Pure and Applied Functional Analysis
Volume 8, Number 4, 2023, 1039–1054

Yokohama Publishers
ISSN 2189-3764 ONLINE JOURNAL

AN EXTENDED CONVERGENCE FRAMEWORK APPLIED TO COMPLEMENTARITY SYSTEMS WITH DEGENERATE AND NONISOLATED SOLUTIONS

ANDREAS FISCHER AND NICO STRASDAT

Abstract. Some classes of nonlinear complementarity systems, like optimality conditions for generalized Nash equilibrium problems, typically have nonisolated solutions. A reformulation of those systems as a constrained or unconstrained system of equations is often done by means of a nonsmooth complementarity function. Degenerate solutions then lead to points where the reformulated system is nonsmooth. Newton-type methods can have difficulties close to a nonisolated and degenerate solution. For this case, it is known that the LP-Newton method or a constrained Levenberg–Marquardt method may show local superlinear convergence provided that the complementarity function is piecewise linear. These results rely on error bounds for active pieces of the reformulation. We prove that a related result can be obtained for the Fischer–Burmeister complementarity function on the basis of a somewhat different Index Error Bound Condition. To this end, a new convergence framework is developed that allows significantly larger steps. Then, by a sophisticated analysis of the constrained Levenberg–Marquardt method and a corresponding choice of the regularization parameter, local superlinear convergence to a solution with an R-order of 4/3 is shown.

1. Introduction

For problems that can be written as a constrained equation

$$t(z) = 0 \quad \text{subject to} \quad z \in \Omega, \tag{1.1}$$

we suggest a framework for the local analysis of iterative methods. The function $t : \mathbb{R}^N \to [0, \infty)$ and the nonempty closed set $\Omega \subseteq \mathbb{R}^N$ are suitably given. Let us further consider some sequence $\{z^k\} \subset \mathbb{R}^N$ generated by

$$z^{k+1} := z^+(z^k), \quad k = 0, 1, 2, \dots,$$

where $z^+ : \Omega \to \Omega$ denotes some mapping and $z^0 \in \Omega$ lies in a sufficiently small neighborhood of a solution. Then, under some circumstances, the new framework provides results on the convergence of $\{z^k\}$ to a solution of (1.1) and a superlinear convergence rate for this sequence. Of course, the function t and the mapping z^+ have to satisfy certain assumptions, which will be specified later on. In contrast to several Newton-type methods with superlinear convergence properties that were

2020 *Mathematics Subject Classification.* 65K05, 49M15, 90C33, 90C30.
Key words and phrases. Complementarity system, degenerate solution, nonisolated solution, error bound, Newton-type method, large steps, Levenberg–Marquardt method, superlinear convergence.

designed for nonsmooth systems of equations with nonisolated solutions [4,5,9,17,18, 21], for smooth systems with nonisolated solutions [1,12,19,28,30], and for problems with nonunique multipliers arising from optimization and variational problems [13, 16, 24, 26, 31, 32], the new framework enables significantly larger steps. Note that the previous references are just a selection.

To apply the framework to a particular class of problems, we consider reformulations of the complementarity system

$$F(x) = 0,\ a(x) \geq 0,\ b(x) \geq 0,\ a(x)^\top b(x) = 0 \tag{1.2}$$

for given functions $F : \mathbb{R}^n \to \mathbb{R}^m$ and $a, b : \mathbb{R}^n \to \mathbb{R}^p$, which are assumed to be continuously differentiable with locally Lipschitz continuous Jacobians. Note that a solution x^* of (1.2) is called *degenerate* if $a_i(x^*) = b_i(x^*) = 0$ for at least one index $i \in \mathcal{N} := \{1, \ldots, p\}$. The reformulation of (1.2) as a constrained system of equations is done by means of a complementarity function (C-function for short) $\phi : \mathbb{R}^2 \to \mathbb{R}$ satisfying

$$\phi(c, d) = 0 \quad \text{if and only if} \quad c \geq 0,\ d \geq 0,\ cd = 0. \tag{1.3}$$

Our aim is to design a constrained Levenberg–Marquardt (LM) method with superlinear convergence in the neighborhood of a degenerate *and* nonisolated solution. The new convergence framework will turn out as a key to achieve this goal, whereas existing methods with the convergence property just mentioned are based on reformulations as constrained systems of equations based on the C-function $\phi_{\min} : \mathbb{R}^2 \to \mathbb{R}$ given by

$$\phi_{\min}(c, d) := \min\{a, b\},$$

see [9, 10, 21]. Instead of this or another piecewise linear C-function, we would like to use the Fischer–Burmeister (FB) C-function $\varphi : \mathbb{R}^2 \to \mathbb{R}$ with

$$\varphi(c, d) := c + d - \sqrt{c^2 + d^2}.$$

This function [14, 15] is not piecewise smooth. We also note that analyzing the behavior of Newton-type methods based on φ instead of a piecewise linear C-function is not only of theoretical interest. Rather, φ and corresponding reformulations might be an alternative for obtaining global convergence of Newton-type methods since φ^2 is continuously differentiable, see recent globalizations of related algorithms in [11, 25] based on φ and [6, 22, 23] for $\phi_{\min}$.

Complementarity systems with nonisolated solutions arise from numerous applications, for instance in Karush–Kuhn–Tucker (KKT) systems for constrained optimization with nonunique multipliers (see citations above), generalized Nash equilibrium problems [20, 27], or classes of optimization problems with a disjunctive structure of constraints [3, 29].

According to the discussion before, we reformulate the complementarity system (1.2) by means of the FB C-function φ. In addition, we use slack variables and consider the constrained system

$$T(z) := \begin{pmatrix} H(z) \\ \Phi(u, v) \end{pmatrix} = 0, \quad z = (x, u, v) \in \mathbb{R}^n \times \mathbb{R}^p_+ \times \mathbb{R}^p_+ \tag{1.4}$$

with $H : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}^{m+2p}$ and $\Phi : \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}^p$ given by

$$H(z) := \begin{pmatrix} F(x) \\ a(x) - u \\ b(x) - v \end{pmatrix} \quad \text{and} \quad \Phi(u,v) := \begin{pmatrix} \varphi(u_1, v_1) \\ \vdots \\ \varphi(u_p, v_p) \end{pmatrix}.$$

Since (1.3) is valid for any C-function, each solution of (1.4) uniquely corresponds to a solution of system (1.2), and vice versa. This would remain true, if the constraint $z = (x, u, v) \in \mathbb{R}^n \times \mathbb{R}^p_+ \times \mathbb{R}^p_+$ within (1.4) is removed. However, this (or a related) constraint turned out to be important for superlinear convergence results with starting points close to degenerate and nonisolated solutions, if the reformulation is based on the min-function [9,10,21]. This constraint will also become a main ingredient to obtain local superlinear convergence for the LM method presented below.

A recent attempt to use the FB C-function φ for the solution of KKT systems with superlinear convergence was given in [2] by means of specially designed constrained LM subproblems. However, it was just possible to prove superlinear convergence if the source of nonisolated solutions lies is the nonuniqueness of Lagrange multipliers.

The current paper suggests a constrained LM method for the solution of the constrained nonsmooth system (1.4). A superlinear rate of convergence is shown under assumptions weaker than those in [2], and which do not restrict the source of nonisolatedness.

The paper is organized as follows. The new convergence framework is presented and analyzed in Section 2. Then, Section 3 continues with preliminaries for the constrained LM method, i.e., we first provide the LM subproblems on which our Newton-type algorithm is based on. Further, the Index Error Bound Condition as main assumption and basic results are given for later use. In Section 4, we analyze a single LM step. This and the convergence framework of Section 2 are then used to prove superlinear convergence of the constrained LM method with an R-order of 4/3 under suitable assumptions in Section 5.

Throughout the paper, $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^l$ and $\mathcal{B}(s^*, \delta)$ is the closed Euclidean ball around $s^* \in \mathbb{R}^l$ with radius $\delta > 0$. For some nonempty set $S \subseteq \mathbb{R}^l$, the distance of $w \in \mathbb{R}^l$ to S is defined by $\operatorname{dist}[w, S] := \inf_{s \in S} \|s - w\|$. Further notation will be introduced when needed.

2. General convergence framework

The following general theorem will be a main tool for the analysis of the constrained LM method described in the next section. Since the theorem does not depend on a particular algorithm or problem, it may be helpful for the design and convergence analysis of other algorithms as well.

To describe a general iteration sequence, the theorem makes use of a nonempty closed set $\Omega \subseteq \mathbb{R}^N$ and of a mapping $z^+ : \Omega \to \Omega$. Then, given $z^0 \in \Omega$,

$$z^{k+1} := z^+(z^k), \quad k = 0, 1, 2, \ldots \tag{2.1}$$

defines the sequence $\{z^k\} \subset \Omega$ generated by the algorithm one is interested in.

Theorem 2.1. *Let $t : \mathbb{R}^N \to [0,\infty)$ be a continuous function, $\Omega \subseteq \mathbb{R}^N$ be a nonempty closed set, and $z^* \in Z := \{z \in \Omega \mid t(z) = 0\}$. Moreover, let $z^+ : \Omega \to \Omega$ denote some mapping. Suppose that $C \geq 1$, $\delta > 0$, and $\sigma > 1$ exist so that the conditions*

$$\begin{array}{lll} \text{(a)} \quad \|z^+(z) - z\| & \leq & Ct(z), \\ \text{(b)} \quad t(z^+(z)) & \leq & Ct(z)^\sigma \end{array}$$

are satisfied for all $z \in \mathcal{B}(z^*, \delta) \cap \Omega$.

Then, there exists $\varepsilon > 0$ so that, for any $z^0 \in \mathcal{B}(z^*, \varepsilon) \cap \Omega$, the sequence $\{z^k\}$ defined by (2.1) is contained in the set $\mathcal{B}(z^*, \delta) \cap \Omega$ and converges to some $\hat{z} \in Z$ with an R-order of at least σ.

If, in addition, the function t is Hölder continuous on $\mathcal{B}(z^*, \delta)$, i.e., if there exist $L_0 > 0$ and $\theta \in (0,1]$ so that

$$|t(z) - t(\tilde{z})| \leq L_0 \|z - \tilde{z}\|^\theta \quad \textit{for all } z, \tilde{z} \in \mathcal{B}(z^*, \delta), \tag{2.2}$$

then, for each $\nu \in \mathbb{N}$, there exists $c_\nu > 0$ so that

$$\|z^{k+\nu} - \hat{z}\| \leq c_\nu \|z^k - \hat{z}\|^{\theta\sigma^\nu} \tag{2.3}$$

holds for all $k \in \mathbb{N}$ large enough. Thus, if $\theta\sigma^\nu > 1$, the sequence $\{z^k\}$ converges to $\hat{z}$ with the ν-step Q-order of at least $\theta\sigma^\nu$.

Proof. Let us first define $\varkappa := \frac{1}{\sigma - 1} > 0$ and

$$r := \min\left\{\exp\left(-\frac{\ln(2)}{\ln(\sigma)}\right), \frac{\delta}{4C^{1-\varkappa}}\right\} \in (0,1). \tag{2.4}$$

Since t is continuous and $t(z^*) = 0$, there exists some $\varepsilon \in (0, \frac{\delta}{2}]$ such that

$$t(z^0) \leq rC^{-\varkappa} \tag{2.5}$$

is satisfied for all $z^0 \in \mathcal{B}(z^*, \varepsilon) \cap \Omega$.

Subsequently, we show by induction that $z^k \in \mathcal{B}(z^*, \delta) \cap \Omega$ for all $k \in \mathbb{N}$. By the assumption on ε, we have $z^0 \in \mathcal{B}(z^*, \delta) \cap \Omega$. Now, suppose that $z^l \in \mathcal{B}(z^*, \delta) \cap \Omega$ for all $l \in \{0, \ldots, k\}$ for some fixed $k \in \mathbb{N}$. Then, $z^{k+1} \in \mathcal{B}(z^*, \delta) \cap \Omega$ has to be shown. Since $z^{k+1} \in \Omega$ follows from the definition of the function z^+, just $\|z^{k+1} - z^*\| \leq \delta$ needs a proof. Using condition (b) and the geometric sum formula, we obtain

$$\begin{array}{rcl} t(z^l) & \leq & Ct(z^{l-1})^\sigma \\ & \leq & C(Ct(z^{l-2})^\sigma)^\sigma = C^{1+\sigma} t(z^{l-2})^{\sigma^2} \\ & \vdots & \\ & \leq & C^{1+\sigma+\cdots+\sigma^{\nu-1}} t(z^{l-\nu})^{\sigma^\nu} = C^{\frac{\sigma^\nu - 1}{\sigma - 1}} t(z^{l-\nu})^{\sigma^\nu} \\ & \leq & C^{\varkappa\sigma^\nu} t(z^{l-\nu})^{\sigma^\nu} \end{array} \tag{2.6}$$

for all $\nu \in \{0, \ldots, l\}$. Furthermore, from (2.5), $r \in (0,1)$ according to (2.4), and

$$\sigma^l \geq 1 + \ln(\sigma^l) = 1 + l\ln(\sigma), \tag{2.7}$$

it follows that

$$\left(C^\varkappa t(z^0)\right)^{\sigma^l} \leq r^{\sigma^l} \leq r^{1 + l\ln(\sigma)} = r\left(r^{\ln(\sigma)}\right)^l. \tag{2.8}$$

Using condition (a), (2.6) with $\nu = l$, (2.8), and (2.4), we obtain

$$\begin{aligned}\|z^{l+1} - z^l\| &= \|z^+(z^l) - z^l\| \\ &\le Ct(z^l) \\ &\le C^{1-\varkappa}\left(C^{\varkappa}t(z^0)\right)^{\sigma^l} \\ &\le rC^{1-\varkappa}\left(r^{\ln(\sigma)}\right)^l \\ &\le rC^{1-\varkappa}2^{-l}\end{aligned} \tag{2.9}$$

for all $l \in \{1, \ldots, k\}$. This and the triangle inequality imply

$$\|z^{k+1} - z^*\| \le \sum_{l=0}^{k} \|z^{l+1} - z^l\| + \|z^0 - z^*\| \le rC^{1-\varkappa}\sum_{l=0}^{k} 2^{-l} + \|z^0 - z^*\|.$$

By means of the geometric series formula and (2.4), we get

$$\|z^{k+1} - z^*\| \le 2rC^{1-\varkappa} + \|z^0 - z^*\| \le \delta,$$

i.e., $z^{k+1} \in \mathcal{B}(z^*, \delta) \cap \Omega$. Thus, $z^k \in \mathcal{B}(z^*, \delta) \cap \Omega$ for all $k \in \mathbb{N}$. In particular, (2.6) and (2.9) hold for all $l \in \mathbb{N}$. This, the triangle inequality, and the geometric sum formula lead to

$$\|z^{k+\nu} - z^k\| \le \sum_{l=0}^{\nu-1} \|z^{k+l+1} - z^{k+l}\| \le 2^{-k}rC^{1-\varkappa}\sum_{l=0}^{\nu-1} 2^{-l} \le 2^{1-k}rC^{1-\varkappa}$$

for any positive $k, \nu \in \mathbb{N}$. Therefore, $\{z^k\} \subset \mathcal{B}(z^*, \delta) \cap \Omega$ is a Cauchy sequence and converges to some $\hat{z} \in \mathcal{B}(z^*, \delta) \cap \Omega$. Moreover, (2.9) implies $t(z^k) \le rC^{-\varkappa}2^{-k}$ for all $k \in \mathbb{N}$, i.e., the sequence $\{t(z^k)\}$ converges to zero. Since t is continuous, $\hat{z} \in Z$ follows.

We now show that the sequence $\{z^k\}$ converges to $\hat{z}$ with an R-order of σ. Applying the triangle inequality and condition (a), we get

$$\|z^{k+\mu} - z^k\| \le \sum_{l=0}^{\mu-1} \|z^{k+l+1} - z^{k+l}\| \le C\sum_{l=k}^{k+\mu-1} t(z^l) \tag{2.10}$$

for any $k \in \mathbb{N}$ and any positive $\mu \in \mathbb{N}$. Using (2.6) with $\nu := l - k$, the right-hand side of (2.10) can be further estimated by

$$C\sum_{l=k}^{k+\mu-1} t(z^l) \le C\sum_{l=k}^{k+\mu-1} (C^{\varkappa}t(z^k))^{\sigma^{l-k}} = C\sum_{l=0}^{\mu-1}(C^{\varkappa}t(z^k))^{\sigma^l}. \tag{2.11}$$

As the sequence $\{t(z^k)\}$ converges to zero, there exists some index k_0 such that $C^{\varkappa}t(z^k) \le 1/2$ for all $k \ge k_0$. Thus, exploiting (2.7), we have, for all $k \ge k_0$,

$$C\sum_{l=0}^{\mu-1}(C^{\varkappa}t(z^k))^{\sigma^l} \le C\sum_{l=0}^{\mu-1}(C^{\varkappa}t(z^k))^{1+l\ln(\sigma)} \le C^{\varkappa+1}t(z^k)\sum_{l=0}^{\mu-1}\left(\frac{1}{2^{\ln(\sigma)}}\right)^l.$$

Further, since $2^{-\ln(\sigma)} \in (0,1)$,

$$C\sum_{l=0}^{\mu-1}(C^{\varkappa}t(z^k))^{\sigma^l} \leq \frac{C^{\varkappa+1}}{1-2^{-\ln(\sigma)}}t(z^k) = Kt(z^k)$$

holds for all positive $\mu \in \mathbb{N}$, all $k \in \mathbb{N}$ with $k \geq k_0$, and with some $K > 0$ suitably defined. Thus, taking into account (2.10) and (2.11), it follows that

$$\|\hat{z} - z^k\| = \lim_{\mu\to\infty}\|z^{k+\mu} - z^k\| \leq C\sum_{l=0}^{\mu-1}(C^{\varkappa}t(z^k))^{\sigma^l} \leq Kt(z^k) \tag{2.12}$$

for all $k \geq k_0$. As the sequence $\{t(z^k)\}$ converges to zero with a Q-order of at least σ, the sequence $\{z^k\}$ converges to $\hat{z}$ with an R-order of σ.

To complete the proof, let us finally suppose that the function t is Hölder continuous according to (2.2). Then, if we replace k in (2.12) by $k+\nu$ and l in (2.6) by $k+\nu$, we get

$$\|z^{k+\nu} - \hat{z}\| \leq Kt(z^{k+\nu}) \leq KC^{\varkappa\sigma^\nu}t(z^k)^{\sigma^\nu} \leq KC^{\varkappa\sigma^\nu}(L_0\|z^k - \hat{z}\|^\theta)^{\sigma^\nu}.$$

for all $k \in \mathbb{N}$ sufficiently large. With $c_\nu := KC^{\varkappa\sigma^\nu}{L_0}^{\sigma^\nu}$, this yields (2.3). □

Lemma 2.9 in [19] can somehow be regarded as a predecessor of the above theorem. However, the lemma is significantly more restricted and does not allow convergence results like in Section 5.

3. The constrained Levenberg–Marquardt method

As already announced, our aim is to exploit the framework from the previous section for the local convergence analysis of a constrained LM method. In Subsection 3.1 below, this LM method is described, whereas Subsections 3.2 and 3.3 provide the Index Error Bound Condition and basic assertions needed for the analysis of a single step of the constrained LM method in Section 4.

With regard to the reformulated complementarity system in (1.4), let us first specify $N := n + 2p$ and

$$\Omega := \{z = (x,u,v) \mid x \in \mathbb{R}^N,\ u \in \mathbb{R}^p_+,\ v \in \mathbb{R}^p_+\}.$$

Moreover, let the solution set of system (1.4) be denoted by

$$\mathcal{Z} := \{z = (x,u,v) \in \Omega \mid T(z) = 0\}.$$

3.1. **Levenberg–Marquardt subproblems.**

We now describe the subproblems of the constrained LM method for treating problem (1.4). Obviously, the mapping T is not differentiable at points $z = (x,u,v)$ with $u_i = v_i = 0$ for at least one index $i \in \mathcal{N} = \{1,\dots,n\}$. Nevertheless, T is locally Lipschitz continuous. Hence, Clarke's generalized Jacobian $\partial T(z)$ is well-defined for any z. Thus, as substitute for the Jacobian of T, a mapping $G : \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^p \to \mathbb{R}^{(m+3p)\times(n+2p)}$ with $G(z) \in \partial T(z)$ is used, which is defined by

$$G(z) := \begin{pmatrix} F'(x) & 0 & 0 \\ a'(x) & -I & 0 \\ b'(x) & 0 & -I \\ 0 & \mathrm{diag}(\alpha_i) & \mathrm{diag}(\beta_i) \end{pmatrix},$$

where $(\alpha_i, \beta_i) \in \partial\varphi(u_i, v_i)$ for $i \in \mathcal{N}$.

For any given $z \in \Omega$, we employ the following LM subproblem to compute the new point z^+

$$\underset{z^+}{\text{minimize}}\ \psi(z^+, z) \quad \text{subject to} \quad z^+ \in \Omega \tag{3.1}$$

with

$$\psi(z^+, z) := \|T(z) + G(z)(z^+ - z)\|^2 + \lambda(z)\|z^+ - z\|^2, \tag{3.2}$$

where the regularization parameter is given by

$$\lambda(z) := \begin{cases} \|T(z)\|^\gamma, & \text{if } z \in \Omega \setminus \mathcal{Z}, \\ 1, & \text{if } z \in \mathcal{Z}, \end{cases} \tag{3.3}$$

for some fixed $\gamma > 0$. Note that different choices of γ for the local superlinear convergence of constrained LM methods were investigated. Values of $\gamma = 2$ in [28] and $\gamma \in [1, 2]$ in [1] yield to local quadratic convergence by different proof techniques. In our case, the choice of γ will turn out to be more crucial.

By construction, the objective function $\psi(\cdot, z)$ in (3.1) is uniformly convex. Hence, problem (3.1) has a unique solution for each $z \in \Omega$, which we denote by $z^+(z)$. The mapping $z^+ : \Omega \to \Omega$ is the basis of our LM method and will be applied in the framework of Section 2. Properties of this mapping will be investigated in detail in Section 4.

3.2. **Index Error Bound Condition.** In this subsection, the Index Error Bound Condition is introduced. It is the fundamental assumption in our setting.

To start with, let us define the mappings

$$T_{IJ}(z) := \begin{pmatrix} H(z) \\ u_I \\ v_J \end{pmatrix}$$

and the corresponding sets of zeros of T_{IJ} in Ω

$$\mathcal{Z}_{IJ} := \{z \in \Omega \mid T_{IJ}(z) = 0\} \quad \text{for } I, J \subseteq \mathcal{N}.$$

Let $z^* = (x^*, u^*, v^*) \in \mathcal{Z}$ denote some fixed solution of problem (1.2). For any $w \in \mathbb{R}^p$, let further the index set $I_0(w) := \{i \in \mathcal{N} \mid w_i = 0\}$ be defined.

Assumption 1 (Index Error Bound Condition)**.** There exists $\omega > 0$ and $\delta > 0$ so that, for all (I, J) satisfying $I \subseteq I_0(u^*)$, $J \subseteq I_0(v^*)$, and $\mathcal{N} = I \cup J$,

$$\omega\, \mathrm{dist}[z, \mathcal{Z}_{IJ}] \le \|T_{IJ}(z)\|$$

holds for all $z \in \mathcal{B}(z^*, \delta) \cap \Omega$.

Note that, under Assumption 1, $\mathcal{Z}_{IJ}$ is a subset of $\mathcal{Z}$. We would further like to mention that Assumption 1 implies the Constrained Error Bound Condition as stated in Proposition 3.1.

Proposition 3.1. *Let Assumption 1 be satisfied. Then, the Constrained Error Bound Condition holds at z^*, i.e., there exist $\omega_E > 0$ and $\delta_E > 0$ so that*

$$\omega_E\, \mathrm{dist}[z, \mathcal{Z}] \le \|T(z)\| \quad \textit{for all } z \in \mathcal{B}(z^*, \delta_E) \cap \Omega. \tag{3.4}$$

This can be seen by means of Proposition 4 in [21]. The Constrained Error Bound Condition is an important ingredient for proving local superlinear convergence of Newton-type methods for constrained *smooth* systems of equations having nonisolated solutions, see [1, 28].

3.3. **Basic assertions.** Here, we present some basic results that will be helpful in Section 4 to analyze a single step of the constrained LM method. The first statement is a simple consequence of Taylor's formula, the assumed smoothness of the functions F, a, b, and the (global) Lipschitz continuity of φ.

Lemma 3.2. *There exists* $L > 0$ *so that the inequalities*

(a) $\|H(z) + H'(z)(\tilde{z} - z) - H(\tilde{z})\| \le L\|\tilde{z} - z\|^2$ *and*

(b) $\|T(z)\| \le L \operatorname{dist}[z, \mathcal{Z}]$

are satisfied for all $z, \tilde{z} \in \mathcal{B}(z^*, 1)$.

The next lemma summarizes some basic properties of the C-function φ.

Lemma 3.3. *Let* $c, d \in \mathbb{R}$, $\tilde{c}, \tilde{d} \in \mathbb{R}_+$, *and* $(\alpha, \beta) \in \partial\varphi(c, d)$ *be arbitrarily chosen. Then, the following assertions are valid:*

$$
\begin{array}{llcl}
\text{(a)} & c^2 + d^2 > 0 & \Rightarrow & \varphi'(c,d) = (\alpha, \beta) = \left(1 - \frac{c}{\sqrt{c^2+d^2}}, 1 - \frac{d}{\sqrt{c^2+d^2}}\right), \\
\text{(b)} & c^2 + d^2 > 0 & \Rightarrow & \varphi(c,d) = \varphi'(c,d)\begin{pmatrix} c \\ d \end{pmatrix}, \\
\text{(c)} & c > 0 & \Rightarrow & \alpha \le \dfrac{d^2}{2c^2}, \\
\text{(d)} & d > 0 & \Rightarrow & \beta \le \dfrac{c^2}{2d^2}, \\
\text{(e)} & c \ge 0,\ d \ge 0 & \Rightarrow & \varphi(c,d) \ge 0, \\
\text{(f)} & \alpha \ge 0 & \text{and} & \beta \ge 0, \\
\text{(g)} & (1-\alpha)^2 + (1-\beta)^2 \le 1, & & \\
\text{(h)} & \varphi(\tilde{c}, \tilde{d}) & \le & \varphi(c,d) + \begin{pmatrix} \alpha \\ \beta \end{pmatrix}^\top \begin{pmatrix} \tilde{c} - c \\ \tilde{d} - d \end{pmatrix}.
\end{array}
$$

Proof. Assertions (a)–(f) can be found in [2, Lemma 3.2], (g) follows from (a) by the definition of Clarke's Jacobian, see also [8, Proposition 3.1], for example.

We thus prove assertion (h) only. If $c^2 + d^2 > 0$, then assertion (b) yields

$$
\varphi(c,d) + \begin{pmatrix} \alpha \\ \beta \end{pmatrix}^\top \begin{pmatrix} \tilde{c} - c \\ \tilde{d} - d \end{pmatrix} = \alpha\tilde{c} + \beta\tilde{d}. \tag{3.5}
$$

This clearly holds for $c = d = 0$ as well. In addition, by the Cauchy–Schwarz inequality and assertion (g), we get

$$
\begin{aligned}
\varphi(\tilde{c}, \tilde{d}) - \alpha\tilde{c} - \beta\tilde{d} &= (1-\alpha)\tilde{c} + (1-\beta)\tilde{d} - \sqrt{\tilde{c}^2 + \tilde{d}^2} \\
&\le \sqrt{(1-\alpha)^2 + (1-\beta)^2}\,\sqrt{\tilde{c}^2 + \tilde{d}^2} - \sqrt{\tilde{c}^2 + \tilde{d}^2} \\
&= \sqrt{\tilde{c}^2 + \tilde{d}^2}\left(\sqrt{(1-\alpha)^2 + (1-\beta)^2} - 1\right) \\
&\le 0.
\end{aligned}
$$

In combination with (3.5), this shows that assertion (h) is valid. □

4. Analysis of a single Levenberg–Marquardt step

The first lemma of the section is a means to construct a special solution $z^\diamond \in \mathcal{Z}$ that is related to some neighboring point $z \in \Omega$. This special solution is defined as a projection of the neighboring point onto a particular subset of the solution set $\mathcal{Z}$.

To formulate and prove the lemma, we use the index set

$$I_\rho(w) := \{i \in \mathcal{N} \mid w_i \le \rho\} \quad \text{for } w \in \mathbb{R}^p \text{ and } \rho \in \mathbb{R}.$$

Lemma 4.1. *Let Assumption 1 be satisfied. Then, there exist $\delta_P > 0$, $\rho_P > 0$, and $C_P > 0$ so that, for all $z = (x,u,v) \in \mathcal{B}(z^*, \delta_P) \cap \Omega$ and all $\rho \in \mathbb{R}$ with* $\operatorname{dist}[z, \mathcal{Z}] \le \rho \le \rho_P$*, there is $z^\diamond \in \mathcal{Z}$ with the properties*

(a) $\|z^\diamond - z\| \le C_P \rho$,

(b) $u_i^\diamond = 0$ *for all* $i \in I_\rho(u)$ *and* $v_i^\diamond = 0$ *for all* $i \in I_\rho(v)$.

Proof. We first define

$$\rho_P := \frac{1}{2} \min\{\min\{u_i^* \mid i \in \mathcal{N} \text{ with } u_i^* > 0\}, \min\{v_i^* \mid i \in \mathcal{N} \text{ with } v_i^* > 0\}\},$$

$$\delta_P := \min\left\{1, \delta, \frac{1}{2}\rho_P\right\} \quad \text{and} \quad C_P := \frac{\sqrt{L^2 + 2p}}{\omega}.$$

Let $z \in \mathcal{B}(z^*, \delta_P) \cap \Omega$ and $\rho \in \mathbb{R}$ with $\operatorname{dist}[z, \mathcal{Z}] \le \rho \le \rho_P$ be arbitrarily given. By definition of δ_P, we particularly have $z \in \mathcal{B}(z^*, \delta) \cap \Omega$ with $\delta > 0$ from Assumption 1.

Subsequently, we are going to use Assumption 1 with

$$I := I_\rho(u) \quad \text{and} \quad J := I_\rho(v).$$

To this end, we show that $I \subseteq I_0(u^*)$ and $J \subseteq I_0(v^*)$ with $\mathcal{N} = I \cup J$.

By assumption, the solution set $\mathcal{Z}$ is closed and not empty. Hence, there is $z^\perp \in \mathcal{Z}$ with $\|z^\perp - z\| = \operatorname{dist}[z, \mathcal{Z}]$. Since $z^\perp \in \mathcal{Z}$, we have $u_i^\perp = 0$ or $v_i^\perp = 0$ for all $i \in \mathcal{N}$. In particular,

$$u_i = u_i - u_i^\perp \le \operatorname{dist}[z, \mathcal{Z}] \le \rho \quad \text{or} \quad v_i = v_i - v_i^\perp \le \operatorname{dist}[z, \mathcal{Z}] \le \rho$$

holds for all $i \in \mathcal{N}$. Thus, $\mathcal{N} = I_\rho(u) \cup I_\rho(v)$ follows. Moreover, because of

$$u_i^* \le u_i + \delta_P \le \rho + \delta_P \le \rho_P + \delta_P \le \frac{3}{2}\rho_P \le \frac{3}{4} \min\{u_i^* \mid i \in \mathcal{N} \text{ with } u_i^* > 0\},$$

we have $u_i^* = 0$ for all $i \in I_\rho(u)$. Thus, $I_\rho(u) \subseteq I_0(u^*)$. In the same way, $I_\rho(v) \subseteq I_0(v^*)$ can be shown. As a consequence, we obtain $T_{IJ}(z^*) = 0$, i.e., $z^* \in \mathcal{Z}_{IJ}$. Hence, $\mathcal{Z}_{IJ}$ is not empty. As $\mathcal{Z}_{IJ}$ is also closed, there is some $z^\diamond \in \mathcal{Z}_{IJ}$ with $\|z^\diamond - z\| = \operatorname{dist}[z, \mathcal{Z}_{IJ}]$. Recall that $\mathcal{Z}_{IJ} \subseteq \mathcal{Z}$. Consequently, $z^\diamond$ belongs to $\mathcal{Z}$ as well.

Since $z^\diamond \in \mathcal{Z}_{IJ}$, property (b) holds due to the definition of $\mathcal{Z}_{IJ}$. Moreover, by Assumption 1, it follows that

$$\begin{aligned} \omega^2 \|z^\diamond - z\|^2 &\le \|T_{IJ}(z)\|^2 \\ &= \|H(z)\|^2 + \sum_{i \in I} u_i^2 + \sum_{i \in J} v_i^2 \qquad (4.1) \\ &\le \|H(z)\|^2 + 2p\rho^2. \end{aligned}$$

By Lemma 3.2 (b), we have

$$\|H(z)\| = \|H(z) - H(z^{\perp})\| \leq L\|z - z^{\perp}\| = L\,\mathrm{dist}[z, \mathcal{Z}] \leq L\rho.$$

This, (4.1), and the definition of C_P yield that $z^{\diamond}$ satisfies property (a). ☐

The construction of the reference solution $z^{\diamond}$ in the previous lemma will be important to prove that the optimal value of the LM subproblem (3.1) is sufficiently small if z is close enough to z^*.

Theorem 4.2. *Let Assumption 1 be satisfied. Then, there exist $C_V > 0$ and $\delta_V > 0$ so that, for any $z \in \mathcal{B}(z^*, \delta_V) \cap \Omega$, the unique solution $z^+(z)$ of subproblem* (3.1) *satisfies*

$$\psi(z^+(z), z) \leq C_V\,\mathrm{dist}[z, \mathcal{Z}]^{\min\{8/3,\,(4+\gamma)/2\}}. \tag{4.2}$$

Proof. First recall that the parameter $\gamma > 0$ occurs in the definition (3.3) of the regularization parameter $\lambda(z)$ within the LM subproblem. With

$$\eta := \max\{2/3,\, 1 - \gamma/4\} \in [2/3, 1)$$

and $\delta_P > 0$, $\rho_P > 0$, $C_P > 0$ from Lemma 4.1, we set

$$\delta_V := \min\left\{\rho_P^{1/\eta}, \delta_P, \frac{1}{2}, \frac{1}{(2C_P)^{1/\eta}}\right\} \quad \text{and} \quad \rho := \mathrm{dist}[z, \mathcal{Z}]^{\eta}.$$

This implies

$$\mathrm{dist}[z, \mathcal{Z}] \leq \rho = \mathrm{dist}[z, \mathcal{Z}]^{\eta} \leq \|z - z^*\|^{\eta} \leq \delta_V^{\eta} \leq \rho_P,$$

i.e., the assertions of Lemma 4.1 are true for $\rho = \mathrm{dist}[z, \mathcal{Z}]^{\eta}$ obtained for any $z \in \Omega \cap \mathcal{B}(z^*, \delta_V)$. Furthermore, we have

$$\rho \leq \delta_V^{\eta} \leq \frac{1}{2C_P}. \tag{4.3}$$

Let $z^{\diamond}$ be defined according to Lemma 4.1. Then, it follows from the definition of $\lambda(z)$ in (3.3), Lemma 4.1(a), and Lemma 3.2 (b) that

$$\lambda(z)\|z^{\diamond} - z\|^2 \leq \|T(z)\|^{\gamma} C_P^2 \rho^2 \leq L^{\gamma} C_P^2\,\mathrm{dist}[z, \mathcal{Z}]^{\gamma + 2\eta}. \tag{4.4}$$

Now, we consider

$$\|T(z) + G(z)(z^{\diamond} - z)\|^2 = \|H(z) + H'(z)(z^{\diamond} - z)\|^2 + \sum_{i=1}^{p} \tau_i^2, \tag{4.5}$$

where we defined $\tau_i := \alpha_i u_i^{\diamond} + \beta_i v_i^{\diamond}$ with $(\alpha_i, \beta_i) \in \partial\varphi(u_i, v_i)$ for $i \in \mathcal{N}$. Since Lemma 4.1(a) and (4.3) yield

$$\|z^{\diamond} - z^*\| \leq \|z^{\diamond} - z\| + \|z - z^*\| \leq C_P\rho + \delta_V \leq 1,$$

the first summand in the right-hand side of (4.5) can be estimated using Lemma 3.2(a) and Lemma 4.1(a) as

$$\|H(z) + H'(z)(z^{\diamond} - z)\| \leq L\|z^{\diamond} - z\|^2 \leq LC_P^2\rho^2 = LC_P^2\,\mathrm{dist}[z, \mathcal{Z}]^{2\eta}. \tag{4.6}$$

To estimate the second summand, we consider the value of τ_i for $i \in \mathcal{N}$. There are three possible cases:

1) If $u_i \le \rho$ and $v_i \le \rho$, we easily obtain $u_i^\diamond = v_i^\diamond = 0$ and hence $\tau_i = 0$ by part (b) of Lemma 4.1.
2) If $u_i > \rho$, it follows that $v_i \le \operatorname{dist}[z, \mathcal{Z}] \le \rho$, i.e., $v_i^\diamond = 0$. Moreover, by Lemma 4.1(a), we get

$$u_i^\diamond - u_i \le |u_i^\diamond - u_i| \le \|z^\diamond - z\| \le C_P \rho < C_P u_i,$$

i.e., $u_i^\diamond \le (1 + C_P) u_i$. Taking into account Lemma 3.3(c), we get

$$\tau_i = \alpha_i u_i^\diamond \le \frac{1}{2} \frac{v_i^2}{u_i^2} u_i^\diamond \le \frac{1 + C_P}{2\rho} \operatorname{dist}[z, \mathcal{Z}]^2 = \frac{1 + C_P}{2} \operatorname{dist}[z, \mathcal{Z}]^{2-\eta}.$$

3) If $v_i > \rho$, the same estimate on τ_i as in the previous case follows by Lemma 3.3 (d) using an analogous calculation.

Putting the results of all three cases together, it follows that

$$\sum_{i=1}^{p} \tau_i^2 \le \frac{p}{4} (1 + C_P)^2 \operatorname{dist}[z, \mathcal{Z}]^{4-2\eta}. \tag{4.7}$$

Then, taking into account (3.2), (4.5), (4.6), (4.7), and (4.4), we obtain

$$\begin{aligned} \psi(z^\diamond, z) \quad \le \quad & L^2 C_P^4 \operatorname{dist}[z, \mathcal{Z}]^{4\eta} + \tfrac{p}{4} (1 + C_P)^2 \operatorname{dist}[z, \mathcal{Z}]^{4-2\eta} + \\ & + L^\gamma C_P^2 \operatorname{dist}[z, \mathcal{Z}]^{\gamma+2\eta}. \end{aligned}$$

Thus, with $C_V := L^2 C_P^4 + p \frac{(1+C_P)^2}{4} + L^\gamma C_P^2$, we have

$$\psi(z^+(z), z) \le \psi(z^\diamond, z) \le C_V \operatorname{dist}[z, \mathcal{Z}]^{\min\{4\eta, 4-2\eta, \gamma+2\eta\}}.$$

To analyze the exponent, we consider two cases. The first one is $\gamma \in (0, \frac{4}{3})$, which implies $\eta = 1 - \frac{\gamma}{4}$ and

$$\min\{4\eta, 4 - 2\eta, \gamma + 2\eta\} = \min\left\{4 - \gamma, 2 + \frac{\gamma}{2}, 2 + \frac{\gamma}{2}\right\} = 2 + \frac{\gamma}{2}.$$

In the second case, $\gamma \ge \frac{4}{3}$, we obtain $\eta = \frac{2}{3}$ and

$$\min\{4\eta, 4 - 2\eta, \gamma + 2\eta\} = \min\left\{\frac{8}{3}, \frac{8}{3}, \gamma + \frac{4}{3}\right\} = \frac{8}{3}.$$

The results for the two cases of γ show that (4.2) is valid. □

Theorem 4.3. *Let Assumption 1 be satisfied and* $\gamma \in (0, \frac{8}{3})$*. Then, there exist* $C_S \ge 1$ *and* $\delta_S > 0$ *so that, for any* $z \in \mathcal{B}(z^*, \delta_S) \cap \Omega$*, the unique solution* $z^+(z)$ *of subproblem* (3.1) *satisfies*

$$\begin{array}{lrcl} \text{(a)} & \|z^+(z) - z\| & \le & C_S \|T(z)\|^{\min\{(8-3\gamma)/6,\, 1-\gamma/4\}} \\ \text{(b)} & \|T(z^+(z))\| & \le & C_S \|T(z)\|^{\min\{(4+\gamma)/4,\, -\gamma+8/3\}} \end{array}$$

Proof. With $C_V > 0$ and $\delta_V > 0$ from Theorem 4.2 and $\omega_E > 0$ from Proposition 3.1 (implied by Assumption 1), let us define

$$\mu := \min\left\{\frac{4}{3} - \frac{\gamma}{2}, 1 - \frac{\gamma}{4}\right\}, \quad \delta_S := \min\left\{\delta_V, \left(\sqrt{C_V \omega_E^{-\gamma}} + 1\right)^{-1/\mu}\right\},$$

and

$$C_S := \max\left\{1, \sqrt{C_V\omega_E^{-(\gamma+2\mu}}, LC_V\omega_E^{-(\gamma+2\mu)} + \omega_E^{-(\mu+\gamma/2)}(p+1)\sqrt{C_V}\right\}.$$

According to subproblem (3.1), $z \in \mathcal{Z}$ implies $z^+(z) = z \in \mathcal{Z}$, i.e., assertions (a) and (b) are satisfied. Therefore, we are going to check these assertions for an arbitrarily chosen $z \in (\mathcal{B}(z^*, \delta_S) \cap \Omega) \setminus \mathcal{Z}$.

From Theorem 4.2 and the definition of ψ in (3.2), we have

$$\begin{aligned} \psi(z^+(z), z) &= \|T(z) + G(z)(z^+(z) - z)\|^2 + \lambda(z)\|z^+(z) - z\|^2 \\ &\le C_V \operatorname{dist}[z, \mathcal{Z}]^{\min\{8/3,\, 2+\gamma/2\}}. \end{aligned} \tag{4.8}$$

With the regularization parameter $\lambda(z) = \|T(z)\|^\gamma$ for $z \in \Omega \setminus \mathcal{Z}$ according to (3.3) and Proposition 3.1, it follows that

$$\begin{aligned} \|z^+(z) - z\|^2 &\le \lambda(z)^{-1} C_V \operatorname{dist}[z, \mathcal{Z}]^{\min\{8/3,\, 2+\gamma/2\}} \\ &\le \|T(z)\|^{-\gamma} C_V \operatorname{dist}[z, \mathcal{Z}]^{\min\{8/3,\, 2+\gamma/2\}} \\ &\le C_V\omega_E^{-\gamma} \operatorname{dist}[z, \mathcal{Z}]^{\min\{8/3,\, 2+\gamma/2\}-\gamma} \\ &= C_V\omega_E^{-\gamma} \operatorname{dist}[z, \mathcal{Z}]^{2\mu}. \end{aligned} \tag{4.9}$$

Hence, taking into account the definition of C_S, assertion (a) must hold.

To prove assertion (b), we first note that, by the triangle inequality,

$$\begin{aligned} \|H(z^+(z))\| \le\ & \|H(z) + H'(z)(z^+(z) - z)\| + \\ & + \|H(z^+(z)) - (H(z) + H'(z)(z^+(z) - z))\| \end{aligned} \tag{4.10}$$

is valid. Using (4.8), we see that the first summand on the right is bounded, i.e.,

$$\begin{aligned} \|H(z) + H'(z)(z^+(z) - z)\| &\le \|T(z) + G(z)(z^+(z) - z)\| \\ &\le \sqrt{C_V} \operatorname{dist}[z, \mathcal{Z}]^{\min\{4/3,\, 1+\gamma/4\}}. \end{aligned} \tag{4.11}$$

By (4.9) and $\operatorname{dist}[z, \mathcal{Z}] \le \|z - z^*\| \le \delta_S \le \delta_S^\mu$, we get

$$\|z^+(z) - z^*\| \le \|z^+(z) - z\| + \|z - z^*\| \le \left(\sqrt{C_V\omega_E^{-\gamma}} + 1\right)\delta_S^\mu \le 1.$$

That is $z^+(z) \in \mathcal{B}(z^*, 1)$. Thus, we can apply Lemma 3.2(a) to estimate the second summand on the right-hand side of (4.10) and obtain

$$\|H(z^+(z)) - (H(z) + H'(z)(z^+(z) - z))\| \le L\|z^+(z) - z\|^2$$

This and (4.9) yield

$$\|H(z^+(z)) - (H(z) + H'(z)(z^+(z) - z))\| \le LC_V\omega_E^{-\gamma} \operatorname{dist}[z, \mathcal{Z}]^{2\mu}. \tag{4.12}$$

Moreover, due to Lemma 3.3 (h) and (4.8), one has

$$\begin{aligned} & 0 \le \varphi(u_i^+(z), v_i^+(z)) \le (T(z) + G(z)(z^+(z) - z))_{m+2p+i} \\ & \quad \le \sqrt{C_V} \operatorname{dist}[z, \mathcal{Z}]^{\min\{4/3,\, 1+\gamma/4\}} \end{aligned} \tag{4.13}$$

for all $i \in \mathcal{N}$. Summarizing (4.10), (4.11), (4.12), and (4.13) provides

$$
\begin{aligned}
\|T(z^+(z))\| &\leq \|H(z^+(z))\| + \textstyle\sum_{i=1}^p \varphi(u_i^+(z), v_i^+(z)) \\
&\leq (1+p)\sqrt{C_V}\,\mathrm{dist}[z,\mathcal{Z}]^{\min\{4/3,\,1+\gamma/4\}} + \\
&\quad + LC_V\omega_E^{-\gamma}\,\mathrm{dist}[z,\mathcal{Z}]^{2\mu} \\
&\leq C_S\|T(z)\|^{\min\{1+\gamma/4,\,8/3-\gamma\}},
\end{aligned}
\tag{4.14}
$$

where it has been taken into account that

$$
\min\left\{\frac{4}{3}, 1+\frac{\gamma}{4}, 2\mu\right\} = \min\left\{\frac{4}{3}, 1+\frac{\gamma}{4}, \frac{8}{3}-\gamma, 2-\frac{\gamma}{2}\right\} = \min\left\{1+\frac{\gamma}{4}, \frac{8}{3}-\gamma\right\}
$$

holds for all $\gamma \in (0, \frac{8}{3})$. Hence, assertion (b) is valid. □

5. Local convergence of the Levenberg–Marquardt method

In this section, we provide results on the local convergence of the constrained LM method with subproblems defined in (3.1). This is done by means of the general convergence framework, see Theorem 2.1. Therefore, we first specify the quantities used in this theorem for the case of the LM method we are interested in. Recall that the mapping T is given by (1.4) and that

$$
N = n + 2p, \quad \Omega = \mathbb{R}^n \times \mathbb{R}_+^p \times \mathbb{R}_+^p, \quad \text{and} \quad \mathcal{Z} = \{z \in \Omega \mid T(z) = 0\}
$$

have been already defined in the beginning of Section 3. The sequence $\{z^k\} \subset \Omega$ generated by the constrained LM method is obtained according to

$$
z^{k+1} := z^+(z^k), \quad k = 0, 1, 2, \ldots, \tag{5.1}
$$

where $z^0 \in \Omega$ is a starting point sufficiently close to some $z^* \in \mathcal{Z}$ and the mapping $z^+ : \Omega \to \Omega$, as described at the end of Subsection 3.1, is based on the solution of subproblem (3.1). For any $z \in \Omega\backslash\mathcal{Z}$, the subproblem makes use of the regularization parameter $\lambda(z) = \|T(z)\|^\gamma$ with some $\gamma > 0$, cf. (3.3). Finally, we define the function $t : \mathbb{R}^N \to [0, \infty)$ by

$$
t(z) := \|T(z)\|^{2/3}. \tag{5.2}
$$

Thus, we see that the solution set $\mathcal{Z}$ of problem (1.4) is the same as the set $Z = \{z \in \Omega \mid t(z) = 0\}$ used in Theorem 2.1.

We are now in the position to apply the latter theorem together with the results of Theorem 4.3 to the sequence $\{z^k\}$ as defined in (5.1).

Theorem 5.1. *Let Assumption 1 be satisfied and $\gamma = 4/3$. Then, there exists $\varepsilon > 0$ such that, for any $z^0 \in \mathcal{B}(z^*, \varepsilon) \cap \Omega$, the sequence $\{z^k\}$ generated by the constrained LM method according to (5.1) converges to some $\hat{z} \in \mathcal{Z}$ with*

(a) *an R-order of 4/3 and*
(b) *a ν-step Q-order of $(2/3)(4/3)^\nu$ for any fixed $\nu \in \mathbb{N}$ with $\nu \geq 2$.*

In addition, the sequences $\{\|T(z^k)\|\}$ and $\{\mathrm{dist}[z^k, \mathcal{Z}]\}$ converge to 0 with a Q-order of 4/3.

Proof. By Theorem 4.3 and (5.2), there exist $C_S \geq 1$ and $\delta_S > 0$ so that

$$\|z^+(z) - z\| \leq C_S \|T(z)\|^{2/3} = C_S t(z),$$
$$t(z^+(z)) = \|T(z^+(z))\|^{2/3} \leq \left(C_S \|T(z)\|^{4/3}\right)^{2/3} \leq C_S t(z)^{4/3}$$

for all $z \in \mathcal{B}(z^*, \delta_S) \cap \Omega$. Hence, conditions (a) and (b) in Theorem 2.1 are satisfied with $C := C_S$, $\delta := \delta_S$, and $\sigma = 4/3$. Finally, the function t as defined in (5.2) is Hölder continuous on $\mathcal{B}(z^*, \delta)$ with some $L > 0$ and $\theta = \frac{2}{3}$. To see this, we assume without loss of generality that $\|T(\tilde{z})\| \geq \|T(z)\| > 0$ and keep in mind that $z, \tilde{z} \in \mathcal{B}(z^*, \delta)$. This implies

$$\tfrac{1}{\|T(\tilde{z})\|^\theta} \left| \|T(\tilde{z})\|^\theta - \|T(z)\|^\theta \right| = 1 - \tfrac{\|T(z)\|^\theta}{\|T(\tilde{z})\|^\theta} \leq 1 - \tfrac{\|T(z)\|}{\|T(\tilde{z})\|} \leq \left| 1 - \tfrac{\|T(z)\|}{\|T(\tilde{z})\|} \right|^\theta.$$

It follows that

$$|t(\tilde{z}) - t(z)| = \left| \|T(\tilde{z})\|^\theta - \|T(z)\|^\theta \right| \leq \left| \|T(\tilde{z})\| - \|T(z)\| \right|^\theta \leq L_0 \|\tilde{z} - z\|^\theta,$$

since the mapping T is locally Lipschitz continuous.

Hence, applying Theorem 2.1 yields the existence of $\varepsilon > 0$ such that the sequence $\{z^k\}$ generated according to (5.1) remains in $\mathcal{B}(z^*, \delta_S) \cap \Omega$, converges to some $\hat{z} \in \mathcal{Z}$, and has the claimed convergence properties. ☐

6. Final remarks

We have seen that the new convergence framework in Section 2 together with the Index Error Bound Condition, a sophisticated analysis of a single step of the constrained LM method, and accordingly the choice of the regularization parameter provides local superlinear convergence of the method even if it is started close to a nonisolated *and* degenerate solution. This result is new since the method is not based on a piecewise smooth reformulation of the complementarity system (1.2) but on a reformulation with the FB C-function. The obtained superlinear R-order of 4/3 indicates that the convergence framework is indeed more general than previous approaches. In particular, for the constrained LM method investigated here, the framework allows steps $z^+(z) - z$ whose length is bounded by $C_0 \operatorname{dist}[z, \mathcal{Z}]^{2/3}$, with some $C_0 > 0$. In contrast to this, the step length of usual Newton-type methods is bounded by $C_0 \operatorname{dist}[z, \mathcal{Z}]$. It can be easily verified that this framework is able to recover many known results, in particular those for complementarity systems with solutions that are both noninsolated and degenerate, especially the Q-quadratic convergence of the constrained LM method [9] and of the LP-Newton method [10, 21], both based on reformulations with the min C-function. Also, an R-order of 3/2 in the recent approach [2] can be recovered. Note that this approach was based on the FB C-function but restricted to an isolated primal solution.

We hope that the convergence framework in Section 2 will be helpful to analyze local convergence of the LP-Newton method based on the FB C-function or on certain other C-functions, which do not lead to piecewise smooth systems of equations.

Acknowledgements

This research was supported by the DFG (Deutsche Forschungsgemeinschaft) [grant number 290762516] and partly by the Volkswagen Foundation [grant number 97 775].

References

[1] R. Behling and A. Fischer, *A unified local convergence analysis of inexact constrained Levenberg–Marquardt methods*, Optimization Letters **6** (2012), 927–940.

[2] R. Behling, A. Fischer, K. Schönefeld and N. Strasdat, *A special complementarity function revisited*, Optimization **68** (2019), 65–79.

[3] O. Burdakov, C. Kanzow and A. Schwartz, *Mathematical programs with cardinality constraints: Reformulation by complementarity-type constraints and a regularization method*, SIAM Journal on Optimization **26** (2016), 397–425.

[4] D. Dorsch, H. Th. Jongen and V. Shikhman, *On Structure and Computation of Generalized Nash Equilibria*, SIAM Journal on Optimization **23** (2013), 452–474.

[5] A. Dreves, *Improved error bound and a hybrid method for generalized Nash equilibrium problems*, Computational Optimization and Applications **65** (2016), 431–448.

[6] A. Dreves, F. Facchinei, A. Fischer and M. Herrich, *A new error bound result for Generalized Nash Equilibrium Problems and its algorithmic application*, Computational Optimization and Applications **59** (2014), 63–84.

[7] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems*, Springer, New York, 2003.

[8] F. Facchinei and J. Soares, *A new merit function for nonlinear complementarity problems and a related algorithm*, SIAM Journal on Optimization **7** (1997), 225–247.

[9] F. Facchinei, A. Fischer and M. Herrich, *A family of Newton methods for nonsmooth constrained systems with nonisolated solutions*, Mathematical Methods of Operations Research **77** (2013), 433–443

[10] F. Facchinei, A. Fischer and M. Herrich, *An LP-Newton method: nonsmooth equations, KKT systems, and nonisolated solutions*, Mathematical Programming **146** (2014), 1–36.

[11] F. Facchinei, C. Kanzow, S. Karl and S. Sagratella, *The semismooth Newton method for the solution of quasi-variational inequalities*, Computational Optimization and Applications **62** (2015), 85–109.

[12] J. Y. Fan and Y. X. Yuan, *On the quadratic convergence of the Levenberg–Marquardt method without nonsingularity assumption*, Computing **74** (2005), 23–39.

[13] D. Fernández and M. Solodov, *Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems*, Mathematical Programming **125** (2010), 47–73.

[14] A. Fischer, *A special Newton-type optimization method*, Optimization **24** (1992), 269–284.

[15] A. Fischer, *An NCP-function and its use for the solution of complementarity problems*, in: Recent Advances in Nonsmooth Optimization, D.-Z. Du et al. (eds.), World Scientific, Singapore, 1995, pp. 88–105.

[16] A. Fischer, *Modified Wilson's method for nonlinear programs with nonunique multipliers*, Mathematics of Operations Research **24** (1999), 699–727.

[17] A. Fischer, *Local behavior of an iterative framework for generalized equations with nonisolated solutions*, Mathematical Programming **94** (2002), 91–124.

[18] A. Fischer and M. Herrich, *Newton-type methods for Fritz John systems of generalized Nash equilibrium problems*, Pure and Applied Functional Analysis **3** (2018), 587–602.

[19] A. Fischer, P. K. Shukla and M. Wang, *On the inexactness level of robust Levenberg–Marquardt methods*, Optimization **59** (2010), 273–287.

[20] A. Fischer, M. Herrich and K. Schönefeld, *Generalized Nash equilibrium problems – recent advances and challenges*, Pesquisa Operacional **34** (2014), 521–558.

[21] A. Fischer, M. Herrich, A. F. Izmailov and M. V. Solodov, *Convergence conditions for Newton-type methods applied to complementarity systems with nonisolated solutions*, Computational Optimization and Applications **63** (2016), 425–459.

[22] A. Fischer, M. Herrich, A. F. Izmailov and M. V. Solodov, *A globally convergent LP-Newton method*, SIAM Journal on Optimization **26** (2016), 2012–2033.

[23] A. Fischer, M. Herrich, A. F. Izmailov, W. Scheck and M. V. Solodov, *A globally convergent LP-Newton method for piecewise smooth constrained equations: escaping nonstationary accumulation points*, Computational Optimization and Applications **69** (2018), 325–349.

[24] W. W. Hager, *Stabilized Sequential Quadratic Programming*, Computational Optimization and Applications **12** (1999), 253–273.

[25] F. Harder, P. Mehlitz and G. Wachsmuth, *Reformulation of the M-stationarity conditions as a system of discontinuous equations and its solution by a semismooth Newton method*, SIAM Journal on Optimization **31** (2021), 1459–1488.

[26] A. F. Izmailov and M. V. Solodov, *Stabilized SQP revisited*, Mathematical Programming **133** (2012), 93–120.

[27] C. Kanzow and A. Schwartz, *Spieltheorie*, Birkhäuser, Cham, 2018.

[28] C. Kanzow, N. Yamashita and M. Fukushima, *Levenberg–Marquardt methods with strong local convergence properties for solving nonlinear equations with convex constraints*, Journal of Computational and Applied Mathematics **172** (2004), 375–397.

[29] J. V. Outrata, M. Kocvara and J. Zowe, *Nonsmooth Approach to Optimization Problems with Equilibrium Constraints: Theory, Applications and Numerical Results*, Kluwer Academic Publishers, Dordrecht, 1998.

[30] N. Yamashita and M. Fukushima, *On the rate of convergence of the Levenberg–Marquardt method*, in: Topics in Numerical Analysis. Computing Supplementa, G. Alefeld and X. Chen (eds.), vol 15. Springer, Vienna, 2001, pp. 239–249.

[31] S. J. Wright, *Superlinear convergence of a stabilized SQP method to a degenerate solution*, Computational Optimization and Applications **11** (1998), 253–275.

[32] S. J. Wright, *An algorithm for degenerate nonlinear programming with rapid local convergence*, SIAM Journal on Optimization **15** (2005), 673–696.

[33] J.-L. Zhang, *On the convergence properties of the Levenberg–Marquardt method*, Optimization **52** (2003), 739–756.

Manuscript received January 19 2022
revised August 7 2022

A. Fischer
Faculty of Mathematics, Technische Universität Dresden, 01062 Dresden, Germany
E-mail address: andreas.fischer@tu-dresden.de

N. Strasdat
Faculty of Mathematics, Technische Universität Dresden, 01062 Dresden, Germany
E-mail address: nico.strasdat@tu-dresden.de