Mathematical Programming manuscript No.
(will be inserted by the editor)

Damián Fernández · Mikhail Solodov

Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems

Received: November 2007 / Revised: September 2008

Abstract The stabilized version of the sequential quadratic programming algorithm (sSQP) had been developed in order to achieve fast convergence despite possible degeneracy of constraints of optimization problems, when the Lagrange multipliers associated to a solution are not unique. Superlinear convergence of sSQP had been previously established under the strong second-order sufficient condition for optimality (without any constraint qualification assumptions). We prove a stronger superlinear convergence result than the above, assuming the usual second-order sufficient condition only. In addition, our analysis is carried out in the more general setting of variational problems, for which we introduce a natural extension of sSQP techniques. In the process, we also obtain a new error bound for Karush-Kuhn-Tucker systems for variational problems that holds under an appropriate second-order condition.

Keywords Stabilized sequential quadratic programming · Karush-Kuhn-Tucker system · variational inequality · Newton methods · superlinear convergence · error bound

Mathematics Subject Classification (2000) 90C30 · 65K05

The second author is supported in part by CNPq Grants 301508/2005-4, 471267/2007-4, by PRONEX–Optimization, and by FAPERJ Grant E-26/151.942/2004.

Damián Fernández
IMPA, Estrada Dona Castorina 110, Rio de Janeiro, RJ 22460-320, Brazil. Currently at IMECC-UNICAMP, State University of Campinas CP 6065, 13081-970 Campinas SP, Brazil. E-mail: dfernan@impa.br

Mikhail Solodov
IMPA – Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil. E-mail: solodov@impa.br

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:

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

where

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

and $\mathcal{T}(x; D)$ is the (standard) tangent cone to the set D at the point $x \in D$. For most results of the paper, we shall assume that on the set of interest,

$$F \text{ is once and } g \text{ is twice continuously differentiable.} \tag{2}$$

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

$$F(x) = f'(x), \quad x \in \mathbb{R}^n, \tag{3}$$

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

$$\min\ f(x) \quad \text{subject to} \quad x \in D. \tag{4}$$

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

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

In the absence of convexity, however, the meaningful form of a local variational condition (in particular, the one consistent with optimality conditions for (4)) is given by (1).

To motivate our development consider, for the moment, the optimization problem (4). Iterations of the fundamental sequential quadratic programming method (SQP, e.g., [1]) for (4) consist of solving subproblems of the form

$$\begin{array}{ll} \min\limits_{y \in \mathbb{R}^n} & \langle f'(x^k), y - x^k\rangle + \frac{1}{2}\langle L''_{xx}(x^k, \mu^k)(y - x^k), y - x^k\rangle \\ \text{s.t.} & g(x^k) + g'(x^k)(y - x^k) \leq 0, \end{array}$$

where

$$L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}, \quad L(x, \mu) = f(x) + \langle \mu, g(x)\rangle,$$

is the Lagrangian of (4), and $(x^k, \mu^k) \in \mathbb{R}^n \times \mathbb{R}^m_+$ is the current primal-dual iterate. Let $\bar{x} \in \mathbb{R}^n$ be a local solution of (4), and let $\mathcal{M}(\bar{x})$ be the set of Lagrange multipliers associated to $\bar{x}$. The minimal conditions [2] which guarantee that the SQP method outlined above is locally well-defined and superlinearly convergent are the existence and uniqueness of the Lagrange multiplier $\bar{\mu}$ associated to $\bar{x}$ (also known as the strict Mangasarian-Fromovitz constraint qualification) and the second-order sufficient condition (SOSC)

$$\langle L''_{xx}(\bar{x}, \bar{\mu})d,\ d\rangle > 0 \quad \forall\, d \in \mathcal{C}(\bar{x}) \setminus \{0\}, \tag{5}$$

where

$$\begin{aligned}\mathcal{C}(\bar{x}) &= \{d \in \mathbb{R}^n \mid \langle f'(\bar{x}), d\rangle = 0,\ \langle g_i'(\bar{x}), d\rangle \le 0\ \forall i \in \mathcal{I}(\bar{x})\} \\ &= \{d \in \mathbb{R}^n \mid \langle g_i'(\bar{x}), d\rangle = 0\ \forall i \in \mathcal{I}_+(\bar{x},\bar{\mu}),\ \langle g_i'(\bar{x}), d\rangle \le 0\ \forall i \in \mathcal{I}_0(\bar{x},\bar{\mu})\},\end{aligned} \tag{6}$$

is the critical cone of (4) at $\bar{x}$, with

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

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

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

being the set of strongly and weakly active constraints, respectively.

We emphasize that convergence of SQP requires certain regularity of constraints (specifically, the strict Mangasarian-Fromovitz constraint qualification).

To deal with the case when constraint qualifications may be violated (and multiplier associated to the primal solution of the optimization problem (4) may not be unique), a stabilized version of SQP (sSQP) has been introduced in [17]. This method can be stated [14] in the form of solving subproblems

$$\begin{aligned}\min_{(y,\lambda)\in\mathbb{R}^n\times\mathbb{R}^m}\ & \langle f'(x^k), y - x^k\rangle + \tfrac{1}{2}\langle L''_{xx}(x^k,\mu^k)(y - x^k), y - x^k\rangle + \tfrac{\sigma(x^k,\mu^k)}{2}\|\lambda\|^2 \\ \text{s.t.}\ & g(x^k) + g'(x^k)(y - x^k) - \sigma(x^k,\mu^k)(\lambda - \mu^k) \le 0,\end{aligned} \tag{7}$$

where $(x^k, \mu^k) \in \mathbb{R}^n \times \mathbb{R}^m_+$ is again the current primal-dual iterate, while the dual stabilization parameter $\sigma(x^k, \mu^k) > 0$ is some computable quantity measuring violation of optimality conditions for (4) by the point (x^k, μ^k). As is easy to see, unlike in SQP, the subproblems (7) are always feasible regardless of constraint qualifications. In [17], superlinear convergence of sSQP has been established under the Mangasarian-Fromovitz constraint qualification (MFCQ, which is equivalent to the nonemptiness and compactness of the multiplier set $\mathcal{M}(\bar{x})$), SOSC (5) for all $\bar{\mu} \in \mathcal{M}(\bar{x})$, and the assumption that the initial dual iterate μ^0 is close enough to a multiplier $\bar{\mu}$ such that $\bar{\mu}_{\mathcal{I}(\bar{x})} > 0$ (in particular, strict complementarity is assumed). In [18, 19], superlinear convergence of sSQP has been shown without strict complementarity, under MFCQ and the strong second-order sufficient condition (SSOSC)

$$\langle L''_{xx}(\bar{x},\bar{\mu})d,\ d\rangle > 0 \quad \forall\, d \in \mathcal{C}^+(\bar{x},\bar{\mu}) \setminus \{0\}, \tag{8}$$

assumed for all $\bar{\mu} \in \mathcal{M}(\bar{x})$, where

$$\mathcal{C}^+(\bar{x},\bar{\mu}) = \{d \in \mathbb{R}^n \mid \langle g_i'(\bar{x}), d\rangle = 0\ \forall i \in \mathcal{I}_+(\bar{x},\bar{\mu})\}.$$

Superlinear convergence had also been shown under the sole assumption of SSOSC (8) for *some* $\bar{\mu} \in \mathcal{M}(\bar{x})$, provided that μ^0 is close enough to such $\bar{\mu}$ [7]; see also [5]. In fact, it was posed as an open question in [5, p. 117] whether or not some condition weaker than SSOSC can be used to prove sSQP convergence when no constraint qualifications are assumed. In this paper, we answer this question is the affirmative. We show that if the starting point is close to $(\bar{x}, \bar{\mu})$ satisfying SOSC (5), then the sSQP method is well-defined and converges superlinearly. Moreover, our development is carried out for the variational setting, in which sSQP for optimization is a special case.

As other local algorithms for optimization that had been proven to be superlinearly convergent under SOSC only, we mention [10] and [20]. The method of [20], in particular, takes an sSQP-like step for an equality-constrained problem, performing separately identification of active constraints. The approach of [10] is also based on active constraints identification and a reduction to local equality-constrained phase, but this local phase is not related to sSQP.

Let us now go back to the variational problem (1). In this context, a natural extension of sSQP is the following iterative procedure, which is obtained from the variational formulation of optimality conditions for (7). To this end, define

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

Let $(x^k, \mu^k) \in \mathbb{R}^n \times \mathbb{R}^m_+$ be the current primal-dual approximation to a solution of (1), and define

$$\Phi_k : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n \times \mathbb{R}^m, \quad \Phi_k(y, \lambda) = \begin{bmatrix} F(x^k) + \Psi'_x(x^k, \mu^k)(y - x^k) \\ \sigma(x^k, \mu^k)\lambda \end{bmatrix},$$

and

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

where $\sigma(x^k, \mu^k) > 0$ is the dual stabilization parameter.

Consider *affine* variational subproblems of the form

$$\text{Find } (y, \lambda) \in \Delta_k \quad \text{s.t.} \quad \langle \Phi_k(y, \lambda), (z, \nu) - (y, \lambda) \rangle \geq 0 \quad \forall\, (z, \nu) \in \Delta_k. \quad (9)$$

As can be easily seen, in the optimization case (3) the variational subproblem (9) is precisely the first-order (primal) necessary optimality condition for the sSQP subproblem (7). Thus this framework contains sSQP for optimization as a special case. Note that the framework makes good sense also in the variational setting, as solving the fully nonlinear problem (1) is replaced by solving a sequence of fully affine subproblems (9) (the mapping Φ_k is affine and the set Δ_k is polyhedral). As in sSQP, the feasible set in (9) is always nonempty. We shall prove that under a suitable second-order condition, the method outlined above is locally well-defined and converges superlinearly to a solution of the Karush-Kuhn-Tucker (KKT) system for (1), which is

$$\begin{gathered} 0 = \Psi(x, \mu) = F(x) + g'(x)^\top \mu, \\ 0 \leq \mu \perp g(x) \leq 0, \end{gathered} \quad (10)$$

where $\mu \perp g(x)$ means that $\langle \mu, g(x) \rangle = 0$. We make the standing assumption that the KKT system (10) has a primal-dual solution (in fact, if the constraints are degenerate, there are many dual solutions associated to the same primal solution). The setting of existence of multipliers, while not assuming any specific constraint qualifications that are sufficient for this, is common when dealing with degenerate problems, e.g., [18,5,10,20,12,11]. In optimization, the combination of SOSC with the existence of multipliers is related to the Guignard constraint qualification (GCQ) [6]. The latter amounts to the existence of a multiplier for every objective function for which the point under consideration is a local minimizer (the feasible set is fixed). However, there is no equivalence, as for a given problem a multiplier may exist when GCQ does not hold (i.e., there may exist another objective function that has the point in question as its minimizer on the given set, but for which there are no multipliers, e.g., Example 1 below).

The rest of the paper is organized as follows. In Section 2, we recall the general iterative framework of Fischer [5] that will be used to prove superlinear convergence of our algorithm. We note that in [5], the general framework has been applied to the method of proximally-regularized linearizations of monotone mixed complementarity problems (MCP), and to sSQP for KKT systems arising from optimization. Compared to the first item, our iterations are different (regularization is in the dual space only), and we do not assume any monotonicity or convexity. Compared to the second item, we cover KKT systems that include variational problems, and prove superlinear convergence under SOSC instead of SSOSC employed in [5]. In Section 3, we prove that subproblems (9) are locally solvable if $\sigma(\cdot)$ provides a local error bound [15,9] on the distance to the solution set of the KKT system (10). In Section 4, among other things, we derive a suitable error bound. The results of Sections 3 and 4 show that the assumptions of [5], stated in Section 2, are verified, which implies superlinear convergence of the method given by (9). Convergence results are formally stated in Section 5.

Some words about our notation. We use $\langle\cdot,\cdot\rangle$ to denote the Euclidean inner product, $\|\cdot\|$ the associated norm, and B the unit ball (the space 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 $\mathcal{I}$. We use I to denote the identity matrix (the dimension is always clear from the context). 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 $\operatorname{dist}(z,S) = \inf_{s\in S}\|z-s\|$. Then $\Pi_S(z) = \{s \in S \mid \operatorname{dist}(z,S) = \|z-s\|\}$ is the set of all points in S that have minimal distance to z. For a cone $K \subset \mathbb{R}^l$, its (positive) dual is $K^* = \{u \in \mathbb{R}^l \mid \langle u,v\rangle \geq 0\, \forall v \in K\}$. Recall that a matrix $M \in \mathbb{R}^{l\times l}$ is said to be copositive on a cone $K \subset \mathbb{R}^l$ if $\langle Mv,v\rangle \geq 0$ for all $v \in K$, and strictly copositive if this inequality is strict for all $v \in K \setminus \{0\}$.

2 Fischer's general iterative framework

Let $G : \mathbb{R}^q \to \mathbb{R}^l$ be a continuous map, $\mathcal{T}$ be a closed set-valued map from $\mathbb{R}^q$ to $\mathbb{R}^l$, and consider the generalized equation (GE)

$$\text{Find } w \in \mathbb{R}^q \quad \text{such that} \quad 0 \in G(w) + \mathcal{T}(w). \tag{11}$$

Denote by Σ_* the (nonempty) solution set of (11).

Consider a class of methods that, given $s \in \mathbb{R}^q$, generate the next iterate by solving a subproblem of the form

$$\text{Find } w \in \mathbb{R}^q \quad \text{such that} \quad 0 \in \mathcal{A}(w,s) + \mathcal{T}(w), \tag{12}$$

where $\mathcal{A}(\cdot, s)$ is an approximation of $G(\cdot)$ around s (in [5], $\mathcal{A}(\cdot, s)$ can be set-valued; in our application it will be point-to-point). Note that when $\mathcal{T}(\cdot)$ is the normal cone associated to a closed convex set and $\mathcal{A}(\cdot, s)$ is the standard linearization of $G(\cdot)$ at the point s, then (12) reduces to an iteration of the classical Josephy-Newton method [13].

Denote by

$$Z(s) = \{w \in \mathbb{R}^q \mid 0 \in \mathcal{A}(w,s) + \mathcal{T}(w)\}$$

the solution set of (12). In local convergence analyses it is standard to assume that the distance between two consecutive iterates is not too large (without very strong assumptions, subproblems (12) may have other solutions that are far from a given solution of (11) that is being approximated; those solutions are irrelevant for the local analysis and should be excluded). To this end, define

$$Z_c(s) = \{w \in Z(s) \mid \|w - s\| \le c \operatorname{dist}(s, \Sigma_*)\},$$

where $c \in [1, +\infty)$ is arbitrary but fixed, and consider the iterative scheme

$$w^0 \in \mathbb{R}^q, \quad w^{k+1} \in Z_c(w^k),\ k = 0, 1, \dots \tag{13}$$

Then the following holds (see [5, Theorem 1]).

Theorem 1 *Let Σ_* be the (nonempty) solution set of* (11)*. Assume the following three properties:*

1. **(Upper Lipschitz-continuity of the solution set of GE)**
There exist numbers $\varepsilon_1, \gamma, t > 0$ such that, with $Q = \Sigma_0 + \varepsilon_1 B$, it holds that

$$\Sigma(p) \cap Q \subseteq \Sigma_* + t\|p\|B \quad \forall p \in \gamma B,$$

where $\Sigma_0 \neq \emptyset$ is a closed subset of Σ_*, and

$$\Sigma(p) = \{w \in \mathbb{R}^q \mid 0 \in G(w) + \mathcal{T}(w) + p\}.$$

2. **(Precision of approximation of $G(\cdot)$ by $\mathcal{A}(\cdot, s)$)**
There exists $\varepsilon_2 > 0$ such that

$$\sup\{\|R(w,s)\| : w \in s + c\, dist(s, \Sigma_*)B\} \le o\,(dist(s, \Sigma_*)) \quad \forall s \in \Sigma_0 + \varepsilon_2 B,$$

where $R(w,s) = G(w) - \mathcal{A}(w,s)$.

3. **(Solvability of subproblems)**
There exists $\varepsilon_3 > 0$ such that $Z_c(s) \neq \emptyset$ for all $s \in \Sigma_0 + \varepsilon_3 B$.

Then there exists $\varepsilon > 0$ such that for any $w^0 \in \Sigma_0 + \varepsilon B$, the iterates generated according to (13) *are well defined and converge superlinearly to some $w^* \in \Sigma_*$. Furthermore, the convergence is of order β if the function $o(\cdot)$ in Item 2 satisfies*

$$o(t) \le c_0 t^\beta \quad \forall t \in [0, 1],$$

for some $c_0 > 0$ and $\beta > 1$ (in particular, convergence is quadratic if $\beta = 2$).

To relate the proposed iterative scheme (9) to the framework above, define

$$G(x,\mu) = \begin{bmatrix} \Psi(x,\mu) \\ -g(x) \end{bmatrix}, \quad \mathcal{T}(x,\mu) = \begin{bmatrix} 0 \\ \mathcal{N}(\mu) \end{bmatrix}, \tag{14}$$

where

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

is the normal cone to the nonnegative orthant $\mathbb{R}^m_+$ at $\mu \in \mathbb{R}^m$. Let $w = (x,\mu) \in \mathbb{R}^n \times \mathbb{R}^m = \mathbb{R}^q$. Then the KKT system (10) for problem (1) is equivalent to solving the generalized equation (11) with G and $\mathcal{T}$ given by (14).

Since subproblem (9) of our method is an affine VI, it is equivalent to solving the KKT system of finding $(y,\lambda,\nu) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ such that

$$\begin{aligned} &0 = F(x^k) + \Psi'_x(x^k,\mu^k)(y - x^k) + g'(x^k)^\top \nu, \\ &0 = \sigma(x^k,\mu^k)\lambda - \sigma(x^k,\mu^k)\nu, \\ &0 \le \nu \perp [g(x^k) + g'(x^k)(y - x^k) - \sigma(x^k,\mu^k)(\lambda - \mu^k)] \le 0. \end{aligned}$$

Noting that $\lambda = \nu$ by the second relation, the above is then equivalent to finding $(y,\lambda) \in \mathbb{R}^n \times \mathbb{R}^m$ such that

$$\begin{aligned} 0 &= F(x^k) + \Psi'_x(x^k,\mu^k)(y - x^k) + g'(x^k)^\top \lambda \\ &= \Psi(x^k,\mu^k) + \Psi'_x(x^k,\mu^k)(y - x^k) + g'(x^k)^\top(\lambda - \mu^k), \\ 0 &\le \lambda \perp [g(x^k) + g'(x^k)(y - x^k) - \sigma(x^k,\mu^k)(\lambda - \mu^k)] \le 0. \end{aligned} \tag{15}$$

Letting now $w = (x,\mu) \in \mathbb{R}^n \times \mathbb{R}^m = \mathbb{R}^q$, $s \in \mathbb{R}^q$,

$$\mathcal{A}(w,s) = G(s) + \Big(G'(s) + \Lambda(s)\Big)(w - s), \quad \Lambda(s) = \begin{bmatrix} 0 & 0 \\ 0 & \sigma(s)I \end{bmatrix},$$

where G is defined in (14), we obtain that solving (15) (and thus (9)) is equivalent to solving GE subproblems (12).

The rest of the paper proves that problem (11) and subproblem (12), corresponding to problem (10) and subproblem (15), respectively, satisfy the assumptions of Theorem 1. The hard part is to prove, under a (weak) second-order condition only, the upper Lipschitz-continuity of the solution set of the KKT system (10) and, especially, solvability of subproblems (15) (Assumptions 1 and 3 of Theorem 1).

Assumption 2 is easily seen to be satisfied, because

$$\begin{aligned} \|R(w,s)\| &= \|G(w) - \mathcal{A}(w,s)\| \\ &= \|G(w) - G(s) - (G'(s) + \Lambda(s))(w - s)\| \\ &\le \left\| \int_0^1 [G'(s + t(w - s)) - G'(s)](w - s)dt \right\| + \|\Lambda(s)(w - s)\| \\ &\le \left(\int_0^1 \|G'(s + t(w - s)) - G'(s)\|dt + \sigma(s) \right) \|w - s\|, \end{aligned}$$

which implies that

$$\|R(w,s)\| \leq o\left(\operatorname{dist}(s,\Sigma_*)\right)$$

when $w \in s + c\operatorname{dist}(s,\Sigma_*)B$ and

$$\sigma(s) \leq L_1 \operatorname{dist}(s,\Sigma_*)$$

for some $L_1 > 0$. The latter inequality holds for any reasonable residual $\sigma(\cdot)$ of the KKT system (by the Lipschitz-continuity); this will be made evident in Section 4.

Note also that if, in addition, the derivatives F' and g'' are Lipschitz-continuous, then so is G', and we have that

$$\|R(w,s)\| \leq L_2 \operatorname{dist}(s,\Sigma_*)^2. \tag{16}$$

3 Solvability of subproblems

We next prove that KKT subproblems of the form (15) (which are equivalent to affine variational subproblems (9)) are locally solvable if a certain second-order condition holds, and if the dual regularization parameters $\sigma(x^k,\mu^k)$ are of the order of the distance to the solution set of the KKT system (10) for problem (1). A specific computable way of choosing such parameters will be discussed in Section 4.

Let $\bar{x}$ be a solution of the variational problem (1), and let

$$\mathcal{M}(\bar{x}) = \{\mu \in \mathbb{R}^m \mid (\bar{x},\mu) \text{ solves } (10)\}$$

be the associated (nonempty) set of Lagrange multipliers. Let the sets of active, strongly active and weakly active constraints ($\mathcal{I} = \mathcal{I}(\bar{x})$, $\mathcal{I}_+(\bar{x},\mu)$ and $\mathcal{I}_0(\bar{x},\mu)$, respectively) be defined as in Section 1.

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 (10) if

$$\langle \Psi_x'(\bar{x},\bar{\mu})u,u\rangle > 0 \quad \forall\, u \in \mathcal{C}(\bar{x};D,F)\backslash\{0\}, \tag{17}$$

where

$$\begin{aligned}\mathcal{C}(\bar{x};D,F) &= \{u \in \mathbb{R}^n \mid \langle F(\bar{x}),u\rangle = 0,\ \langle g_i'(\bar{x}),u\rangle \leq 0\ \forall\, i \in \mathcal{I}(\bar{x})\} \\ &= \left\{ u \in \mathbb{R}^n \;\middle|\; \begin{array}{ll} \langle g_i'(\bar{x}),u\rangle = 0 & \forall\, i \in \mathcal{I}_+(\bar{x},\mu) \\ \langle g_i'(\bar{x}),u\rangle \leq 0 & \forall\, i \in \mathcal{I}_0(\bar{x},\mu) \end{array} \right\}.\end{aligned} \tag{18}$$

(As is well known, the second equality above does not depend on the choice of $\mu \in \mathcal{M}(\bar{x})$.) In the case of the optimization problem (4), $\mathcal{C}(\bar{x};D,F)$ is the standard critical cone (6) at $\bar{x}$, and (17) is the standard second-order condition (5) which is sufficient for optimality of the point $\bar{x}$.

As already mentioned, we assume also that the function $\sigma(\cdot)$ satisfies the error bound property. As Lemma 2 in Section 4 shows that under SOC (17) (in fact, under the more general SOC (51)) the primal part $\bar{x}$ of the solution is locally unique, we can write our error bound in the following form: there exist a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\mu})$ and constants $\beta_2 \geq \beta_1 > 0$ such that for all $(x, \mu) \in \mathcal{V}$ it holds that

$$\beta_1 \Big(\|x - \bar{x}\| + \mathrm{dist}(\mu, \mathcal{M}(\bar{x})) \Big) \leq \sigma(x, \mu) \leq \beta_2 \Big(\|x - \bar{x}\| + \mathrm{dist}(\mu, \mathcal{M}(\bar{x})) \Big). \quad (19)$$

More details on a computable choice of $\sigma(\cdot)$ will be given in Section 4.

We start with extending SOC (17) from the copositivity property of the matrix in a primal cone to uniform positivity, in a neighbourhood of the point $(\bar{x}, \bar{\mu})$, of a certain function in a certain parametric primal-dual cone.

Proposition 1 *Let F and g satisfy the smoothness assumptions* (2)*. Suppose that SOC* (17) *holds at $(\bar{x}, \bar{\mu})$ and that σ satisfies the right-most inequality in* (19)*. Then there exist a constant $\gamma_1 > 0$ and a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\mu})$ such that for all $(x, \mu) \in \mathcal{V}$ it holds that*

$$\langle \Psi'_x(x, \mu)u, u \rangle + \sigma(x, \mu)\|v\|^2 \geq \gamma_1 \Big(\|u\|^2 + \sigma(x, \mu)\|v\|^2 \Big) \quad \forall (u, v) \in K(x, \mu), \quad (20)$$

where

$$K(x, \mu) = \left\{ (u, v) \in \mathbb{R}^n \times \mathbb{R}^{|\mathcal{I}|} \;\middle|\; \begin{array}{ll} \langle g'_i(x), u \rangle = \sigma(x, \mu)v_i, & i \in \mathcal{I}_+(\bar{x}, \bar{\mu}) \\ \langle g'_i(x), u \rangle \leq \sigma(x, \mu)v_i, & i \in \mathcal{I}_0(\bar{x}, \bar{\mu}) \end{array} \right\}. \quad (21)$$

Proof Suppose the contrary, i.e., that there exist $\{(x^k, \mu^k)\} \to (\bar{x}, \bar{\mu})$ and $(u^k, v^k) \in K(x^k, \mu^k)$ such that

$$\langle \Psi'_x(x^k, \mu^k)u^k, u^k \rangle + \sigma_k \|v^k\|^2 < \frac{1}{k}(\|u^k\|^2 + \sigma_k\|v^k\|^2), \quad (22)$$

where $\sigma_k = \sigma(x^k, \mu^k)$. Evidently, (22) subsumes that $(u^k, v^k) \neq 0$. Let $\eta_k = \|(u^k, \sqrt{\sigma_k} v^k)\| > 0$. Passing onto a subsequence, if necessary, we can assume that

$$\frac{1}{\eta_k} \begin{bmatrix} u^k \\ \sqrt{\sigma_k} v^k \end{bmatrix} \to \begin{bmatrix} \bar{u} \\ \bar{w} \end{bmatrix} \neq 0. \quad (23)$$

Observe that since $\sigma_k \to 0$ by the right-most inequality in (19), while $\sqrt{\sigma_k} v^k / \eta_k$ is bounded, it holds that

$$\sigma_k \frac{v^k}{\eta_k} = \sqrt{\sigma_k} \frac{\sqrt{\sigma_k} v^k}{\eta_k} \to 0. \quad (24)$$

Since $K(x^k, \mu^k)$ is a cone, we have that $(u^k/\eta_k, v^k/\eta_k) \in K(x^k, \mu^k)$. Dividing now relations in (21) by η_k and passing onto the limit, taking into account (24) we obtain that

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

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

On the other hand, dividing (22) by η_k^2 and taking limits, we have that

$$\langle \Psi'_x(\bar{x}, \bar{\mu})\bar{u}, \bar{u} \rangle + \|\bar{w}\|^2 \leq 0. \quad (25)$$

This shows that $\langle \Psi'_x(\bar{x}, \bar{\mu})\bar{u}, \bar{u} \rangle \leq 0$ for $\bar{u} \in \mathcal{C}(\bar{x}; D, F)$. Hence, $\bar{u} = 0$. Now from (25) we have that $\bar{w} = 0$ also, in contradiction with (23). □

Corollary 1 *Let F and g satisfy the smoothness assumptions* (2). *Suppose that SOC (17) holds at $(\bar{x}, \bar{\mu})$ and that σ satisfies the right-most inequality in* (19). *Then there exists a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\mu})$ such that the matrix*

$$\begin{bmatrix} \Psi'_x(x,\mu) & g'_{\mathcal{I}}(x)^\top \\ -g'_{\mathcal{I}}(x) & \sigma(x,\mu)I \end{bmatrix} \tag{26}$$

is nonsingular for all $(x, \mu) \in \mathcal{V}$ such that $\sigma(x, \mu) > 0$.

Proof By Proposition 1, there exists a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\mu})$ such that (20) holds. Let $(x, \mu) \in \mathcal{V}$, $\sigma(x, \mu) > 0$, and suppose that (u, v) is a vector in the kernel of the matrix given in (26), i.e.,

$$0 = \Psi'_x(x,\mu)u + g'_{\mathcal{I}}(x)^\top v, \tag{27}$$

$$0 = -g'_{\mathcal{I}}(x)u + \sigma(x,\mu)v. \tag{28}$$

By (28) we have that $\langle g'_i(x), u\rangle = \sigma(x,\mu)v_i$ for all $i \in \mathcal{I}$. This shows that $(u, v) \in K(x, \mu)$ defined in (21). Also, multiplying (28) by v we have

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

Multiplying by u both sides in (27), we then obtain that

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

Then, by (20), we have that $0 \geq \gamma_1(\|u\|^2 + \sigma(x,\mu)\|v\|^2)$. Hence, $u = 0$ and $v = 0$, implying that the matrix in (26) is nonsingular. □

Our proof of existence of solutions of subproblems is done in two steps. We start with showing that a certain part of KKT subproblem (15) has a solution. We shall make use of the existence result in [4, Theorem 2.5.10]. More specifically, we shall need a consequence of [4, Theorem 2.5.10], which we state as follows.

Theorem 2 *Let K be a closed convex cone in $\mathbb{R}^l$ and $M \in \mathbb{R}^{l\times l}$. Suppose that $d = 0$ is the unique solution of the generalized complementarity problem*

$$K \ni d \perp Md \in K^*, \tag{29}$$

and that M is copositive on K.

Then for all $q \in \mathbb{R}^l$, the generalized complementarity problem of finding $d \in \mathbb{R}^l$ such that

$$K \ni d \perp Md + q \in K^*$$

has a nonempty compact solution set.

Clearly, if M is strictly copositive on K then (29) has the origin as the unique solution, and all the assumptions of Theorem 2 hold.

Proposition 2 *Let F and g satisfy the smoothness assumptions* (2)*. Suppose that SOC* (17) *holds at* $(\bar{x}, \bar{\mu})$ *and that* σ *satisfies the right-most inequality in* (19)*. Then there exists a neighborhood* $\mathcal{V}$ *of* $(\bar{x}, \bar{\mu})$ *such that for all* $(x, \mu) \in \mathcal{V}$ *with* $\sigma(x, \mu) > 0$*, the mixed complementarity problem of finding* $(y, \lambda_{\mathcal{I}}) \in \mathbb{R}^n \times \mathbb{R}^{|\mathcal{I}|}$ *such that*

$$\begin{aligned} &0 = F(x) + \Psi'_x(x, \mu)(y - x) + g'_{\mathcal{I}}(x)^\top \lambda_{\mathcal{I}}, \\ &0 = g_i(x) + \langle g'_i(x), y - x\rangle - \sigma(x, \mu)(\lambda_i - \mu_i), \ i \in \mathcal{I}_+(\bar{x}, \bar{\mu}), \\ &0 \le \lambda_i \perp g_i(x) + \langle g'_i(x), y - x\rangle - \sigma(x, \mu)(\lambda_i - \mu_i) \le 0, \ i \in \mathcal{I}_0(\bar{x}, \bar{\mu}), \end{aligned} \tag{30}$$

has a nonempty compact solution set.

Proof Define

$$M = \begin{bmatrix} \Psi'_x(x, \mu) & 0 \\ 0 & \sigma(x, \mu)I \end{bmatrix}, \quad q = \begin{bmatrix} F(x) - \Psi'_x(x, \mu)x \\ 0 \end{bmatrix},$$

$$b_i = g_i(x) - \langle g'_i(x), x\rangle + \sigma(x, \mu)\mu_i, \quad i \in \mathcal{I},$$

and the $|\mathcal{I}| \times (n + |\mathcal{I}|)$ matrix A with rows given by

$$a^i = \begin{bmatrix} g'_i(x) \\ -\sigma(x, \mu)e^i \end{bmatrix},$$

where $e^i \in \mathbb{R}^{|\mathcal{I}|}$ is the i-th vector of the canonical basis. With this notation, it can be seen that (30) is equivalent to solving the following affine variational inequality:

$$\text{Find } \bar{z} \in Q \quad \text{s.t.} \quad \langle M\bar{z} + q, z - \bar{z}\rangle \ge 0 \quad \forall\, z \in Q, \tag{31}$$

where

$$Q = \{z \in \mathbb{R}^n \times \mathbb{R}^{|\mathcal{I}|} \mid A_{\mathcal{I}_+} z + b_{\mathcal{I}_+} = 0, A_{\mathcal{I}_0} z + b_{\mathcal{I}_0} \le 0\},$$

$\mathcal{I}_+ = \mathcal{I}_+(\bar{x}, \bar{\mu})$, $\mathcal{I}_0 = \mathcal{I}_0(\bar{x}, \bar{\mu})$.

Let $(\tilde{u}, \tilde{v}_{\mathcal{I}})$ be the unique solution of the linear system

$$\begin{bmatrix} \Psi'_x(x, \mu) & g'_{\mathcal{I}}(x)^\top \\ -g'_{\mathcal{I}}(x) & \sigma(x, \mu)I \end{bmatrix} \begin{bmatrix} u \\ v_{\mathcal{I}} \end{bmatrix} = \begin{bmatrix} -F(x) - g'_{\mathcal{I}}(x)^\top \mu_{\mathcal{I}} \\ g_{\mathcal{I}}(x) \end{bmatrix},$$

which exists due Corollary 1. Define $\tilde{z} = (x + \tilde{u}, \mu_{\mathcal{I}} + \tilde{v}_{\mathcal{I}})$. For each $i \in \mathcal{I}$ we then have that

$$\begin{aligned} \langle a^i, \tilde{z}\rangle &= \langle g'_i(x), x\rangle - \sigma(x, \mu)\mu_i + \langle g'_i(x), \tilde{u}\rangle - \sigma(x, \mu)\tilde{v}_i \\ &= \langle g'_i(x), x\rangle - \sigma(x, \mu)\mu_i - g_i(x) \\ &= -b_i. \end{aligned}$$

In particular, $\tilde{z} \in Q$ and all the constraints defining the polyhedral set Q are active at $\tilde{z}$. Note that, in the adopted notation, the cone $K = K(x, \mu)$ defined in (21) can be written as

$$K = \{d \in \mathbb{R}^n \times \mathbb{R}^{|\mathcal{I}|} \mid A_{\mathcal{I}_+} d = 0, A_{\mathcal{I}_0} d \le 0\}.$$

Hence,

$$Q = \{z \in \mathbb{R}^n \times \mathbb{R}^{|\mathcal{I}|} \mid A_{\mathcal{I}_+}(z - \tilde{z}) = 0, A_{\mathcal{I}_0}(z - \tilde{z}) \leq 0\} = \tilde{z} + K.$$

We can then write (31) in the following form:

$$\text{Find } \bar{d} \in K \quad \text{s.t.} \quad \langle M\bar{d} + M\tilde{z} + q, d - \bar{d}\rangle \geq 0 \quad \forall\, d \in K,$$

which is the generalized complementarity problem

$$K \ni \bar{d} \perp M\bar{d} + M\tilde{z} + q \in K^*. \tag{32}$$

Furthermore, the copositivity property (20)-(21) with $K = K(x, \mu)$, can be written in the form

$$\langle Md, d\rangle \geq \gamma_1 \langle Ed, d\rangle \quad \forall\, d \in K, \tag{33}$$

where

$$E = \begin{bmatrix} I & 0 \\ 0 & \sigma(x,\mu)I \end{bmatrix}.$$

By Proposition 1, there exists a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\mu})$ such that (33) holds for all $(x, \mu) \in \mathcal{V}$. This shows that if $\sigma(x, \mu) > 0$ then M is strictly copositive on the cone K. Now Theorem 2 implies that (32) has a nonempty compact solution set. □

We next show that the step given by solving the system (30), which is part of our subproblem (15), satisfies the localization property appearing in the iterative framework of Section 2.

Proposition 3 *Let F and g satisfy the smoothness assumptions* (2)*. Suppose that SOC* (17) *holds at $(\bar{x}, \bar{\mu})$ and that σ satisfies both inequalities in* (19)*. Then there exist a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\mu})$ and a constant $\gamma_3 > 0$ such that for all $(x, \mu) \in \mathcal{V} \cap (\mathbb{R}^n \times \mathbb{R}^m_+)$ with $\sigma(x, \mu) > 0$, it holds that*

$$\left\| \begin{bmatrix} y - x \\ \lambda_{\mathcal{I}} - \mu_{\mathcal{I}} \end{bmatrix} \right\| \leq \gamma_3 \sigma(x, \mu),$$

where $(y, \lambda_{\mathcal{I}})$ is any solution of (30).

Proof Suppose the contrary, i.e., that there exists a sequence $\{(x^k, \mu^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m_+$ such that

$$(x^k, \mu^k) \to (\bar{x}, \bar{\mu}) \text{ and } \eta_k = \left\| \begin{bmatrix} y^k - x^k \\ \lambda^k_{\mathcal{I}} - \mu^k_{\mathcal{I}} \end{bmatrix} \right\| > k\sigma_k,$$

where $\sigma_k = \sigma(x^k, \mu^k) > 0$ and $(y^k, \lambda^k_{\mathcal{I}})$ satisfies

$$0 = F(x^k) + \Psi'_x(x^k, \mu^k)(y^k - x^k) + g'_{\mathcal{I}}(x^k)^\top \lambda^k_{\mathcal{I}}, \tag{34}$$

$$0 = g_i(x^k) + \langle g'_i(x^k), y^k - x^k\rangle - \sigma_k(\lambda^k_i - \mu^k_i),\ i \in \mathcal{I}_+(\bar{x}, \bar{\mu}), \tag{35}$$

$$0 \leq \lambda^k_i \perp g_i(x^k) + \langle g'_i(x^k), y^k - x^k\rangle - \sigma_k(\lambda^k_i - \mu^k_i) \leq 0, i \in \mathcal{I}_0(\bar{x}, \bar{\mu}). \tag{36}$$

By the assumption above,

$$\frac{\sigma_k}{\eta_k} < \frac{1}{k} \to 0. \tag{37}$$

Note first that, by (19), it holds that

$$\|g_{\mathcal{I}}(x^k)\| = \|g_{\mathcal{I}}(x^k) - g_{\mathcal{I}}(\bar{x})\| \le c_1\|x^k - \bar{x}\| \le c_2\sigma_k, \tag{38}$$

$$\|g'_{\mathcal{I}}(x^k) - g'_{\mathcal{I}}(\bar{x})\| \le c_3\|x^k - \bar{x}\| \le c_4\sigma_k. \tag{39}$$

Denote $\hat{\mu}^k = \Pi_{\mathcal{M}(\bar{x})}(\mu^k)$. Since $\hat{\mu}_i^k = 0$ for all $i \notin \mathcal{I}$, we have that

$$\begin{aligned}\|F(x^k) + g'_{\mathcal{I}}(x^k)^\top \mu^k_{\mathcal{I}}\| &= \|F(x^k) + g'_{\mathcal{I}}(x^k)^\top \mu^k_{\mathcal{I}} - F(\bar{x}) - g'_{\mathcal{I}}(\bar{x})^\top \hat{\mu}^k_{\mathcal{I}}\| \\ &\le c_5\left(\|x^k - \bar{x}\| + \|\mu^k_{\mathcal{I}} - \hat{\mu}^k_{\mathcal{I}}\|\right) \\ &\le c_6\sigma_k, \end{aligned} \tag{40}$$

where the first inequality follows from the Lipschitz-continuity of the functions involved, and the last follows from (19).

Taking a subsequence, if necessary, we can assume that

$$\frac{1}{\eta_k}\begin{bmatrix} y^k - x^k \\ \lambda^k_{\mathcal{I}} - \mu^k_{\mathcal{I}} \end{bmatrix} \to \begin{bmatrix} u \\ w \end{bmatrix} \ne 0. \tag{41}$$

Using (34), we have that

$$0 = F(x^k) + g'_{\mathcal{I}}(x^k)^\top \mu^k_{\mathcal{I}} + \Psi'_x(x^k, \mu^k)(y^k - x^k) + g'_{\mathcal{I}}(x^k)^\top(\lambda^k_{\mathcal{I}} - \mu^k_{\mathcal{I}}).$$

Dividing by η_k and taking the limits, using (37) and (40) we obtain that

$$0 = \Psi'_x(\bar{x}, \bar{\mu})u + g'_{\mathcal{I}}(\bar{x})^\top w. \tag{42}$$

By (35) and (36), using also that $\mu^k_{\mathcal{I}} \ge 0$, we have that

$$\langle \lambda^k_{\mathcal{I}}, g_{\mathcal{I}}(x^k) + g'_{\mathcal{I}}(x^k)(y^k - x^k) - \sigma_k(\lambda^k_{\mathcal{I}} - \mu^k_{\mathcal{I}})\rangle = 0,$$

$$\langle \mu^k_{\mathcal{I}}, g_{\mathcal{I}}(x^k) + g'_{\mathcal{I}}(x^k)(y^k - x^k) - \sigma_k(\lambda^k_{\mathcal{I}} - \mu^k_{\mathcal{I}})\rangle \le 0.$$

Hence,

$$\langle \lambda^k_{\mathcal{I}} - \mu^k_{\mathcal{I}}, g_{\mathcal{I}}(x^k) + g'_{\mathcal{I}}(x^k)(y^k - x^k) - \sigma_k(\lambda^k_{\mathcal{I}} - \mu^k_{\mathcal{I}})\rangle \ge 0.$$

Dividing by η_k^2 and taking the limits, using (37) and (38) we obtain that

$$\langle w, g'_{\mathcal{I}}(\bar{x})u\rangle \ge 0. \tag{43}$$

Also, from (35) and (36), dividing by η_k and taking the limits we have that

$$\langle g'_i(\bar{x}), u\rangle = 0,\ i \in \mathcal{I}_+(\bar{x}, \bar{\mu}), \quad \langle g'_i(\bar{x}), u\rangle \le 0,\ i \in \mathcal{I}_0(\bar{x}, \bar{\mu}).$$

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

Multiplying by u in (42) and using (43), we obtain

$$0 \ge \langle \Psi'_x(\bar{x}, \bar{\mu})u, u\rangle,$$

so that SOC (17) implies that $u = 0$. Hence,

$$0 = g'_{\mathcal{I}}(\bar{x})^\top w. \tag{44}$$

Consider the QR-factorization of $g'_{\mathcal{I}}(\bar{x})$, that is

$$g'_{\mathcal{I}}(\bar{x}) = [U\ V] \begin{bmatrix} R \\ 0 \end{bmatrix},$$

where $[U\ V] \in \mathbb{R}^{|\mathcal{I}|\times|\mathcal{I}|}$ is an orthogonal matrix and $R^\top$ has zero kernel (in particular, columns of V give an orthonormal basis for $\ker g'_{\mathcal{I}}(\bar{x})^\top$).

Since

$$g_{\mathcal{I}}(x^k) = g_{\mathcal{I}}(\bar{x}) + g'_{\mathcal{I}}(\bar{x})(x^k - \bar{x}) + O(\|x^k - \bar{x}\|^2) = g'_{\mathcal{I}}(\bar{x})(x^k - \bar{x}) + O(\|x^k - \bar{x}\|^2)$$

and

$$V^\top g'_{\mathcal{I}}(\bar{x}) = 0, \tag{45}$$

we have that

$$V^\top g_{\mathcal{I}}(x^k) = O(\|x^k - \bar{x}\|^2).$$

By (19), we then have that

$$\|V^\top g_{\mathcal{I}}(x^k)\| \le c_7 \sigma_k^2. \tag{46}$$

By (44), we have that $0 = g'_{\mathcal{I}}(\bar{x})^\top w = R^\top U^\top w$. Thus $U^\top w = 0$. Hence,

$$w = UU^\top w + VV^\top w = VV^\top w. \tag{47}$$

Let

$$\mathcal{I}^k = \{i \in \mathcal{I} \mid g_i(x^k) + \langle g'_i(x^k), y^k - x^k\rangle - \sigma_k(\lambda_i^k - \mu_i^k) = 0\}.$$

Evidently, there exists an index set $\mathcal{J}$ such that $\mathcal{I}^k = \mathcal{J}$ for infinitely many indices k. From now on, we consider the subsequence such that $\mathcal{I}^k = \mathcal{J}$, without introducing further subindices.

If $i \notin \mathcal{J}$ then $\lambda_i^k = 0$, so that $\lambda_i^k - \mu_i^k = -\mu_i^k \le 0$. Thus from (41), $w_i \le 0$ for all $i \notin \mathcal{J}$.

Let us define the cone

$$Q = \{\xi \in \mathbb{R}^{|\mathcal{I}|} \mid \xi_i = 0,\ i \in \mathcal{J},\ \xi_i \ge 0,\ i \notin \mathcal{J}\}.$$

Since $w_i \le 0$ for $i \notin \mathcal{J}$, it holds that

$$-w \in Q^*.$$

By (35) and (36), we have that

$$-g_{\mathcal{I}}(x^k) - g'_{\mathcal{I}}(x^k)(y^k - x^k) + \sigma_k(\lambda_{\mathcal{I}}^k - \mu_{\mathcal{I}}^k) \in Q.$$

Multiplying this relation by $V^\top$, dividing by $\eta_k\sigma_k$ and using (45), gives

$$-\frac{V^\top g_{\mathcal{I}}(x^k)}{\eta_k\sigma_k} - \frac{V^\top(g'_{\mathcal{I}}(x^k) - g'_{\mathcal{I}}(\bar{x}))}{\sigma_k}\frac{(y^k - x^k)}{\eta_k} + \frac{V^\top(\lambda_{\mathcal{I}}^k - \mu_{\mathcal{I}}^k)}{\eta_k} \in V^\top Q.$$

Taking the limits, using (46), (39), (37) and the facts that $(y^k - x^k)/\eta_k \to u = 0$ and that the set $V^\top Q$ is closed, we obtain that

$$V^\top w \in V^\top Q.$$

Then there exists $\xi \in Q$ such that $V^\top w = V^\top \xi$. Since $-w \in Q^*$ and $w = VV^\top w$, we conclude that

$$0 \geq \langle w, \xi \rangle = \langle VV^\top w, \xi \rangle = \langle VV^\top \xi, \xi \rangle = \|V^\top \xi\|^2.$$

Thus $V^\top w = V^\top \xi = 0$, so that (47) implies that $w = 0$.

Then $(u, w) = 0$, in contradiction with (41). □

We now extend the solution of (30) to the solution of our subproblem (15), showing also that the needed localization property holds.

Theorem 3 *Let F and g satisfy the smoothness assumptions* (2). *Suppose that SOC* (17) *holds at $(\bar{x}, \bar{\mu})$ and that σ satisfies both inequalities in* (19).

Then there exist a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\mu})$ and a constant $\gamma_4 > 0$ such that for all $(x, \mu) \in \mathcal{V} \cap (\mathbb{R}^n \times \mathbb{R}^m_+)$ with $\sigma(x, \mu) > 0$, there exists $(\bar{y}, \bar{\lambda})$, a solution of the mixed complementarity problem of finding $(y, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m$ such that

$$\begin{aligned} &0 = F(x) + \Psi'_x(x, \mu)(y - x) + g'(x)^\top \lambda, \\ &0 \leq \lambda \perp [g(x) + g'(x)(y - x) - \sigma(x, \mu)(\lambda - \mu)] \leq 0, \end{aligned} \tag{48}$$

satisfying

$$\left\| \begin{bmatrix} \bar{y} - x \\ \bar{\lambda} - \mu \end{bmatrix} \right\| \leq \gamma_4 \sigma(x, \mu). \tag{49}$$

Proof By Proposition 3, there exist a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\mu})$ and a constant $\gamma_3 > 0$ such that

$$\left\| \begin{bmatrix} y - x \\ \lambda_{\mathcal{I}} - \mu_{\mathcal{I}} \end{bmatrix} \right\| \leq \gamma_3 \sigma(x, \mu), \tag{50}$$

for any $(x, \mu) \in \mathcal{V}$ such that $\sigma(x, \mu) > 0$ and any solution $(y, \lambda_{\mathcal{I}})$ of (30).

Set $\bar{y} = y$, $\bar{\lambda}_{\mathcal{I}} = \lambda_{\mathcal{I}}$ and $\bar{\lambda}_i = 0$ for all $i \notin \mathcal{I}$. Evidently, with this choice, (30) implies the first equality in (48), as well as the complementarity conditions in (48) for the indices in $\mathcal{I}_0(\bar{x}, \bar{\mu})$.

For $i \notin \mathcal{I}$, we have that

$$\begin{aligned} g_i(x) + \langle g'_i(x), \bar{y} - x \rangle - \sigma(x, \mu)(\bar{\lambda}_i - \mu_i) &= g_i(x) + \langle g'_i(x), \bar{y} - x \rangle + \sigma(x, \mu)\mu_i \\ &\leq g_i(\bar{x})/2 < 0 \end{aligned}$$

if (x, μ) is sufficiently close to $(\bar{x}, \bar{\mu})$ (so that $\sigma(x, \mu)$ is small enough and, consequently, so is $(\bar{y} - x)$, by (50)). This verifies the complementarity conditions in (48) for the indices not in $\mathcal{I}$.

Given the second relation in (30), it remains to check the nonnegativity of $\bar{\lambda}_i$, $i \in \mathcal{I}_+(\bar{x}, \bar{\mu})$. For $i \in \mathcal{I}_+(\bar{x}, \bar{\mu})$, we have that

$$\bar{\lambda}_i = \mu_i + (\bar{\lambda}_i - \mu_i) \geq \bar{\mu}_i/2 > 0$$

if (x, μ) is sufficiently close to $(\bar{x}, \bar{\mu})$ (so that $\sigma(x, \mu)$ is small enough and, consequently, so is $(\bar{\lambda}_{\mathcal{I}} - \mu_{\mathcal{I}})$, by (50)).

This concludes the proof of the existence of a solution of (48). Finally, let $\hat{\mu} = \Pi_{\mathcal{M}(\bar{x})}(\mu)$. For $i \notin \mathcal{I}$, we have that

$$|\bar{\lambda}_i - \mu_i| = |\mu_i| = |\mu_i - \hat{\mu}_i| \leq \frac{1}{\beta_1}\sigma(x, \mu),$$

by (19). Combining this with (50) proves that (49) holds. □

Theorem 3 establishes that Assumption 3 of Theorem 1 is satisfied for $\Sigma_0 = \{(\bar{x}, \bar{\mu})\}$. In particular, subproblems given by (9) (equivalently, by (15)) are locally solvable, and the distance between consecutive iterates can be bounded above by a measure of violation of KKT conditions for the original problem (1).

4 Upper Lipschitz-continuity of the solution set and a new error bound for KKT systems

This Section verifies Assumption 1 of Theorem 1 under SOC

$$\langle \Psi_x'(\bar{x}, \bar{\mu})u, u\rangle \neq 0 \quad \forall\, u \in \mathcal{C}(\bar{x}; D, F)\backslash\{0\}, \tag{51}$$

which is an extension of (17) used in Section 3 (Note that since the cone $\mathcal{C}(\bar{x}; D, F)$ is convex, (51) means that the inequality holds either with the positive sign for all $u \in \mathcal{C}(\bar{x}; D, F)\backslash\{0\}$, or with the negative sign).

We also show that the so-called natural residual [15]

$$\sigma : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}_+, \quad \sigma(x, \mu) = \left\| \begin{bmatrix} \Psi(x, \mu) \\ \min\{-g(x), \mu\} \end{bmatrix} \right\|, \tag{52}$$

where the minimum is applied component-wise, provides a local error bound (19) for the solution set of the KKT system (10) under SOC (51), see Theorem 4 below. We note that there are a number of related error bound results in the literature for KKT systems of optimization problems. In particular, for optimization, [20, Theorem 3.1] establishes essentially the same result as ours under SOC (17) (i.e., (51) with the positive sign). See also [3, 8, 5, 9] for related work. However, apart from our error bound being valid in the more general variational context, we note that while SOC (51) with the positive sign has the obvious counterpart in SOSC (5) for minimization, SOC (51) with the negative sign has no optimization counterpart (neither for minimization neither for maximization). For this reason, our error bound result is not an obvious translation from the optimization case.

Our result on upper Lipschitz-continuity of the solution set of KKT systems is an extension of the analysis in [8] for optimization to the variational setting.

We start with considering the following problem with affine constraints: find $(x, \mu) \in \mathbb{R}^n \times \mathbb{R}^m$ such that

$$\begin{array}{c} 0 = F(x) + A^\top \mu, \\ 0 \leq \mu \perp Ax + b \leq 0, \end{array} \tag{53}$$

where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. This is the KKT system associated to the variational problem

$$\text{Find } x \in \tilde{D} \quad \text{s.t.} \quad \langle F(x), y - x\rangle \geq 0 \quad \forall\, y \in \tilde{D}, \tag{54}$$

$$\tilde{D} = \{x \in \mathbb{R}^n \mid Ax + b \leq 0\}.$$

We first prove local uniqueness of the primal part of the solution of (53) under SOC (51). Note that in the case of affine constraints, $\Psi'_x(\bar{x}, \bar{\mu}) = F'(\bar{x})$. Our result is an extension of [8, Proposition 1], where the optimization case (4) under the assumption that $F'(\bar{x}) = f''(\bar{x})$ is strictly copositive on the critical cone $\mathcal{C}(\bar{x})$ is considered.

Proposition 4 *Let F be continuously differentiable at a solution $\bar{x}$ of (54) such that*

$$\langle F'(\bar{x})u, u\rangle \neq 0 \quad \forall\, u \in \mathcal{C}(\bar{x}; \tilde{D}, F)\backslash\{0\}.$$

Then there exists a neighborhood $\mathcal{V}$ of $\bar{x}$ such that if $x \in \mathcal{V}$ and (x, μ) is a solution of (53), then $x = \bar{x}$.

Proof Suppose the contrary, i.e., that there exists a sequence $\{(x^k, \mu^k)\}$ of solutions of (53) such that $x^k \to \bar{x}$, $x^k \neq \bar{x}$. Taking a subsequence, if necessary, we can assume that

$$\frac{x^k - \bar{x}}{\|x^k - \bar{x}\|} \to u \neq 0.$$

Using that $\mathcal{I}(x^k) \subset \mathcal{I}(\bar{x})$ for k sufficiently large, we have that if $i \notin \mathcal{I}(\bar{x})$ then $i \notin \mathcal{I}(x^k)$ and, hence, $\mu_i^k = 0$. Thus if $i \notin \mathcal{I}(\bar{x})$ then $\mu_i^k(A\bar{x} + b)_i = 0$ for all k sufficiently large. Since this equality holds trivially for $i \in \mathcal{I}(\bar{x})$, we conclude that

$$\langle \mu^k, A\bar{x} + b\rangle = 0 \tag{55}$$

for all k sufficiently large.

Since (x^k, μ^k) is a solution of (53), we have that

$$\begin{aligned} 0 &= \langle F(x^k) + A^\top \mu^k, x^k - \bar{x}\rangle = \langle F(x^k), x^k - \bar{x}\rangle + \langle \mu^k, A(x^k - \bar{x})\rangle \\ &= \langle F(x^k), x^k - \bar{x}\rangle + \langle \mu^k, Ax^k + b\rangle - \langle \mu^k, A\bar{x} + b\rangle \\ &= \langle F(x^k), x^k - \bar{x}\rangle, \end{aligned} \tag{56}$$

where in the last equation we use (55) and the complementarity condition for (x^k, μ^k). Dividing (56) by $\|x^k - \bar{x}\|$ and taking the limits, we obtain that

$$\langle F(\bar{x}), u\rangle = 0. \tag{57}$$

If $i \in \mathcal{I}(\bar{x})$, then $(A(x^k - \bar{x}))_i = (Ax^k + b)_i \leq 0$. Dividing this inequality by $\|x^k - \bar{x}\|$ and taking the limits, we obtain that

$$(Au)_i \leq 0 \quad \forall\, i \in \mathcal{I}(\bar{x}). \tag{58}$$

Together with (57) this shows that $u \in \mathcal{C}(\bar{x}; \tilde{D}, F)\backslash\{0\}$.

Also, note that

$$\langle F(x^k), u\rangle = -\langle A^\top \mu^k, u\rangle = -\sum_{i\in\mathcal{I}(x^k)} \mu_i^k (Au)_i \geq 0, \tag{59}$$

where in the last inequality we use (58) and the fact that $\mathcal{I}(x^k) \subset \mathcal{I}(\bar{x})$.

Using now (57) and (59) we have that

$$\begin{aligned} 0 = \langle F(\bar{x}), u\rangle &= \langle F(x^k) + F'(x^k)(\bar{x} - x^k), u\rangle + o(\|x^k - \bar{x}\|) \\ &\geq \langle F'(x^k)(\bar{x} - x^k), u\rangle + o(\|x^k - \bar{x}\|). \end{aligned}$$

Dividing this relation by $\|x^k - \bar{x}\|$ and taking the limit, we conclude that

$$0 \leq \langle F'(\bar{x})u, u\rangle.$$

On the other hand, using (56), the fact that $\bar{x}$ is a solution of the variational problem (54) while $x^k \in \tilde{D}$, we have

$$\begin{aligned} 0 = \langle F(x^k), x^k - \bar{x}\rangle &= \langle F(\bar{x}), x^k - \bar{x}\rangle + \langle F'(\bar{x})(x^k - \bar{x}), x^k - \bar{x}\rangle + o(\|x^k - \bar{x}\|^2) \\ &\geq \langle F'(\bar{x})(x^k - \bar{x}), x^k - \bar{x}\rangle + o(\|x^k - \bar{x}\|^2). \end{aligned}$$

Dividing this relation by $\|x^k - \bar{x}\|^2$ and taking the limit, we obtain that

$$0 \geq \langle F'(\bar{x})u, u\rangle.$$

Hence,

$$\langle F'(\bar{x})u, u\rangle = 0$$

for $u \in \mathcal{C}(\bar{x}; \tilde{D}, F)\backslash\{0\}$, in contradiction with the assumption. □

Let now $\tilde{F}(x) = Mx + q$, where $M \in \mathbb{R}^{n\times n}$ and $q \in \mathbb{R}^n$. Consider the KKT system: find $(x, \mu) \in \mathbb{R}^n \times \mathbb{R}^m$ such that

$$\begin{aligned} 0 &= Mx + q + A^\top \mu, \\ 0 &\leq \mu \perp Ax + b \leq 0, \end{aligned} \tag{60}$$

associated to the affine variational problem

$$\text{Find } x \in \tilde{D} \quad \text{s.t.} \quad \langle \tilde{F}(x), y - x\rangle \geq 0 \quad \forall\, y \in \tilde{D}.$$

Define

$$N(x, \mu) = \begin{bmatrix} Mx + A^\top \mu \\ -Ax \end{bmatrix} \text{ and } \psi = \begin{bmatrix} q \\ b \end{bmatrix},$$

so that (60) is equivalent to the generalized equation

$$\psi \in N(x, \mu) + \mathcal{T}(x, \mu), \tag{61}$$

where $\mathcal{T}$ is defined in (14).

The following is an extension of [8, Lemma 1], where M is assumed to be symmetric and strictly copositive on $\mathcal{C}(\bar{x}; \tilde{D}, \tilde{F})$, to variational setting. Once Proposition 4 has been established, the argument is essentially the same as in [8]; we include it for completeness.

Lemma 1 *Suppose that $(\bar{x}, \bar{\mu})$ is a solution of (61) for $\bar{\psi}$ and that*

$$\langle Mu, u\rangle \neq 0 \quad \forall\, u \in \mathcal{C}(\bar{x}; \tilde{D}, \tilde{F})\backslash\{0\}.$$

Then there exist $\beta > 0$ and neighborhoods $\mathcal{V}$ of $\bar{x}$ and $\mathcal{U}$ of $\bar{\psi}$ such that if (x, μ) is a solution of (61) for $\psi \in \mathcal{U}$ and $x \in \mathcal{V}$, then

$$\|x - \bar{x}\| \leq \beta\|\psi - \bar{\psi}\|.$$

Proof As is well known [16], the function $\mathcal{F}(x, \mu) = N(x, \mu) + \mathcal{T}(x, \mu)$ and its inverse

$$\mathcal{F}^{-1}(\psi) = \{\omega \in \mathbb{R}^n \times \mathbb{R}^m \mid 0 \in \mathcal{F}(\omega) - \psi\},$$

are polyhedral multifunctions. Furthermore, the function $\mathcal{P}$ such that $\mathcal{P}(x, \mu) = x$ is polyhedral, and so is the composition $\mathcal{P} \circ \mathcal{F}^{-1}$.

By [16, Proposition 1], polyhedral multifunctions are locally upper Lipschitzian at every point, and the Lipschitz constant is independent of the point. Thus there exist a constant $\beta > 0$ and a neighborhood $\mathcal{U}$ of $\bar{\psi}$ such that

$$\mathcal{P} \circ \mathcal{F}^{-1}(\psi) \subset \mathcal{P} \circ \mathcal{F}^{-1}(\bar{\psi}) + \beta\|\psi - \bar{\psi}\|B \quad \forall\, \psi \in \mathcal{U}. \tag{62}$$

Since $\mathcal{P} \circ \mathcal{F}^{-1}(\psi)$ is the set of x-components of solutions of (60), by Proposition 4 there exists a neighborhood $\mathcal{V}$ of $\bar{x}$ such that

$$\mathcal{P} \circ \mathcal{F}^{-1}(\bar{\psi}) \cap \mathcal{V} = \{\bar{x}\}.$$

Let $\rho = \operatorname{dist}(\bar{x}, \mathcal{P} \circ \mathcal{F}^{-1}(\bar{\psi})\backslash\{\bar{x}\})$, and choose $\mathcal{V}$ smaller if necessary so that $\mathcal{V} \subset \{\bar{x}\} + \frac{\rho}{3}B$. Choose $\mathcal{U}$ sufficiently small so that

$$\{\bar{x}\} + \beta\|\psi - \bar{\psi}\|B \subset \mathcal{V} \quad \forall\, \psi \in \mathcal{U}.$$

If $\psi \in \mathcal{U}$ and $x \in \mathcal{P} \circ \mathcal{F}^{-1}(\psi) \cap \mathcal{V}$, we obtain from (62) that there exist $\hat{x} \in \mathcal{P} \circ \mathcal{F}^{-1}(\bar{\psi})$ and $p \in B$ such that $x = \hat{x} + \beta\|\psi - \bar{\psi}\|p$. But then

$$\|\bar{x} - \hat{x}\| = \big\|\bar{x} - x + \beta\|\psi - \bar{\psi}\|p\big\| \leq \|x - \bar{x}\| + \beta\|\psi - \bar{\psi}\| \leq \frac{\rho}{3} + \frac{\rho}{3} < \rho,$$

implying that $\hat{x} = \bar{x}$. Hence, $x = \bar{x} + \beta\|\psi - \bar{\psi}\|p$ for some $p \in B$, i.e.,

$$\|x - \bar{x}\| \leq \beta\|\psi - \bar{\psi}\|.$$

□

Thus for our main problem (11) we can state the following error estimates, that verify the upper Lipschitz-continuity property of the solution set of KKT systems. The argument is the same as in [8, Lemma 2]; we include it for completeness.

Lemma 2 *Let F be differentiable and g twice differentiable at $\bar{x}$, and suppose that there exists $\bar{\mu} \in \mathcal{M}(\bar{x})$ such that $(\bar{x}, \bar{\mu})$ satisfies SOC (51).*

Then there exist a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\mu})$ and constants $\gamma, \tau > 0$ such that for every $(x, \mu) \in \mathcal{V}$ and for each $p \in \gamma B$ satisfying

$$0 \in G(x, \mu) + \mathcal{T}(x, \mu) + p, \tag{63}$$

where G and $\mathcal{T}$ are defined in (14)*, it holds that*

$$\|x - \bar{x}\| + \|\mu - \Pi_{\mathcal{M}(\bar{x})}(\mu)\| \leq \tau\|p\|.$$

Proof Consider the affine variational problem (61) with

$$M = \Psi'_x(\bar{x}, \bar{\mu}) \quad \text{and} \quad A = g'(\bar{x}).$$

Let $\mathcal{F}(x, \mu) = N(x, \mu) + \mathcal{T}(x, \mu)$.

Define $\psi^1 = N(x, \mu) - G(x, \mu) - p$, where p satisfies (63). Then $(x, \mu) \in \mathcal{F}^{-1}(\psi^1)$. Define $\psi^2 = N(\bar{x}, \mu) - G(\bar{x}, \mu)$. Since $F(\bar{x}) + A^\top \bar{\mu} = 0$, we have that

$$\psi^2 = \begin{bmatrix} M\bar{x} + A^\top \mu - F(\bar{x}) - A^\top \mu \\ -A\bar{x} + g(\bar{x}) \end{bmatrix} = \begin{bmatrix} M\bar{x} + A^\top \bar{\mu} \\ -A\bar{x} \end{bmatrix} + \begin{bmatrix} 0 \\ g(\bar{x}) \end{bmatrix}.$$

As $g(\bar{x}) \in \mathcal{N}(\bar{\mu})$, this shows that $(\bar{x}, \bar{\mu}) \in \mathcal{F}^{-1}(\psi^2)$.

By the differentiability assumptions, ψ^1 is close to ψ^2 when (x, μ) is close to $(\bar{x}, \bar{\mu})$ and p is close to 0. Consequently, by choosing $\mathcal{V}$ and γ sufficiently small, Lemma 1 gives us the estimate

$$\|x - \bar{x}\| \leq \beta_0 \|\psi^1 - \psi^2\| = \beta_0 \|G(x, \mu) - G(\bar{x}, \mu) - (N(x, \mu) - N(\bar{x}, \mu)) + p\|, \tag{64}$$

for all $(x, \mu) \in \mathcal{V}$ and $p \in \gamma B$.

Given any $\varepsilon > 0$, using the differentiability assumptions and taking $\mathcal{V}$ sufficiently small, we obtain that

$$\|G(x, \mu) - G(\bar{x}, \mu) - N(x - \bar{x}, 0)\| \leq \varepsilon \|x - \bar{x}\| \quad \forall\, (x, \mu) \in \mathcal{V}.$$

Combining this with (64), we have that

$$\|x - \bar{x}\| \leq \beta_0 \varepsilon \|x - \bar{x}\| + \beta_0 \|p\|.$$

Thus, taking $\varepsilon < 1/\beta_0$, we obtain

$$\|x - \bar{x}\| \leq \beta_1 \|p\|. \tag{65}$$

where $\beta_1 = \frac{\beta_0}{1 - \varepsilon \beta_0}$. Consider the decomposition $p = (u, v) \in \mathbb{R}^n \times \mathbb{R}^m$. If $i \notin \mathcal{I}$ then $g_i(\bar{x}) < 0$. Thus we can take $\mathcal{V}$ and γ small enough, so that $g_i(x) - v_i < 0$. From (63) we have that $g(x) - v \in \mathcal{N}(\mu)$. Hence, $\mu_i = 0$ for all $i \notin \mathcal{I}$ and $\mu_i \geq 0$ for $i \in \mathcal{I}$. Since

$$\mathcal{M}(\bar{x}) = \{\nu \in \mathbb{R}^m \mid F(\bar{x}) + g'(\bar{x})^\top \nu = 0;\ \nu_i \geq 0, i \in \mathcal{I};\ \nu_i = 0, i \notin \mathcal{I}\},$$

by Hoffman's error bound for linear systems we obtain that

$$\|\mu - \hat{\mu}\| \leq \beta_2 \|F(\bar{x}) + g'(\bar{x})^\top \mu\| = \beta_2 \|\Psi(\bar{x}, \mu)\|, \tag{66}$$

where $\hat{\mu} = \Pi_{\mathcal{M}(\bar{x})}(\mu)$.

From (63), we have that $\Psi(x, \mu) + u = 0$. Then using the differentiability assumptions and taking $\mathcal{V}$ smaller if necessary, we have

$$\|\Psi(\bar{x}, \mu)\| \leq \|\Psi(x, \mu)\| + \|\Psi(\bar{x}, \mu) - \Psi(x, \mu)\| \leq \|u\| + \beta_3 \|x - \bar{x}\|.$$

Since $\|u\| \leq \|p\|$, using (65), we obtain

$$\|\Psi(\bar{x}, \mu)\| \leq (1 + \beta_1 \beta_3) \|p\|.$$

Combining this with (65) and (66) gives

$$\|x - \bar{x}\| + \|\mu - \hat{\mu}\| \leq \tau \|p\|,$$

for $\tau = \beta_1 + \beta_2(1 + \beta_1 \beta_3)$. □

This result verifies Assumption 1 of Theorem 1 for $\Sigma_0 = \{(\bar{x}, \bar{\mu})\}$. Moreover, by [5, Theorem 2], it now also follows that the natural residual (52) provides a valid local error bound for the KKT system (10) (as for the rightmost inequality in (19), it follows from Lipschitz-continuity of the functions involved and the fact that $\sigma(\bar{x}, \bar{\mu}) = 0$ for any $\bar{\mu} \in \mathcal{M}(\bar{x})$.) Specifically, we have the following.

Theorem 4 *Let F be differentiable and g twice differentiable at $\bar{x}$, and suppose there exists $\bar{\mu} \in \mathcal{M}(\bar{x})$ such that $(\bar{x}, \bar{\mu})$ satisfies SOC (51).*

Then there exist a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\mu})$ and constants $\beta_2 \geq \beta_1 > 0$ such that for all $(x, \mu) \in \mathcal{V}$ the function σ defined in (52) *satisfies the error bound* (19).

We note that Theorem 4 gives the first error bound for KKT systems in variational context that does not subsume some regularity-type assumptions about the constraints. Concerning the relations with error bounds for optimization, e.g., [20, Theorem 3.1], we point out that SOC (51) with the negative sign does not have a counterpart in any sufficiency condition for optimization problems. We refer the reader to [9] for a detailed discussion and comparisons of error bounds for KKT systems.

5 Convergence Results

The results established in Sections 3 and 4 complete the proof of superlinear convergence of our method, that we formalize as follows.

Theorem 5 *Let F and g satisfy the smoothness assumptions* (2)*, and suppose that there exists $\bar{\mu} \in \mathcal{M}(\bar{x})$ such that $(\bar{x}, \bar{\mu})$ satisfies SOC (17).*

Then there exist a neighborhood $\mathcal{V}$ of $(\bar{x}, \bar{\mu})$ such that for any $(x^0, \mu^0) \in \mathcal{V} \cap (\mathbb{R}^n \times \mathbb{R}^m_+)$, the iterates defined by (13) *are well defined and converge superlinearly to $(\bar{x}, \mu)$, where μ is some element of $\mathcal{M}(\bar{x})$. Furthermore, the convergence rate is quadratic if F' and g'' are Lipschitz-continuous in a neighborhood of $\bar{x}$.*

We illustrate our convergence result with the following example.

Example 1 Consider the optimization problem

$$\begin{array}{ll} \min & x_1 x_2 - x_2^2/2 \\ \text{s.t.} & x_2^2 \leq 0, \\ & -2x_1 + x_2 \leq 0, \\ & x_1 - 2x_2 \leq 0. \end{array} \tag{67}$$

It can be seen that $\bar{x} = (0, 0)$ is the unique solution of this problem, and that the associated set of Lagrange multipliers is given by

$$\mathcal{M}(\bar{x}) = \{(\mu_1, \mu_2, \mu_3) \in \mathbb{R}^3 \mid \mu_1 \geq 0,\ \mu_2 = \mu_3 = 0\}.$$

In particular, MFCQ does not hold (incidentally, it can be seen that GCQ mentioned in Section 1 also does not hold). Furthermore, SOSC (5) holds at

Table 1 Distance to solution on last 5 iterations

$\|x-\bar{x}\| + \mathrm{dist}(\mu, \mathcal{M}(\bar{x}))$
1.5891e-001
1.0599e-002
7.3688e-005
6.3242e-009
8.4865e-017

$(\bar{x}, \bar{\mu})$ for any $\bar{\mu} \in \mathcal{M}(\bar{x})$ with $\bar{\mu}_1 > 0$, but SSOSC (8) is not satisfied for any multiplier.

We have written a Matlab implementation of sSQP, using the built-in subroutine `quadprog` for solving subproblems (7). Experiments were performed choosing random starting points $x_i^0 \in [-1/2, 1/2]$, $i = 1, 2$, and $\mu_j^0 \in [0, 1]$, $j = 1, 2, 3$. The stopping criteria was $\sigma(x^k, \mu^k) < 10^{-15}$.

In about 10% of the cases, the sequence converged linearly to $(\bar{x}, \bar{\mu})$ with $\bar{\mu}_1 = 0$ (SOSC is not valid at this solution). Such cases appear to correspond to the choices of starting points that are not close enough to a solution (so that Theorem 5 does not apply). About 3% of the starting points produced unsolvable subproblems at the first iteration (for the same reason as above – starting points not being close enough to a solution). All the remaining runs converged superlinearly to a primal-dual solution satisfying SOSC. Table 1 shows the average values of $\|x^k - \bar{x}\| + \mathrm{dist}(\mu^k, \mathcal{M}(\bar{x}))$ for the last 5 iterations in the cases of convergence to a primal-dual solution satisfying SOSC.

The approach presented here can be used also to prove the uniqueness of solutions of subproblems (15), extending the result for optimization under SSOSC (8) obtained in [7] (see also [5]). In our case, for this purpose we shall assume the stronger version of SOC (17), i.e., that

$$\langle \Psi_x'(\bar{x}, \bar{\mu})u, u \rangle > 0 \quad \forall\, u \in \mathcal{C}^+(\bar{x}, \bar{\mu}) \setminus \{0\}, \tag{68}$$

where

$$\mathcal{C}^+(\bar{x}, \bar{\mu}) = \{u \in \mathbb{R}^n \mid \langle g_i'(\bar{x}), u \rangle = 0 \ \forall i \in \mathcal{I}_+(\bar{x}, \bar{\mu})\}.$$

Theorem 6 *Suppose that the assumptions of Theorem 5 are satisfied, with SOC* (17) *replaced by SSOC* (68)*.*

Then all the assertions of Theorem 5 hold and, in addition, solutions of subproblems (15) *are locally unique.*

Proof We shall provide the main steps, indicating the changes needed in the preceding analysis.

Regarding the proof of Proposition 1, it can be seen that under SSOC (68) there exists a constant $\gamma_2 > 0$ such that for all (x, μ) in a neighborhood of $(\bar{x}, \bar{\mu})$ it holds that

$$\langle \Psi_x'(x, \mu)u, u \rangle + \sigma(x, \mu)\|v\|^2 \geq \gamma_2 \Big(\|u\|^2 + \sigma(x, \mu)\|v\|^2 \Big) \quad \forall (u, v) \in K^+(x, \mu), \tag{69}$$

where

$$K^+(x,\mu) = \left\{(u,v) \in \mathbb{R}^n \times \mathbb{R}^{|\mathcal{I}|} \mid \langle g_i'(x), u\rangle = \sigma(x,\mu)v_i,\ i \in \mathcal{I}_+(\bar{x},\bar{\mu})\right\}. \tag{70}$$

Since $\mathcal{C}(\bar{x}; D, F) \subset \mathcal{C}^+(\bar{x},\bar{\mu})$, we have that SOC (17) and, thus, Propositions 1 and 2 hold. In particular, in the proof of Proposition 2, the generalized complementarity problem

$$\text{Find } \bar{d} \quad \text{s.t.} \quad K \ni \bar{d} \perp M\bar{d} + M\tilde{z} + q \in K^*,$$

where K is given by (21), has a nonempty compact solution set. Let d^1 and d^2 be solutions of this complementarity problem. Then

$$\begin{aligned}\langle M(d^1 - d^2), d^1 - d^2\rangle &= \langle Md^1 + M\tilde{z} + q - (Md^2 + M\tilde{z} + q), d^1 - d^2\rangle \\ &= -\langle Md^1 + M\tilde{z} + q, d^2\rangle - \langle Md^2 + M\tilde{z} + q, d^1\rangle \\ &\le 0. \end{aligned} \tag{71}$$

Since $K^+(x,\mu)$ is a subspace and $d^1, d^2 \in K \subset K^+(x,\mu)$, we have that

$$d^1 - d^2 \in K^+(x,\mu).$$

Since (69) implies that M is strictly copositive on $K^+(x,\mu)$, from (71) we conclude that $d^1 - d^2 = 0$. Hence, the mixed complementarity problem (30) has the unique solution.

Let us now show that under SSOC (68), for (x,μ) sufficiently close to $(\bar{x},\bar{\mu})$ we have that $(\bar{y},\bar{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^m_+$, where $\bar{\lambda}_i = 0, i \notin \mathcal{I}$ and $(\bar{y},\bar{\lambda}_{\mathcal{I}})$ is the solution of (30), is the unique solution of (48) satisfying (49). By Theorem 3, $(\bar{y},\bar{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^m_+$ defined in this way is a solution of (48) satisfying (49). Conversely, if $(\bar{y},\bar{\lambda}) \in \mathbb{R}^n \times \mathbb{R}^m_+$ is a solution of (48) satisfying (49), and if (x,μ) is sufficiently close to $(\bar{x},\bar{\mu})$, we have that

$$g_i(x) + \langle g_i'(x), \bar{y} - x\rangle - \sigma(x,\mu)(\bar{\lambda}_i - \mu_i) < 0 \quad \forall\, i \notin \mathcal{I},$$

$$\bar{\lambda}_i > 0 \quad \forall\, i \in \mathcal{I}_+(\bar{x},\bar{\mu}).$$

Then by the complementarity conditions in (48), we obtain that

$$\bar{\lambda}_i = 0 \quad \forall\, i \notin \mathcal{I},$$

$$g_i(x) + \langle g_i'(x), \bar{y} - x\rangle - \sigma(x,\mu)(\bar{\lambda}_i - \mu_i) = 0 \quad \forall\, i \in \mathcal{I}_+(\bar{x},\bar{\mu}).$$

Hence, $(\bar{y},\bar{\lambda}_{\mathcal{I}})$ is a solution of (30), which has been established to be unique. □

Acknowledgment. We would like to thank the Associate Editor and the two anonymous referees for their constructive suggestions.

References

1. Boggs, B., Tolle, J.: Sequential quadratic programming. Acta Numerica **4**, 1–51 (1996)
2. Bonnans, J.F.: Local analysis of Newton-type methods for variational inequalities and nonlinear programming. Applied Mathematics and Optimization **29**, 161–186 (1994)
3. Facchinei, F., Fischer, A., Kanzow C.: On the accurate identification of active constraints. SIAM Journal on Optimization **9**, 14–32 (1999)
4. Facchinei, F., Pang, J.S.: Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer–Verlag, New York, NY (2003)
5. Fischer, A.: Local behavior of an iterative framework for generalized equations with nonisolated solutions. Mathematical Programming **94**, 91–124 (2002)
6. Guignard, M.: Generalized Kuhn-Tucker conditions for mathematical programming problems in a Banach space. SIAM Journal on Control, **7**, 232–241 (1969)
7. Hager, W.: Stabilized sequential quadratic programming. Computational Optimization and Applications **12**, 253–273 (1999)
8. Hager, W., Gowda, M.: Stability in the presence of degeneracy and error estimation. Mathematical Programming **85**, 181–192 (1999)
9. Izmailov, A., Solodov, M.: Karush-Kuhn-Tucker systems: regularity conditions, error bounds and a class of Newton-type methods. Mathematical Programming **95**, 631–650 (2003)
10. Izmailov, A., Solodov, M.: Newton-type methods for optimization problems without constraint qualifications. SIAM Journal on Optimization **16**, 210–228 (2004)
11. Izmailov, A.F., Solodov, M.V.: Examples of dual behaviour of Newton-type methods on optimization problems with degenerate constraints. Computational Optimization and Applications (2007). DOI 10.1007/s10589-007-9074-4
12. Izmailov, A.F., Solodov, M.V.: On attraction of Newton-type iterates to multipliers violating second-order sufficiency conditions. Mathematical Programming **117**, 271–304 (2009)
13. Josephy, N.H.: Newton's method for generalized equations. Technical Summary Report 1965, Mathematics Research Center, University of Wisconsin (1979)
14. Li, D.H., Qi, L.: A stabilized SQP method via linear equations. Tech. Rep. Applied mathematics technical report AMR00/5, The University of New South Wales (2000)
15. Pang, J.S.: Error bounds in mathematical programming. Mathematical Programming **79**, 299–332 (1997)
16. Robinson, S.: Some continuity properties of polyhedral multifunctions. Mathematical Programming Study **14**, 206–214 (1981)
17. Wright, S.: Superlinear convergence of a stabilized SQP method to a degenerate solution. Computational Optimization and Applications **11**, 253–275 (1998)
18. Wright, S.: Modifying SQP for degenerate problems. SIAM Journal on Optimization **13**, 470–497 (2002)
19. Wright, S.: Constraint identification and algorithm stabilization for degenerate nonlinear programs. Mathematical Programming **95**, 137–160 (2003)
20. Wright, S.: An algorithm for degenerate nonlinear programming with rapid local convergence. SIAM Journal on Optimization **15**, 673–696 (2005)