Comput Optim Appl (2016) 63:425–459
DOI 10.1007/s10589-015-9782-0

Convergence conditions for Newton-type methods applied to complementarity systems with nonisolated solutions

Andreas Fischer[1] · **Markus Herrich**[1] · **Alexey F. Izmailov**[2] · **Mikhail V. Solodov**[3]

Received: 15 May 2015 / Published online: 20 August 2015
© Springer Science+Business Media New York 2015

Abstract We consider a class of Newton-type methods that are designed for the difficult case when solutions need not be isolated, and the equation mapping need not be differentiable at the solutions. We show that the only structural assumption needed for rapid local convergence of those algorithms applied to PC^1-equations is the piecewise error bound, i.e., a local error bound holding for the branches of the solution set resulting from partitions of the bi-active complementarity indices. The latter error bound is implied by various piecewise constraint qualifications, including relatively weak ones. We apply our results to KKT systems arising from optimization or variational problems, and from generalized Nash equilibrium problems. In the first case, we show convergence if the dual part of the solution is a noncritical Lagrange multiplier, and in the second case convergence follows under a relaxed constant rank condition. In both cases, previously available results are improved.

✉ Andreas Fischer
andreas.fischer@tu-dresden.de

Markus Herrich
markus.herrich@tu-dresden.de

Alexey F. Izmailov
izmaf@ccas.ru

Mikhail V. Solodov
solodov@impa.br

[1] Department of Mathematics, Technische Universität Dresden, 01062 Dresden, Germany

[2] OR Department, VMK Faculty, Lomonosov Moscow State University, MSU, Uchebniy Korpus 2, Leninskiye Gory, 119991 Moscow, Russia

[3] IMPA – Instituto Nacional de Matemática Pura e Aplicada, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ 22460-320, Brazil

Keywords Complementarity condition · KKT system · Error bound · Generalized Nash equilibrium problem · LP-Newton method · Levenberg–Marquardt method

1 Introduction

The main contributions of this paper are concerned with local analysis of Newton-type methods for solving the following system of equations and inequalities:

$$a(z) = 0, \quad b(z) \geq 0, \quad c(z) \geq 0, \ d(z) \geq 0, \ c(z)^\top d(z) = 0, \tag{1}$$

where the functions $a : \mathbb{R}^n \to \mathbb{R}^p$, $b : \mathbb{R}^n \to \mathbb{R}^q$, $c : \mathbb{R}^n \to \mathbb{R}^r$, and $d : \mathbb{R}^n \to \mathbb{R}^r$ are differentiable with locally Lipschitz-continuous derivatives. We call the problem (1) *complementarity system* because of the complementarity constraints appearing therein (given by the last three relations). There are many problems which are naturally stated as (1), or can be reduced to this form. Some examples are: (mixed) complementarity problems and Karush–Kuhn–Tucker (KKT) systems arising from optimization, variational problems, or generalized Nash equilibrium problems; see [10, Chapter 1], [21, Chapter 1], and also Sects. 6 and 7 below.

A good number of efficient techniques for solving problem (1) and its special cases go first through reformulating it as a constrained system of equations:

$$F(z) = 0 \quad \text{s.t.} \quad z \in \Omega, \tag{2}$$

with a continuous function $F : \mathbb{R}^n \to \mathbb{R}^m$ and a nonempty and closed set $\Omega \subseteq \mathbb{R}^n$. There are many ways of casting (1) as a constrained system of Eq. (2), depending on the choice of the function F *and* of the set Ω. In this paper, we shall mainly consider

$$F(z) := \begin{pmatrix} a(z) \\ \min\{c(z), d(z)\} \end{pmatrix} = 0 \quad \text{s.t.} \quad z \in \Omega, \tag{3}$$

where

$$\Omega := \left\{ z \in \mathbb{R}^n \mid b(z) \geq 0, \ c(z) \geq 0, \ d(z) \geq 0 \right\}, \tag{4}$$

and the minimum in (3) is taken componentwise. Under our standing assumptions above, the set Ω is obviously closed. It is also easy to see that every solution of (1) is a solution of (3), and vice versa. Note also that, since the min-operation fully characterizes complementarity (including the signs), the inclusion of the constraints $c(z) \geq 0$ and $d(z) \geq 0$ in the definition of Ω is not necessary to state the problems' equivalence. However, local quadratic convergence of the Newtonian methods in [6,7] (which are also the subject of this paper) cannot be guaranteed without including those constraints into the definition of Ω. This will become clear in the subsequent sections; see also [7]. The same concerns putting the constraint $b(z) \geq 0$ into Ω instead of writing it as $\min\{0, b(z)\} = 0$ and incorporating this term into the function F. In this respect, it should also be noted that if any of the functions b, c or d are nonlinear, constrained subproblems of Newton-type methods applied to (3), with Ω given by (4), would be (generally) impractical to solve. In that case, a further slacks reformulation

of the nonlinear constraints in (4) is performed, moving the resulting equality into the function F and thus obtaining a polyhedral set Ω; see [6,7,20], and also Sects. 4 and 5 below.

Next, it should be emphasized that, unless the restrictive strict complementarity condition holds at a solution z^* of problem (1) (i.e., $c(z^*) + d(z^*) > 0$), the function F given in (3) cannot be expected to be differentiable at z^*. Furthermore, we are interested here in the general cases, where z^* also need not be an isolated solution of the problem. This combination of difficulties (nondifferentiability and degeneracy) makes for a very challenging case, with most approaches not applicable; see [7] for a detailed discussion.

We next recall the methods in [6,7] for solving constrained Eq. (2), to be considered in the sequel. Eventually, our main focus will be on the case of piecewise continuously differentiable (PC^1-) functions F, and in particular on the reformulation (3)–(4) of the general complementarity system (1), as well as the special case of KKT systems arising from optimization or variational problems and generalized Nash equilibrium problems. The methods of [6,7] have very strong local convergence properties, even when solutions are not isolated and the function is not differentiable at solutions. It should be mentioned that even for the case of KKT systems, other Newton-type methods for degenerate problems need assumptions that imply that the primal part of the solution is isolated [11,12,16–19,29]; see also [21, Chapter 7].

The linear-programming-Newton method (LP-Newton method from now on), proposed in [7], is the following iterative procedure. Let $s \in \Omega$ be a given iterate, and let $\|\cdot\|$ be an arbitrary but fixed vector norm, where the space is clear from the context. Then, the new iterate is the z-part of a solution of the following optimization problem:

$$\begin{aligned} \min_{z,\gamma} \ \gamma \ \text{s.t.} \quad & z \in \Omega, \\ & \|F(s) + G(s)(z - s)\| \le \gamma \, \|F(s)\|^2, \\ & \|z - s\| \le \gamma \, \|F(s)\|, \end{aligned} \tag{5}$$

where $G(s)$ is a suitable substitute for the derivative of F at s (the precise conditions imposed on G will be specified in due time). It can be shown that, if s is not a solution of (2), the subproblem (5) always has a solution; see [7, Proposition 1]. (Note that [7] uses the additional constraint $\gamma \ge 0$ to define the feasible set in (5). However, this inequality is redundant, as in (5) it is clearly satisfied automatically if s is not a solution of (1). And if s is a solution, the method would simply stop before the subproblem (5) needs to be solved.) If the infinity norm $\|\cdot\|_\infty$ is used and if Ω is a polyhedral set, then (5) is a linear program (LP). This is the reason for the name "LP-Newton method". It is shown in [7] that a sequence generated by this algorithm converges locally quadratically to a solution of (2), under certain mild assumptions (recalled in Sect. 2 below), which in particular imply neither the differentiability of F nor the local uniqueness of a solution of (2).

In [6], it was shown that the constrained version of the Levenberg–Marquardt method for (2) converges under the same mild assumptions as those required for the LP-Newton scheme. Again, let $s \in \Omega$ be a given iterate. Then, the constrained Levenberg–Marquardt method (see [2,22]) computes the next iterate as a solution of the following optimization subproblem:

$$\min_z \quad \|F(s) + G(s)(z-s)\|^2 + \mu(s)\|z-s\|^2 \quad \text{s.t.} \quad z \in \Omega, \tag{6}$$

with an appropriate *regularization* parameter $\mu(s) > 0$, and $G(s)$ being again a suitable substitute for the derivative of F at s. If the Euclidean norm $\|\cdot\|_2$ is used and if Ω is a polyhedral set, then (6) is a strongly convex quadratic program (QP), having a unique solution.

The main goal of this paper is to establish conditions under which the methods described above converge quadratically to a solution of (3)–(4), and therefore of the complementarity system (1). We start with recalling in Sect. 2 the assumptions required in [7] for fast local convergence in the context of the general constrained equation (2). Then, in Sect. 3, we consider the case of a PC^1-function F, of which the function in (3) is a special case. For PC^1-functions, we develop conditions which are sufficient for the assumptions in [7]. A condition based on error bounds for some of the selection functions defining F was already introduced in [7], and it was shown there that this condition implies one of the convergence assumptions. The first main result in our paper is that the same condition, in addition to a suitable choice of the feasible set Ω, is actually sufficient for all of the convergence assumptions to hold. A related result for a slacks reformulation of (2) is proved in Sect. 4. In Sect. 5, we go back to the complementarity system (1), and transfer the main result for PC^1-functions to the constrained reformulation (3)–(4) of (1). In particular, we show that the methods under consideration converge quadratically to a solution of (1) under the only assumption that the complementarity system satisfies a mild piecewise local error bound condition. In Sect. 6, the special case of KKT systems arising from optimization or variational problems is considered, and it is shown that the noncriticality assumption on the Lagrange multiplier [15,19] is sufficient for convergence in this case. Sections 7 and 8 consider different reformulations of KKT systems associated to generalized Nash equilibrium problems, where we exhibit sufficient conditions for the required error bounds in that setting. Specifically, a result from [20] based on the full row rank of a certain matrix is recovered. In addition, a more general result is presented, showing that even the weaker constant rank of certain matrices is sufficient.

Throughout the paper, we assume that the solution set Z of problem (2) is not empty and that $z^* \in Z$ is an arbitrary but fixed solution. We denote by JH the Jacobian of a differentiable function H. Throughout the paper, $\|\cdot\|$ stands for an arbitrary but fixed vector norm in a finite dimensional space, always clear from the context. The specific Euclidean norm is denoted by $\|\cdot\|_2$ and the infinity norm by $\|\cdot\|_\infty$. The distance from a point $z \in \mathbb{R}^n$ to a nonempty set $U \subseteq \mathbb{R}^n$ is given by $\mathrm{dist}[z, U] := \inf\{\|s-z\| \mid s \in U\}$. By $\mathscr{B}_\delta(z) := \{s \in \mathbb{R}^n \mid \|s-z\| \le \delta\}$ we denote the closed ball around $z \in \mathbb{R}^n$ with radius $\delta > 0$.

The following constraint qualifications (CQs) for systems of equations and inequalities will play some role in the sequel. To this end, consider the system

$$R(z) = 0, \quad S(z) \ge 0 \tag{7}$$

with continuously differentiable functions $R : \mathbb{R}^n \to \mathbb{R}^{m_1}$ and $S : \mathbb{R}^n \to \mathbb{R}^{m_2}$. Let $\bar{z}$ denote an arbitrary but fixed solution of (7). We indicate by $\mathscr{S}_0$ the index set of those functions S_i which are active at $\bar{z}$, i.e.,

$$\mathscr{S}_0 := \{i \in \{1, \ldots, m_2\} \mid S_i(\bar{z}) = 0\}.$$

The *linear independence constraint qualification* (LICQ for short) holds at $\bar{z}$ if the matrix

$$\begin{pmatrix} JR(\bar{z}) \\ JS_{\mathscr{S}_0}(\bar{z}) \end{pmatrix}$$

has full row rank. Here, $S_{\mathscr{S}_0}$ consists of those components of S whose indices belong to $\mathscr{S}_0$. System (7) satisfies the *Mangasarian–Fromovitz constraint qualification* (MFCQ for short) at $\bar{z}$ if the matrix $JR(\bar{z})$ has full row rank and there exists $h \in \mathbb{R}^n$ such that

$$JR(\bar{z})h = 0 \quad \text{and} \quad JS_{\mathscr{S}_0}(\bar{z})h > 0$$

hold. The MFCQ was introduced in [26]. The next constraint qualification which we would like to recall is the *relaxed constant rank constraint qualification* (RCRCQ for short) introduced in [24]. This condition is satisfied at $\bar{z}$ if there exists $\varepsilon > 0$ such that, for each index set $\mathscr{S} \subseteq \mathscr{S}_0$, the matrices

$$\begin{pmatrix} JR(z) \\ JS_{\mathscr{S}}(z) \end{pmatrix}$$

have the same rank for all $z \in \mathscr{B}_\varepsilon(\bar{z})$. It is well known that LICQ implies both MFCQ and RCRCQ. However, MFCQ neither is implied by nor implies RCRCQ [27].

Finally, we say that the *local error bound condition* for system (7) is satisfied at $\bar{z}$ if there exist $\varepsilon > 0$ and $\omega > 0$ such that

$$\operatorname{dist}[z, \mathscr{Z}] \leq \omega(\|R(z)\| + \|\min\{0, S(z)\}\|)$$

holds for all $z \in \mathscr{B}_\varepsilon(\bar{z})$, where $\mathscr{Z}$ denotes the solution set of (7). This condition will play a central role in this paper. It is shown in [28] that the local error bound condition is implied by MFCQ. Moreover, the RCRCQ also implies the local error bound condition. The latter is established in [27] and also in [25]. It should also be mentioned that for twice continuously differentiable functions, the relaxed constant positive linear dependence condition is weaker than RCRCQ, but also implies the error bound [1]. As we do not assume twice differentiability in this paper, we shall not state or use this weaker CQ.

2 A general framework for the analysis of local convergence

In this section, we recall and discuss the assumptions required for local convergence of the LP-Newton method, described by (5), for solving constrained Eq. (2). These assumptions were given in [7]. Recall, however, that the same set of assumptions also guarantees local convergence of the constrained Levenberg–Marquardt method described by (6). This was established in [6].

Let a function $G : \mathbb{R}^n \to \mathbb{R}^{m\times n}$ be given. The choice of G will be specified in the next section, where F is assumed to be a PC^1 function. For the moment, it is enough to say that G is the Jacobian of F, or a suitable substitute if F is not differentiable.

It is shown in [6,7] that the algorithms in question converge locally quadratically to a solution of (2) if the following four assumptions are satisfied, with some suitable $\delta > 0$.

Assumption 1 (*Lipschitzian growth of the residual*) There exists $L > 0$ such that

$$\|F(s)\| \le L\mathrm{dist}[s, Z]$$

holds for all $s \in \mathscr{B}_\delta(z^*) \cap \Omega$.

Assumption 2 (*Error bound*) There exists $\omega > 0$ such that

$$\mathrm{dist}[s, Z] \le \omega\|F(s)\|$$

holds for all $s \in \mathscr{B}_\delta(z^*) \cap \Omega$.

Assumption 3 (*Nonemptyness of approximate solution sets*) There exists $\Gamma > 0$ such that the set

$$\mathscr{F}(s, \Gamma) := \left\{z \in \Omega \mid \|F(s) + G(s)(z - s)\| \le \Gamma\|F(s)\|^2,\ \|z - s\| \le \Gamma\|F(s)\|\right\}$$

is nonempty for all $s \in \mathscr{B}_\delta(z^*) \cap \Omega$.

Assumption 4 (*Quality of the approximate solution set*) There exists $\kappa > 0$ such that

$$z \in \left\{z \in \Omega \mid \|F(s) + G(s)(z - s)\| \le \alpha^2,\ \|z - s\| \le \alpha\right\}$$

implies

$$\|F(z)\| \le \kappa\alpha^2$$

for all $s \in (\mathscr{B}_\delta(z^*) \cap \Omega)\backslash Z$ and all $\alpha \in [0, \delta]$.

Note that Assumption 3 above is different from its formulation in [7]. There, it is required that, for some $\tilde{\Gamma} \ge 1$, the optimal value $\gamma(s)$ of (5) is bounded above by $\tilde{\Gamma}$ for all $s \in \mathscr{B}_\delta(z^*) \cap \Omega$. Under the natural convention that for our subproblem (5) it holds that $\gamma(s) = -\infty$ if $F(s) = 0$, Assumption 3 and its counterpart in [7] are equivalent. Indeed, if the set $\mathscr{F}(s, \Gamma)$ is nonempty for some $s \in \mathscr{B}_\delta(z^*) \cap \Omega$, and z is an element of $\mathscr{F}(s, \Gamma)$, then (z, γ) with $\gamma := \Gamma$ is feasible for problem (5). Therefore, $\gamma(s) \le \Gamma \le \tilde{\Gamma}$ with $\tilde{\Gamma} := \max\{1, \Gamma\}$ holds for all $s \in \mathscr{B}_\delta(z^*) \cap \Omega$. Conversely, if $\gamma(s) \le \tilde{\Gamma}$ is valid for some $s \in \mathscr{B}_\delta(z^*) \cap \Omega$, then the z-part of every solution of (5) belongs to $\mathscr{F}(s, \Gamma)$ with $\Gamma := \tilde{\Gamma}$. Thus, the set $\mathscr{F}(s, \Gamma)$ is nonempty for all $s \in \mathscr{B}_\delta(z^*) \cap \Omega$.

At the end of this section, let us recall the main local convergence results of the LP-Newton and the constrained Levenberg–Marquardt method. The following theorem follows by [7, Theorem 1] and [6, Corollary 2].

Theorem 1 *Let Assumptions 1–4 be satisfied. Then, there is $r > 0$ such that any infinite sequence $\{z^k\}$ generated by the LP-Newton method, described by (5), or by the constrained Levenberg–Marquardt method, described by (6), with a starting point $z^0 \in \mathscr{B}_r(z^*) \cap \Omega$ converges quadratically to some solution of (2).*

3 Discussion of the assumptions for PC^1-functions

In this section, we consider the case of F being a PC^1-function. We shall provide sufficient conditions for Assumptions 1–4 to hold in this setting. Our main result is Theorem 2 below which shows that all the needed assumptions (i.e., Assumptions 1–4) are satisfied if certain error bound conditions hold, and if the set Ω is appropriately chosen.

A function $F : \mathbb{R}^n \to \mathbb{R}^m$ is called *piecewise continuously differentiable* (PC^1) if F is continuous and if there are continuously differentiable functions $F^1, \ldots, F^\ell : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$F(z) \in \{F^1(z), \ldots, F^\ell(z)\}$$

holds for all $z \in \mathbb{R}^n$. The functions $F^1, \ldots, F^\ell$ are called *selection functions*. For $z \in \mathbb{R}^n$, we denote by $\mathscr{A}(z)$ the index set of all selection functions which are active at z, i.e.,

$$\mathscr{A}(z) := \left\{ i \in \{1, \ldots, \ell\} \mid F(z) = F^i(z) \right\}.$$

We assume that the derivatives of the selection functions are locally Lipschitz-continuous. We also assume that $G : \mathbb{R}^n \to \mathbb{R}^{m \times n}$ is such that $G(z) \in \{JF^i(z) \mid i \in \mathscr{A}(z)\}$ holds for all $z \in \mathbb{R}^n$, which is clearly natural.

Obviously, the function $F : \mathbb{R}^n \to \mathbb{R}^{p+r}$ defined in (3) is a PC^1-function (with $\ell = 2^r$ selection functions $F^1, \ldots, F^{2^r} : \mathbb{R}^n \to \mathbb{R}^{p+r}$). Due to our smoothness assumptions on a, c, and d, the selection functions $F^1, \ldots, F^{2^r}$ are differentiable and have locally Lipschitz-continuous derivatives.

Next, we discuss Assumptions 1–4 in the context of PC^1-functions. It is clear that Assumption 1 is satisfied, because a PC^1-function is locally Lipschitz-continuous [14, Theorem 2.1].

Proposition 1 *Any PC^1-function F is locally Lipschitz-continuous. In particular, Assumption 1 is satisfied with arbitrary $\delta > 0$.*

Our next objective is to provide sufficient conditions for the Assumptions 2 and 3 to hold. To this end, consider the following.

Condition 1 (Error bounds in terms of residuals of active pieces) There exist $\delta_1 > 0$ and $\omega_1 > 0$ such that

$$\operatorname{dist}[s, Z] \leq \omega_1 \left\| F^i(s) \right\|$$

holds for all $i \in \mathscr{A}(z^*)$ and all $s \in \mathscr{B}_{\delta_1}(z^*) \cap \Omega$.

Proposition 2 below shows that Condition 1 implies Assumption 2. Roughly speaking, Condition 1 requires that every selection function which is active at z^*, provides a local overestimate for the distance to the solution set of problem (2).

Proposition 2 *Let Condition 1 be satisfied. Then, Assumption 2 holds for $\delta > 0$ sufficiently small.*

Proof Let $\delta \in (0, \delta_1]$ be small enough so that $\mathscr{A}(s) \subseteq \mathscr{A}(z^*)$ holds for all $s \in \mathscr{B}_\delta(z^*)$ (such δ exists, due to the continuity of F). Now, for $s \in \mathscr{B}_\delta(z^*) \cap \Omega$ and $i \in \mathscr{A}(s)$, we have $F(s) = F^i(s)$, and Condition 1 yields

$$\operatorname{dist}[s, Z] \leq \omega_1 \|F^i(s)\| = \omega_1 \|F(s)\|.$$

Hence, Assumption 2 is satisfied with $\omega := \omega_1$. □

Note that the differentiability of the selection functions is not needed in the last proof. The assertion of Proposition 2 remains valid if the selection functions are only continuous.

We next show that Condition 1 does not imply Assumption 3, even if F and Ω are given by (3) and (4).

Example 1 Consider the following constrained system of equations:

$$F(z) := F(x, y) := \min\{x^2 - y, -2y\} = 0 \quad \text{s.t.} \quad z = (x, y)^\top \in \Omega \tag{8}$$

with

$$\Omega := \{(x, y)^\top \in \mathbb{R}^2 \mid x^2 - y \geq 0,\ -2y \geq 0\} = \{(x, y)^\top \in \mathbb{R}^2 \mid y \leq 0\}.$$

The function F is a PC^1-function with the selection functions

$$F^1(z) := F^1(x, y) := x^2 - y \quad \text{and} \quad F^2(z) := F^2(x, y) := -2y.$$

Let $G : \mathbb{R}^2 \to \mathbb{R}^{1\times 2}$ be any function satisfying $G(z) \in \{JF^i(z) \mid i \in \mathscr{A}(z)\}$. Obviously, $z^* := (0, 0)^\top$ is a solution of (8) with $\mathscr{A}(z^*) = \{1, 2\}$. The solution set of (8) is

$$Z = \{(x, y)^\top \in \mathbb{R}^2 \mid \min\{x^2 - y, -2y\} = 0\} = \{(x, y)^\top \in \mathbb{R}^2 \mid y = 0\}.$$

Throughout the rest of this example, the Euclidean vector norm is used. The distance from $s = (x, y)^\top$ to Z is equal to $|y|$ for all $s \in \mathbb{R}^2$. Taking into account that $y \le 0$ for $s = (x, y)^\top \in \Omega$, we have

$$\operatorname{dist}[s, Z] = |y| = -y \le \begin{cases} x^2 - y = |F^1(s)| \\ -2y \quad\;\; = |F^2(s)| \end{cases}$$

for all $s \in \Omega$. Therefore, Condition 1 is satisfied with arbitrary $\delta_1 > 0$ and $\omega_1 := 1$.

However, to see that Assumption 3 is not valid at z^*, let us consider the sequence $\{s^k\} \subset \Omega$ given by $s^k := \left(\frac{1}{k}, -\frac{2}{k^2}\right)^\top$. Obviously, this sequence converges to z^*. For every $k \ge 1$ we have

$$F^1(s^k) = \frac{3}{k^2} < \frac{4}{k^2} = F^2(s^k).$$

Therefore, $\mathscr{A}(s^k) = \{1\}$ holds for all $k \in \mathbb{N}$ and thus,

$$F(s^k) = F^1(s^k) = \frac{3}{k^2} \quad \text{and} \quad G(s^k) = JF^1(s^k) = \left(\frac{2}{k}, -1\right).$$

Suppose that Assumption 3 is satisfied at z^* with some constants $\Gamma > 0$ and $\delta > 0$. Then, for k sufficiently large, there is $z^k = (x_k, y_k)^\top \in \mathscr{F}(s^k, \Gamma)$. Due to $\|z^k - s^k\|_2 \le \Gamma |F(s^k)|$ we particularly have

$$\left| x_k - \frac{1}{k} \right| \le \Gamma |F(s^k)| = \frac{3}{k^2}\Gamma.$$

Using this, $|F(s^k) + G(s^k)(z^k - s^k)| \le \Gamma |F(s^k)|^2$, and $y_k \le 0$, we obtain

$$\frac{9}{k^4}\Gamma \ge \left| \frac{3}{k^2} + \frac{2}{k}\left(x_k - \frac{1}{k}\right) - \left(y_k + \frac{2}{k^2}\right) \right| \ge \left| -y_k + \frac{1}{k^2} \right| - \frac{2}{k}\left| x_k - \frac{1}{k} \right| \ge \frac{1}{k^2} - \frac{6}{k^3}\Gamma.$$

This inequality cannot be satisfied for k sufficiently large. Thus, we have a contradiction and Assumption 3 does not hold. Note that the assertions of this example stay true for any vector norm. □

Next, for $i \in \{1, \ldots, \ell\}$, let Z_i denote the set of all zeros of the selection function F^i in Ω, i.e.,

$$Z_i := \left\{ z \in \Omega \mid F^i(z) = 0 \right\}.$$

For $i \in \mathscr{A}(z^*)$, it is clear that the set Z_i is nonempty since z^* belongs to Z_i. Condition 2 below requires that every selection function F^i which is active at z^* provides an overestimate for the distance to Z_i for all points in a certain neighborhood of z^*. Note that, unlike Condition 1, Condition 2 refers to the distance to Z_i, not to Z.

Condition 2 (Individual error bounds on Ω for active pieces) There exist $\delta_2 > 0$ and $\omega_2 > 0$ such that

$$\operatorname{dist}[s, Z_i] \leq \omega_2 \left\| F^i(s) \right\|$$

holds for all $i \in \mathscr{A}(z^*)$ and all $s \in \mathscr{B}_{\delta_2}(z^*) \cap \Omega$.

It will turn out that Condition 2 implies Assumption 3. Moreover, it also implies Assumption 2 if the following property is additionally satisfied: the set Ω excludes all zeros of the selection functions (being active at z^*) which are not zeros of F, at least in a certain neighborhood of z^*. In other words, it is required that every zero of any selection function being active at z^*, which is sufficiently close to z^* and which belongs to Ω, is also a zero of F. Formally, we state this requirement as follows.

Ω-property We say that the *Ω-property* holds at z^* if there exists $\delta_\Omega > 0$ such that

$$Z_i \cap \mathscr{B}_{\delta_\Omega}(z^*) \subseteq Z \quad \text{for all } i \in \mathscr{A}(z^*).$$

In the case of F given by (3), the Ω-property means that the set Ω excludes all zeros of the selection functions for which some components of $c(\bar{z})$ or $d(\bar{z})$ are negative. In particular, this is guaranteed if Ω is defined by (4). The next proposition shows that Condition 2 and the Ω-property, together, imply Condition 1.

Proposition 3 *Let Condition 2 and the Ω-property at z^* be satisfied. Then, Condition 1 holds.*

Proof For $\delta_1 > 0$ with $\delta_1 \leq \min\{\delta_2, \frac{1}{2}\delta_\Omega\}$ let us take any $s \in \mathscr{B}_{\delta_1}(z^*) \cap \Omega$, and let $i \in \mathscr{A}(z^*)$ be arbitrary but fixed. Moreover, let $\bar{s}$ be an element of Z_i with

$$\operatorname{dist}[s, Z_i] = \|s - \bar{s}\|. \tag{9}$$

Note that Z_i is nonempty, because $i \in \mathscr{A}(z^*)$ implies $z^* \in Z_i$. Further, Z_i is closed due to the continuity of F^i, and the closedness of Ω. Condition 2 yields

$$\|s - \bar{s}\| \leq \omega_2 \|F^i(s)\|. \tag{10}$$

Since

$$\|\bar{s} - z^*\| \leq \|s - \bar{s}\| + \|s - z^*\| \leq 2\|s - z^*\| \leq 2\delta_1 \leq \delta_\Omega,$$

the Ω-property ensures $\bar{s} \in Z$. By (10), we obtain

$$\operatorname{dist}[s, Z] \leq \|s - \bar{s}\| \leq \omega_2 \|F^i(s)\|.$$

Thus, Condition 1 is satisfied with $\omega_1 := \omega_2$. □

Now, we prove a relation between Condition 2 and Assumptions 2 and 3.

Proposition 4 *Let Condition 2 be satisfied. Then, for $\delta > 0$ sufficiently small, the following assertions are valid.*

(i) *Assumption 2 holds if the Ω-property at z^* is additionally satisfied.*
(ii) *Assumption 3 holds.*

Proof (i) Since Condition 2 and the Ω-property are satisfied, Condition 1 holds as well due to Proposition 3. The validity of Assumption 2 for $\delta > 0$ sufficiently small follows by Proposition 2.

(ii) Let $\delta \in (0, \delta_2]$ be small enough so that $\mathscr{A}(s) \subseteq \mathscr{A}(z^*)$ holds for all $s \in \mathscr{B}_\delta(z^*)$. Moreover, let $C > 0$ be a constant such that

$$\|F^i(s) + JF^i(s)(z - s) - F^i(z)\| \leq C\|z - s\|^2 \tag{11}$$

is valid for all $i \in \mathscr{A}(z^*)$ and all $s, z \in \mathscr{B}_{2\delta}(z^*)$. Such a constant exists since the selection functions are differentiable and have locally Lipschitz-continuous Jacobians.

For any $s \in \mathscr{B}_\delta(z^*) \cap \Omega$ let i be an element of $\mathscr{A}(s)$ with $G(s) = JF^i(s)$. Moreover, let $\bar{s}$ be an element of Z_i such that (9) is satisfied. As in the proof of the last proposition we obtain (10) by Condition 2. Due to $i \in \mathscr{A}(s)$ this yields

$$\|s - \bar{s}\| \leq \omega_2 \|F(s)\|. \tag{12}$$

Because of

$$\|\bar{s} - z^*\| \leq \|s - \bar{s}\| + \|s - z^*\| \leq 2\|s - z^*\| \leq 2\delta,$$

we can apply (11) and (12) and obtain

$$\begin{aligned} \|F(s) + G(s)(\bar{s} - s)\| &= \|F^i(s) + JF^i(s)(\bar{s} - s) - F^i(\bar{s})\| \\ &\leq C\|s - \bar{s}\|^2 \\ &\leq C\omega_2^2 \|F(s)\|^2. \end{aligned}$$

Since s was arbitrarily chosen, Assumption 3 is valid with $\Gamma := \max\{\omega_2, C\omega_2^2\}$. □

Note that item (ii) of Proposition 4 had been established in [7, Corollary 3]. However, the proof there is indirect. In particular, nothing was specified about the value of δ. The simpler and direct proof presented above seems useful.

We next consider Assumption 4. To this end, we introduce the following.

Condition 3 (Residual bounded above by residuals of active pieces) There exist $\delta_3 > 0$ and $K > 0$ such that

$$\|F(s)\| \leq K\|F^i(s)\|$$

holds for all $i \in \mathscr{A}(z^*)$ and all $s \in \mathscr{B}_{\delta_3}(z^*) \cap \Omega$.

We show that Condition 3 is sufficient for Assumption 4 to hold.

Proposition 5 *Let Condition 3 be satisfied. Then, Assumption 4 holds for $\delta > 0$ sufficiently small.*

Proof Let $\delta \in (0, \frac{1}{2}\delta_3]$ be small enough so that $\mathscr{A}(s) \subseteq \mathscr{A}(z^*)$ holds for all $s \in \mathscr{B}_\delta(z^*)$. Since all selection functions are differentiable and have locally Lipschitz-continuous Jacobians, there is $C > 0$ such that (11) holds for all $i \in \mathscr{A}(z^*)$ and all $s, z \in \mathscr{B}_{2\delta}(z^*)$. Now let $\alpha \in [0, \delta]$ and $s \in (\mathscr{B}_\delta(z^*) \cap \Omega)\backslash Z$ be arbitrarily chosen and let i be an element of $\mathscr{A}(s)$ such that $G(s) = JF^i(s)$ holds. Moreover, let $z \in \Omega$ be a point for which the inequalities

$$\|F(s) + G(s)(z - s)\| \leq \alpha^2 \quad \text{and} \quad \|z - s\| \leq \alpha \tag{13}$$

are satisfied. Then,

$$\|z - z^*\| \leq \|z - s\| + \|s - z^*\| \leq \alpha + \delta \leq 2\delta$$

holds. Using (11) and (13), we obtain

$$\|F^i(z)\| \leq \|F^i(s) + \underbrace{JF^i(s)}_{=G(s)}(z - s)\| + C\|z - s\|^2 \leq (1 + C)\alpha^2.$$

This together with Condition 3 yields

$$\|F(z)\| \leq (1 + C)K\alpha^2.$$

Therefore, Assumption 4 is valid with $\kappa := (1 + C)K$. □

It can be easily seen that Condition 3 implies the Ω-property at z^*, i.e., that the zeros of the selection functions which are not zeros of F, do not belong to Ω (at least in a sufficiently small neighborhood of z^*). This assertion is proven in the next proposition.

Proposition 6 *If Condition 3 is satisfied, then the Ω-property holds at z^*.*

Proof Let $\delta_\Omega \in (0, \delta_3]$, $i \in \mathscr{A}(z^*)$, and $s \in Z_i \cap \mathscr{B}_{\delta_\Omega}(z^*)$ be arbitrarily chosen. Then, by $s \in Z_i$, Condition 3 yields

$$\|F(s)\| \leq K\|F^i(s)\| = 0,$$

so that $F(s) = 0$ and, thus, $s \in Z$ follow. □

However, the Ω-property is not sufficient for Assumption 4 to hold. This can be shown by a simple example.

Example 2 Let us consider the function $F : \mathbb{R}^2 \to \mathbb{R}$ defined by $F(z) := F(x, y) := \min\{x, y\}$, which is a PC^1-function with the selection functions $F^1(x, y) := x$ and $F^2(x, y) := y$. Let $G : \mathbb{R}^2 \to \mathbb{R}^{1\times 2}$ be any function satisfying $G(z) \in \{JF^i(z) \mid$

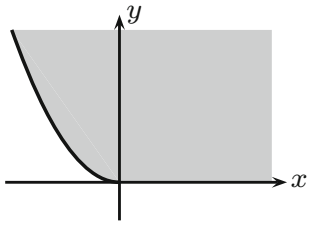

Fig. 1 The set Ω from Example 2

$i \in \mathcal{A}(z)\}$. Throughout this example, we use the Euclidean vector norm. Note that the choice of the norm does not influence the validity of our assumptions and conditions, respectively, it might only change the values of some constants. Let us define Ω by

$$\Omega := \mathbb{R}^2_+ \cup \{(x, y)^\top \in \mathbb{R}^2 \mid y \geq x^2\}.$$

Figure 1 shows an illustration of Ω. Obviously, the points on the negative half axes do not belong to Ω, i.e., the zeros of F^1 and F^2 which are not zeros of F are excluded. Therefore, the Ω-property holds at $z^* := (0, 0)^\top$ (for arbitrary $\delta_\Omega > 0$). However, Assumption 4 is not satisfied at z^* for this choice of Ω. In order to show this, let us assume the contrary, i.e., that Assumption 4 is valid with some constants $\delta > 0$ and $\kappa > 0$. Furthermore, let us define sequences $\{s^k\} \subset \Omega \backslash Z$, $\{z^k\} \subset \Omega$, and $\{\alpha_k\} \subset (0, \infty)$ as follows: $s^k := (\frac{2}{k^2}, \frac{1}{k^2})^\top$, $z^k := (-\frac{1}{k}, \frac{1}{k^2})^\top$, $\alpha_k := \frac{3}{k}$.

Obviously, the sequence $\{s^k\}$ converges to z^* and the sequence $\{\alpha_k\}$ to zero. Therefore, for sufficiently large k, we have $\alpha_k \leq \delta$ and $s^k \in (\mathcal{B}_\delta(z^*) \cap \Omega) \backslash Z$. Moreover,

$$\|z^k - s^k\|_2 = \frac{1}{k} + \frac{2}{k^2} \leq \frac{3}{k} = \alpha_k \tag{14}$$

is satisfied for all $k \geq 1$. Furthermore, $\mathcal{A}(s^k) = \{2\}$ holds for all k and therefore

$$F(s^k) = F^2(s^k) = \frac{1}{k^2}, \quad G(s^k) = JF^2(s^k) = (0, 1).$$

This implies

$$|F(s^k) + G(s^k)(z^k - s^k)| = \left|\frac{1}{k^2} + \frac{1}{k^2} - \frac{1}{k^2}\right| = \frac{1}{k^2} < \frac{9}{k^2} = \alpha_k^2 \tag{15}$$

for all $k \geq 1$. Taking into account (14) and (15), Assumption 4 yields

$$|F(z^k)| \leq \kappa \alpha_k^2 = \kappa \cdot \frac{9}{k^2}$$

for all sufficiently large $k \geq 1$. On the other hand, $F(z^k) = -\frac{1}{k}$ holds for all k which provides a contradiction. Therefore, Assumption 4 is not satisfied. □

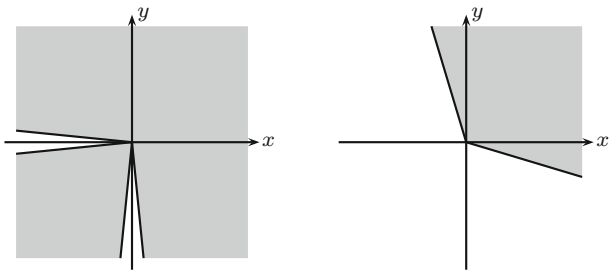

Fig. 2 The set Ω_ε from Example 3—on the *left* for $\varepsilon < 1$ and on the *right* for $\varepsilon > 1$

Next, we present two examples where Condition 3 (and therefore Assumption 4) are satisfied. In particular, it will turn out that Assumption 4 holds for system (3) with Ω defined by (4).

Example 3 We consider the function from Example 2 again, i.e., $F(z) := F(x, y) := \min\{x, y\}$. As before, $F^1(x, y) := x$ and $F^2(x, y) := y$ denote the selection functions. For $\varepsilon > 0$, define

$$\Omega_\varepsilon := \{(x, y)^\top \in \mathbb{R}^2 \mid |y| \geq \varepsilon|x| \text{ if } x < 0,\ |x| \geq \varepsilon|y| \text{ if } y < 0\}.$$

Figure 2 shows illustrations of Ω_ε for two different values of ε.

For all $(x, y)^\top \in \Omega_\varepsilon$ we have, using the Euclidean norm,

$$|F(x, y)| = |\min\{x, y\}| \leq \max\left\{\frac{1}{\varepsilon}, 1\right\} |x| = K|F^1(x, y)| \quad \text{and}$$
$$|F(x, y)| = |\min\{x, y\}| \leq \max\left\{\frac{1}{\varepsilon}, 1\right\} |y| = K|F^2(x, y)|,$$

where $K := \max\left\{\frac{1}{\varepsilon}, 1\right\}$. Therefore, Condition 3 is satisfied with this K and an arbitrary $\delta_3 > 0$. Condition 3 stays true for $\Omega_+ := \mathbb{R}^2_+$. By Proposition 5 we know that Condition 3 implies Assumption 4. □

Example 4 Let us consider the PC^1-function $F : \mathbb{R}^n \to \mathbb{R}^r$ defined by

$$F(z) := \min\{c(z), d(z)\}$$

with given continuously differentiable functions $c, d : \mathbb{R}^n \to \mathbb{R}^r$. As in the previous examples, the Euclidean vector norm is used. Once again, note that the choice of the norm does not influence the validity of Condition 3. For

$$\Omega \subseteq \Omega_+ := \{z \in \mathbb{R}^n \mid c(z) \geq 0,\ d(z) \geq 0\}$$

Condition 3 is satisfied at every solution of the system $F(z) = 0$ s.t. $z \in \Omega$ because

$$|F_i(z)| = |\min\{c_i(z), d_i(z)\}| = \min\{|c_i(z)|, |d_i(z)|\} \tag{16}$$

holds for all $i \in \{1, \ldots, r\}$ and all $z \in \Omega$. By a similar argument it can be seen that Condition 3 and, thus, Assumption 4 are valid for problem (3) with Ω given by (4).

Condition 3 stays true at every solution of the system $F(z) = 0$ s.t. $z \in \Omega$, if Ω is chosen according to

$$\Omega \subseteq \Omega_0 := \{z \in \mathbb{R}^n \mid c(z) + d(z) \geq 0\}.$$

In order to show this, let $z \in \Omega$ and $i \in \{1, \ldots, r\}$ be arbitrary but fixed. Without loss of generality, we assume that $c_i(z) \leq d_i(z)$ holds. If $c_i(z)$ is nonnegative, we have

$$|F_i(z)| = |\min\{c_i(z), d_i(z)\}| = |c_i(z)| = \min\{|c_i(z)|, |d_i(z)|\}.$$

Otherwise, if $c_i(z) < 0$ holds, then $d_i(z) \geq |c_i(z)|$ must be valid due to the choice of Ω and therefore we obtain

$$|F_i(z)| = |\min\{c_i(z), d_i(z)\}| = |c_i(z)| = \min\{|c_i(z)|, |d_i(z)|\},$$

too. Thus, Condition 3 and therefore Assumption 4 are satisfied. This recovers [7, Corollary 5]. □

We next show that Condition 3 (and therefore Assumption 4) are implied by Condition 2, together with the Ω-property, i.e., with a suitable choice of Ω.

Proposition 7 *Let Condition 2 and the Ω-property at z^* be satisfied. Then, Condition 3 is valid and therefore, Assumption 4 holds for $\delta > 0$ sufficiently small.*

Proof Choose $\delta_3 \in (0, \min\{\delta_2, \frac{1}{2}\delta_\Omega\}]$. By Proposition 1, there is $l > 0$ such that

$$\|F(z) - F(s)\| \leq l\|z - s\| \tag{17}$$

holds for all $s, z \in \mathscr{B}_{2\delta_3}(z^*)$. Now, let $i \in \mathscr{A}(z^*)$ and $s \in \mathscr{B}_{\delta_3}(z^*) \cap \Omega$ be arbitrarily chosen. Moreover, let $\bar{s}$ be an element of Z_i such that $\|s - \bar{s}\| = \operatorname{dist}[s, Z_i]$ holds. Taking into account Condition 2,

$$\|s - \bar{s}\| \leq \omega_2 \|F^i(s)\| \tag{18}$$

follows. Furthermore, since z^* also belongs to Z_i, we have

$$\|\bar{s} - z^*\| \leq \|s - \bar{s}\| + \|s - z^*\| \leq 2\|s - z^*\| \leq 2\delta_3 \leq \delta_\Omega.$$

Therefore, the Ω-property yields that $\bar{s}$ is a zero of F, too. Using this, (17), and (18), we obtain

$$\|F(s)\| = \|F(s) - F(\bar{s})\| \leq l\|s - \bar{s}\| \leq l\omega_2 \|F^i(s)\|.$$

Hence, Condition 3 is valid with $K := l\omega_2$. Assumption 4 is satisfied due to Proposition 5. □

Combining now the assertions of Propositions 1, 4, and 7, we state the main result of this section. Specifically, Condition 2 and the Ω-property are sufficient for Assumptions 1–4 to hold and, therefore, for the quadratic convergence of the Newton-type methods discussed above.

Theorem 2 *Let F be a PC^1-function. Suppose that Condition 2 and the Ω-property at z^* are satisfied. Then, Assumptions 1–4 hold.*

Let us summarize the results of this section. We introduced conditions and proved relations to the convergence assumptions which we recalled in Sect. 2. In particular, it was shown that the validity of individual local error bound conditions on Ω for the active pieces (Condition 2) is sufficient for Assumptions 1–4 to hold if additionally Ω excludes all zeros of the selection functions which are not zeros of F. We also would like to mention two very influencing papers. The first is [23] for solving systems of equations with PC^1-functions under the nonsingularity assumption for all the elements of $G(z^*)$. The second paper [22] provides conditions for the quadratic convergence of the constrained Levenberg–Marquardt method if the function F is smooth but may have nonisolated zeros. The results in the current paper weaken the classical convergence conditions from [22,23] but still enable to obtain superlinear convergence.

Figure 3 illustrates the relations which were proved in this section. Moreover, relations between Condition 2 and Conditions 4 and 5 are already shown in Fig. 3. The latter conditions will be introduced in Sects. 4 and 5. Condition 4 is an extension of Condition 2. It will turn out that Condition 4, together with the Ω-property, does not only imply Assumptions 1–4 for system (2) but also the counterparts of Assumptions 1–4 for the reformulation of (2) with slack variables. Such a reformulation is advisable from the computational point of view; see the discussion at the beginning of Sect. 4. Condition 5 is equivalent to Condition 4 for problem (3) with Ω from (4).

4 Slacks reformulations of inequalities

The feasible set Ω often involves inequality constraints. For instance, in our problem (3), the inequalities $b(z) \geq 0$, $c(z) \geq 0$, and $d(z) \geq 0$ are present. If any of those functions is nonlinear, Ω need not be a polyhedral set so that the subproblems of the LP-Newton as well as of the constrained Levenberg–Marquardt method would not be computationally acceptable (unless further structure is present and can be exploited). To alleviate this complication, slack variables can be used to reduce the problem to that with simple (bound) constraints.

Consider the problem (2) with Ω given by

$$\Omega := \{z \in \mathbb{R}^n \mid \alpha(z) \geq 0,\ \ \beta(z) \geq 0\},$$

with an affine function $\alpha : \mathbb{R}^n \to \mathbb{R}^{k_1}$ and a nonlinear function $\beta : \mathbb{R}^n \to \mathbb{R}^{k_2}$. The latter is assumed to be differentiable with a locally Lipschitz-continuous derivative. Then, from the algorithmic point of view, it is advisable to consider the following reformulation of (2):

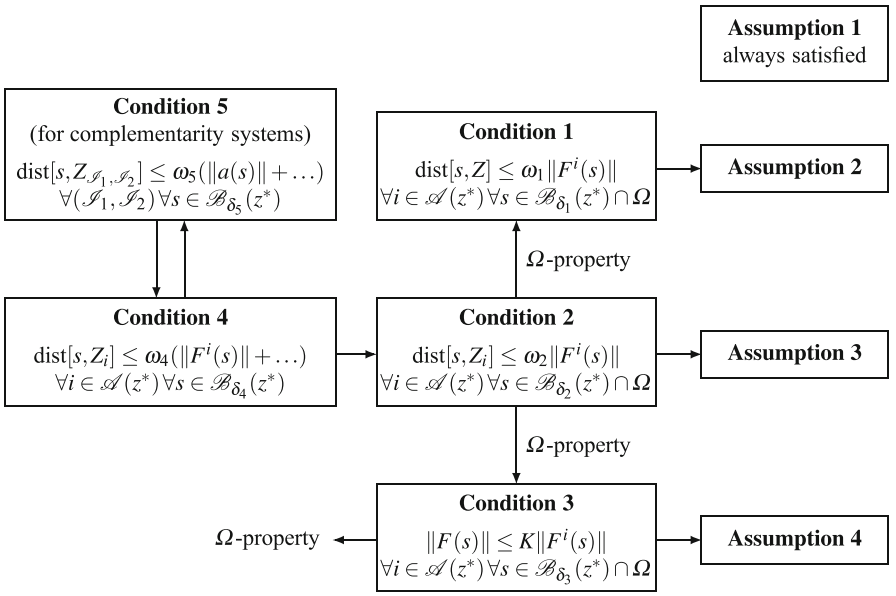

Fig. 3 Scheme on relations of the conditions and assumptions

$$\hat{F}(z, w) := \begin{pmatrix} F(z) \\ \beta(z) - w \end{pmatrix} = 0 \quad \text{s.t.} \quad (z, w) \in \hat{\Omega}, \tag{19}$$

where $\hat{\Omega}$ is given by

$$\hat{\Omega} := \{(z, w) \in \mathbb{R}^n \times \mathbb{R}^{k_2} \mid \alpha(z) \ge 0, \ \ w \ge 0\}.$$

Obviously, a point $\bar{z}$ is a solution of (2) if and only if $(\bar{z}, \bar{w})$ with $\bar{w} := \beta(\bar{z})$ solves (19). By $\hat{Z}$ we denote the solution set of (19). As before, by z^* an arbitrary but fixed solution of (2) is indicated and by (z^*, w^*) with $w^* := \beta(z^*)$ the counterpart for problem (19). The advantage of problem (19) compared to (2) is that $\hat{\Omega}$ is polyhedral.

Assuming again that F is a PC^1-function with the selection functions $F^1, \ldots, F^\ell$, the function $\hat{F}$ is also a PC^1-function with the selection functions

$$\hat{F}^i(z, w) := \begin{pmatrix} F^i(z) \\ \beta(z) - w \end{pmatrix},$$

$i = 1, \ldots, \ell$. Obviously, $\mathscr{A}(z, w) = \mathscr{A}(z)$ holds for all $(z, w) \in \mathbb{R}^n \times \mathbb{R}^{k_2}$. For every $i = 1, \ldots, \ell$ we denote the set of all zeros of $\hat{F}^i$ belonging to $\hat{\Omega}$ by $\hat{Z}_i$. Note that at least for $i \in \mathscr{A}(z^*)$ the set $\hat{Z}_i$ is nonempty since (z^*, w^*) belongs to it.

The counterparts of Assumptions 1–4 for problem (19) are implied, according to Theorem 2, by the counterparts of Condition 2 and the Ω-property for (19). It is easy to see that the counterpart of the Ω-property for (19) is satisfied if and only if the Ω-property itself holds at z^*. The counterpart of Condition 2 for problem (19) says that there are $\hat{\delta}_2 > 0$ and $\hat{\omega}_2 > 0$ such that

$$\operatorname{dist}[(s, w), \hat{Z}_i] \leq \hat{\omega}_2 \|\hat{F}^i(s, w)\|$$

holds for all $i \in \mathscr{A}(z^*)$ and all $(s, w) \in \mathscr{B}_{\hat{\delta}_2}(z^*, w^*) \cap \hat{\Omega}$. Unfortunately, this counterpart of Condition 2 for (19) is in general not implied by Condition 2 itself because the former requires that the inequality also holds for certain points s where some components of $\beta(s)$ might be negative. That was not needed in Condition 2.

This motivates to consider the following Condition 4 for problem (2) which is an extension of Condition 2.

Condition 4 (Individual error bounds for active pieces) There exist $\delta_4 > 0$ and $\omega_4 > 0$such that

$$\operatorname{dist}[s, Z_i] \leq \omega_4 \left(\|F^i(s)\| + \|\min\{0, \alpha(s)\}\| + \|\min\{0, \beta(s)\}\| \right)$$

holds for all $i \in \mathscr{A}(z^*)$ and all $s \in \mathscr{B}_{\delta_4}(z^*)$.

Obviously, for all $s \in \Omega$, the inequality in Condition 4 coincides with the inequality in Condition 2. Thus, Condition 4 implies Condition 2. But, unlike Condition 2, Condition 4 makes a requirement also for points which do not belong to Ω. The counterpart of Condition 4 for the constrained system with slack variables (19) can be stated as follows. There exist $\hat{\delta}_4 > 0$ and $\hat{\omega}_4 > 0$ such that

$$\operatorname{dist}[(s, w), \hat{Z}_i] \leq \hat{\omega}_4 \left(\|F^i(s)\| + \|\beta(s) - w\| + \|\min\{0, \alpha(s)\}\| + \|\min\{0, w\}\| \right) \tag{20}$$

holds for all $i \in \mathscr{A}(z^*)$ and all $(s, w) \in \mathscr{B}_{\hat{\delta}_4}(z^*, w^*)$. The next proposition shows that Condition 4 is valid if and only if its counterpart for system (19) is satisfied.

Proposition 8 *Condition 4 is satisfied if and only if there exist $\hat{\delta}_4 > 0$ and $\hat{\omega}_4 > 0$ such that (20) holds for all $i \in \mathscr{A}(z^*)$ and all $(s, w) \in \mathscr{B}_{\hat{\delta}_4}(z^*, w^*)$.*

Proof Let $l > 0$ be a Lipschitz constant of the function β on $\mathscr{B}_\varepsilon(z^*)$ with a sufficiently large $\varepsilon > 0$, i.e., the inequality

$$\|\beta(s) - \beta(z)\|_2 \leq l\|s - z\|_2 \tag{21}$$

holds for all $s, z \in \mathscr{B}_\varepsilon(z^*)$. Assume that Condition 4 is satisfied and let a vector $(s, w) \in \mathscr{B}_{\delta_4}(z^*, w^*)$ and an index $i \in \mathscr{A}(z^*)$ be arbitrarily chosen. In particular, $s \in \mathscr{B}_{\delta_4}(z^*)$ holds. Let $\bar{s} \in Z_i$ be a point with the property

$$\|s - \bar{s}\|_2 = \operatorname{dist}[s, Z_i].$$

Then, Condition 4 implies

$$\|s - \bar{s}\|_2 \leq \omega_4 \left(\|F^i(s)\|_2 + \|\min\{0, \alpha(s)\}\|_2 + \|\min\{0, \beta(s)\}\|_2 \right). \tag{22}$$

To estimate the right-hand side of the latter inequality, let $j \in \{1, \ldots, k_2\}$ be arbitrary but fixed. If $\beta_j(s) \geq 0$ holds, we obviously have

$$|\min\{0, \beta_j(s)\}| \leq |\beta_j(s) - w_j| + |\min\{0, w_j\}|.$$

Otherwise, this inequality also holds, because

$$\begin{aligned}|\min\{0, \beta_j(s)\}| = -\beta_j(s) + w_j - w_j \leq |\beta_j(s) - w_j| - w_j \leq |\beta_j(s)\\ - w_j| + |\min\{0, w_j\}|.\end{aligned}$$

Since $j \in \{1, \ldots, k_2\}$ was arbitrarily chosen this implies

$$\|\min\{0, \beta(s)\}\|_2 \leq \|\beta(s) - w\|_2 + \|\min\{0, w\}\|_2. \tag{23}$$

Combining (22) and (23) yields

$$\|s - \bar{s}\|_2 \leq \omega_4 \left(\|F^i(s)\|_2 + \|\beta(s) - w\|_2 + \|\min\{0, \alpha(s)\}\|_2 + \|\min\{0, w\}\|_2\right). \tag{24}$$

Now, let us set $\bar{w} := \beta(\bar{s})$. Obviously, $(\bar{s}, \bar{w}) \in \hat{Z}_i$ holds and therefore

$$\operatorname{dist}[(s, w), \hat{Z}_i] \leq \|s - \bar{s}\|_2 + \|w - \bar{w}\|_2 \tag{25}$$

follows. Using the triangle inequality and (21), we obtain

$$\|w - \bar{w}\|_2 \leq \|w - \beta(s)\|_2 + \|\beta(s) - \beta(\bar{s})\|_2 \leq \|w - \beta(s)\|_2 + l\|s - \bar{s}\|_2.$$

Thus, (25) yields

$$\operatorname{dist}[(s, w), \hat{Z}_i] \leq (l + 1)\|s - \bar{s}\|_2 + \|\beta(s) - w\|_2.$$

This together with (24) implies

$$\begin{aligned}\operatorname{dist}[(s, w), \hat{Z}_i] \leq (1 + (l + 1)\omega_4)(\|F^i(s)\|_2 + \|\beta(s) - w\|_2\\ + \|\min\{0, \alpha(s)\}\|_2 + \|\min\{0, w\}\|_2).\end{aligned}$$

Therefore, (20) is satisfied with $\hat{\delta}_4 := \delta_4$ and $\hat{\omega}_4 := 1 + (l + 1)\omega_4$.

Now we assume that there are $\hat{\delta}_4 > 0$ and $\hat{\omega}_4 > 0$ such that (20) is valid for all $i \in \mathscr{A}(z^*)$ and all $(s, w) \in \mathscr{B}_{\hat{\delta}_4}(z^*, w^*)$. Let us define δ_4 according to $\delta_4 := \frac{\hat{\delta}_4}{1+l}$ and let a vector $s \in \mathscr{B}_{\delta_4}(z^*)$ and an index $i \in \mathscr{A}(z^*)$ be arbitrarily chosen. We set $w := \beta(s)$. Taking into account $(z^*, w^*) \in \hat{Z}_i$ and (21) we obtain

$$\|w - w^*\|_2 = \|\beta(s) - \beta(z^*)\|_2 \leq l\|s - z^*\|_2 \leq l\delta_4.$$

This implies

$$\|s - z^*\|_2 + \|w - w^*\|_2 \leq (1 + l)\delta_4 = \hat{\delta}_4$$

and therefore $(s, w) \in \mathscr{B}_{\hat{\delta}_4}(z^*, w^*)$. Let $(\bar{s}, \bar{w}) \in \hat{Z}_i$ be a point with the property

$$\left\| \begin{pmatrix} s - \bar{s} \\ w - \bar{w} \end{pmatrix} \right\|_2 = \operatorname{dist}[(s, w), \hat{Z}_i].$$

Obviously, $\bar{s} \in Z_i$ holds. Furthermore, (20) and the definition of w yield

$$\begin{aligned} \operatorname{dist}[s, Z_i] &\leq \|s - \bar{s}\|_2 \\ &\leq \operatorname{dist}[(s, w), \hat{Z}_i] \\ &\leq \hat{\omega}_4(\|F^i(s)\|_2 + \underbrace{\|\beta(s) - w\|_2}_{=0} + \|\min\{0, \alpha(s)\}\|_2 + \|\min\{0, w\}\|_2) \\ &= \hat{\omega}_4(\|F^i(s)\|_2 + \|\min\{0, \alpha(s)\}\|_2 + \|\min\{0, \beta(s)\}\|_2). \end{aligned}$$

Therefore, Condition 4 is satisfied with $\delta_4 := \frac{\hat{\delta}_4}{1+l}$ and $\omega_4 := \hat{\omega}_4$. □

The essence of Proposition 8 is the following. Every condition which implies Condition 4 is also sufficient for its counterpart for system (19) to hold and therefore, assuming that the Ω-property is additionally satisfied, guarantees that the Newton-type methods described in the Introduction applied to (19) converge locally quadratically to a solution; see Theorem 1. We emphasize that the Q-quadratic convergence rate obtained this way is in the aggregated variable (z, w), which includes slacks. In principle, it does not imply the same convergence rate in the original variable z separately, or even monotonic decrease of the distance from the z-iterates to their limit. However, it implies R-quadratic rate of convergence in z variable, which still means that this convergence is fast.

In the next section we specify Condition 4 for problem (3) with Ω defined according to (4). In Sects. 6–8, we present conditions implying Condition 4 for special problem classes.

5 Individual error bounds for active pieces in case of complementarity systems

In this section, we analyze Condition 4 above for the case when the constrained equation arises from the complementarity system (1). As already mentioned, in this setting F defined in (3) is a PC^1-function, and the Newtonian methods employ G with $G(s) \in \{JF^i(s) \mid i \in \mathscr{A}(s)\}$, $s \in \mathbb{R}^n$. Let us define the index sets related to "activity" of the selection functions at z^*:

$$\mathscr{I}_c := \mathscr{I}_c(z^*) := \{i \in \{1, \ldots, r\} \mid 0 = c_i(z^*) < d_i(z^*)\},$$
$$\mathscr{I}_d := \mathscr{I}_d(z^*) := \{i \in \{1, \ldots, r\} \mid 0 = d_i(z^*) < c_i(z^*)\},$$
$$\mathscr{I}_0 := \mathscr{I}_0(z^*) := \{i \in \{1, \ldots, r\} \mid 0 = c_i(z^*) = d_i(z^*)\}.$$

Obviously, these index sets are a partition of the set $\{1, \ldots, r\}$, i.e., they are pairwise disjoint and their union is the set $\{1, \ldots, r\}$. It can be easily seen that a selection function F^i is active at z^* (i.e., $i \in \mathscr{A}(z^*)$) if and only if there is a partition $(\mathscr{I}_1, \mathscr{I}_2)$ of $\mathscr{I}_0$ such that F^i has, after some row permutations, the same structure as the function $F^{\mathscr{I}_1,\mathscr{I}_2}$ defined by

$$F^{\mathscr{I}_1,\mathscr{I}_2}(z) := \begin{pmatrix} a(z) \\ c_{\mathscr{I}_c \cup \mathscr{I}_1}(z) \\ d_{\mathscr{I}_d \cup \mathscr{I}_2}(z) \end{pmatrix}. \tag{26}$$

We define the set of all zeros of $F^{\mathscr{I}_1,\mathscr{I}_2}$ belonging to Ω by $Z_{\mathscr{I}_1,\mathscr{I}_2}$, i.e.,

$$Z_{\mathscr{I}_1,\mathscr{I}_2} = \{z \in \mathbb{R}^n \mid F^{\mathscr{I}_1,\mathscr{I}_2}(z) = 0,\ b(z) \geq 0,\ c(z) \geq 0,\ d(z) \geq 0\}.$$

Condition 5 below was already considered in [20] and is equivalent to Condition 4 for problem (3) with Ω given by (4).

Condition 5 (Piecewise error bound) There exist $\delta_5 > 0$ and $\omega_5 > 0$ such that

$$\begin{aligned}\operatorname{dist}[s, Z_{\mathscr{I}_1,\mathscr{I}_2}] \leq \omega_5 \big(&\|a(s)\| + \|c_{\mathscr{I}_c \cup \mathscr{I}_1}(s)\| + \|d_{\mathscr{I}_d \cup \mathscr{I}_2}(s)\| + \|\min\{0, b(s)\}\| \\ &+ \|\min\{0, c_{\mathscr{I}_d \cup \mathscr{I}_2}(s)\}\| + \|\min\{0, d_{\mathscr{I}_c \cup \mathscr{I}_1}(s)\}\|\big)\end{aligned}$$

holds for all partitions $(\mathscr{I}_1, \mathscr{I}_2)$ of $\mathscr{I}_0$ and for all $s \in \mathscr{B}_{\delta_5}(z^*)$.

In other words, Condition 5 requires that, for all partitions $(\mathscr{I}_1, \mathscr{I}_2)$ of $\mathscr{I}_0$, the local error bound condition is satisfied at z^* for the following system:

$$\begin{aligned} &a(z) = 0,\ c_{\mathscr{I}_c \cup \mathscr{I}_1}(z) = 0,\ d_{\mathscr{I}_d \cup \mathscr{I}_2}(z) = 0, \\ &b(z) \geq 0,\ c_{\mathscr{I}_d \cup \mathscr{I}_2}(z) \geq 0,\ d_{\mathscr{I}_c \cup \mathscr{I}_1}(z) \geq 0. \end{aligned} \tag{27}$$

The following theorem is an application of Theorem 2 to problem (3). Recall that the definition of Ω by (4) implies the Ω-property for system (3); thus, this property need not be required explicitly in the following theorem.

Theorem 3 *Let F and Ω be given by (3) and (4). If Condition 5 is satisfied, then Assumptions 1–4 hold.*

As is explained in [20], if, for each partition $(\mathscr{I}_1, \mathscr{I}_2)$ of $\mathscr{I}_0$, some constraint qualification implying the local error bound condition for (27) is satisfied at z^*, then Condition 5 holds. For example, if MFCQ or RCRCQ holds at z^* for system (27) for each partition $(\mathscr{I}_1, \mathscr{I}_2)$ of $\mathscr{I}_0$, then Condition 5 is valid. In [20], it is actually stated that Condition 5 implies Condition 1 for system (3) and so Assumption 2. However, the relation between Condition 5 and Assumption 3 was not analyzed in [20].

By Theorems 1 and 3 we know that the LP-Newton method as well as the constrained Levenberg–Marquardt method converge locally quadratically to a solution of (3) if Condition 5 is satisfied. It follows from the last section that the local quadratic convergence is preserved if Condition 5 [which is Condition 4 for system (3)] holds and one of those methods is applied to the following reformulation with slack variables:

$$\hat{F}(z,u,v,w) := \begin{pmatrix} a(z) \\ \min\{c(z), d(z)\} \\ b(z) - u \\ c(z) - v \\ d(z) - w \end{pmatrix} = 0 \quad \text{s.t.} \quad (z,u,v,w) \in \hat{\Omega}$$

with $\hat{\Omega} := \mathbb{R}^n \times \mathbb{R}^q_+ \times \mathbb{R}^{2r}_+$. Moreover, if we replace problem (3) by

$$\tilde{F}(z,u,v,w) := \begin{pmatrix} a(z) \\ \min\{v, w\} \\ b(z) - u \\ c(z) - v \\ d(z) - w \end{pmatrix} = 0 \quad \text{s.t.} \quad (z,u,v,w) \in \hat{\Omega}, \tag{28}$$

we find that the counterpart of Condition 4 for problem (28) is equivalent to Condition 4. This follows easily from Proposition 8 and the fact that, for any $(z, v, w) \in \mathbb{R}^n \times \mathbb{R}^{2r}$ and any $i \in \{1, \ldots, r\}$, we have

$$|\min\{c_i(z), d_i(z)\}| \le |c_i(z) - v_i| + |d_i(z) - w_i| + |\min\{v_i, w_i\}|$$

and

$$|\min\{v_i, w_i\}| \le |c_i(z) - v_i| + |d_i(z) - w_i| + |\min\{c_i(z), d_i(z)\}|.$$

Thus, the LP-Newton method and the constrained Levenberg–Marquardt method applied to system (28) exhibit local quadratic convergence if Condition 4 is satisfied.

6 KKT systems of optimization or variational problems

Consider the problem of solving the KKT system

$$\Theta(x) + Jg(x)^\top \lambda = 0, \quad \lambda \ge 0, \quad g(x) \le 0, \quad \lambda^\top g(x) = 0, \tag{29}$$

where $\Theta : \mathbb{R}^n \to \mathbb{R}^n$ is differentiable and has a locally Lipschitz-continuous derivative, and $g : \mathbb{R}^n \to \mathbb{R}^m$ is twice differentiable and has locally Lipschitz-continuous second-order derivatives. Equality constraints can easily be incorporated; we do not consider them here in order to simplify the presentation. For origins of problems with

the primal-dual structure like (29), see [10, Chapter 1] and [21, Chapter 1]. Let us define the function $\Psi : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ by

$$\Psi(x,\ \lambda) := \Theta(x) + Jg(x)^\top \lambda.$$

Moreover, for a given solution $(x^*,\ \lambda^*) \in \mathbb{R}^n \times \mathbb{R}^m$ of the KKT system (29), we define the index sets

$$\begin{array}{ll} \mathscr{A} := \{i \in \{1,\ \ldots,\ m\} \mid g_i(x^*) = 0\}, & \mathscr{N} := \{i \in \{1,\ \ldots,\ m\} \mid g_i(x^*) < 0\}, \\ \mathscr{A}_+ := \{i \in \mathscr{A} \mid \lambda_i^* > 0\}, & \mathscr{A}_0 := \{i \in \mathscr{A} \mid \lambda_i^* = 0\}. \end{array}$$

The set $\mathscr{A}$ consists of the indices of those constraints being active at x^*. The sets $\mathscr{A}$ and $\mathscr{N}$ partition the set $\{1,\ \ldots,\ m\}$. The critical cone of the system (29) at x^* is

$$\mathscr{C}(x^*) := \{\xi \in \mathbb{R}^n \mid Jg_{\mathscr{A}}(x^*)\xi \leq 0,\ \Theta(x^*)^T\xi = 0\}.$$

For x^* fixed and any choice of λ^* such that (x^*, λ^*) satisfies (29) it holds that

$$\mathscr{C}(x^*) = \{\xi \in \mathbb{R}^n \mid Jg_{\mathscr{A}_+}(x^*)\xi = 0,\ Jg_{\mathscr{A}_0}(x^*)\xi \leq 0\}, \tag{30}$$

see, e.g., [21, p. 32].

Local quadratic convergence of the LP-Newton method applied to an appropriate reformulation of the KKT system (29) is established in [7, Corollary 6] (thus, the same also holds for the constrained Levenberg–Marquardt method; see [6]). The only assumption required in this result is the following second-order condition (SOC):

$$\xi^\top J_x\Psi(x^*,\ \lambda^*)\xi \neq 0 \quad \text{for all } \xi \in \mathscr{C}(x^*)\backslash\{0\}, \tag{31}$$

where J_x denotes the partial Jacobian with respect to the variable x. We next show that quadratic convergence is in fact guaranteed under the noncriticality of the dual part λ^* of the solution $(x^*,\ \lambda^*)$ of (29). Noncriticality is a weaker assumption than SOC (31).

We note that the result on local quadratic convergence under noncriticality can also be derived using [7, Theorem 5] and the local primal-dual error bound which is known to be implied for KKT systems by noncriticality (see, e.g., [21, Proposition 1.43]). Here, we give a different argument which is, in a sense, more direct: we demonstrate that Condition 5 implying the whole set of Assumptions 1–4 holds under noncriticality of the Lagrange multiplier. Observe that the error bound itself is an immediate consequence of Proposition 9 below and of [20, Lemma 1].

In the terminology of [21, Definition 1.41] (see also [15,19]), a multiplier λ^* associated to a primal solution x^* of the KKT system (29), is called *noncritical* if there exists no pair $(\xi,\ \eta) \in \mathbb{R}^n \times \mathbb{R}^m$, with $\xi \neq 0$, satisfying

$$\begin{array}{l} J_x\Psi(x^*,\ \lambda^*)\xi + Jg(x^*)^\top\eta = 0, \\ Jg_{\mathscr{A}_+}(x^*)\xi = 0, \\ \eta_i \geq 0,\ \ Jg_i(x^*)\xi \leq 0,\ \ \eta_i Jg_i(x^*)\xi = 0,\ \ i \in \mathscr{A}_0, \\ \eta_{\mathscr{N}} = 0. \end{array} \tag{32}$$

It can be easily shown that if SOC (31) holds, then λ^* is automatically noncritical; see [15,19,21] for details. Also, it should be emphasized that usually, the class of noncritical multipliers is much wider than the class of multipliers satisfying SOC.

To start, we write (29) in the form of (1) with $p := n$, $q := 0$, $r := m$, and

$$z := (x,\ \lambda) \in \mathbb{R}^{n+m}, \quad a(z) := \Psi(x,\ \lambda), \quad c(z) := -g(x), \quad d(z) := \lambda.$$

In Sect. 5, we partitioned the set $\{1,\ \ldots,\ r\}$ into $\mathscr{I}_c$, $\mathscr{I}_d$, and $\mathscr{I}_0$. Here, for a fixed solution $z^* = (x^*,\ \lambda^*)$, these sets are $\mathscr{I}_c = \mathscr{A}_+$, $\mathscr{I}_d = \mathscr{N}$, $\mathscr{I}_0 = \mathscr{A}_0$. Therefore, Condition 5 in the context of KKT systems means that there exist $\delta_5 > 0$ and $\omega_5 > 0$ such that

$$\begin{aligned}\operatorname{dist}[s,\ Z_{\mathscr{I}_1,\,\mathscr{I}_2}] \le \omega_5(&\|\Psi(x,\ \lambda)\| + \|g_{\mathscr{A}_+\cup\mathscr{I}_1}(x)\| + \|\lambda_{\mathscr{N}\cup\mathscr{I}_2}\| \\ &+ \|\min\{0,\ -g_{\mathscr{N}\cup\mathscr{I}_2}(x)\}\| + \|\min\{0,\ \lambda_{\mathscr{A}_+\cup\mathscr{I}_1}\}\|)\end{aligned} \tag{33}$$

holds for all partitions $(\mathscr{I}_1,\ \mathscr{I}_2)$ of $\mathscr{A}_0$ and for all $s = (x,\ \lambda) \in \mathscr{B}_{\delta_5}(z^*)$, where by $Z_{\mathscr{I}_1,\,\mathscr{I}_2}$ we denote the solution set of the "branch" system

$$\begin{aligned}&\Psi(x,\ \lambda) &&= 0,\\ &g_{\mathscr{A}_+\cup\mathscr{I}_1}(x) &&= 0,\ \lambda_{\mathscr{N}\cup\mathscr{I}_2} = 0,\\ &g_{\mathscr{N}\cup\mathscr{I}_2}(x) &&\le 0,\ \lambda_{\mathscr{A}_+\cup\mathscr{I}_1} \ge 0.\end{aligned} \tag{34}$$

Proposition 9 *Let $z^* = (x^*,\ \lambda^*)$ be a solution of the KKT system (29), and let the multiplier λ^* be noncritical. Then, there exist $\delta_5 > 0$ and $\omega_5 > 0$ such that the inequality*

$$\begin{aligned}\|x - x^*\| + \mathit{dist}[\lambda, \mathscr{M}_{\mathscr{I}_1,\mathscr{I}_2}(x^*, \lambda^*)] \le \omega_5(&\|\Psi(x,\lambda)\| + \|g_{\mathscr{A}_+\cup\mathscr{I}_1}(x)\| + \|\lambda_{\mathscr{N}\cup\mathscr{I}_2}\| \\ &+ \|\min\{0, -g_{\mathscr{I}_2}(x)\}\| + \|\min\{0, \lambda_{\mathscr{I}_1}\}\|)\end{aligned} \tag{35}$$

holds for all partitions $(\mathscr{I}_1,\ \mathscr{I}_2)$ of $\mathscr{A}_0$ and all $(x,\ \lambda) \in \mathscr{B}_{\delta_5}(z^*)$, where $\mathscr{M}_{\mathscr{I}_1,\,\mathscr{I}_2}(x^*, \lambda^*)$ is the solution set of the system (34) with respect to λ, for $x = x^*$.

Proof Since the number of different partitions of $\mathscr{A}_0$ is finite, it is sufficient to prove the needed property for an arbitrary but fixed partition $(\mathscr{I}_1,\ \mathscr{I}_2)$.

For any $(x,\ \lambda) \in \mathbb{R}^n \times \mathbb{R}^m$, let $\rho_{\mathscr{I}_1,\,\mathscr{I}_2}(x,\ \lambda)$ stand for the expression in brackets on the right-hand side of (35). We first prove the primal estimate, i.e., the existence of $\omega_5 > 0$ such that

$$\|x - x^*\| \le \omega_5 \rho_{\mathscr{I}_1,\,\mathscr{I}_2}(x,\ \lambda) \tag{36}$$

holds for $(x,\ \lambda)$ close enough to $(x^*,\ \lambda^*)$. We argue by contradiction. Suppose that the assertion above is not true, which means the existence of a sequence $\{(x^k,\ \lambda^k)\} \subset \mathbb{R}^n \times \mathbb{R}^m$ convergent to $(x^*,\ \lambda^*)$, with $x^k \ne x^*$ for all k, and such that

$$\rho_{\mathscr{I}_1,\,\mathscr{I}_2}(x^k,\ \lambda^k) = o(\|x^k - x^*\|) \tag{37}$$

as $k \to \infty$. By the definition of $\rho_{\mathscr{I}_1,\,\mathscr{I}_2}(x,\,\lambda)$, the latter immediately implies

$$Jg_{\mathscr{A}_+\cup\mathscr{I}_1}(x^*)(x^k - x^*) = g_{\mathscr{A}_+\cup\mathscr{I}_1}(x^k) + o(\|x^k - x^*\|), \tag{38}$$

$$\lambda^k_{\mathscr{N}\cup\mathscr{I}_2} = o(\|x^k - x^*\|) \tag{39}$$

as $k \to \infty$. Furthermore, since the number of different partitions of $\mathscr{I}_1$ and $\mathscr{I}_2$ is finite, passing onto a subsequence, if necessary, we can assume (without loss of generality) that there exist a partition $(\mathscr{I}_1^1,\ \mathscr{I}_1^2)$ of $\mathscr{I}_1$, and a partition $(\mathscr{I}_2^1,\ \mathscr{I}_2^2)$ of $\mathscr{I}_2$, such that

$$g_{\mathscr{I}_2^1}(x^k) \le 0, \quad g_{\mathscr{I}_2^2}(x^k) > 0, \tag{40}$$

$$\lambda^k_{\mathscr{I}_1^1} \ge 0, \quad \lambda^k_{\mathscr{I}_1^2} < 0 \tag{41}$$

for all k. Then, according to the definition of $\rho_{\mathscr{I}_1,\mathscr{I}_2}(x,\lambda)$ and (37),

$$-g_{\mathscr{I}_2^2}(x^k) = \min\{0,\ -g_{\mathscr{I}_2^2}(x^k)\} = o(\|x^k - x^*\|), \tag{42}$$

$$\lambda^k_{\mathscr{I}_1^2} = \min\{0,\ \lambda^k_{\mathscr{I}_1^2}\} = o(\|x^k - x^*\|) \tag{43}$$

follow for $k \to \infty$. Moreover, employing (37) again, we derive

$$\begin{aligned}
0 &= \Psi(x^k, \lambda^k) + o(\|x^k - x^*\|)\\
&= J_x\Psi(x^*, \lambda^*)(x^k - x^*) + Jg(x^*)^\top(\lambda^k - \lambda^*) + o(\|x^k - x^*\|)\\
&= J_x\Psi(x^*, \lambda^*)(x^k - x^*) + Jg_{\mathscr{A}_+}(x^*)^\top(\lambda^k - \lambda^*)_{\mathscr{A}_+} + Jg_{\mathscr{I}_1^1}(x^*)^\top\lambda^k_{\mathscr{I}_1^1} + o(\|x^k - x^*\|),
\end{aligned}$$

where the third equality is by (39) and (43). Taking into account the first inequality in (41), we then conclude that

$$-Jg_{\mathscr{A}_+}(x^*)^\top\mathbb{R}^{|\mathscr{A}_+|} - Jg_{\mathscr{I}_1^1}(x^*)^\top\mathbb{R}_+^{|\mathscr{I}_1^1|} \ni J_x\Psi(x^*,\,\lambda^*)(x^k - x^*) + o(\|x^k - x^*\|) \tag{44}$$

as $k \to \infty$. Passing onto a further subsequence, if necessary, we can assume that the sequence $\{(x^k - x^*)/\|x^k - x^*\|\}$ converges to some $\xi \in \mathbb{R}^n$, $\|\xi\| = 1$. Since the set on the left-hand side of (44) is a closed cone (as a sum of linear subspaces and a polyhedral cone), dividing the right-hand side of this expression by $\|x^k - x^*\|$ and passing onto the limit as $k \to \infty$, we then obtain the inclusion

$$J_x\Psi(x^*,\,\lambda^*)\xi \in -Jg_{\mathscr{A}_+}(x^*)^\top\mathbb{R}^{|\mathscr{A}_+|} - Jg_{\mathscr{I}_1^1}(x^*)^\top\mathbb{R}_+^{|\mathscr{I}_1^1|}.$$

This means the existence of $\eta \in \mathbb{R}^m$ such that

$$\begin{aligned}
&J_x\Psi(x^*,\,\lambda^*)\xi + Jg(x^*)^\top\eta = 0,\\
&\eta_{\mathscr{I}_1^1} \ge 0, \quad \eta_{\mathscr{N}\cup\mathscr{I}_1^2\cup\mathscr{I}_2} = 0.
\end{aligned} \tag{45}$$

Furthermore, by similar manipulations, from (38), the first inequality in (40), and (42), we derive that

$$Jg_{\mathscr{A}_+\cup\mathscr{I}_1\cup\mathscr{I}_2^2}(x^*)\xi = 0, \quad Jg_{\mathscr{I}_2^1}(x^*)\xi \leq 0. \tag{46}$$

Combining (45) with (46) evidently yields that $(\xi,\ \eta)$ is a solution of the system (32), which contradicts the assumption that λ^* is a noncritical multiplier. This shows that (36) holds for all (x, λ) close to (x^*, λ^*) with some $\omega_5 > 0$.

To establish the remaining dual estimate, i.e.,

$$\operatorname{dist}[\lambda,\ \mathscr{M}_{\mathscr{I}_1,\,\mathscr{I}_2}(x^*, \lambda^*)] \leq \omega_5 \rho_{\mathscr{I}_1,\,\mathscr{I}_2}(x,\ \lambda)$$

for $(x,\ \lambda)$ close enough to $(x^*,\ \lambda^*)$ (perhaps with a larger ω_5), observe that the set $\mathscr{M}_{\mathscr{I}_1,\,\mathscr{I}_2}(x^*, \lambda^*)$ is given by linear constraints. The needed result is obtained by Hoffman's Lemma (e.g., [3, Theorem 2.200]), and by the primal estimate (36). □

Since (35) evidently implies (33), we conclude by Proposition 9 that noncriticality of the multiplier λ^* implies Condition 5. Therefore, according to Theorems 1 and 3, the methods considered in this paper converge quadratically under the assumption that the starting point is close enough to a solution $(x^*,\ \lambda^*)$ of the KKT system (29) with λ^* being a noncritical multiplier. This distinguishes these methods from, say, the stabilized Newton method for variational problems and the stabilized sequential quadratic programming for optimization, for which SOC (31) cannot be replaced by noncriticality when there are active inequality constraints (see [19] and [21, Section 7.2]).

7 KKT systems of generalized Nash equilibrium problems

The aim of this section is to find conditions which are sufficient for Condition 5 to hold if the complementarity system (1) arises from a KKT system of a generalized Nash equilibrium problem (GNEP for short). In particular, we will recover a result from [20] which says that the full row rank of a certain matrix implies Condition 5. Moreover, this result is extended by showing that Condition 5 is even implied by the constant rank of certain matrices.

We consider a GNEP with N players $\nu = 1, \ldots, N$ where the constraints of the players are described by inequalities. By n_ν the number of variables of player ν is denoted and by $n := \sum_{\nu=1}^{N} n_\nu$ the number of all variables. For the sake of simplicity, a GNEP with shared constraints only is considered so that the optimization problem of the ν-th player is given by

$$\min_{x^\nu} \quad \theta_\nu(x^\nu, x^{-\nu}) \quad \text{s.t.} \quad g(x^\nu, x^{-\nu}) \leq 0. \tag{47}$$

It is assumed that the functions $\theta_1, \ldots, \theta_N : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are twice differentiable and have locally Lipschitz-continuous second-order derivatives. The notation $(x^\nu, x^{-\nu})$ for a strategy vector $x \in \mathbb{R}^n$ is typically used in the context of GNEPs to emphasize the variables x^ν of the ν-th player and the variables $x^{-\nu}$ of all the rival players. For more information on GNEPs see [9,13] and references therein.

The KKT system associated with the ν-th player's optimization problem (47) is given by

$$(J_{x^\nu}\theta_\nu(x))^\top + (J_{x^\nu}g(x))^\top\lambda^\nu = 0, \quad \lambda^\nu \geq 0, \quad g(x) \leq 0, \quad (\lambda^\nu)^\top g(x) = 0$$

with some multiplier vector $\lambda^\nu \in \mathbb{R}^m$. By concatenating the KKT conditions of all players we obtain the KKT system of the GNEP:

$$\Theta(x) + B(x)\lambda = 0, \quad \lambda \geq 0, \quad g(x) \leq 0, \quad (\lambda^\nu)^\top g(x) = 0, \ \nu = 1, \ldots, N, \quad (48)$$

where

$$\lambda := \begin{pmatrix} \lambda^1 \\ \vdots \\ \lambda^N \end{pmatrix}, \quad \Theta(x) := \begin{pmatrix} (J_{x^1}\theta_1(x))^\top \\ \vdots \\ (J_{x^N}\theta_N(x))^\top \end{pmatrix},$$

and

$$B(x) := \begin{pmatrix} J_{x^1}g(x) & 0 & \cdots & 0 \\ 0 & J_{x^2}g(x) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & J_{x^N}g(x) \end{pmatrix}^\top.$$

In order to shorten the notation, we use from now on block$(\cdot)$ to denote a block diagonal matrix. Therefore, we can write for instance

$$B(x) = \text{block}\left((J_{x^\nu}g(x))^\top\right)_{\nu=1}^N.$$

By $z^* = (x^*, \lambda^*)$, an arbitrary but fixed solution of (48) is denoted. Let the index sets $\mathcal{N}$ and $\mathcal{A}$ be defined as in Sect. 6, and let us set

$$\mathcal{A}_+ := \{i \in \mathcal{A} \mid \exists \nu \in \{1, \ldots, N\} : \lambda_i^{*,\nu} > 0\}, \quad \mathcal{A}_0 := \mathcal{A}\backslash\mathcal{A}_+.$$

The sets $\mathcal{A}_+$ and $\mathcal{A}_0$ are a partition of $\mathcal{A}$. An index $i \in \{1, \ldots, m\}$ belongs to $\mathcal{A}_+$ if and only if g_i is active at x^* and the corresponding multiplier $\lambda_i^{*,\nu}$ of at least one player ν is strictly positive. The set $\mathcal{A}_0$ consists of the indices of those active constraints where the corresponding multipliers of all players are equal to zero. Moreover, for every $\nu \in \{1, \ldots, N\}$, we set

$$\mathcal{A}_+^\nu := \{i \in \mathcal{A} \mid \lambda_i^{*,\nu} > 0\}, \quad \mathcal{A}_0^\nu := \{i \in \mathcal{A} \mid \lambda_i^{*,\nu} = 0\}.$$

Obviously, for any ν, the sets $\mathcal{A}_+^\nu$ and $\mathcal{A}_0^\nu$ partition the set $\mathcal{A}$.

System (48) can be written as problem (1) with $p := n$, $q := 0$, $r := Nm$, and

$$z := (x, \lambda), \quad a(z) := \Theta(x) + B(x)\lambda, \quad c(z) := (-g(x))_{\nu=1}^{N}, \quad d(z) := \lambda.$$

In Sect. 5, depending on the fixed solution z^*, we partitioned the set $\{1, \ldots, r\}$ into $\mathscr{I}_c$, $\mathscr{I}_d$, and $\mathscr{I}_0$. By means of these index sets we stated Condition 5 for problem (1). In order to formulate Condition 5 for system (48), we now partition, for each player ν, the set $\{1, \ldots, m\}$ into $\mathscr{I}_c^\nu = \mathscr{A}_+^\nu$, $\mathscr{I}_d^\nu = \mathscr{N}$, and $\mathscr{I}_0^\nu = \mathscr{A}_0^\nu$.

It is not difficult to see that Condition 5 is satisfied in the context of GNEPs if and only if there are $\delta_5 > 0$ and $\omega_5 > 0$ such that the following inequality is satisfied for all partitions $(\mathscr{I}_1^\nu, \mathscr{I}_2^\nu)$ of the set $\mathscr{A}_0^\nu$ $(\nu = 1, \ldots, N)$ and all $s = (x, \lambda) \in \mathscr{B}_{\delta_5}(x^*, \lambda^*)$:

$$\begin{aligned} \operatorname{dist}[s, Z_{\{(\mathscr{I}_1^\nu, \mathscr{I}_2^\nu)\}_{\nu=1}^N}] \le \omega_5 \Bigg(& \|\Theta(x) + B(x)\lambda\| + \|g_{\mathscr{A}_+ \cup \mathscr{I}_1}(x)\| \\ & + \|\min\{0, -g_{\mathscr{N} \cup \mathscr{I}_2}(x)\}\| \\ & + \sum_{\nu=1}^{N} \left(\|\lambda^\nu_{\mathscr{N} \cup \mathscr{I}_2^\nu}\| + \|\min\{0, \lambda^\nu_{\mathscr{A}_+^\nu \cup \mathscr{I}_1^\nu}\}\| \right) \Bigg), \end{aligned} \tag{49}$$

where $\mathscr{I}_1$ and $\mathscr{I}_2$ are defined by

$$\mathscr{I}_1 := \left(\bigcup_{\nu=1}^{N} \mathscr{I}_1^\nu \right) \setminus \mathscr{A}_+ \quad \text{and} \quad \mathscr{I}_2 := \left(\bigcup_{\nu=1}^{N} \mathscr{I}_2^\nu \right) \setminus (\mathscr{A}_+ \cup \mathscr{I}_1), \tag{50}$$

respectively. By $Z_{\{(\mathscr{I}_1^\nu, \mathscr{I}_2^\nu)\}_{\nu=1}^N}$ the solution set of the system

$$\begin{aligned} & \Theta(x) + B(x)\lambda = 0, \ g_{\mathscr{A}_+ \cup \mathscr{I}_1}(x) = 0, \ g_{\mathscr{N} \cup \mathscr{I}_2}(x) \le 0, \\ & \lambda^\nu_{\mathscr{N} \cup \mathscr{I}_2^\nu} = 0, \ \lambda^\nu_{\mathscr{A}_+^\nu \cup \mathscr{I}_1^\nu} \ge 0, \quad \nu = 1, \ldots, N \end{aligned} \tag{51}$$

is indicated. Note that $Z_{\{(\mathscr{I}_1^\nu, \mathscr{I}_2^\nu)\}_{\nu=1}^N}$ is nonempty since z^* belongs to it. In other words, (49) requires that the system (51) of equations and inequalities satisfies the local error bound condition at $z^* = (x^*, \lambda^*)$. In the rest of this section we look for sufficient conditions for these error bound conditions and therefore for Condition 5 to hold. At first it turns out that the full row rank of a certain matrix implies the validity of Condition 5. This result was already proved in [20, Theorem 1] for the case of two players.

Condition 6 The following matrix has full row rank:

$$\left(\begin{array}{c|c} J_x\,(\Theta(x^*) + B(x^*)\lambda^*) & \operatorname{block}\left((J_{x^\nu} g_{\mathscr{A}_+^\nu}(x^*))^\top \right)_{\nu=1}^{N} \\ \hline Jg_{\mathscr{A}}(x^*) & 0 \end{array} \right),$$

where $J_{x^\nu} g_{\mathscr{A}_+^\nu}(x^*)$ consists of those rows of $J_{x^\nu} g(x^*)$ whose indices belong to $\mathscr{A}_+^\nu$.

Theorem 4 *Let Condition 6 be satisfied. Then, Condition 5 holds at z^*.*

We omit the proof of Theorem 4 because it is an immediate consequence of the next result which shows that even the constant rank of certain matrices in a neighborhood of (x^*, λ^*) is sufficient for Condition 5 to hold.

Condition 7 There exists $\delta_7 > 0$ such that, for each tuple $(\mathscr{K}, \mathscr{K}^1, \ldots, \mathscr{K}^N)$ of subsets $\mathscr{K} \subseteq \mathscr{A}_0$ and $\mathscr{K}^\nu \subseteq \mathscr{A}_0^\nu$ $(\nu = 1, \ldots, N)$, the matrices

$$\left(\begin{array}{c|c} J_x\left(\Theta(x) + B(x)\lambda\right) & \operatorname{block}\left(\left(J_{x^\nu} g_{\mathscr{A}_+^\nu \cup \mathscr{K}^\nu}(x)\right)^\top\right)_{\nu=1}^N \\ \hline Jg_{\mathscr{A}_+ \cup \mathscr{K}}(x) & 0 \end{array}\right)$$

have the same rank for all $(x, \lambda) \in \mathscr{B}_{\delta_7}(x^*, \lambda^*)$.

Note that Condition 7 is implied by Condition 6. In fact, Condition 6 implies that the matrix from Condition 7 has full row rank at z^* for every $(N+1)$-tuple $(\mathscr{K}, \mathscr{K}^1, \ldots, \mathscr{K}^N)$ of subsets $\mathscr{K} \subseteq \mathscr{A}_0$ and $\mathscr{K}^\nu \subseteq \mathscr{A}_0^\nu$ $(\nu = 1, \ldots, N)$. Due to the continuity of all functions involved in the matrices from Condition 7, the rows stay linearly independent for all points (x, λ) in a sufficiently small neighborhood of (x^*, λ^*).

Theorem 5 *Let Condition 7 be satisfied. Then, Condition 5 holds at z^*.*

Proof Let, for each player $\nu = 1, \ldots, N$, a partition $(\mathscr{I}_1^\nu, \mathscr{I}_2^\nu)$ of $\mathscr{A}_0^\nu$ be arbitrarily chosen and let $\mathscr{I}_1$ and $\mathscr{I}_2$ be defined according to (50). We are going to show that the system (51) satisfies RCRCQ at z^*; see Sect. 1 for details on RCRCQ. For our system (51), RCRCQ requires that, for each $(N+1)$-tuple $(\mathscr{L}, \mathscr{L}^1, \ldots, \mathscr{L}^N)$ of index sets $\mathscr{L} \subseteq \mathscr{I}_2$ and $\mathscr{L}^\nu \subseteq \mathscr{I}_1^\nu$ $(\nu = 1, \ldots, N)$, the matrices

$$\left(\begin{array}{c|c|c} J_x\left(\Theta(x) + B(x)\lambda\right) & \operatorname{block}\left(\left(J_{x^\nu} g_{\mathscr{N} \cup \mathscr{I}_2^\nu \cup \mathscr{L}^\nu}(x)\right)^\top\right)_{\nu=1}^N & C(x) \\ \hline Jg_{\mathscr{A}_+ \cup \mathscr{K}}(x) & 0 & 0 \\ \hline 0 & \operatorname{block}\left(I_{|\mathscr{N} \cup \mathscr{I}_2^\nu \cup \mathscr{L}^\nu|}\right)_{\nu=1}^N & 0 \end{array}\right) \tag{52}$$

have the same rank for all (x, λ) in a sufficiently small neighborhood of (x^*, λ^*), where the index sets $\mathscr{K}, \mathscr{K}^1, \ldots, \mathscr{K}^N$ are defined by $\mathscr{K} := \mathscr{I}_1 \cup \mathscr{L}$ and $\mathscr{K}^\nu := \mathscr{I}_1^\nu \setminus \mathscr{L}^\nu$, and the matrix $C(x)$ is given by $C(x) := \operatorname{block}\left(\left(J_{x^\nu} g_{\mathscr{A}_+^\nu \cup \mathscr{K}^\nu}(x)\right)^\top\right)_{\nu=1}^N$.

Generally, if the rank of some matrix $A_1 \in \mathbb{R}^{\rho \times \sigma}$ equals $k \in \mathbb{N}$, then the rank of the matrix

$$A_2 := \left(\begin{array}{c|c} A_1 & \alpha \\ \hline 0 & 1 \end{array}\right) \in \mathbb{R}^{(\rho+1)\times(\sigma+1)}$$

equals $k+1$ for any vector $\alpha \in \mathbb{R}^{\rho}$, because the last column of A_2 is obviously linearly independent of the other columns. Therefore, the maximal number of linear independent columns is increased by one. Thus, the rank of A_2 is also increased by one compared to the rank of A_1. With these observations we obtain that the rank of the matrix in (52) equals the rank of the matrix

$$\left(\begin{array}{c|c} J_x\left(\Theta(x)+B(x)\lambda\right) & C(x) \\ \hline Jg_{\mathscr{A}_+\cup\mathscr{K}}(x) & 0 \end{array}\right) \tag{53}$$

plus $\sum_{\nu=1}^{N}|\mathscr{N}\cup\mathscr{I}_2^{\nu}\cup\mathscr{L}^{\nu}|$. So, in particular, the matrices in (52) have the same rank for all (x,λ) near (x^*,λ^*) if and only if the matrices in (53) have this property. The matrices in (53) actually have the same rank for all $(x,\lambda)\in\mathscr{B}_{\delta_7}(x^*,\lambda^*)$ due to Condition 7. Therefore, the RCRCQ is satisfied for system (51) at z^*. Hence, the local error bound condition (49) also holds for this system since RCRCQ implies the local error bound condition; see Sect. 1 for details. Since the partitions $(\mathscr{I}_1^{\nu},\mathscr{I}_2^{\nu})$ of $\mathscr{A}_0^{\nu}$ $(\nu=1,\ldots,N)$ were arbitrarily chosen, Condition 5 is satisfied. □

Let us summarize the results of this section. We stated Condition 6 and Condition 7 which were shown to be sufficient for Condition 5 to hold. Condition 7 is implied by Condition 6. Taking into account Theorems 1 and 3, the LP-Newton method as well as the constrained Levenberg–Marquardt method, which we recalled in the Introduction, converge locally with a Q-quadratic rate to a solution of (48) if Condition 7 is satisfied.

It can be easily shown that Condition 5 implies [8, Assumption 2]. Condition 7 is in particular satisfied if the objectives θ_ν of the players are quadratic functions and the constraint functions g_i are affine. Thus, particularly in that case Condition 5 holds. From this result [8, Theorem 8] can be recovered. Moreover, in our setting the full row rank of the matrix in Condition 6 implies the full row rank assumption in [8, Theorem 9] where strict complementarity is additionally required. Hence, Theorem 4 improves [8, Theorem 9] since we do not require strict complementarity.

8 An error bound result for another reformulation of the KKT system of a GNEP

We still consider a GNEP with N players where the optimization problem of the ν-th player is given by (47). However, in this section the following smooth and constrained reformulation of the KKT system (48) is considered:

$$H(z):=H(x,\lambda,w):=\begin{pmatrix}\Theta(x)+B(x)\lambda\\ g(x)+w\\ w\circ\lambda^1\\ \vdots\\ w\circ\lambda^N\end{pmatrix}=0 \quad \text{s.t.} \quad z=(x,\lambda,w)\in\Omega, \tag{54}$$

where $w \circ \lambda^\nu$ denotes the Hadamard product of the vectors $w \in \mathbb{R}^m$ and $\lambda^\nu \in \mathbb{R}^m$, i.e., $w \circ \lambda^\nu = (w_i \lambda_i^\nu)_{i=1}^m$, and Ω is defined by $\Omega := \mathbb{R}^n \times \mathbb{R}_+^{Nm} \times \mathbb{R}_+^m$. A reformulation similar to (54) was used in [5]. In contrast to (54), different slack variables for repeated constraints were introduced there. Throughout this section, $z^* = (x^*, \lambda^*, w^*)$ denotes again an arbitrary but fixed solution of (54). In [5], a hybrid algorithm for the solution of (54) is described which, under appropriate assumptions, is both globally and locally quadratically convergent. The local part of this hybrid method is the LP-Newton method from [7] which we recalled in Sect. 1. The advantage of the smooth reformulation (54) of the KKT system is that Assumption 4 is satisfied at z^* and Assumption 3 holds if Assumption 2 is valid. This follows from [7, Corollary 1]. Therefore, Assumption 2 is the only assumption which is left to guarantee. Assumption 2 for system (54) at z^* says that there are $\delta > 0$ and $\omega > 0$ such that

$$\operatorname{dist}[s, Z] \leq \omega \|H(s)\|$$

holds for all $s = (x, \lambda, w) \in \mathscr{B}_\delta(z^*) \cap \Omega$. Of course, Z denotes the solution set of the constrained system (54).

In [5], a condition implying Assumption 2 is provided. However, this condition is in general even stronger than Condition 6 from the last section. The aim of this section is to present a weaker condition which is related to Condition 5 and which is sufficient for Assumption 2 to hold. Let the index sets $\mathscr{N}$, $\mathscr{A}$, $\mathscr{A}_+$, $\mathscr{A}_0$, $\mathscr{A}_0^\nu$, and $\mathscr{A}_+^\nu$ depending on the fixed solution z^* be defined as in Sect. 7.

Condition 8 The set $\mathscr{A}_0$ is empty, i.e., for each active constraint the multiplier of at least one player is strictly positive.

Assuming that Condition 8 is satisfied, we will show that Assumption 2 for system (54) is equivalent to Assumption 2 for the constrained system $H_{\min}(z) = 0$ s.t. $z \in \Omega$ with $H_{\min}$ defined by

$$H_{\min}(z) := H_{\min}(x, \lambda, w) := \begin{pmatrix} \Theta(x) + B(x)\lambda \\ g(x) + w \\ \min\{w, \lambda^1\} \\ \vdots \\ \min\{w, \lambda^N\} \end{pmatrix}.$$

Obviously, (54) and $H_{\min}(z) = 0$ s.t. $z \in \Omega$ have the same solution set Z.

Proposition 10 *Let Condition 8 be satisfied. Then, the following assertions are equivalent.*

(i) *There exist $\delta > 0$ and $\omega > 0$ such that*

$$\operatorname{dist}[s, Z] \leq \omega \|H(s)\|$$

holds for all $s = (x, \lambda, w) \in \mathscr{B}_\delta(z^*) \cap \Omega$.

(ii) *There exist* $\tilde{\delta} > 0$ *and* $\tilde{\omega} > 0$ *such that*

$$dist[s, Z] \leq \tilde{\omega} \|H_{\min}(s)\|$$

holds for all $s = (x, \lambda, w) \in \mathscr{B}_{\tilde{\delta}}(z^*) \cap \Omega$.

Proof We show that constants $\varepsilon > 0$, $C_1 > 0$, and $C_2 > 0$ exist such that

$$C_1 \|H(s)\|_\infty \leq \|H_{\min}(s)\|_\infty \leq C_2 \|H(s)\|_\infty \tag{55}$$

holds for all $s \in \mathscr{B}_\varepsilon(z^*) \cap \Omega$. Obviously, this suffices to prove the proposition. We first set

$$\mu := \frac{1}{2} \min \left\{ \left\{ w_i^* \mid i \in \mathscr{N} \right\} \cup \left\{ \lambda_i^{*,\nu} \mid i \in \mathscr{A}_+^\nu,\ \nu \in \{1, \ldots, N\} \right\} \right\},$$
$$\hat{\mu} := 2 \max \left\{ \left\{ w_i^* \mid i \in \mathscr{N} \right\} \cup \left\{ \lambda_i^{*,\nu} \mid i \in \mathscr{A}_+^\nu,\ \nu \in \{1, \ldots, N\} \right\} \right\},$$

and choose some $\varepsilon > 0$ which is small enough such that the following relations hold for all $s = (x, \lambda, w) \in \mathscr{B}_\varepsilon(z^*) \cap \Omega$:

$$\begin{array}{ll}
\hat{\mu} \geq w_i \geq \mu > 0, & \text{for all } i \in \mathscr{N}, \\
w_i \leq \mu, & \text{for all } i \in \mathscr{A}, \\
\hat{\mu} \geq \lambda_i^\nu \geq \mu > 0, & \text{for all } i \in \mathscr{A}_+^\nu \text{ and all } \nu \in \{1, \ldots, N\}, \\
\lambda_i^\nu \leq \mu, & \text{for all } i \in \mathscr{A}_0^\nu \cup \mathscr{N} \text{ and all } \nu \in \{1, \ldots, N\}.
\end{array}$$

Now, let $s = (x, \lambda, w) \in \mathscr{B}_\varepsilon(z^*) \cap \Omega$ be arbitrarily chosen. At first, we prove the right inequality in (55) by considering the components of $H_{\min}(s)$. Obviously,

$$|\Theta_i(x) + (B(x)\lambda)_i| \leq \|H(s)\|_\infty \tag{56}$$

holds for all $i = 1, \ldots, n$ and

$$|g_i(x) + w_i| \leq \|H(s)\|_\infty \tag{57}$$

is valid for all $i = 1, \ldots, m$. Let some $i \in \mathscr{N}$ and $\nu \in \{1, \ldots, N\}$ be given. Taking into account the choice of ε, we have

$$|\min\{w_i, \lambda_i^\nu\}| = |\lambda_i^\nu| = \frac{1}{|w_i|} |w_i \lambda_i^\nu| \leq \frac{1}{\mu} \|H(s)\|_\infty. \tag{58}$$

Moreover, for $i \in \mathscr{A}$ we know from Condition 8 and the choice of ε that there is $\nu(i)$ such that $\lambda_i^{\nu(i)} \geq \mu > 0$. Therefore, due to $w, \lambda^\nu \in \mathbb{R}_+^m$,

$$|\min\{w_i, \lambda_i^\nu\}| = \min\{w_i, \lambda_i^\nu\} \leq |w_i| = \frac{1}{|\lambda_i^{\nu(i)}|} |w_i \lambda_i^{\nu(i)}| \leq \frac{1}{\mu} \|H(s)\|_\infty \tag{59}$$

holds for $i \in \mathscr{A}$ and all $\nu \in \{1, \ldots, N\}$. By (56)–(59), we obtain

$$\|H_{\min}(s)\|_\infty \leq \max\left\{1, \frac{1}{\mu}\right\} \|H(s)\|_\infty. \tag{60}$$

Next, we show that the left inequality of (55) is valid. First, note that

$$|\Theta_i(x) + (B(x)\lambda)_i| \leq \|H_{\min}(s)\|_\infty \tag{61}$$

holds for $i = 1, \ldots, n$ and that

$$|g_i(x) + w_i| \leq \|H_{\min}(s)\|_\infty \tag{62}$$

is valid for $i = 1, \ldots, m$. Taking into account the choice of ε, we obtain

$$|w_i \lambda_i^\nu| \leq \hat{\mu} |\min\{w_i, \lambda_i^\nu\}| \leq \hat{\mu} \|H_{\min}(s)\|_\infty. \tag{63}$$

for $i \in \{1, \ldots, m\}$ and $\nu \in \{1, \ldots, N\}$. Combining (61)–(63) yields

$$\|H(s)\|_\infty \leq \max\{1, \hat{\mu}\} \|H_{\min}(s)\|_\infty. \tag{64}$$

Since s was arbitrarily chosen, (60) and (64) show that (55) holds with

$$C_1 := \min\{1,\ 1/\hat{\mu}\} \quad \text{and} \quad C_2 := \max\{1,\ 1/\mu\}.$$

This completes the proof. □

Proposition 10 tells us that, if Condition 8 holds, any condition implying item (ii) of this proposition also yields that Assumption 2 is satisfied for system (54). This significantly generalizes Theorem 2 in [4], at least for GNEPs with shared constraints. Taking into account the results of the last sections, we know that item (ii) of Proposition 10 is implied by Condition 5 or its equivalent formulation in the context of GNEPs from the last section. In fact, we showed that Condition 5 does not only imply Assumption 2 but also the counterpart of Assumption 2 for the constrained system with slack variables; see the discussion at the end of Sect. 5.

9 Concluding remarks

We demonstrated that for a certain class of Newton-type methods for PC^1-equations with possibly nonisolated solutions, the only structural assumption required for local quadratic convergence is the piecewise error bound, i.e., a local error bound for every branch of the solution set. The latter property is implied by various relatively weak piecewise CQs, for example. When applied to KKT systems arising from optimization or variational problems, our results imply local quadratic convergence under the assumption that the dual part of the solution is a noncritical Lagrange multiplier. For generalized Nash equilibrium problems, our results yield local quadratic convergence

under the relaxed constant rank condition. In all these cases, previous local convergence theories have been improved.

Acknowledgments Research of the Alexey F. Izmailov and Mikhail V. Solodov is supported in part by the Russian Foundation for Basic Research Grant 14-01-00113, by the Russian Science Foundation Grant 15-11-10021, by CNPq Grants 302637/2011-7 and PVE 401119/2014-9, by PRONEX–Optimization, and by FAPERJ.

References

1. Andreani, R., Haeser, G., Schuverdt, M.L., Silva, P.J.S.: A relaxed constant positive linear dependence constraint qualification and applications. Math. Program. **135**, 255–273 (2012)
2. Behling, R., Fischer, A.: A unified local convergence analysis of inexact constrained Levenberg–Marquardt methods. Optim. Lett. **6**, 927–940 (2012)
3. Bonnans, J.F., Shapiro, A.: Perturbation Analysis of Optimization Problems. Springer, New York (2000)
4. Dreves, A.: Improved error bound and a hybrid method for generalized Nash equilibrium problems. Comput. Optim. Appl. (2014). doi:10.1007/s10589-014-9699-z
5. Dreves, A., Facchinei, F., Fischer, A., Herrich, M.: A new error bound result for generalized Nash equilibrium problems and its algorithmic application. Comput. Optim. Appl. **59**, 63–84 (2014)
6. Facchinei, F., Fischer, A., Herrich, M.: A family of Newton methods for nonsmooth constrained systems with nonisolated solutions. Math. Methods Oper. Res. **77**, 433–443 (2013)
7. Facchinei, F., Fischer, A., Herrich, M.: An LP-Newton method: nonsmooth equations, KKT systems, and nonisolated solutions. Math. Program. **146**, 1–36 (2014)
8. Facchinei, F., Fischer, A., Piccialli, V.: Generalized Nash equilibrium problems and Newton methods. Math. Program. **117**, 163–194 (2009)
9. Facchinei, F., Kanzow, C.: Generalized Nash equilibrium problems. Ann. Oper. Res. **175**, 177–211 (2010)
10. Facchinei, F., Pang, J.-S.: Finite-Dimensional Variational Inequalities and Complementarity Problems. Springer, New York (2003)
11. Fernández, D., Solodov, M.: Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems. Math. Program. **125**, 47–73 (2010)
12. Fischer, A.: Modified Wilson's method for nonlinear programs with nonunique multipliers. Math. Oper. Res. **24**, 699–727 (1999)
13. Fischer, A., Herrich, M., Schönefeld, K.: Generalized Nash equilibrium problems—recent advances and challenges. Pesquisa Operacional **34**, 521–558 (2014)
14. Hager, W.W.: Lipschitz continuity for constrained processes. SIAM J. Control Optim. **17**, 321–338 (1979)
15. Izmailov, A.F., Kurennoy, A.S., Solodov, M.V.: A note on upper Lipschitz stability, error bounds, and critical multipliers for Lipschitz-continuous KKT systems. Math. Program. **142**, 591–604 (2013)
16. Izmailov, A.F., Kurennoy, A.S., Solodov, M.V.: The Josephy-Newton method for semismooth generalized equations and semismooth SQP for optimization. Set Valued Var. Anal. **21**, 17–45 (2013)
17. 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)
18. Izmailov, A.F., Solodov, M.V.: Newton-type methods for optimization problems without constraint qualifications. SIAM J. Optim. **15**, 210–228 (2004)
19. Izmailov, A.F., Solodov, M.V.: Stabilized SQP revisited. Math. Program. **133**, 93–120 (2012)
20. Izmailov, A.F., Solodov, M.V.: On error bounds and Newton-type methods for generalized Nash equilibrium problems. Comput. Optim. Appl. **59**, 201–218 (2014)
21. Izmailov, A.F., Solodov, M.V.: Newton-Type Methods for Optimization and Variational Problems. Springer Series in Operations Research and Financial Engineering. Springer International Publishing, Switzerland (2014)
22. Kanzow, C., Yamashita, N., Fukushima, M.: Levenberg–Marquardt methods for constrained nonlinear equations with strong local convergence properties. J. Comput. Appl. Math. **172**, 375–397 (2004)

23. Kojima, M., Shindo, S.: Extensions of Newton and quasi-Newton methods to systems of PC^1 equations. J. Oper. Res. Soc. Jpn. **29**, 352–374 (1986)
24. Kyparisis, J.: Sensitivity analysis for nonlinear programs and variational inequalities with nonunique multipliers. Math. Oper. Res. **15**, 286–298 (1990)
25. Lu, S.: Relation between the constant rank and the relaxed constant rank constraint qualifications. Optimization **61**, 555–666 (2012)
26. Mangasarian, O.L., Fromovitz, S.: The Fritz John necessary optimality conditions in the presence of equality and inequality constraints. J. Math. Anal. Appl. **17**, 37–47 (1967)
27. Minchenko, L., Stakhovski, S.: On relaxed constant rank regularity condition in mathematical programming. Optimization **60**, 429–440 (2011)
28. Robinson, S.M.: Stability theory for systems of inequalities, part II: differentiable nonlinear systems. SIAM J. Numer. Anal. **13**, 497–513 (1976)
29. Wright, S.J.: An algorithm for degenerate nonlinear programming with rapid local convergence. SIAM J. Optim. **15**, 673–696 (2005)