb{R}^m$ such that*

$$\begin{aligned} &0 = F(x) + M(y - x) + g'(x)^\top \nu, \\ &0 \leq \nu \perp \left[g(x) + g'(x)(y - x) - \sigma(x, \mu)(\nu - \mu)\right] \leq 0, \end{aligned} \tag{25}$$

satisfying

$$\left\| \begin{bmatrix} \bar{y} - x \\ \bar{\nu} - \mu \end{bmatrix} \right\| \leq c\, \sigma(x, \mu). \tag{26}$$

Proof The same as [8, Theorem 3]. □

4 Convergence of the primal-dual sequence

The convergence will be proved adapting the ideas presented in [3] and [9]. To this end, along this section we will write the main problem and subproblems as generalized equations. Is important to remark that, as has been said in Sect. 2, the SOC (13) implies the calmness condition (12).

Let us define $w = (x, \mu) \in \mathbb{R}^n \times \mathbb{R}^m = \mathbb{R}^q$ and

$$H_\sigma(w, M) = \begin{bmatrix} M & g'(x)^\top \\ -g'(x) & \sigma(w) I \end{bmatrix}. \tag{27}$$

Then, any point $\bar{z} = (\bar{y}, \bar{\nu})$ satisfying (25) and (26) must belong into the set

$$\mathcal{S}_{c,\sigma}(w, M) = \left\{ z \in \mathbb{R}^q \,\middle|\, \begin{array}{c} 0 \in G(w) + H_\sigma(w, M)(z - w) + \mathcal{N}_{\mathbb{R}^n \times \mathbb{R}^m_+}(z) \\ \text{and} \quad \|z - w\| \leq c\, \sigma(w) \end{array} \right\}, \tag{28}$$

where G was defined in (10).

In order to prove the following partial result, and hence our convergence result, we strength the smoothness of the involved functions. This degree of smoothness is the standard for the quasi-Newton methods presented in the literature.

Proposition 4 *Let F' and g'' be Lipschitz-continuous at $\bar{x}$. Suppose that SOC (13) holds at $\bar{w} = (\bar{x}, \bar{\mu})$ and σ is a function satisfying (11). Then there exist constants $\varepsilon_2, \delta_2 > 0$ such that if $\|w - \bar{w}\| < \varepsilon_2$, $\|M - \Psi_x'(\bar{w})\| < \delta_2$ and $z \in \mathcal{S}_{c,\sigma}(w, M)$, it holds that*

$$dist(z, \Sigma(0)) \leq \frac{1}{2} dist(w, \Sigma(0)).$$

Proof Since σ satisfies (11) and $\bar{w} \in \Sigma(0)$, then

$$\sigma(w) \leq \beta_2 \text{dist}(w, \Sigma(0)) \leq \beta_2 \|w - \bar{w}\|. \tag{29}$$

For any matrix

$$H = \begin{bmatrix} M_{11} & M_{12}^\top \\ -M_{12} & 0 \end{bmatrix},$$

define $\|\cdot\|_\sharp$ so that $\|H\|_\sharp = \|M_{11}\| + \|M_{12}\|$ and let $c_1 > 0$ satisfies $\|\cdot\| \leq c_1 \|\cdot\|_\sharp$. Hence, using that $\|g'(x) - g'(\bar{x})\| \leq c_2 \|x - \bar{x}\|$ for some $c_2 > 0$, we have

$$\begin{aligned} \|H_\sigma(w, M) - G'(\bar{w})\| &\leq c_1 \|M - \Psi_x'(\bar{x}, \bar{\mu})\| + c_1 c_2 \|x - \bar{x}\| + \sigma(w) \\ &\leq c_1 \|M - \Psi_x'(\bar{x}, \bar{\mu})\| + (c_1 c_2 + \beta_2) \|w - \bar{w}\|. \end{aligned} \tag{30}$$

Using that $z \in \mathcal{S}_{c,\sigma}(w, M)$, we obtain

$$\begin{aligned} \|z - \bar{w}\| &\leq \|z - w\| + \|w - \bar{w}\| \\ &\leq c\,\sigma(w) + \|w - \bar{w}\| \leq (c\beta_2 + 1)\|w - \bar{w}\|. \end{aligned} \tag{31}$$

Now, by the Lipschitz-continuity of G' at $\bar{w}$, there exist $c_3 > 0$ such that for any $t \in [0, 1]$,

$$\begin{aligned} \|G'(w + t(z - w)) - G'(\bar{w})\| &\leq c_3(\|z - w\| + \|w - \bar{w}\|) \\ &\leq c_3(c\beta_2 + 1)\|w - \bar{w}\|. \end{aligned} \tag{32}$$

Let $p = G(w) + H_\sigma(w, M)(z - w) - G(z)$, hence from (28) and (9), we conclude that $z \in \Sigma(p)$. Also, we have

$$\|p\| \leq \left[\int_0^1 \|G'(w + t(z - w)) - G'(\bar{w})\| + \|G'(\bar{w}) - H_\sigma(w, M)\| \right] \|z - w\|.$$

Take

$$\varepsilon_2 \leq \min\left\{\frac{1}{(1+c\beta_2)6\tau c_3 c\beta_2}, \frac{1}{(\beta_2 + c_1 c_2)6\tau c\beta_2}, \frac{\bar{\varepsilon}}{1+c\beta_2}, 2\tau\gamma\right\},$$
$$\delta_2 \leq \frac{1}{6\tau c_1 c\beta_2},$$

with $\bar{\varepsilon}$, τ, γ satisfying (12). Thus, using (30, 32) and the fact that $\|w - \bar{w}\| < \varepsilon_2$ and $\|M - \Psi'_x(\bar{x}, \bar{\mu})\| < \delta_2$, we obtain

$$\begin{aligned} \|p\| &< \frac{1}{2\tau c\beta_2}\|z - w\| \leq \frac{1}{2\tau}\operatorname{dist}(w, \Sigma(0)) \\ &\leq \frac{1}{2\tau}\|w - \bar{w}\| \leq \gamma. \end{aligned} \tag{33}$$

Using the definition of ε_2 in (31), we have

$$\|z - \bar{w}\| < \bar{\varepsilon}$$

Hence, since $z \in \Sigma(p) \cap (\bar{w} + \bar{\varepsilon}B)$ and $p \in \gamma B$, using (12) and (33) it follows that

$$\operatorname{dist}(z, \Sigma(0)) \leq \tau\|p\| \leq \frac{1}{2}\operatorname{dist}(w, \Sigma(0)).$$

□

4.1 Updating the matrices

Before proving the consistency and convergence of the method, we must specify how to generate the sequence of matrices $\{M_k\}$. To this end, let us consider a subspace $\mathcal{X} \subseteq \mathbb{R}^{n\times n}$ such that

$$\Psi'_x(x, \mu) \in \mathcal{X}.$$

Let us provided $\mathcal{X}$ with the topology induced by the inner product $\langle\!\langle \cdot, \cdot \rangle\!\rangle$. Given a function $\varphi : \mathcal{X} \to [-\infty, +\infty]$, strictly convex and differentiable on the interior of $\operatorname{dom}\varphi = \{X \in \mathcal{X} \mid \varphi(X) < \infty\}$, we define the *Bregman "distance"* between matrices as

$$D_\varphi : \mathcal{X} \times \operatorname{int}(\operatorname{dom}\varphi) \to [0, \infty],$$

such that

$$D_\varphi(X, Y) = \varphi(X) - \varphi(Y) - \langle\!\langle \varphi'(Y), X - Y \rangle\!\rangle.$$

The function D_φ is not a distance in the sense of metric topology, because it is neither symmetric (unless φ is quadratic) nor satisfies the triangle inequality. Nonetheless, D_φ satisfies the following generalized Pythagorean inequality: for any $X, Y, Z \in \operatorname{int}(\operatorname{dom}\varphi)$,

$$D_\varphi(X, Z) = D_\varphi(X, Y) + D_\varphi(Y, Z) - \langle\!\langle \varphi'(Z) - \varphi'(Y), X - Y \rangle\!\rangle\,. \tag{34}$$

Given a linear subspace $\mathcal{L} \subset \mathcal{X}$ such that $\mathcal{L} \cap \operatorname{int}(\operatorname{dom}\varphi) \neq \emptyset$, the proposed quasi-Newton update will be defined as the unique solution $\Pi_{\varphi,\mathcal{L}}(M)$ of the problem

$$\min_{N\in\mathcal{L}} D_\varphi(N, M). \tag{35}$$

Since $\mathcal{L}$ is a subspace, by the optimality conditions of this problem and (34), it can be seen that

$$D_\varphi(N, M) = D_\varphi(N, \Pi_{\varphi,\mathcal{L}}(M)) + D_\varphi(\Pi_{\varphi,\mathcal{L}}(M), M), \tag{36}$$

for any $N \in \mathcal{L}$.

In order to guarantee that (35) has a unique solution we assume that the function φ is a *Bregman/Legendre function* (see [1]). For completeness, we introduce the following technical definitions. If $\operatorname{dom}\varphi \neq \emptyset$ and φ never takes the value $-\infty$, then we say that φ is proper. A proper convex function φ is called *essentially smooth* if it is differentiable on $\operatorname{int}(\operatorname{dom}\varphi) \neq \emptyset$ and $\|\varphi'(X_k)\| \to +\infty$ for every sequence $\{X_k\} \subset \operatorname{int}(\operatorname{dom}\varphi)$ that approaches a boundary point of $\operatorname{dom}\varphi$. A proper convex function φ is a *Legendre function* if it is essentially smooth and strictly convex on every convex subset of $\operatorname{int}(\operatorname{dom}\varphi)$. The conjugate of φ is the function $\varphi^* : \mathcal{X} \to (-\infty, +\infty]$ defined by $\varphi^*(Y) = \sup_{X\in\mathcal{X}}\{\langle\!\langle X, Y \rangle\!\rangle - \varphi(X)\}$.

Hence, by [1, Definition 5.2], φ is a *Bregman/Legendre function* if φ is a Legendre function and the following properties hold:

- BL0. $\operatorname{dom}\varphi^*$ is open.
- BL1. $D_\varphi(X, \cdot)$ is coercive for all $X \in \operatorname{dom}\varphi \backslash \operatorname{int}(\operatorname{dom}\varphi)$.
- BL2. $X \in \operatorname{dom}\varphi \backslash \operatorname{int}(\operatorname{dom}\varphi)$, $\{Y_k\} \subset \operatorname{int}(\operatorname{dom}\varphi)$, $Y_k \to Y \in \operatorname{bd}(\operatorname{dom}\varphi)$ and $\{D_\varphi(X, Y_k)\}$ bounded imply that $D_\varphi(Y, Y_k) \to 0$ (and hence $Y \in \operatorname{dom}\varphi$).
- BL3. $\{X_k\}, \{Y_k\} \subset \operatorname{int}(\operatorname{dom}\varphi)$, $X_k \to X \in \operatorname{dom}\varphi \backslash \operatorname{int}(\operatorname{dom}\varphi)$, $Y_k \to Y \in \operatorname{dom}\varphi \backslash \operatorname{int}(\operatorname{dom}\varphi)$ and $D_\varphi(X_k, Y_k) \to 0$ imply that $X = Y$.

Hence, if φ is a Bregman/Legendre function, we will define

$$M_{k+1} = \Pi_{\varphi,\mathcal{L}_k}(M_k),$$

where

$$\mathcal{L}_k = \left\{ M \in \mathcal{X} \mid M\left(x^{k+1} - x^k\right) = \Psi\left(x^{k+1}, \mu^{k+1}\right) - \Psi\left(x^k, \mu^{k+1}\right)\right\}.$$

Example 1 If $\mathcal{L} = \{M \in \mathcal{X} \mid Ms = y\}$, we obtain the following quasi-Newton updates:

1. Let $\mathcal{X} = \mathbb{R}^{n\times n}$, $\langle\!\langle N, M \rangle\!\rangle = \operatorname{tr}(M^\top N)$ and $\varphi(X) = \frac{1}{2}\|X\|^2$.
   By definition, it can be seen that φ is a Legendre function. Since $\operatorname{dom}\varphi = \operatorname{dom}\varphi^* = \mathbb{R}^{n\times n}$ is open, by [1, Theorem 5.6] we have that φ is a Bregman/Legendre function. Hence, (35) can be written as
$$\min_{N\in\mathcal{L}} \frac{1}{2}\|N - M\|^2.$$
   From [6], we have that
$$\Pi_{\varphi,\mathcal{L}}(M) = M + \frac{(y - Ms)s^\top}{s^\top s}.$$
   Hence, we obtain the Broyden update [4].
2. Let $\mathcal{X} = \mathbb{S}^n$ be the space of symmetric matrices with the induced inner product $\langle\!\langle N, M \rangle\!\rangle = \operatorname{tr}(MN)$ and φ as before.
   We have again that φ is a Bregman/Legendre function. From [6],
$$\Pi_{\varphi,\mathcal{L}}(M) = M + \frac{(y - Ms)s^\top + s(y - Ms)^\top}{s^\top s} - \frac{\langle s, y - Ms\rangle}{(s^\top s)^2} ss^\top.$$
   That is, the Powell symmetric Broyden update [20].
3. Let $\mathcal{X} = \mathbb{S}^n$, $\langle\!\langle N, M \rangle\!\rangle = \operatorname{tr}(MN)$ and $\varphi(X) = -\log(\det(X))$ if $X \in \{Y \in \mathbb{S}^n \mid Y \text{ is positive definite}\}$, $\varphi(X) = +\infty$ otherwise. In [1, Example 7.17] was proven that this function is Bregman/Legendre. Since $\varphi'(X) = -X^{-1}$ for $X \in \operatorname{dom}\varphi$, problem (35) can be written as
$$\min_{N\in\mathcal{L}\cap\operatorname{dom}\varphi} \operatorname{tr}(M^{-1}N) - \log(\det(N)) + \text{const}$$
   From [10] we have that
$$\Pi_{\varphi,\mathcal{L}}(M) = M - \frac{Mss^\top M}{s^\top Ms} + \frac{yy^\top}{y^\top s}.$$
   Then, the Broyden–Fletcher–Goldfarb–Shanno update can also be used with this approach. For a short proof of this formula (and related) see [13].

4.2 Convergence result

Since we are dealing with problems where the multiplier associated to the primal solution may not be unique, a priori, the dual sequence $\{\mu^k\}$ can converge to a multiplier that violates the second-order condition, cornerstone of the analysis of the previous section. To avoid this behavior, note that when $(\bar{x}, \bar{\mu})$ satisfies the SOC (13), we have that there exists a constant $\varepsilon_{\bar{\mu}} > 0$ such that

$$(\bar{x}, \hat{\mu}) \text{ satisfies SOC (13) for all } \hat{\mu} \in \bar{\mu} + \varepsilon_{\bar{\mu}} B. \tag{37}$$

With this safeguard, we can state the following result.

Lemma 1 *Let F' and g'' be Lipschitz-continuous at $\bar{x}$ and let φ be a Bregman/Legendre function. Suppose that SOC (13) holds at $\bar{w} = (\bar{x}, \bar{\mu})$, $\Psi'_x(\bar{x}, \bar{\mu}) \in int(dom\, \varphi)$ and σ is a function satisfying (11). Then there exist constants $\eta, \zeta > 0$ such that if*

$$w^0 \in (\bar{w} + \eta B) \cap \left(\mathbb{R}^n \times \mathbb{R}^m_+\right) \ \textit{and} \ \ M_0 \in \mathcal{X},\ \left\|M_0 - \Psi'_x(\bar{x}, \bar{\mu})\right\| < \zeta, \tag{38}$$

then:

(i) *The sequence generated by $w^{k+1} \in \mathcal{S}_{c,\sigma}(w^k, M_k)$, $M_{k+1} = \Pi_{\varphi,\mathcal{L}_k}(M_k)$ is well-defined.*
(ii) *$\{w^k\}$ converges to some point $\hat{w} = (\bar{x}, \hat{\mu})$ satisfying SOC (13).*
(iii) *$\left\{D_\varphi(\Psi'_x(\bar{x}, \hat{\mu}), M_k)\right\}$ converges.*

Proof Let $c, \varepsilon_1, \delta_1 > 0$ be the constants given by Theorem 1, $\varepsilon_2, \delta_2 > 0$ given in Proposition 4 and $\varepsilon_{\bar{\mu}} > 0$ given in (37).

By the Lipschitz-continuity of F' and g'' at $\bar{x}$ we have that there exists $l_1 > 1$ such that

$$\left\|\Psi'_x(y, \nu) - \Psi'_x(\bar{x}, \mu)\right\| \le \frac{l_1}{2}(\|y - \bar{x}\| + \|\nu - \mu\|), \tag{39}$$

for any $(y, \nu), (x, \mu) \in (\bar{x}, \bar{\mu}) + \varepsilon_{\bar{\mu}} B$.

Let us fix $\gamma_0 > 0$ and define $\Gamma(N) = \{M \in \mathcal{X} \mid D_\varphi(N, M) \le \gamma_0\}$ if $N \in \text{dom}\, \varphi$ and $\Gamma(N) = \emptyset$ otherwise. Since φ is Bregman/Legendre and $\Psi'_x(\bar{x}, \bar{\mu}) \in \text{int}(\text{dom}\, \varphi)$, by [1, Theorem 3.7(vi)] we have that $D_\varphi(\Psi'_x(\bar{x}, \bar{\mu}), \cdot)$ is coercive and then $\Gamma(\Psi'_x(\bar{x}, \bar{\mu}))$ is compact. By [1, Theorem 3.8(i)] we have that $\Gamma(\Psi'_x(\bar{x}, \bar{\mu})) \subset \text{int}(\text{dom}\, \varphi)$. Then, using the convexity of dom φ, there exists a convex compact set Ω such that $\Gamma(\Psi'_x(\bar{x}, \bar{\mu})) + \delta B \subset \Omega \subset \text{int}(\text{dom}\, \varphi)$, for some $\delta > 0$. As in [21, Example 5.8], we obtain that the set-valued map Γ is outer semicontinuous at $\Psi'_x(\bar{x}, \bar{\mu})$. Then, there exists $\delta_3 > 0$ such that

$$\Gamma(N) \subset \Omega \quad \text{if} \quad \left\|N - \Psi'_x(\bar{x}, \bar{\mu})\right\| < \delta_3. \tag{40}$$

The compactness of Ω implies that $c_0 = \sup_{M,N \in \Omega} \|M - N\| < +\infty$, and guarantee the existence of $L > 0$ such that

$$\left\|\varphi'(M) - \varphi'(N)\right\| \le L\|M - N\| \quad \text{for all } M, N \in \Omega. \tag{41}$$

Hence, by definition of D_φ, the convexity of Ω and (41) we have that

$$\begin{aligned} D_\varphi(N, M) &= \int_0^1 \langle\!\langle \varphi'(M + t(N - M)) - \varphi'(M), N - M \rangle\!\rangle\, dt \\ &\le \frac{L}{2}\|N - M\|^2 \quad \text{for all } M, N \in \Omega. \end{aligned} \tag{42}$$

Thus, by (34) we conclude that for any $N, M, \tilde{M} \in \Omega$,

$$D_\varphi(N, M) \le D_\varphi(N, \tilde{M}) + D_\varphi(\tilde{M}, M) + L \left\| M - \tilde{M} \right\| \left\| N - \tilde{M} \right\|$$
$$\le D_\varphi(\tilde{M}, M) + \frac{3}{2} L c_0 \left\| N - \tilde{M} \right\|, \qquad (43)$$

where in the first inequality we use (41) and in the second we use (42).

Let us define $\tilde{\varepsilon} = \min \left\{ \varepsilon_1, \delta_1, \varepsilon_2, \delta_2, \varepsilon_3, \frac{\delta_3}{l_1}, \varepsilon_{\bar{\mu}}, \sqrt{\frac{\gamma_0}{L(l_1^2+1)}} \right\}$.

Using [1, Theorem 3.9(iii)] we obtain the existence of $\delta_4 > 0$ such that

$$\left\| M - \Psi'_x(\bar{x}, \bar{\mu}) \right\| < \tilde{\varepsilon} \quad \text{if} \quad D_\varphi(\Psi'_x(\bar{x}, \bar{\mu}), M) < \delta_4. \qquad (44)$$

Taking into account the constant $\beta_2 > 0$ given in (11), define

$$l_2 = \frac{3}{2} L c_0, \quad l_3 = l_1 l_2 (c\beta_2 + 2), \quad l_4 = l_1 l_2 (c\beta_2 + 1) + 2 l_3 \qquad (45)$$

and take

$$\eta = \min \left\{ \frac{\tilde{\varepsilon}}{1 + 2c\beta_2}, \frac{\delta_4}{2 l_4} \right\}, \quad \zeta = \min \left\{ \tilde{\varepsilon}, \sqrt{\frac{\delta_4}{L}} \right\}.$$

If

$$\left\| w^0 - \bar{w} \right\| < \eta, \quad \left\| M_0 - \Psi'_x(\bar{x}, \bar{\mu}) \right\| < \zeta,$$

by Theorem 1 we have that there exists $w^1 \in \mathcal{S}_{c,\sigma}\left(w^0, M_0\right)$.

In order to use an induction argument, suppose that for $j = 1, \ldots, k$ we have

$$\left\| w^j - \bar{w} \right\| < \tilde{\varepsilon}, \quad \left\| M_j - \Psi'_x(\bar{x}, \bar{\mu}) \right\| < \tilde{\varepsilon}, \quad w^{j+1} \in \mathcal{S}_{c,\sigma}\left(w^j, M_j\right). \qquad (46)$$

This holds trivially for $j = 0$ by the definition of η and ζ.

Hence, for any $j \in \{0, \ldots, k\}$ we have that $\|w^{j+1} - w^j\| \le c\sigma(w^j) \le c\beta_2 \mathrm{dist}(w^j, \Sigma(0))$ and $\mathrm{dist}(w^{j+1}, \Sigma(0)) \le \frac{1}{2}\mathrm{dist}(w^j, \Sigma(0))$, by Proposition 4. Thus

$$\left\| w^{k+1} - w^0 \right\| \le \sum_{i=0}^{k} \left\| w^{i+1} - w^i \right\| \le c\beta_2 \sum_{i=0}^{k} \mathrm{dist}(w^i, \Sigma(0))$$
$$\le c\beta_2 \sum_{i=0}^{k} \frac{1}{2^i} \mathrm{dist}(w^0, \Sigma(0)) \le 2c\beta_2 \mathrm{dist}(w^0, \Sigma(0))$$
$$\le 2c\beta_2 \left\| w^0 - \bar{w} \right\|, \qquad (47)$$

which implies that

$$\left\| w^{k+1} - \bar{w} \right\| \leq \left\| w^{k+1} - w^0 \right\| + \left\| w^0 - \bar{w} \right\| \leq (2c\beta_2 + 1) \left\| w^0 - \bar{w} \right\| < \tilde{\varepsilon}. \quad (48)$$

Note that, for any $j \in \{0, \ldots, k\}$ and $t \in [0, 1]$ we have

$$\left\| x^j + t\left(x^{j+1} - x^j\right) - \bar{x} \right\| \leq \max\left\{ \left\| x^{j+1} - \bar{x} \right\|, \left\| x^j - \bar{x} \right\| \right\} < \tilde{\varepsilon}.$$

Let $\hat{\mu}^j = \Pi_{\mathcal{M}(\bar{x})}(\mu^j)$, then

$$\max\left\{ \left\| \hat{\mu}^{j+1} - \bar{\mu} \right\|, \left\| \hat{\mu}^j - \bar{\mu} \right\| \right\} \leq \max\left\{ \left\| \mu^{j+1} - \bar{\mu} \right\|, \left\| \mu^j - \bar{\mu} \right\| \right\} < \tilde{\varepsilon}.$$

Define

$$N_j = \int_0^1 \Psi'_x\left(x^j + t\left(x^{j+1} - x^j\right), \mu^{j+1}\right) dt. \quad (49)$$

By (39)

$$\left\| N_j - \Psi'_x(\bar{x}, \bar{\mu}) \right\| < l_1 \tilde{\varepsilon} \leq \delta_3,$$

which implies that $\Gamma(N_j) \subset \Omega$ by (40). Hence, $N_j \in \Gamma(N_j) \subset \Omega \subset \operatorname{int}(\operatorname{dom}\varphi)$. Using (39) we also obtain

$$\begin{aligned} \left\| N_j - \Psi'_x\left(\bar{x}, \hat{\mu}^j\right) \right\| &\leq l_1 \left(\left\| w^{j+1} - w^j \right\| + \operatorname{dist}(w^j, \Sigma(0)) \right) \\ &\leq l_1(c\beta_2 + 1)\operatorname{dist}(w^j, \Sigma(0)), \end{aligned} \quad (50)$$

and

$$\begin{aligned} \left\| N_j - \Psi'_x\left(\bar{x}, \hat{\mu}^{j+1}\right) \right\| &\leq l_1 \left(\operatorname{dist}(w^{j+1}, \Sigma(0)) + \frac{1}{2}\operatorname{dist}(w^j, \Sigma(0)) \right) \\ &\leq l_1 \operatorname{dist}(w^j, \Sigma(0)). \end{aligned} \quad (51)$$

Since $\|M_j - \Psi'_x(\bar{x}, \bar{\mu})\| < \tilde{\varepsilon} < \delta_3$, from (40) we have that $M_j \in \Gamma(M_j) \subset \Omega \subset \operatorname{int}(\operatorname{dom}\varphi)$. Then by (42),

$$\begin{aligned} D_\varphi(N_j, M_j) &\leq \frac{L}{2} \left\| N_j - M_j \right\|^2 \\ &\leq L\left(\left\| N_j - \Psi'_x(\bar{x}, \bar{\mu}) \right\|^2 + \left\| M_j - \Psi'_x(\bar{x}, \bar{\mu}) \right\|^2 \right) \\ &\leq L(l_1^2 + 1)\tilde{\varepsilon}^2 \leq \gamma_0. \end{aligned}$$

From the fact that $M_{j+1} = \Pi_{\varphi,\mathcal{L}_j}(M_j)$ and $N_j \in \mathcal{L}_j$, by (36) we have that $D_\varphi(N_j, M_{j+1}) \leq D_\varphi(N_j, M_j) \leq \gamma_0$. Thus, $M_{j+1} \in \Gamma(N_j) \subset \Omega$.

As before, using (40) we have that $\Psi'_x(\bar{x}, \hat{\mu}^{j+1}), \Psi'_x(\bar{x}, \hat{\mu}^j) \in \Omega$. Hence, by (43) and (45) we obtain that

$$D_\varphi(\Psi'_x(\bar{x}, \hat{\mu}^{j+1}), M_{j+1}) \leq D_\varphi(N_j, M_{j+1}) + l_2 \left\| N_j - \Psi'_x(\bar{x}, \hat{\mu}^{j+1}) \right\|,$$

and

$$D_\varphi(N_j, M_j) \leq D_\varphi(\Psi'_x(\bar{x}, \hat{\mu}^j), M_j) + l_2 \left\| N_j - \Psi'_x(\bar{x}, \hat{\mu}^j) \right\|.$$

Then, since $D_\varphi(N_j, M_{j+1}) \leq D_\varphi(N_j, M_j)$, by (50, 51) and (45) we obtain the following bounded deterioration condition

$$D_\varphi(\Psi'_x(\bar{x}, \hat{\mu}^{j+1}), M_{j+1}) \leq D_\varphi(\Psi'_x(\bar{x}, \hat{\mu}^j), M_j) + l_3 \mathrm{dist}(w^j, \Sigma(0)), \tag{52}$$

for any $j \in \{0, \ldots, k\}$.

Hence, it holds that

$$\begin{aligned} D_\varphi(\Psi'_x(\bar{x}, \hat{\mu}^{k+1}), M_{k+1}) &\leq D_\varphi(\Psi'_x(\bar{x}, \hat{\mu}^0), M_0) + l_3 \sum_{j=0}^{k} \mathrm{dist}(w^j, \Sigma(0)) \\ &\leq D_\varphi(\Psi'_x(\bar{x}, \hat{\mu}^0), M_0) + 2 l_3 \mathrm{dist}(w^0, \Sigma(0)). \end{aligned} \tag{53}$$

Using (39, 43) and (45) we have

$$D_\varphi(\Psi'_x(\bar{x}, \hat{\mu}^0), M_0) \leq D_\varphi(\Psi'_x(\bar{x}, \bar{\mu}), M_0) + \frac{l_1}{2} l_2 \left\| \hat{\mu}^0 - \bar{\mu} \right\|, \tag{54}$$

and

$$D_\varphi(\Psi'_x(\bar{x}, \bar{\mu}), M_{k+1}) \leq D_\varphi(\Psi'_x(\bar{x}, \hat{\mu}^{k+1}), M_{k+1}) + \frac{l_1}{2} l_2 \left\| \hat{\mu}^{k+1} - \bar{\mu} \right\|. \tag{55}$$

Since $\|\hat{\mu}^{k+1} - \bar{\mu}\| \leq (2c\beta_2 + 1)\|w^0 - \bar{w}\|$ (by (48)) and $\mathrm{dist}(w^0, \Sigma(0)) \leq \|w^0 - \bar{w}\|$, combining (53, 54, 55) and using (42, 45) we then have

$$\begin{aligned} D_\varphi(\Psi'_x(\bar{x}, \bar{\mu}), M_{k+1}) &\leq D_\varphi(\Psi'_x(\bar{x}, \bar{\mu}), M_0) + l_4 \left\| w^0 - \bar{w} \right\| \\ &< \frac{L}{2}\zeta^2 + l_4 \eta \leq \delta_4. \end{aligned} \tag{56}$$

Hence, by (44) we obtain that $\left\| M_{k+1} - \Psi'_x(\bar{x}, \bar{\mu}) \right\| < \tilde{\varepsilon}$. Now, since $\left\| w^{k+1} - \bar{w} \right\| < \tilde{\varepsilon}$ (by (48)), Theorem 1 guarantee the existence of $w^{k+2} \in \mathcal{S}_{c,\sigma}(w^{k+1}, M_{k+1})$. Then (46) holds for all j, proving (i).

Since (46) holds for all j, we have that $\{w^k\}$ is bounded, and hence, it has at least one accumulation point $\hat{w}$. Using that $\mathrm{dist}(w^k, \Sigma(0)) \leq 2^{-k}\mathrm{dist}(w^0, \Sigma(0))$ for all k we obtain that any accumulation point belongs to $\Sigma(0)$.

Let us show that the entire sequence converges to $\hat{w}$. With an argument similar to that used in (47), it can be seen that for any $k_0 < k_1$ we have

$$\left\| w^{k_1} - w^{k_0} \right\| \leq 2c\beta_2 \mathrm{dist}(w^{k_0}, \Sigma(0)). \tag{57}$$

By contradiction, suppose that $\{w^k\}$ has another accumulation point w^*, and let $\hat{\varepsilon} = \|\hat{w} - w^*\|$. Choose k_0, k_1 large enough so that

$$\left\| w^{k_0} - \hat{w} \right\| \leq \frac{\hat{\varepsilon}}{4}, \quad \left\| w^{k_1} - w^* \right\| \leq \frac{\hat{\varepsilon}}{4}, \quad \mathrm{dist}(w^{k_0}, \Sigma(0)) \leq \frac{\hat{\varepsilon}}{8c\beta_2}.$$

Then

$$\left\| w^{k_1} - w^{k_0} \right\| \leq 2c\beta_2 \mathrm{dist}(w^{k_0}, \Sigma(0)) \leq \frac{\hat{\varepsilon}}{4},$$

and

$$\left\| w^{k_1} - w^{k_0} \right\| \geq \|\hat{w} - w^*\| - \left\| w^{k_0} - \hat{w} \right\| - \left\| w^{k_1} - w^* \right\| \geq \frac{\hat{\varepsilon}}{2}.$$

Hence $\{w^k\}$ converges to $\hat{w} = (\bar{x}, \hat{\mu})$ and $\|\hat{\mu} - \bar{\mu}\| \leq \|\hat{w} - \bar{w}\| < \tilde{\varepsilon} \leq \varepsilon_{\bar{\mu}}$, which proves (ii).

To show (iii), from (52) we deduce that for any $k > j$

$$\begin{aligned} D_\varphi(\Psi'_x(\bar{x}, \hat{\mu}^k), M_k) &\leq D_\varphi(\Psi'_x(\bar{x}, \hat{\mu}^j), M_j) + l_3 \sum_{i=0}^{k-j-1} \mathrm{dist}(w^{j+i}, \Sigma(0)) \\ &\leq D_\varphi(\Psi'_x(\bar{x}, \hat{\mu}^j), M_j) + l_4 \mathrm{dist}(w^j, \Sigma(0)), \end{aligned}$$

where we use that $2l_3 \leq l_4$. Since $\Psi'_x(\bar{x}, \hat{\mu}) \in \Omega$, by an argument similar to that used in (54) and (55), we obtain

$$D_\varphi(\Psi'_x(\bar{x}, \hat{\mu}^j), M_j) \leq D_\varphi(\Psi'_x(\bar{x}, \hat{\mu}), M_j) + l_4 \left\| \hat{\mu}^j - \hat{\mu} \right\|,$$

and

$$D_\varphi(\Psi'_x(\bar{x}, \hat{\mu}), M_k) \leq D_\varphi(\Psi'_x(\bar{x}, \hat{\mu}^k), M_k) + l_4 \left\| \hat{\mu}^k - \hat{\mu} \right\|,$$

where we use that $l_1 l_2 \leq l_4$. Thus

$$\begin{aligned} D_\varphi(\Psi'_x(\bar{x}, \hat{\mu}), M_k) \leq\ & D_\varphi(\Psi'_x(\bar{x}, \hat{\mu}), M_j) \\ & + l_4 \left(\left\| \hat{\mu}^j - \hat{\mu} \right\| + \left\| \hat{\mu}^k - \hat{\mu} \right\| + \mathrm{dist}(w^j, \Sigma(0)) \right). \end{aligned}$$

Since $\hat{\mu}^k = \Pi_{\mathcal{M}(\bar{x})}(\mu^k) \to \Pi_{\mathcal{M}(\bar{x})}(\hat{\mu}) = \hat{\mu}$, then

$$\limsup_{k\to\infty} D_\varphi(\Psi'_x(\bar{x}, \hat{\mu}), M_k) \le D_\varphi(\Psi'_x(\bar{x}, \hat{\mu}), M_j) + l_4 \left(\left\| \hat{\mu}^j - \hat{\mu} \right\| + \mathrm{dist}(w^j, \Sigma(0)) \right).$$

Taking limit in the right-hand side, we deduce

$$\limsup_{k\to\infty} D_\varphi(\Psi'_x(\bar{x}, \hat{\mu}), M_k) \le \liminf_{k\to\infty} D_\varphi(\Psi'_x(\bar{x}, \hat{\mu}), M_k).$$

□

Remark 1 In the proof, we show that the sequence of matrices satisfies a kind of *bounded deterioration principle* (52), similar to that introduced in [5].

Remark 2 The Bregman/Legendre functions in Example 1 are essentially locally strongly convex (see [11]). In such case, we have that $D_\varphi(N, M) \ge \bar{\gamma}\|M - N\|^2$ for some $\bar{\gamma} > 0$ and any M, N near $\Psi'_x(\bar{x}, \bar{\mu})$. Then, the compactness of Γ, the fact that $\Gamma(\Psi'_x(\bar{x}, \bar{\mu})) \subset \mathrm{int}(\mathrm{dom}\,\varphi)$ and (44) follows inmediatly, without use the tools given by [1].

Using the ideas of the analysis developed in [12], we obtain the following.

Corollary 2 *The sequences $\{w^k\}$ and $\{M_k\}$ generated according Lemma 1 satisfy*

$$\lim_{k\to\infty} \|M_k - M_{k+1}\| = 0, \tag{58}$$

and

$$(\Psi'_x(\bar{x}, \hat{\mu}) - M_k)(x^{k+1} - x^k) = o\left(\left\| x^{k+1} - x^k \right\| \right). \tag{59}$$

Proof Let N_k be defined as in (49). By (34) we have that for any $M \in \Omega$

$$|D_\varphi(N_k, M) - D_\varphi(\Psi'_x(\bar{x}, \hat{\mu}), M)| \le D_\varphi(N_k, \Psi'_x(\bar{x}, \hat{\mu})) + Lc_0\|N_k - \Psi'_x(\bar{x}, \hat{\mu})\|.$$

Since $N_k \to \Psi'_x(\bar{x}, \hat{\mu})$, using (iii) in Lemma 1 we conclude that $D_\varphi(N_k, M_{k+1})$ and $D_\varphi(N_k, M_k)$ converge to the same limit. Then by (36),

$$D_\varphi(M_{k+1}, M_k) = D_\varphi(N_k, M_k) - D_\varphi(N_k, M_{k+1}) \to 0. \tag{60}$$

By contradiction, suppose that there exist $\varepsilon > 0$ and an index set $\mathcal{K}$ such that

$$\|M_k - M_{k+1}\| > \varepsilon \quad \text{for all } k \in \mathcal{K}. \tag{61}$$

Since $\{M_k\} \subset \Omega$, we have accumulation points $\bar{M}, \hat{M} \in \Omega$ and an index set $\mathcal{K}_1 \subset \mathcal{K}$ such that

$$M_k \xrightarrow[k\in\mathcal{K}_1]{} \bar{M}, \qquad M_{k+1} \xrightarrow[k\in\mathcal{K}_1]{} \hat{M}.$$

Using (60) we have that $D_\varphi(\hat{M}, \bar{M}) = 0$ with $\bar{M}, \hat{M} \in \Omega \subset \operatorname{int}(\operatorname{dom}\varphi)$. Then, the strict convexity of φ on Ω implies that $\bar{M} = \hat{M}$.

Hence

$$\|M_k - M_{k+1}\| \underset{k \in \mathcal{K}_1}{\longrightarrow} 0,$$

in contradiction with (61). Thus, (58) holds.

By the definition of $\mathcal{L}_k$ we obtain that

$$\begin{aligned} M_{k+1}\left(x^{k+1} - x^k\right) &= \Psi\left(x^{k+1}, \mu^{k+1}\right) - \Psi\left(x^k, \mu^{k+1}\right) \\ &= \Psi\left(x^{k+1}, \hat{\mu}\right) - \Psi\left(x^k, \hat{\mu}\right) + \left(g'\left(x^{k+1}\right) - g'\left(x^k\right)\right)^\top \left(\mu^{k+1} - \hat{\mu}\right) \\ &= \Psi'_x(\bar{x}, \hat{\mu})\left(x^{k+1} - x^k\right) + o\left(\left\|x^{k+1} - x^k\right\|\right). \end{aligned}$$

Hence,

$$\begin{aligned} \left(\Psi'_x(\bar{x}, \hat{\mu}) - M_k\right)\left(x^{k+1} - x^k\right) &= (M_{k+1} - M_k)\left(x^{k+1} - x^k\right) + o\left(\left\|x^{k+1} - x^k\right\|\right) \\ &= o\left(\left\|x^{k+1} - x^k\right\|\right). \end{aligned}$$

□

Without introduce further assumptions, we can show that the convergence in (ii) Lemma 1 occur with superlinear rate.

Theorem 2 *Let F' and g'' be Lipschitz-continuous at $\bar{x}$ and let φ be a Bregman/Legendre function. Suppose that SOC (13) holds at $\bar{w} = (\bar{x}, \bar{\mu})$, $\Psi'_x(\bar{x}, \bar{\mu}) \in$ int(dom φ), σ is a function satisfying (11) and the sequences $\{M_k\}$ and $\{w^k\}$ were generated by*

$$w^{k+1} \in \mathcal{S}_{c,\sigma}(w^k, M_k) \quad \text{and} \quad M_{k+1} = \Pi_{\varphi, \mathcal{L}_k}(M_k),$$

with w^0 and M_0 satisfying (38) for constants $\eta, \zeta > 0$ given by Lemma 1. Then $\left\{\operatorname{dist}(w^k, \Sigma(0))\right\}$ converges superlinearly to 0 and $\{w^k\}$ converges superlinearly to $\hat{w}$.

Proof By Lemma 1, we have that $\{w^k\}$ converges to $\hat{w}$. Using (59), it can be seen that

$$\left(G'(\hat{w}) - H_\sigma(w^k, M_k)\right)\left(w^{k+1} - w^k\right) = o\left(\left\|w^{k+1} - w^k\right\|\right),$$

where G was defined in (10) and H_σ in (27). Since F' and g'' are Lipschitz-continuous, then

$$G\left(w^{k+1}\right) - G\left(w^k\right) - G'\left(\hat{w}\right)\left(w^{k+1} - w^k\right) = o\left(\left\|w^{k+1} - w^k\right\|\right).$$

Hence,

$$p^k = G\left(w^k\right) + H_\sigma\left(w^k, M_k\right)\left(w^{k+1} - w^k\right) - G\left(w^{k+1}\right)$$
$$= o\left(\left\|w^{k+1} - w^k\right\|\right).$$

By the definition of the sequence we have that

$$0 \in G(w^{k+1}) + \mathcal{N}_{\mathbb{R}^n \times \mathbb{R}^m_+}(w^{k+1}) + p^k.$$

Using SOC (13) we obtain that the multifunction Σ defined in (9) satisfies the condition (12). Thus, there exists $\tau > 0$ such that for k large enough

$$w^{k+1} \in \Sigma(p^k) \subseteq \Sigma(0) + \tau\|p^k\|,$$

which implies

$$\mathrm{dist}(w^{k+1}, \Sigma(0)) = o\left(\left\|w^{k+1} - w^k\right\|\right). \tag{62}$$

From (57), taking $k_0 = k + 1$ and letting k_1 goes to infinity, he have

$$\left\|w^{k+1} - \hat{w}\right\| \leq 2c\beta_2 \mathrm{dist}(w^{k+1}, \Sigma(0)).$$

Hence, $\|w^{k+1} - \hat{w}\| = o(\|w^{k+1} - w^k\|)$, implying that

$$\left\|w^{k+1} - \hat{w}\right\| = o\left(\left\|w^k - \hat{w}\right\|\right).$$

From (57), taking $k_1 = k + 1$ and $k_0 = k$, we have

$$\|w^{k+1} - w^k\| \leq 2c\beta_2 \mathrm{dist}(w^k, \Sigma(0)).$$

Hence, using (62),

$$\mathrm{dist}(w^{k+1}, \Sigma(0)) = o\left(\mathrm{dist}(w^k, \Sigma(0))\right). \tag{63}$$

□

5 Comments about primal convergence

To be coherent with the previous analysis, we will say some words about the primal convergence of the method without introduce any constraint qualification assumption nor other additional hypothesis. As has been commented in Sect. 2, the SOC (13) at $(\bar{x}, \bar{\mu})$ implies that the natural residual provides a local error bound at this point and also the isolatedness of the primal solution $\bar{x}$. Hence, there exists a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\mu})$ such that for all $(x, \mu) \in \mathcal{V}$,

$$\left(\|x - \bar{x}\|^2 + \operatorname{dist}(\mu, \mathcal{M}(\bar{x}))^2\right)^{\frac{1}{2}} = \operatorname{dist}((x, \mu), \Sigma(0)) \tag{64}$$

and

$$\beta_0 \operatorname{dist}((x, \mu), \Sigma(0)) \leq \left\| \begin{bmatrix} \Psi(x, \mu) \\ \min\{-g(x), \mu\} \end{bmatrix} \right\|, \tag{65}$$

for some $\beta_0 > 0$.

Using this property, we conclude that the convergence rate of the primal sequence is at least two-step superlinear.

Proposition 5 *Under the hypotheses of Theorem 2, the sequence $\{x^k\}$, primal part of $\{w^k\}$, satisfies*

$$\lim_{k\to\infty} \frac{\left\|x^{k+1} - \bar{x}\right\|}{\max\left\{\left\|x^k - \bar{x}\right\|, \left\|x^{k-1} - \bar{x}\right\|\right\}} = 0.$$

Proof By Theorem 2, we have that $(x^k, \mu^k) \to (\bar{x}, \hat{\mu})$ and (63) holds. Note that shrinking $\varepsilon_{\bar{\mu}}$ in (37), we can obtain $(\bar{x}, \hat{\mu})$ in a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\mu})$ where (64) and (65) are valid.

Let $\eta_k = \max\left\{\left\|x^k - \bar{x}\right\|, \left\|x^{k-1} - \bar{x}\right\|\right\}$.

Since $0 = F(x^{k-1}) + M_{k-1}(x^k - x^{k-1}) + g'(x^{k-1})^\top \mu^k$ and $M_k \in \mathcal{L}_{k-1}$, we have

$$\begin{aligned} M_k\left(x^k - x^{k-1}\right) &= \Psi\left(x^k, \mu^k\right) - \Psi\left(x^{k-1}, \mu^k\right) \\ &= \Psi\left(x^k, \mu^k\right) + M_{k-1}\left(x^k - x^{k-1}\right). \end{aligned}$$

Hence, using (58), we conclude that

$$\left\|\Psi\left(x^k, \mu^k\right)\right\| = \left\|(M_k - M_{k-1})\left(x^k - x^{k-1}\right)\right\| = o(\eta_k).$$

For $i \notin \mathcal{I}$ and k large enough, by continuity, we have $g_i(x^k) < 0$ and

$$g_i\left(x^{k-1}\right) + \left\langle g_i'\left(x^{k-1}\right), x^k - x^{k-1}\right\rangle - \sigma_{k-1}\left(\mu^k - \mu^{k-1}\right) \leq \frac{1}{2} g_i(\bar{x}) < 0.$$

Hence, $\mu_i^k = 0$ by complementarity. Obtaining that $\min\left\{-g_i(x^k), \mu_i^k\right\} = 0$ for $i \notin \mathcal{I}$. Thus, there exists $L > 0$ independent of k such that

$$\begin{aligned}\left\|\min\left\{-g(x^k),\mu^k\right\}\right\| &= \left\|\min\left\{-g_{\mathcal{I}}(x^k),\mu^k_{\mathcal{I}}\right\}\right\|\\ &= \left\|\mu^k_{\mathcal{I}} - \Pi_{\mathbb{R}^{|\mathcal{I}|}_+}\left(\mu^k_{\mathcal{I}} + g_{\mathcal{I}}(x^k)\right)\right\|\\ &\leq \left\|g_{\mathcal{I}}(\bar{x}) - g_{\mathcal{I}}(x^k)\right\|\\ &\leq L\eta_k,\end{aligned}$$

where we use that $\mu^k_{\mathcal{I}} \in \mathbb{R}^{|\mathcal{I}|}_+$ and $g_{\mathcal{I}}(\bar{x}) = 0$.

By the SOC (13), we have the existence of $L_0 > 0$ such that

$$\operatorname{dist}(w^k, \Sigma(0)) \leq \frac{1}{\beta_0}\left\|\begin{bmatrix}\Psi(x^k,\mu^k)\\ \min\left\{-g(x^k),\mu^k\right\}\end{bmatrix}\right\| \leq L_0\eta_k,$$

for k large enough. Hence,

$$\frac{\left\|x^{k+1}-\bar{x}\right\|}{\eta_k} \leq \frac{\operatorname{dist}(w^{k+1},\Sigma(0))}{\eta_k} = \frac{\operatorname{dist}(w^{k+1},\Sigma(0))}{\operatorname{dist}(w^k,\Sigma(0))}\frac{\operatorname{dist}(w^k,\Sigma(0))}{\eta_k} \to 0,$$

where we use (63). □

As can be seen, in the previous result we only use the equality part in (7), without using the complementarity part, where the stabilization is done. The addition of this dual information creates a strong dependency between primal and dual iteration, given some difficulties in the study of the rate of convergence of the primal sequence.

6 Numerical example

We illustrate our convergence result with the following example.

$$\begin{array}{ll}\min & 8(x_1+2)^2 + x_2^2\\ \text{s.t.} & x_1^3 - x_1^2 - x_2^2 + 2 \leq 0,\\ & -x_1 - 3x_2 - 1 \leq 0,\\ & -x_1 + 3x_2 - 1 \leq 0.\end{array}$$

This problem has a unique solution $\bar{x} = (-1, 0)$ (the unique feasible point) with an associated Lagrange multipliers set given by

$$\mathcal{M}(\bar{x}) = \left\{\left(\alpha, \tfrac{5}{2}\alpha + 8, \tfrac{5}{2}\alpha + 8\right) \mid \alpha \geq 0\right\}.$$

It can be seen that MFCQ does not hold and that $\Psi'_x(\bar{x},\bar{\mu})$ is a symmetric and positive definite matrix for any $\bar{\mu} \in \left\{\left(\alpha, \tfrac{5}{2}\alpha + 8, \tfrac{5}{2}\alpha + 8\right) \mid 2 > \alpha \geq 0\right\}$.

Also, after some algebraics, it can be proved that if

$$x^k \in \mathcal{B} := \left\{x \in \mathbb{R}^2 \mid x_1 \leq 0,\ -\left(\tfrac{9}{2}x_1^2 - 3x_1\right) \leq x_2 \leq \tfrac{9}{2}x_1^2 - 3x_1\right\}$$

Table 1 Error on last 5 iterations

$\sigma\left(x^k, \mu^k\right)$	$\left\Vert x^k - \bar{x}\right\Vert$
1.6854e−01	2.6551e−02
3.1049e−02	4.1603e−03
2.0858e−03	2.2209e−04
3.2931e−05	2.0981e−06
2.4906e−08	1.7831e−09

with $x^k \neq \bar{x}$, then

$$\left\{y \in \mathbb{R}^2 \mid g(x^k) + g'(x^k)(y - x^k) \leq 0\right\} = \emptyset.$$

Since $\bar{x}$ is an interior point of $\mathcal{B}$, we have that the standard SQP method fails to solve this problem (unless $x^k = \bar{x}$ for some k).

We have written an `Octave` implementation of a quasi-Newton stabilized SQP method with BFGS update, using the built-in subroutine `qp` for solving subproblems (6). Experiments were performed choosing 100 random starting points $x_1^0 \in [-2, 0]$, $x_2^0 \in [-1, 1]$, $\mu_1^0 \in [0, 2]$ and $\mu_j^0 \in [8, 13]$, $j = 2, 3$. The stopping criteria was $\sigma(x^k, \mu^k) < 10^{-7}$.

The convergence of the primal-dual sequence was superlinear in all the cases. Surprisingly, the convergence of the primal sequence was also superlinear.

Table 1 shows the average values of $\sigma(x^k, \mu^k)$ and $\|x^k - \bar{x}\|$ for the last 5 iterations.

7 Final remarks

In this paper we have demonstrated that the sSQP method preserves their good convergence properties even when the exact higher order derivatives are replaced by quasi-Newton approximations. Note that as in the exact case, the method converges without any constraint qualification assumptions. To the best of our knowledge, no bounded deterioration property for degenerate constraints is available in the literature, thus, (52) seems to be the first of this kind. We also provided a unified framework to study quasi-Newton updates through its associated Bregman/Legendre function. The condition $\Psi'_x(\bar{x}, \bar{\mu}) \in \operatorname{int}(\operatorname{dom}\varphi)$ in Lemma 1 say that, for optimization problems, the Hessian of the Lagrangian function must be positive definite in order to guarantee the convergence of the method with the BFGS update. This is a strong condition usually not fulfilled in practice. To avoid this condition we have to change $\operatorname{dom}\varphi$ and therefore, its associated quasi-Newton update. It remains to investigate what are the quasi-Newton updates associated to those Bregman/Legendre functions present in the literature and to see which one provides a computationally plausible updating formula.

Acknowledgments This research was supported by FAPESP (Grant 2008/00062-8) and developed in the Department of Applied Mathematics, IMECC-UNICAMP, State University of Campinas, Campinas SP, Brazil. I would like to thank the anonymous referees for their valuable comments and a thorough reading of the manuscript. I also thank to Profesor J.M. Martínez for discussions about the BFGS update.

References

1. Bauschke, H.H., Borwein, J.M.: Legendre functions and the method of random Bregman projections. J. Convex Anal. **4**(1), 27–67 (1997)
2. Boggs, P.T., Tolle, J.W., Wang, P.: On the local convergence of quasi-Newton methods for constrained optimization. SIAM J. Control Optim. **20**(2), 161–171 (1982)
3. Bonnans, J.F.: Local analysis of Newton-type methods for variational inequalities and nonlinear programming. Appl. Math. Optim. **29**(2), 161–186 (1994)
4. Broyden, C.G.: A class of methods for solving nonlinear simultaneous equations. Math. Comp. **19**, 577–593 (1965)
5. Broyden, C.G., Dennis, J.E. Jr., Moré, J.J.: On the local and superlinear convergence of quasi-Newton methods. J. Inst. Math. Appl. **12**, 223–245 (1973)
6. Dennis, J.E. Jr., Schnabel, R.B.: Least change secant updates for quasi-Newton methods. SIAM Rev. **21**(4), 443–459 (1979)
7. Facchinei, F., Pang, J.S.: Finite-dimensional variational inequalities and complementarity problems. vol. I, Springer, New York (2003)
8. Fernández, D., Solodov, M.: Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems. Math. Program. **125**(1, Ser. A), 47–73 (2010)
9. Fischer, A.: Local behavior of an iterative framework for generalized equations with nonisolated solutions. Math. Program. **94**(1, Ser. A), 91–124 (2002)
10. Fletcher, R.: A new variational result for quasi-Newton formulae. SIAM J. Optim. **1**(1), 18–21 (1991)
11. Goebel, R., Rockafellar, R.T.: Local strong convexity and local Lipschitz continuity of the gradient of convex functions. J. Convex Anal. **15**(2), 263–270 (2008)
12. Grzegórski, S.M.: Orthogonal projections on convex sets for Newton-like methods. SIAM J. Numer. Anal. **22**(6), 1208–1219 (1985)
13. Guler, O., Gurtuna, F., Shevchenko, O.: Duality in quasi-Newton methods and new variational characterizations of the DFP and BFGS updates. Optim. Methods Softw. **24**(1), 45–62 (2009)
14. Hager, W.W.: Stabilized sequential quadratic programming. Comput. Optim. Appl. **12**(1–3), 253–273 (1999)
15. Hager, W.W., Gowda, M.S.: Stability in the presence of degeneracy and error estimation. Math. Program. **85**(1, Ser. A), 181–192 (1999)
16. Izmailov, A.F.: Solution sensitivity for Karush-Kuhn-Tucker systems with non-unique Lagrange multipliers. Optimization **59**(5–6), 747–775 (2010)
17. Josephy, N.: Newton's Method for Generalized Equations. Technical Summary Report 1965, Mathematics Research Center, University of Wisconsin, Madison, Wisconsin (1979)
18. Josephy, N.: Quasi–Newton Methods for Generalized Equations. Technical Summary Report 1966, Mathematics Research Center, University of Wisconsin, Madison, Wisconsin (1979)
19. Li, D.H., Qi, L.: A stabilized SQP Method Via Linear Equations. Technical Report Applied mathematics technical report AMR00/5, The University of New South Wales (2000)
20. Powell, M.J.D.: A new algorithm for unconstrained optimization. In: Nonlinear Programming (Proceedings of Symposium University of Wisconsin, Madison, Wisconsin, 1970), pp. 31–65. Academic Press, New York (1970)
21. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis. Springer, Berlin (1998)
22. Wright, S.J.: Superlinear convergence of a stabilized SQP method to a degenerate solution. Comput. Optim. Appl. **11**(3), 253–275 (1998)
23. Wright, S.J.: Modifying SQP for degenerate problems. SIAM J. Optim. **13**(2), 470–497 (2002)
24. Wright, S.J.: Constraint identification and algorithm stabilization for degenerate nonlinear programs. Math. Program. **95**(1, Ser. B), 137–160 (2003)
25. Wright, S.J.: An algorithm for degenerate nonlinear programming with rapid local convergence. SIAM J. Optim. **15**(3), 673–696 (2005)