

On local convergence of sequential quadratically-constrained quadratic-programming type methods, with an extension to variational problems

Damián Fernández · Mikhail Solodov

Received: 6 April 2005 / Revised: 26 June 2006 / Published online: 22 September 2007
© Springer Science+Business Media, LLC 2007

Abstract We consider the class of quadratically-constrained quadratic-programming methods in the framework extended from optimization to more general variational problems. Previously, in the optimization case, Anitescu (SIAM J. Optim. 12, 949–978, 2002) showed superlinear convergence of the primal sequence under the Mangasarian-Fromovitz constraint qualification and the quadratic growth condition. Quadratic convergence of the primal-dual sequence was established by Fukushima, Luo and Tseng (SIAM J. Optim. 13, 1098–1119, 2003) under the assumption of convexity, the Slater constraint qualification, and a strong second-order sufficient condition. We obtain a new local convergence result, which complements the above (it is neither stronger nor weaker): we prove primal-dual quadratic convergence under the linear independence constraint qualification, strict complementarity, and a second-order sufficiency condition. Additionally, our results apply to variational problems beyond the optimization case. Finally, we provide a necessary and sufficient condition for superlinear convergence of the primal sequence under a Dennis-Moré type condition.

Keywords Quadratically constrained quadratic programming · Karush-Kuhn-Tucker system · Variational inequality · Quadratic convergence · Superlinear convergence · Dennis-Moré condition

Research of the second author is partially supported by CNPq Grants 300734/95-6 and 471780/2003-0, by PRONEX–Optimization, and by FAPERJ.

D. Fernández · M. Solodov (✉)
Instituto de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil
e-mail: solodov@impa.br

D. Fernández
e-mail: dfernan@impa.br

1 Introduction

Given sufficiently smooth mappings $F : \mathfrak{R}^n \rightarrow \mathfrak{R}^n$ and $g : \mathfrak{R}^n \rightarrow \mathfrak{R}^m$ (precise smoothness requirements will be specified later, within the statements of our convergence results), we consider the following variational problem [7]:

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

where

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

and $T(x; D)$ is the (standard) tangent cone to D at $x \in D$. When for some smooth function $f : \mathfrak{R}^n \rightarrow \mathfrak{R}$ it holds that

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

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

$$\min f(x) \text{ s.t. } x \in D. \tag{3}$$

We consider the following iterative procedure. (As will be seen below, in the case of the optimization problem (3) it reduces to the sequential quadratically-constrained quadratic-programming method, e.g., [1, 8, 19]. In the variational setting, this method appears to be new.) If $x^k \in \mathfrak{R}^n$ is the current iterate, then the next iterate x^{k+1} is obtained as a solution of an approximation of the variational problem (1) of the following form:

$$\text{Find } x \in D_k \text{ s.t. } \langle F_k(x), y - x \rangle \geq 0 \quad \forall y \in (x + T(x; D_k)), \tag{4}$$

where

$$F_k(x) = F(x^k) + F'(x^k)(x - x^k), \quad x \in \mathfrak{R}^n, \\ D_k = \left\{ x \in \mathfrak{R}^n \mid \begin{array}{l} g_i(x^k) + \langle g'_i(x^k), x - x^k \rangle + \frac{1}{2} \langle g''_i(x^k)(x - x^k), x - x^k \rangle \leq 0, \\ i = 1, \dots, m \end{array} \right\},$$

and $T(x; D_k)$ is the tangent cone to D_k at $x \in D_k$. Subproblem (4) can be considered as a “one-step-further” approximation when compared to the classical Josephy-Newton method for variational inequalities [7, 10], where at every step the mapping F is approximated to the first order (as in (4)), but the set D is not being simplified (unlike in (4)). Specifically, given the current iterate x^k , the Josephy-Newton method solves the following subproblem:

$$\text{Find } x \in D \text{ s.t. } \langle F_k(x), y - x \rangle \geq 0 \quad \forall y \in (x + T(x; D)). \tag{5}$$

It is clear that subproblem (4) is structurally simpler than (5) (in (5) the constraints are general nonlinear, while in (4) they are quadratic). Thus, in principle, (4) should be easier to solve. That said, we shall not be concerned here with specific methods for

solving subproblems of the structure of (4) (at the very least, the same techniques as for (5) can be used). In the case of optimization, as discussed below, specific methods are readily available.

For optimization problems (3), an iteration of the sequential quadratically-constrained quadratic-programming method (SQCQP) consists of minimizing a quadratic approximation of the objective function subject to a quadratic approximation of the constraints. Specifically, if $x^k \in \mathfrak{H}^n$ is the current iterate, then the next iterate x^{k+1} is obtained as a solution of the following approximation of the original problem:

$$\min \quad f_k(x) := \langle f'(x^k), x - x^k \rangle + \frac{1}{2} \langle f''(x^k)(x - x^k), x - x^k \rangle \quad \text{s.t.} \quad x \in D_k. \quad (6)$$

Note that taking into account (2), the variational subproblem (4) describes (primal) first-order necessary optimality conditions for (6). Therefore, SQCQP for optimization is a special case in our framework.

As some previous work on SQCQP and related methods, we mention [1, 8, 11, 16, 17, 19, 21]. In the convex case, subproblem (6) can be cast as a second-order cone program [12, 15], which can be solved efficiently by interior-point algorithms (such as [14, 20]). In [1], nonconvex subproblems (6) were also handled quite efficiently by using other nonlinear programming techniques. Even though quadratically constrained subproblems are computationally more difficult than those that are linearly constrained (which is the case for the more traditional SQP methods, [2]), they are manageable by modern computational tools and the extra effort in solving them may be worthwhile. In other words, at least in some situations, one may expect that fewer subproblems will need to be solved, when compared to SQP. Some numerical validation of this can be found in the computational experiments of [1].

In order to guarantee global convergence, SQCQP methods require some modifications to subproblem (6), as well as a linesearch procedure for an adequately chosen penalty function. (See, for example, [8, 19].) But under certain assumptions, locally all those modifications reduce precisely to (6). Moreover, the unit stepsize satisfies the linesearch criteria under very mild conditions [19, Proposition 8] (in particular, no second-order sufficiency is needed for this), which is one of the attractive features of SQCQP. Thus, what is relevant for local convergence analysis is precisely the method given by (6), and this is the subject of this paper (except that we consider the more general variational setting of (4)). Note that, as a consequence of acceptance of the unit stepsize, the Maratos effect [13, 18] does not occur in SQCQP. (We note that Maratos effect can also be avoided in SQP methods that use second-order corrections, or use, for example, an augmented Lagrangian merit function.)

We next survey the previous local rate of convergence results and compare them to ours. As already mentioned, in the variational setting our method appears to be new. Therefore, we limit our discussion to the case of optimization. In [1], local primal superlinear rate of convergence of a trust-region SQCQP method is obtained under the Mangasarian-Fromovitz constraint qualification (MFCQ) and a certain quadratic growth condition. We note that, under MFCQ, quadratic growth is equivalent to the second-order sufficient condition for optimality (SOSC), see [3, Theorem 3.70]. Quadratic convergence of the primal-dual sequence is obtained in [8] (the dual part

of the sequence is formed by the Lagrange multipliers associated to solutions of (6)). The assumptions in [8] are as follows: convexity of f and of g , the Slater condition (equivalent to MFCQ in the convex case) and a strong second-order sufficient condition (implying quadratic growth). This set of assumptions is stronger than in [1], but the assertions in the two papers are different and not comparable to each other. Thus, neither of the two results implies the other one. To complement the picture, in this paper we prove a third local convergence result, which is in the same relation to the two previous ones: it neither follows from them nor implies them. Specifically, we shall establish primal-dual quadratic convergence under the linear independence constraint qualification (LICQ), strict complementarity condition, and SOSC. Compared to [8], our assumptions are essentially different (we do not make any convexity assumptions; while [8] makes weaker regularity assumptions). Our assertions are stronger than in [8], because in addition to primal-dual quadratic convergence we also prove superlinear primal convergence. Compared to [1], our assumptions are more restrictive, of course. But our assertions are stronger as well: we prove quadratic primal-dual convergence *and* superlinear primal convergence instead of superlinear primal convergence only. In addition, we shall exhibit a Dennis-Moré type [6] necessary and sufficient condition for superlinear convergence of the primal sequence in the case when the primal-dual convergence is given.

A few words about our notation. For a matrix M of arbitrary dimensions, M_I denotes the submatrix of M with rows indexed by I . When in matrix notation, vectors are considered columns, and for a vector x we denote by x_I the subvector of x with coordinates indexed by I . By $\langle \cdot, \cdot \rangle$ we denote the Euclidean inner product, with $\| \cdot \|$ being the associated norm (the space will always be clear from the context). We use the notation $\phi(t) = o(t)$ for any function $\phi : \mathfrak{R}_+ \rightarrow \mathfrak{R}^p$ such that $\lim_{t \rightarrow 0} t^{-1} \phi(t) = 0$. For a function $\Psi : \mathfrak{R}^n \times \mathfrak{R}^m \rightarrow \mathfrak{R}^p$, 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 the same point.

If $\Phi : \mathfrak{R}^s \times \mathfrak{R}^p \rightarrow \mathfrak{R}^p$ is Lipschitz continuous in a neighborhood of a point $(\bar{\sigma}, \bar{\xi}) \in \mathfrak{R}^s \times \mathfrak{R}^p$, by $\partial\Phi(\bar{\sigma}, \bar{\xi})$ we denote the *Clarke generalized Jacobian* of Φ at $(\bar{\sigma}, \bar{\xi})$, i.e.,

$$\partial\Phi(\bar{\sigma}, \bar{\xi}) = \text{conv} \left\{ \lim_{l \rightarrow \infty} \Phi'(\sigma^l, \xi^l) \mid (\sigma^l, \xi^l) \rightarrow (\bar{\sigma}, \bar{\xi}), (\sigma^l, \xi^l) \in \mathcal{N}_\Phi \right\},$$

where conv denotes convex hull of a set, and \mathcal{N}_Φ is the set of points at which Φ is differentiable (by Rademacher’s Theorem, Φ is differentiable almost everywhere in a neighborhood of $(\bar{\sigma}, \bar{\xi})$). In the sequel, we shall make use of the following Implicit Function Theorem.

Theorem 1 ([4, p. 256]) *Let $\Phi : \mathfrak{R}^s \times \mathfrak{R}^p \rightarrow \mathfrak{R}^p$ be Lipschitz continuous in a neighborhood of a point $(\bar{\sigma}, \bar{\xi}) \in \mathfrak{R}^s \times \mathfrak{R}^p$ such that $\Phi(\bar{\sigma}, \bar{\xi}) = 0$.*

Suppose that the set of $p \times p$ matrices M , for which there exists a $p \times s$ matrix N such that $[N, M] \in \partial\Phi(\bar{\sigma}, \bar{\xi})$, has full rank.

Then there exist a neighborhood U_0 of $\bar{\sigma}$, a neighborhood Ω_0 of $\bar{\xi}$, and the unique Lipschitz continuous function $\xi : U_0 \rightarrow \Omega_0$ such that $\Phi(\sigma, \xi(\sigma)) = 0$ for all $\sigma \in U_0$.

2 Primal-dual quadratic convergence

As is well-known, under adequate constraint qualifications (which would be the case here), the solution of variational problem (1) is equivalent to the solution of the Karush-Kuhn-Tucker (KKT) system: find $(x, \mu) \in \mathfrak{R}^n \times \mathfrak{R}^m$ such that

$$\begin{aligned}
 F(x) + \sum_{i=1}^m \mu_i g'_i(x) &= 0, \\
 g_i(x) \leq 0, \quad \mu_i \geq 0, \quad \mu_i g_i(x) &= 0, \quad i = 1, \dots, m.
 \end{aligned}
 \tag{7}$$

For the same reason, solutions of subproblem (4) are described by the following mixed complementarity problem [7] in $(x, \mu) \in \mathfrak{R}^n \times \mathfrak{R}^m$:

$$F(x^k) + F'(x^k)(x - x^k) + \sum_{i=1}^m \mu_i (g'_i(x^k) + g''_i(x^k)(x - x^k)) = 0,
 \tag{8}$$

and for all $i = 1, \dots, m$, it holds that

$$g_i(x^k) + \langle g'_i(x^k), x - x^k \rangle + \frac{1}{2} \langle g''_i(x^k)(x - x^k), x - x^k \rangle \leq 0,
 \tag{9}$$

$$\mu_i \geq 0,
 \tag{10}$$

$$\mu_i (g_i(x^k) + \langle g'_i(x^k), x - x^k \rangle + \frac{1}{2} \langle g''_i(x^k)(x - x^k), x - x^k \rangle) = 0.
 \tag{11}$$

Note that in the case of the optimization problem (3), i.e., when (2) holds, the above are precisely the optimality conditions for SQCQP subproblem (6).

Let $(\bar{x}, \bar{\mu}) \in \mathfrak{R}^n \times \mathfrak{R}^m$ be some fixed solution of the KKT system (7), which by virtue of further assumptions will be locally unique.

We say that LICQ holds at \bar{x} if

$$\{g'_i(\bar{x}), i \in I\} \text{ is a linearly independent set,}
 \tag{12}$$

where

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

is the index set of constraints active at $\bar{x} \in D$. Under LICQ, the multiplier $\bar{\mu}$ associated to the given \bar{x} is unique by necessity. We shall also use the following partitioning of I :

$$I_+ = I_+(\bar{x}, \bar{\mu}) = \{i \in I \mid \bar{\mu}_i > 0\}, \quad I_0 = I_0(\bar{x}, \bar{\mu}) = \{i \in I \mid \bar{\mu}_i = 0\} = I \setminus I_+,$$

corresponding to strongly and weakly active constraints, respectively. Define

$$\Psi : \mathfrak{R}^n \times \mathfrak{R}^m \rightarrow \mathfrak{R}^n, \quad \Psi(x, \mu) = F(x) + \sum_{i=1}^m \mu_i g'_i(x),
 \tag{13}$$

which is the mapping appearing in the pure equality part of the KKT system (7). We say that $(\bar{x}, \bar{\mu})$ satisfies the second-order sufficiency condition (SOSC) if

$$\langle \Psi'_x(\bar{x}, \bar{\mu})d, d \rangle \neq 0 \quad \forall d \in K \setminus \{0\}, \tag{14}$$

where

$$K = K(\bar{x}) = \{d \in \mathfrak{R}^n \mid \langle g'_i(\bar{x}), d \rangle \leq 0, i \in I_0; \langle g'_i(\bar{x}), d \rangle = 0, i \in I_+\}. \tag{15}$$

Note that since the cone K is convex, (14) means that the quadratic form has the same nonzero sign for all $d \in K \setminus \{0\}$. The word ‘‘sufficiency’’ should not be taken literally in the setting of a general KKT system; it is used here by analogy with the optimization case, where conditions of this form (with the positive sign) are sufficient for optimality. In the case of the optimization problem corresponding to (2), K is the standard critical cone of (3) at \bar{x} , and

$$\Psi(x, \mu) = L'_x(x, \mu),$$

where

$$L : \mathfrak{R}^n \times \mathfrak{R}^m \rightarrow \mathfrak{R}, \quad L(x, \mu) = f(x) + \sum_{i=1}^m \mu_i g_i(x)$$

is the Lagrangian of (3). Then (14) with the positive sign reduces to the classical second-order sufficient condition for optimality

$$\langle L''_{xx}(\bar{x}, \bar{\mu})d, d \rangle > 0 \quad \forall d \in K \setminus \{0\}.$$

Finally, we say that strict complementarity holds at $(\bar{x}, \bar{\mu})$ if $I_0 = \emptyset$, i.e.,

$$\bar{\mu}_i > 0 \quad \forall i \in I. \tag{16}$$

We are now in position to state our first convergence result. Since we are not making any convexity/monotonicity type assumptions, even under the stated below conditions at $(\bar{x}, \bar{\mu})$, the mixed complementarity problem (8–11) (or the optimization subproblem (6)) may have solutions ‘‘of no interest’’, far from x^k (or \bar{x}). We therefore talk about the specific solution closest to x^k . This is typical in results of this nature.

Theorem 2 *Let $(\bar{x}, \bar{\mu}) \in \mathfrak{R}^n \times \mathfrak{R}^m$ be a solution of the KKT system (7). Suppose that F is differentiable and g is twice differentiable in some neighborhood of \bar{x} , and that the first derivative of F and the second derivative of g are Lipschitz continuous in this neighborhood. Suppose further that LICQ (12), SOSC (14) and the strict complementarity condition (16) are satisfied.*

Then there exists a neighborhood U of \bar{x} such that if $x^k \in U$, then the mixed complementarity problem (8–11) has a solution $(x^{k+1}, \mu^{k+1}) \in \mathfrak{R}^n \times \mathfrak{R}^m$. Moreover, if $x^0 \in U$ and, for each $k \geq 0$, x^{k+1} is the closest to x^k solution of (8–11), then there exists a neighborhood V of $\bar{\mu}$ such that (8–11) defines unique sequence $\{(x^{k+1}, \mu^{k+1})\}$ which stays in $U \times V$ and converges quadratically to $(\bar{x}, \bar{\mu})$.

Proof We first prove existence of a solution for the mixed complementarity problem (8–11), starting with (8) and (11). To this end, we shall apply the Implicit Function Theorem (Theorem 1) to the mapping $\Phi : \mathfrak{R}^n \times \mathfrak{R}^n \times \mathfrak{R}^m \rightarrow \mathfrak{R}^n \times \mathfrak{R}^m$ defined by

$$\Phi(x; y, \mu) = \begin{pmatrix} \Psi(x, \mu) + \Psi'_x(x, \mu)(y - x) \\ \mu_1(g_1(x) + \langle g'_1(x), y - x \rangle + \frac{1}{2}\langle g''_1(x)(y - x), y - x \rangle) \\ \vdots \\ \mu_m(g_m(x) + \langle g'_m(x), y - x \rangle + \frac{1}{2}\langle g''_m(x)(y - x), y - x \rangle) \end{pmatrix}, \tag{17}$$

where Ψ is given by (13). Thinking of $x \in \mathfrak{R}^n$ as a parameter, the system $\Phi(x; y, \mu) = 0$ has $n + m$ equations and $n + m$ unknowns $(y, \mu) \in \mathfrak{R}^n \times \mathfrak{R}^m$.

Since $(\bar{x}, \bar{\mu})$ is a solution of the KKT system (7), we have that $\Phi(\bar{x}; \bar{x}, \bar{\mu}) = 0$. By our smoothness hypotheses on F and g , Φ is Lipschitz continuous in a neighborhood of $(\bar{x}; \bar{x}, \bar{\mu})$. Moreover, since Φ is continuously differentiable with respect to y and μ , it easily follows that $\partial\Phi(\bar{x}; \bar{x}, \bar{\mu})$ is the set of matrices $[N, M]$, where M is given by

$$M = (\Phi'_y, \Phi'_\mu)(\bar{x}; \bar{x}, \bar{\mu}) = \begin{pmatrix} \Psi'_x(\bar{x}, \bar{\mu}) & g'_1(\bar{x}) & g'_2(\bar{x}) & \dots & g'_m(\bar{x}) \\ \bar{\mu}_1 g'_1(\bar{x})^\top & g_1(\bar{x}) & 0 & \dots & 0 \\ \bar{\mu}_2 g'_2(\bar{x})^\top & 0 & g_2(\bar{x}) & \dots & 0 \\ \vdots & \vdots & & \ddots & \vdots \\ \bar{\mu}_m g'_m(\bar{x})^\top & 0 & \dots & 0 & g_m(\bar{x}) \end{pmatrix}, \tag{18}$$

and

$$N \in \text{conv} \left\{ \lim_{l \rightarrow \infty} \Phi'_x(x^l; y^l, \mu^l) \mid (x^l; y^l, \mu^l) \rightarrow (\bar{x}; \bar{x}, \bar{\mu}), (x^l; y^l, \mu^l) \in \mathcal{N}_\Phi \right\}.$$

To apply Theorem 1, it remains to show that M is nonsingular. Suppose that $M \begin{pmatrix} v \\ w \end{pmatrix} = 0$, where $v \in \mathfrak{R}^n$ and $w \in \mathfrak{R}^m$. Then we have

$$\Psi'_x(\bar{x}, \bar{\mu})v + \sum_{i=1}^m w_i g'_i(\bar{x}) = 0, \tag{19}$$

$$\bar{\mu}_i \langle g'_i(\bar{x}), v \rangle + w_i g_i(\bar{x}) = 0, \quad i = 1, \dots, m. \tag{20}$$

Since $g_i(\bar{x}) < 0$ and $\bar{\mu}_i = 0$ for all $i \notin I$; and by the strict complementarity condition (16), $g_i(\bar{x}) = 0$ and $\bar{\mu}_i > 0$ for all $i \in I$, it follows from (20) that

$$\begin{aligned} \langle g'_i(\bar{x}), v \rangle &= 0, & \forall i \in I, \\ w_i &= 0, & \forall i \notin I. \end{aligned} \tag{21}$$

Since strict complementarity means that $I_0 = \emptyset$, from (15) and (21) we have that $v \in K$. Multiplying both sides in (19) by v , we obtain

$$\begin{aligned} 0 &= \langle \Psi'_x(\bar{x}, \bar{\mu})v, v \rangle + \sum_{i \in I} w_i \langle g'_i(\bar{x}), v \rangle + \sum_{i \notin I} w_i \langle g'_i(\bar{x}), v \rangle \\ &= \langle \Psi'_x(\bar{x}, \bar{\mu})v, v \rangle, \end{aligned}$$

where the second equality follows from (21). Since $v \in K$, SOSC (14) implies that $v = 0$. Now by (19) and (21), using also that $v = 0$, we obtain that

$$0 = \sum_{i \in I} w_i g'_i(\bar{x}).$$

Then LICQ (12) implies that $w_i = 0$ for all $i \in I$. Taking into account (21), we conclude that $w = 0$, so that $(v, w) = 0$. Hence, M is nonsingular.

Then, by Theorem 1, there exist a neighborhood U_0 of \bar{x} in \mathfrak{R}^n , a neighborhood Ω_0 of $(\bar{x}, \bar{\mu})$ in $\mathfrak{R}^n \times \mathfrak{R}^m$, and a Lipschitz continuous function $\xi : U_0 \rightarrow \Omega_0$ such that $\Phi(x; \xi(x)) = 0$ for all $x \in U_0$, where $\xi(x) = (y(x), \mu(x))$ and $\xi(\bar{x}) = (\bar{x}, \bar{\mu})$. Furthermore, ξ is unique in the sense that if $\hat{x} \in U_0$, $(\hat{y}, \hat{\mu}) \in \Omega_0$ and $\Phi(\hat{x}; \hat{y}, \hat{\mu}) = 0$, then $(\hat{y}, \hat{\mu}) = \xi(\hat{x})$.

Using the continuity of y and μ at \bar{x} and the strict complementarity condition (16), it follows that the sets

$$U_1 = \left\{ x \in U_0 \mid g_i(x) + \langle g'_i(x), y(x) - x \rangle + \frac{1}{2} \langle g''_i(x)(y(x) - x), y(x) - x \rangle < 0, \forall i \notin I \right\},$$

$$U_2 = \{x \in U_0 \mid \mu_i(x) > 0, \forall i \in I\},$$

are nonempty and open (and they contain \bar{x}). Furthermore, since Ω_0 is a neighborhood of $(\bar{x}, \bar{\mu})$, there exist a neighborhood W of \bar{x} in \mathfrak{R}^n and a neighborhood V of $\bar{\mu}$ in \mathfrak{R}^m such that $W \times V \subset \Omega_0$. Let

$$U_3 = \{x \in U_1 \cap U_2 \mid \xi(x) \in W \times V\}.$$

If $x \in U_3$, then $(y(x), \mu(x)) \in W \times V$, and since $\Phi(x; \xi(x)) = 0$, using the definitions of U_1 and U_2 , we conclude that

$$\begin{aligned} 0 &= g_i(x) + \langle g'_i(x), y(x) - x \rangle + \frac{1}{2} \langle g''_i(x)(y(x) - x), y(x) - x \rangle, \quad \forall i \in I, \\ 0 &= \mu_i(x), \quad \forall i \notin I. \end{aligned} \tag{22}$$

Now, combining $\Phi(x; \xi(x)) = 0$ with (22) and with the definitions of U_1 and U_2 , we obtain that $(y(x), \mu(x))$ is a solution of the mixed complementarity problem (8–11).

Now let $x^k \in U_3, k \geq 0$. We next show that if x^{k+1} is the solution of (8–11) closest to x^k and μ^{k+1} is the associated multiplier, then these are uniquely defined by $x^{k+1} = y(x^k)$ and $\mu^{k+1} = \mu(x^k)$. First, note that the gradients of constraints in (9) that are active at $y(x^k)$ form the set $\{g'_i(x^k) + g''_i(x^k)(y(x^k) - x^k), i \in I\}$. For x^k sufficiently close to \bar{x} , this is a small perturbation of the linearly independent set in the LICQ condition (12). Thus, it is linearly independent itself, which implies that $\mu(x^k)$ is in fact the unique multiplier associated to $y(x^k)$. Taking U_0 sufficiently small (so that U_3 is sufficiently small), it can also be seen that the solution closest to x^k (among all the solutions of (8–11)) is precisely $y(x^k)$, since it is the only solution in W . From now on, $x^k \in U_3, x^{k+1} = y(x^k)$ and $\mu^{k+1} = \mu(x^k)$.

By (8), we have that

$$\begin{aligned}
 0 &= F(x^k) + \sum_{i=1}^m \mu_i^{k+1} g'_i(x^k) \\
 &\quad + \left(F'(x^k) + \sum_{i=1}^m \mu_i^{k+1} g''_i(x^k) \right) (x^{k+1} - x^k) \\
 &= F(x^k) + \sum_{i \in I} \mu_i^k g'_i(x^k) + \sum_{i \in I} (\mu_i^{k+1} - \mu_i^k) g'_i(x^k) \\
 &\quad + \left(F'(x^k) + \sum_{i \in I} \mu_i^k g''_i(x^k) \right) (x^{k+1} - x^k) \\
 &\quad + \sum_{i \in I} (\mu_i^{k+1} - \mu_i^k) g''_i(x^k) (x^{k+1} - x^k), \tag{23}
 \end{aligned}$$

where we have taken into account that $\mu_i^{k+1} = 0$ for all $i \notin I$.

By (22), we also have that

$$0 = g_i(x^k) + \langle g'_i(x^k), x^{k+1} - x^k \rangle + \frac{1}{2} \langle g''_i(x^k)(x^{k+1} - x^k), x^{k+1} - x^k \rangle, \quad \forall i \in I. \tag{24}$$

Defining

$$H : \mathbb{R}^n \times \mathbb{R}^{|I|} \rightarrow \mathbb{R}^n \times \mathbb{R}^{|I|}, \quad H(z) = \begin{pmatrix} F(x) + \sum_{i \in I} \mu_i g'_i(x) \\ g_I(x) \end{pmatrix}, \quad z = (x, \mu_I),$$

relations (23) and (24) can be written as

$$0 = H(z^k) + H'(z^k)(z^{k+1} - z^k) + E_{k,k+1}, \tag{25}$$

where

$$E_{k,k+1} = \begin{pmatrix} \sum_{i \in I} (\mu_i^{k+1} - \mu_i^k) g'_i(x^k) (x^{k+1} - x^k) \\ \frac{1}{2} \langle g''_i(x^k)(x^{k+1} - x^k), x^{k+1} - x^k \rangle, i \in I \end{pmatrix}.$$

Note that (25) is not a Newton equation, as it is not linear with respect to z^{k+1} . However, we shall relate it, a posteriori, to a specially perturbed Newton type iterative process. The rest of the proof makes this precise and establishes the quadratic rate of convergence.

First, note that $H(\bar{z}) = 0$. By a proof similar to that for the nonsingularity of the matrix M defined in (18), it can be seen that the matrix

$$H'(\bar{z}) = \begin{pmatrix} \Psi'_x(\bar{x}, \bar{\mu}) & g'_I(\bar{x})^\top \\ g'_I(\bar{x}) & 0 \end{pmatrix}$$

is nonsingular (in the above formula for $H'(\bar{z})$, we have used the fact that $\bar{\mu}_i = 0$ for all $i \notin I$). Since $H'(\bar{z})$ is nonsingular, there exists a constant $\eta > 0$ such that

$$\bar{z} \in \tilde{U}_4 = \{z \in \mathbb{R}^{n+|I|} \mid \|H'(z)^{-1}\| < \eta\}.$$

Since F' and $g_i'', i = 1, \dots, m$, are Lipschitz continuous functions in a neighborhood of \bar{x} , taking $\rho > 0$ sufficiently small, there exists a constant $L > 0$ such that $\|H'(w) - H'(z)\| \leq L\|w - z\|$ for all $w, z \in B(\bar{z}, \rho)$, where $B(\bar{z}, \rho)$ denotes the open ball in $\mathfrak{R}^{n+|I|}$ with center at \bar{z} and radius ρ .

We next show that if $z^k \in B(\bar{z}, \rho)$ then there exists a constant $c > 0$ such that $\|E_{k,k+1}\| \leq c\|z^{k+1} - z^k\|^2$ for all $k \geq 1$, where $z^k = (x^k, \mu_I^k)$. Since $g_i'', i = 1, \dots, m$, are continuous at \bar{x} , there exists a constant $\gamma > 0$ such that $\|g_i''(x)\| \leq \gamma, i = 1, \dots, m$, for all $x \in \mathfrak{R}^n$ such that $\|x - \bar{x}\| \leq \rho$. Since $z^k \in B(\bar{z}, \rho)$ implies $\|x^k - \bar{x}\| < \rho$, we have that

$$\begin{aligned} \|E_{k,k+1}\| &\leq \sqrt{n}\gamma\|\mu_I^{k+1} - \mu_I^k\| \|x^{k+1} - x^k\| + \frac{\gamma}{2} \sum_{i \in I} \|x^{k+1} - x^k\|^2 \\ &\leq \sqrt{n}\gamma(\max\{\|\mu_I^{k+1} - \mu_I^k\|, \|x^{k+1} - x^k\|\})^2 + \frac{\gamma m}{2} \|x^{k+1} - x^k\|^2 \\ &\leq \sqrt{n}\gamma\|z^{k+1} - z^k\|^2 + \frac{\gamma m}{2} \|x^{k+1} - x^k\|^2 \\ &\leq \gamma(\sqrt{n} + m/2)\|z^{k+1} - z^k\|^2 \\ &= c\|z^{k+1} - z^k\|^2, \end{aligned} \tag{26}$$

where the monotonicity of the norm has been used repeatedly.

Let $r = 1/(2\eta(L + 4c))$, and define $U_5 = \{x \in U_3 \mid \|y(x) - \bar{x}\|^2 + \|\mu(x) - \bar{\mu}\|^2 < r^2\}$, $\tilde{U}_5 = U_5 \times \mathfrak{R}^{|I|}$. Then there exists $\delta > 0$ such that $B(\bar{z}, \delta) \subset \tilde{U}_4 \cap \tilde{U}_5$.

Let $\varepsilon = \min\{\delta, r, \rho\}$, and define

$$U = \{x \in \mathfrak{R}^n \mid \|y(x) - \bar{x}\|^2 + \|\mu(x) - \bar{\mu}\|^2 < \varepsilon^2\}.$$

Then $x^0 \in U$ implies that $\|z^1 - \bar{z}\| < \varepsilon$.

Now, proceeding by induction, we will show that if $\|z^k - \bar{z}\| < \varepsilon$ then $\|z^{k+1} - \bar{z}\| < \varepsilon$. By the construction of the set U , if $\|z^k - \bar{z}\| < \varepsilon$ then the following properties hold:

$$\|H'(z^k)^{-1}\| < \eta, \tag{27}$$

$$\|z^k - \bar{z}\| < r = \frac{1}{2\eta(L + 4c)}, \tag{28}$$

$$\|z^{k+1} - \bar{z}\| < r < \frac{1}{4\eta c}, \tag{29}$$

where in (27) we use that $z^k \in \tilde{U}_4$, (28) holds since $\varepsilon \leq r$, and (29) follows from $x^k \in U_5$. Also, because $x^k \in U_3$, by (25) it follows that

$$z^{k+1} = z^k - H'(z^k)^{-1}(H(z^k) - H(\bar{z}) + E_{k,k+1}),$$

where $H(\bar{z}) = 0$ was also taken into account. We further obtain

$$\begin{aligned} \|z^{k+1} - \bar{z}\| &= \|z^k - \bar{z} - H'(z^k)^{-1}(H(z^k) - H(\bar{z}) + E_{k,k+1})\| \\ &\leq \|H'(z^k)^{-1}\| \|H'(z^k)(z^k - \bar{z}) + H(\bar{z}) - H(z^k) - E_{k,k+1}\| \end{aligned}$$

$$\begin{aligned}
 &\leq \eta \left\| \int_0^1 [H'(z^k) - H'(\bar{z} + t(z^k - \bar{z}))](z^k - \bar{z})dt - E_{k,k+1} \right\| \\
 &\leq \eta \left(L \|z^k - \bar{z}\|^2 \int_0^1 (1-t)dt + c \|z^{k+1} - z^k\|^2 \right) \\
 &\leq \frac{\eta L}{2} \|z^k - \bar{z}\|^2 + 2\eta c \|z^{k+1} - \bar{z}\|^2 + 2\eta c \|z^k - \bar{z}\|^2 \\
 &\leq \eta \left(\frac{L}{2} + 2c \right) \|z^k - \bar{z}\|^2 + \frac{1}{2} \|z^{k+1} - \bar{z}\|,
 \end{aligned}$$

where the second inequality follows from (27) and the Mean-Value Theorem, in the third inequality we use the Lipschitz continuity of H' and (26), for the fourth inequality we use that $\|z^{k+1} - z^k\|^2 \leq 2(\|z^{k+1} - \bar{z}\|^2 + \|z^k - \bar{z}\|^2)$, and the fifth inequality follows from $2\eta c \|z^{k+1} - \bar{z}\| < \frac{1}{2}$, which is ensured by (29).

Now, rearranging terms in the relation above, we deduce that

$$\|z^{k+1} - \bar{z}\| \leq \eta(L + 4c) \|z^k - \bar{z}\|^2. \tag{30}$$

Then, by (28), we have $\|z^{k+1} - \bar{z}\| < \frac{1}{2} \|z^k - \bar{z}\| < \varepsilon$.

In consequence, if $x^0 \in U$ then $(x^{k+1}, \mu^{k+1}) \in U \times V$ for all $k \geq 0$, and since $\mu_i^{k+1} = \bar{\mu}_i = 0$ for all $i \notin I$, we have

$$\|(x^{k+1}, \mu^{k+1}) - (\bar{x}, \bar{\mu})\| = \|z^{k+1} - \bar{z}\| < \frac{1}{2} \|z^k - \bar{z}\| < \dots < \left(\frac{1}{2}\right)^k \|z^1 - \bar{z}\|,$$

so that $\{(x^{k+1}, \mu^{k+1})\}$ converges to $(\bar{x}, \bar{\mu})$. Then, by (30), we conclude that the rate of convergence is quadratic. □

3 Primal superlinear convergence

As is well known, quadratic convergence of $\{(x^k, \mu^k)\}$ to $(\bar{x}, \bar{\mu})$ does not imply even superlinear or linear convergence of $\{x^k\}$ to \bar{x} . Assuming that some type of convergence occurs, we next give necessary and sufficient conditions for superlinear convergence of the primal sequence. This condition is of the Dennis-Moré type [6], and allows for derivative approximations. Specific rules for quasi-Newton updates are certainly of great interest, but these are beyond the scope of this paper. But we note that our analysis covers the situations where computing the derivatives involves computational work and the precision of approximation can be controlled. Such is the case, for example, when the derivatives are approximated by finite-difference procedures. The accuracy parameter can be controlled using estimates for the distance to the solution via error bounds (see [7] for a discussion of error bounds for variational problems and [5, 9] for detailed comparisons in the context of KKT systems specifically). In particular, those estimates give some idea of how precise should be the approximation in order to conform to conditions (37) or (38) below.

Let H_k be some approximation of $F'(x^k)$ and $G_{i,k}$ be some approximation of $g_i''(x^k)$ (of course, this includes the possibility of exact derivatives, as in the setting

of Sect. 2). We consider a sequence $\{(x^k, \mu^k)\}$ generated by the following process. Given $x^k \in \mathfrak{N}^n$, $H_k \in \mathfrak{N}^{n \times n}$ and $G_{i,k} \in \mathfrak{N}^{n \times n}$, $i = 1, \dots, m$, find $(x^{k+1}, \mu^{k+1}) \in \mathfrak{N}^n \times \mathfrak{N}^m$ such that

$$F(x^k) + H_k(x^{k+1} - x^k) + \sum_{i=1}^m \mu_i^{k+1}(g'_i(x^k) + G_{i,k}(x^{k+1} - x^k)) = 0, \tag{31}$$

and for all $i = 1, \dots, m$, it holds that

$$g_i(x^k) + \langle g'_i(x^k), x^{k+1} - x^k \rangle + \frac{1}{2} \langle G_{i,k}(x^{k+1} - x^k), x^{k+1} - x^k \rangle \leq 0, \tag{32}$$

$$\mu_i^{k+1} \geq 0, \tag{33}$$

$$\mu_i^{k+1}(g_i(x^k) + \langle g'_i(x^k), x^{k+1} - x^k \rangle + \frac{1}{2} \langle G_{i,k}(x^{k+1} - x^k), x^{k+1} - x^k \rangle) = 0. \tag{34}$$

In the sequel, we shall consider separately the two possible cases in SOSOC (14) (i.e., when (14) holds with the positive sign and when it holds with the negative sign). Note also that since the cone K is closed, those two cases can be stated as follows: there exists $t > 0$ such that

$$\langle \Psi'_x(\bar{x}, \bar{\mu})d, d \rangle \geq t \|d\|^2 \quad \forall d \in K, \tag{35}$$

or

$$-\langle \Psi'_x(\bar{x}, \bar{\mu})d, d \rangle \geq t \|d\|^2 \quad \forall d \in K. \tag{36}$$

Theorem 3 *Let $(\bar{x}, \bar{\mu}) \in \mathfrak{N}^n \times \mathfrak{N}^m$ be a solution of the KKT system (7). Suppose that F is differentiable and g is twice differentiable in some neighborhood of \bar{x} . Suppose further that a sequence $\{(x^k, \mu^k)\}$, generated according to (31–34) with uniformly bounded $G_{i,k}$, $i = 1, \dots, m$, converges to $(\bar{x}, \bar{\mu})$.*

If $\{x^k\}$ converges superlinearly to \bar{x} then

$$P_K[(\Psi'_x(\bar{x}, \bar{\mu}) - M_k)(x^{k+1} - x^k)] = o(\|x^{k+1} - x^k\|), \tag{37}$$

where $P_K[\cdot]$ denotes the orthogonal projector onto the cone K defined in (15) and

$$M_k = H_k + \sum_{i=1}^m \mu_i^{k+1} G_{i,k}.$$

Conversely, if LICQ (12) and SOSOC (14) are satisfied, then the rate of convergence of $\{x^k\}$ to \bar{x} is superlinear if (35) and (37) hold, or if (36) holds and

$$P_K[(M_k - \Psi'_x(\bar{x}, \bar{\mu}))(x^{k+1} - x^k)] = o(\|x^{k+1} - x^k\|). \tag{38}$$

Proof Denote $d^k = x^{k+1} - x^k$. By (31), we have that

$$0 = F(x^k) + H_k d^k + \sum_{i=1}^m \mu_i^{k+1}(g'_i(x^k) + G_{i,k} d^k) = \Psi(x^k, \mu^{k+1}) + M_k d^k. \tag{39}$$

Also, we have that

$$\begin{aligned} \Psi(x^k, \bar{\mu}) &= \Psi(x^k, \mu^{k+1}) + \sum_{i=1}^m (\bar{\mu}_i - \mu_i^{k+1}) g'_i(x^k) \\ &= \Psi(x^k, \mu^{k+1}) + \sum_{i=1}^m (\bar{\mu}_i - \mu_i^{k+1}) g'_i(\bar{x}) + o(\|x^k - \bar{x}\|) \\ &= -M_k d^k + \sum_{i=1}^m (\bar{\mu}_i - \mu_i^{k+1}) g'_i(\bar{x}) + o(\|x^k - \bar{x}\|), \end{aligned} \tag{40}$$

where the last equality is by (39).

Suppose first that $\{x^k\}$ converges to \bar{x} superlinearly, i.e., $x^{k+1} - \bar{x} = o(\|x^k - \bar{x}\|)$. Since $\Psi(\bar{x}, \bar{\mu}) = 0$, it holds that

$$\begin{aligned} \Psi(x^k, \bar{\mu}) &= \Psi(\bar{x}, \bar{\mu}) + \Psi'_x(\bar{x}, \bar{\mu})(x^k - \bar{x}) + o(\|x^k - \bar{x}\|) \\ &= -\Psi'_x(\bar{x}, \bar{\mu})d^k + \Psi'_x(\bar{x}, \bar{\mu})(x^{k+1} - \bar{x}) + o(\|x^k - \bar{x}\|) \\ &= -\Psi'_x(\bar{x}, \bar{\mu})d^k + o(\|x^k - \bar{x}\|). \end{aligned} \tag{41}$$

Combining (41) and (40), we obtain

$$(\Psi'_x(\bar{x}, \bar{\mu}) - M_k)d^k = \sum_{i=1}^m (\mu_i^{k+1} - \bar{\mu}_i) g'_i(\bar{x}) + o(\|x^k - \bar{x}\|). \tag{42}$$

Taking into account that each $G_{i,k}$ is uniformly bounded and using the continuity argument in (34), we conclude that for all sufficiently large k , it holds that $\mu_i^{k+1} - \bar{\mu}_i = 0, \forall i \notin I$, and $\mu_i^{k+1} - \bar{\mu}_i = \mu_i^{k+1} \geq 0, \forall i \in I_0$. Then, by (42), for all $v \in K$ it holds that

$$\begin{aligned} \langle (\Psi'_x(\bar{x}, \bar{\mu}) - M_k)d^k - o(\|x^k - \bar{x}\|), v \rangle &= \sum_{i=1}^m (\mu_i^{k+1} - \bar{\mu}_i) \langle g'_i(\bar{x}), v \rangle \\ &= \sum_{i \in I} (\mu_i^{k+1} - \bar{\mu}_i) \langle g'_i(\bar{x}), v \rangle \\ &= \sum_{i \in I_0} (\mu_i^{k+1} - \bar{\mu}_i) \langle g'_i(\bar{x}), v \rangle \\ &= \sum_{i \in I_0} \mu_i^{k+1} \langle g'_i(\bar{x}), v \rangle \leq 0, \end{aligned} \tag{43}$$

where we have used that $\langle g'_i(\bar{x}), v \rangle = 0, \forall i \in I_+$, $\langle g'_i(\bar{x}), v \rangle \leq 0, \forall i \in I_0$ (see (15)). By properties of projection operator onto a convex cone, inequality (43) means that

$$P_K[(\Psi'_x(\bar{x}, \bar{\mu}) - M_k)d^k - o(\|x^k - \bar{x}\|)] = 0.$$

Then, by nonexpansiveness of the projection operator, it follows that

$$\begin{aligned} \|P_K[(\Psi'_x(\bar{x}, \bar{\mu}) - M_k)d^k]\| &= \|P_K[(\Psi'_x(\bar{x}, \bar{\mu}) - M_k)d^k - o(\|x^k - \bar{x}\|)] \\ &\quad - P_K[(\Psi'_x(\bar{x}, \bar{\mu}) - M_k)d^k]\| \\ &= o(\|x^k - \bar{x}\|). \end{aligned}$$

It remains to show that $o(\|x^k - \bar{x}\|) = o(\|d^k\|)$. For this, note that

$$\begin{aligned} \frac{o(\|x^k - \bar{x}\|)}{\|d^k\|} &\leq \frac{o(\|x^k - \bar{x}\|)}{\|x^k - \bar{x}\| - \|x^{k+1} - \bar{x}\|} = \frac{o(\|x^k - \bar{x}\|)}{\|x^k - \bar{x}\| - o(\|x^k - \bar{x}\|)} \\ &= \frac{o(\|x^k - \bar{x}\|)/\|x^k - \bar{x}\|}{1 - o(\|x^k - \bar{x}\|)/\|x^k - \bar{x}\|} \rightarrow 0 \quad \text{as } k \rightarrow \infty. \end{aligned}$$

This concludes the proof of (37).

We now prove the sufficiency part, assuming LICQ and SOSC. Denote

$$\Gamma_{i,k} = g_i(x^k) + \langle g'_i(x^k), d^k \rangle + \frac{1}{2} \langle G_{i,k}d^k, d^k \rangle.$$

By the continuity argument (taking also into account uniform boundedness of $G_{i,k}$), $\{\Gamma_{i,k}\}$ converges to $g_i(\bar{x})$, as $k \rightarrow \infty$. Thus for all k sufficiently large, taking into account (34), we have that

$$\begin{aligned} \Gamma_{i,k} &< 0, & \mu_i^{k+1} &= 0, & \forall i \notin I, \\ \Gamma_{i,k} &= 0, & \mu_i^{k+1} &> 0, & \forall i \in I_+. \end{aligned} \tag{44}$$

By the Mean-Value Theorem, for each $i = 1, \dots, m$, there exists a vector $z^{i,k}$ in the line segment joining x^k and \bar{x} , such that

$$g_i(x^k) = g_i(\bar{x}) + \langle g'_i(\bar{x}), x^k - \bar{x} \rangle + \frac{1}{2} \langle g''_i(z^{i,k})(x^k - \bar{x}), x^k - \bar{x} \rangle.$$

Note that $\{z^{i,k}\}$ converges to \bar{x} when $k \rightarrow \infty$. For $i \in I$, we then obtain

$$\begin{aligned} \Gamma_{i,k} &= g_i(\bar{x}) + \langle g'_i(\bar{x}), x^k - \bar{x} \rangle + \frac{1}{2} \langle g''_i(z^{i,k})(x^k - \bar{x}), x^k - \bar{x} \rangle \\ &\quad + \langle g'_i(x^k), d^k \rangle + \frac{1}{2} \langle G_{i,k}d^k, d^k \rangle \\ &= \langle g'_i(\bar{x}), x^{k+1} - \bar{x} \rangle + w_i^k, \end{aligned} \tag{45}$$

where

$$w_i^k = \langle g'_i(x^k) - g'_i(\bar{x}), d^k \rangle + \frac{1}{2} \langle g''_i(z^{i,k})(x^k - \bar{x}), x^k - \bar{x} \rangle + \frac{1}{2} \langle G_{i,k}d^k, d^k \rangle.$$

Clearly,

$$w_i^k = o(\|x^k - \bar{x}\|) + o(\|d^k\|).$$

By LICQ (12), for each k , there exists $u^k \in \mathfrak{N}^n$ such that

$$g'_i(\bar{x})u^k = w^k_i, \quad \text{where } u^k = o(\|x^k - \bar{x}\|) + o(\|d^k\|). \tag{46}$$

Let $v^k = x^{k+1} - \bar{x} + u^k$. Then by (46) and (45), we have

$$\langle g'_i(\bar{x}), v^k \rangle = \langle g'_i(\bar{x}), x^{k+1} - \bar{x} \rangle + w^k_i = \Gamma_{i,k} \quad \forall i \in I. \tag{47}$$

Since $\Gamma_{i,k} = 0, \forall i \in I_+$ (by (44)) and $\Gamma_{i,k} \leq 0, \forall i \in I_0$ (by (32)), relation (47) shows that $v^k \in K$. Since $\Psi(\bar{x}, \bar{\mu}) = 0$, we have that

$$0 = \langle \Psi(\bar{x}, \bar{\mu}), v^k \rangle = \langle F(\bar{x}), v^k \rangle + \sum_{i \in I_+} \bar{\mu}_i \langle g'_i(\bar{x}), v^k \rangle = \langle F(\bar{x}), v^k \rangle.$$

We then obtain

$$\begin{aligned} \langle \Psi(\bar{x}, \mu^{k+1}), v^k \rangle &= \langle F(\bar{x}), v^k \rangle + \sum_{i=1}^m \mu_i^{k+1} \langle g'_i(\bar{x}), v^k \rangle \\ &= \sum_{i \notin I} \mu_i^{k+1} \langle g'_i(\bar{x}), v^k \rangle + \sum_{i \in I} \mu_i^{k+1} \Gamma_{i,k} \\ &= 0, \end{aligned} \tag{48}$$

where we have used (47) for the second equality, and (44) with (34) for the last equality.

Also,

$$\begin{aligned} \Psi(x^{k+1}, \mu^{k+1}) &= \Psi(x^k, \mu^{k+1}) + \Psi'_x(x^k, \mu^{k+1})d^k + o(\|d^k\|) \\ &= (\Psi'_x(x^k, \mu^{k+1}) - M_k)d^k + o(\|d^k\|) \\ &= (\Psi'_x(\bar{x}, \bar{\mu}) - M_k)d^k + (\Psi'_x(x^k, \mu^{k+1}) - \Psi'_x(\bar{x}, \bar{\mu}))d^k + o(\|d^k\|) \\ &= (\Psi'_x(\bar{x}, \bar{\mu}) - M_k)d^k + o(\|d^k\|), \end{aligned} \tag{49}$$

where (39) has been used in the second equality, and the last equality is by the continuity of Ψ'_x . Let $p^k = v^k / \|v^k\|$. Multiplying both sides in (49) by p^k (which is bounded), we conclude that

$$\langle \Psi(x^{k+1}, \mu^{k+1}), p^k \rangle = \langle (\Psi'_x(\bar{x}, \bar{\mu}) - M_k)d^k, p^k \rangle + o(\|d^k\|). \tag{50}$$

On the other hand,

$$\begin{aligned} \langle \Psi(x^{k+1}, \mu^{k+1}), p^k \rangle &= \langle \Psi(\bar{x}, \mu^{k+1}), p^k \rangle \\ &\quad + \langle \Psi'_x(\bar{x}, \mu^{k+1})(x^{k+1} - \bar{x}), p^k \rangle + o(\|x^{k+1} - \bar{x}\|) \\ &= \langle \Psi'_x(\bar{x}, \mu^{k+1})(x^{k+1} - \bar{x}), p^k \rangle + o(\|x^{k+1} - \bar{x}\|) \\ &= \langle \Psi'_x(\bar{x}, \bar{\mu})(x^{k+1} - \bar{x}), p^k \rangle + o(\|x^{k+1} - \bar{x}\|), \end{aligned} \tag{51}$$

where the second equality follows from (48), and the last follows from the continuity of Ψ'_x and boundedness of $\{p^k\}$.

Combining (50) and (51), we conclude that

$$\langle \Psi'_x(\bar{x}, \bar{\mu})(x^{k+1} - \bar{x}), p^k \rangle = \langle (\Psi'_x(\bar{x}, \bar{\mu}) - M_k)d^k, p^k \rangle + o(\|d^k\|) + o(\|x^{k+1} - \bar{x}\|). \tag{52}$$

Suppose now that SOSC holds. Then for the case (36) and (38), by (52) and (46), we have

$$\begin{aligned} t\|v^k\| &\leq -\langle \Psi'_x(\bar{x}, \bar{\mu})v^k, p^k \rangle \\ &= \langle \Psi'_x(\bar{x}, \bar{\mu})(\bar{x} - x^{k+1}), p^k \rangle - \langle \Psi'_x(\bar{x}, \bar{\mu})u^k, p^k \rangle \\ &= \langle (M_k - \Psi'_x(\bar{x}, \bar{\mu}))d^k, p^k \rangle + o(\|d^k\|) + o(\|x^{k+1} - \bar{x}\|) + o(\|x^k - \bar{x}\|) \\ &\leq \langle P_K[(M_k - \Psi'_x(\bar{x}, \bar{\mu}))d^k], p^k \rangle + o(\|d^k\|) + o(\|x^{k+1} - \bar{x}\|) + o(\|x^k - \bar{x}\|) \\ &= o(\|d^k\|) + o(\|x^{k+1} - \bar{x}\|) + o(\|x^k - \bar{x}\|), \end{aligned}$$

where the second inequality follows from the fact that for any closed convex cone C and $v \in C$, it holds that $\langle x, v \rangle \leq \langle P_C[x], v \rangle \forall x$. Similarly, for the case (35) and (37) we obtain

$$\begin{aligned} t\|v^k\| &\leq \langle \Psi'_x(\bar{x}, \bar{\mu})v^k, p^k \rangle \\ &= \langle (\Psi'_x(\bar{x}, \bar{\mu}) - M_k)d^k, p^k \rangle + o(\|d^k\|) + o(\|x^{k+1} - \bar{x}\|) + o(\|x^k - \bar{x}\|) \\ &\leq \langle P_K[(\Psi'_x(\bar{x}, \bar{\mu}) - M_k)d^k], p^k \rangle + o(\|d^k\|) + o(\|x^{k+1} - \bar{x}\|) + o(\|x^k - \bar{x}\|) \\ &= o(\|d^k\|) + o(\|x^{k+1} - \bar{x}\|) + o(\|x^k - \bar{x}\|). \end{aligned}$$

Summarizing, in both cases $v^k = o(\|d^k\|) + o(\|x^{k+1} - \bar{x}\|) + o(\|x^k - \bar{x}\|)$.

Since $x^{k+1} - \bar{x} = v^k - u^k = o(\|x^{k+1} - \bar{x}\|) + o(\|x^k - \bar{x}\|) + o(\|d^k\|)$, there exists a sequence $\{t_k\}$ converging to 0 such that

$$\begin{aligned} \|x^{k+1} - \bar{x}\| &\leq t_k(\|x^{k+1} - \bar{x}\| + \|x^k - \bar{x}\| + \|d^k\|) \\ &\leq 2t_k(\|x^{k+1} - \bar{x}\| + \|x^k - \bar{x}\|). \end{aligned} \tag{53}$$

Since $t_k < 1/2$ for k sufficiently large, rearranging terms in (53) we obtain

$$\frac{\|x^{k+1} - \bar{x}\|}{\|x^k - \bar{x}\|} \leq \frac{2t_k}{1 - 2t_k} \rightarrow 0 \quad \text{as } k \rightarrow \infty.$$

In consequence, $x^{k+1} - \bar{x} = o(\|x^k - \bar{x}\|)$, i.e., $\{x^k\}$ converges superlinearly to \bar{x} . \square

In particular, Theorem 3 shows superlinear convergence of $\{x^k\}$ to \bar{x} in the setting of Theorem 2, where $H_k = F'(x^k)$, $G_{i,k} = g_i''(x^k)$, $i = 1, \dots, m$, so that $M_k = H_k + \sum_{i=1}^m \mu_i^{k+1} G_{i,k} \rightarrow F'(\bar{x}) + \sum_{i=1}^m \bar{\mu}_i g_i''(\bar{x}) = \Psi'_x(\bar{x}, \bar{\mu})$ as $k \rightarrow \infty$. In this case, conditions (37) and (38) are automatically satisfied.

We also note that in the setting of Theorem 2 (or more generally, when cone K is a subspace), we do not have to consider separately the two cases of SOS (35) and (36) and the two cases of the Dennis-Moré condition (37) and (38). Indeed, when K is a subspace, we have $\langle x, v \rangle = \langle P_K[x], v \rangle$ for all $v \in K$. We can further state the SOS (14) as

$$|\langle \Psi'_x(\bar{x}, \bar{\mu})v, v \rangle| \geq t \|v\|^2 \quad \forall v \in K,$$

and modify the corresponding parts of the proof of Theorem 3, as follows.

For the necessity part, note that for any $x \in \mathfrak{R}^n$, there exists the unique decomposition $x = v + v^*$ with $v = P_K[x] \in K$ and $v^* \in K^\perp$. Evidently, changing the sign, one has $-x = -v - v^*$, where $-v = P_K[-x] \in K$ and $-v^* \in K^\perp$. Hence, $\|P_K[x]\| = \|P_K[-x]\|$ for any $x \in \mathfrak{R}^n$. It follows that in this case, conditions (37) and (38) are equivalent.

For the sufficiency part, we have that

$$\begin{aligned} t \|v^k\| &\leq |\langle \Psi'_x(\bar{x}, \bar{\mu})v^k, p^k \rangle| \\ &\leq |\langle (\Psi'_x(\bar{x}, \bar{\mu}) - M_k)d^k, p^k \rangle| + o(\|d^k\|) + o(\|x^{k+1} - \bar{x}\|) + o(\|x^k - \bar{x}\|) \\ &= |\langle P_K[(\Psi'_x(\bar{x}, \bar{\mu}) - M_k)d^k], p^k \rangle| + o(\|d^k\|) + o(\|x^{k+1} - \bar{x}\|) + o(\|x^k - \bar{x}\|) \\ &= o(\|d^k\|) + o(\|x^{k+1} - \bar{x}\|) + o(\|x^k - \bar{x}\|), \end{aligned}$$

and the rest of the proof applies.

4 Concluding remarks

We have established a new result on the quadratic convergence of the primal-dual sequence of the sequential quadratically-constrained quadratic-programming method. A necessary and sufficient characterization of the superlinear convergence of the primal sequence has also been provided. Additionally, the class of methods under consideration has been extended from the optimization setting to the more general variational problems.

References

1. Anitescu, M.: A superlinearly convergent sequential quadratically constrained quadratic programming algorithm for degenerate nonlinear programming. *SIAM J. Optim.* **12**, 949–978 (2002)
2. Boggs, B.T., Tolle, J.W.: Sequential quadratic programming. *Acta Numer.* **4**, 1–51 (1996)
3. Bonnans, J.F., Shapiro, A.: *Perturbation Analysis of Optimization Problems*. Springer, New York (2000)
4. Clarke, F.H.: *Optimization and Nonsmooth Analysis*. SIAM, Philadelphia (1990)
5. Daryina, A.N., Izmailov, A.F., Solodov, M.V.: A class of active-set Newton methods for mixed complementarity problems. *SIAM J. Optim.* **15**, 409–429 (2004)
6. Dennis, J.E., Moré, J.J.: A characterization of superlinear convergence and its application to quasi-Newton methods. *Math. Comput.* **28**, 549–560 (1974)
7. Facchinei, F., Pang, J.-S.: *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer, New York (2003)

8. Fukushima, M., Luo, Z.-Q., Tseng, P.: A sequential quadratically constrained quadratic programming method for differentiable convex minimization. *SIAM J. Optim.* **13**, 1098–1119 (2003)
9. Izmailov, A.F., Solodov, M.V.: Karush-Kuhn-Tucker systems: regularity conditions, error bounds and a class of Newton-type methods. *Math. Program.* **95**, 631–650 (2003)
10. Joseph, N.H.: Newton's method for generalized equations. Technical Summary Report 1965, Mathematics Research Center, University of Wisconsin, Madison, Wisconsin (1979)
11. Kruk, S., Wolkowicz, H.: Sequential, quadratically constrained, quadratic programming for general nonlinear programming. In: Wolkowicz, H., Saigal, R., Vandenberghe, L. (eds.) *Handbook of Semidefinite Programming*, pp. 563–575. Kluwer Academic, Dordrecht (2000)
12. Lobo, M.S., Vandenberghe, L., Boyd, S., Lebret, H.: Applications of second-order cone programming. *Linear Algebra Appl.* **284**, 193–228 (1998)
13. Maratos, N.: Exact penalty function algorithms for finite dimensional and control optimization problems. Ph.D. thesis, Imperial College, University of London (1978)
14. Monteiro, R.D.C., Tsuchiya, T.: Polynomial convergence of primal-dual algorithms for the second-order cone programs based on the MZ-family of directions. *Math. Program.* **88**, 61–83 (2000)
15. Nesterov, Y.E., Nemirovskii, A.S.: *Interior Point Polynomial Methods in Convex Programming: Theory and Applications*. SIAM, Philadelphia (1993)
16. Panin, V.M.: A second-order method for discrete min-max problem. *USSR Comput. Math. Math. Phys.* **19**, 90–100 (1979)
17. Panin, V.M.: Some methods of solving convex programming problems. *USSR Comput. Math. Math. Phys.* **21**, 57–72 (1981)
18. Powell, M.J.D.: Variable metric methods for constrained optimization. In: Bachem, A., Grötschel, M., Korte, B. (eds.) *Mathematical Programming: The State of Art*, pp. 288–311. Springer, Berlin (1983)
19. Solodov, M.V.: On the sequential quadratically constrained quadratic programming methods. *Math. Oper. Res.* **29**, 64–79 (2004)
20. Tsuchiya, T.: A convergence analysis of the scale-invariant primal-dual path-following algorithms for second-order cone programming. *Optim. Methods Softw.* **11**, 141–182 (1999)
21. Wiest, E.J., Polak, E.: A generalized quadratic programming-based phase-I–phase-II method for inequality-constrained optimization. *Appl. Math. Optim.* **26**, 223–252 (1992)