Technische Universität Dresden
Institute of Numerical Mathematics

A Special Complementarity Function Revisited

Roger Behling
Andreas Fischer
Klaus Schönefeld
Nico Strasdat

Technical Report
MATH-NM-06-2017
December 2017

A Special Complementarity Function Revisited

Roger Behling

Department of Mathematics
Federal University of Santa Catarina, Blumenau, Brazil
roger.behling@ufsc.br

Andreas Fischer Klaus Schönefeld Nico Strasdat

Institute of Numerical Mathematics, Faculty of Mathematics
Technische Universität Dresden, 01062 Dresden, Germany
{Andreas.Fischer,Klaus.Schoenefeld,Nico.Strasdat}@tu-dresden.de

December 10, 2017

Abstract. Recently, a local framework of Newton-type methods for constrained systems of equations has been developed which, applied to the solution of Karush-Kuhn-Tucker (KKT) systems, enables local quadratic convergence under conditions that allow nonisolated and degenerate KKT points. This result is based on a reformulation of the KKT conditions as a constrained piecewise smooth system of equations. It is an open question whether a comparable result can be achieved for other (not piecewise smooth) reformulations. It will be shown that this is possible if the KKT system is reformulated by means of the Fischer-Burmeister complementarity function under conditions that allow degenerate KKT points and nonisolated Lagrange multipliers. To obtain this result, novel constrained Levenberg-Marquardt subproblems are introduced which allow significantly longer steps for updating the multipliers. Based on this, a convergence rate of at least 1.5 is shown.

Keywords. Karush-Kuhn-Tucker system; nonunique multipliers; degenerate solution; constrained Levenberg-Marquardt method

1. Introduction

Let us consider the Karush-Kuhn-Tucker (KKT) system

$$\begin{aligned} F(x) + g'(x)^\top u &= 0, \\ g(x) + s &= 0, \\ u \geq 0,\, s \geq 0,\, s^\top u &= 0, \end{aligned} \tag{1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ and $g : \mathbb{R}^n \to \mathbb{R}^m$ are given functions. Throughout, as blanket assumption, let F be differentiable and g be twice differentiable with F' and g'' being locally Lipschitz continuous. By

$$\mathcal{Z} := \{z := (x, s, u) \mid z \text{ solves } (1)\},$$

the set of all KKT points is denoted. A point $z^* = (x^*, s^*, u^*) \in \mathcal{Z}$ is called degenerate if it violates strict complementarity, i.e., if $s_i^* = u_i^* = 0$ holds for at least one index $i \in I := \{1, \dots, m\}$.

As it is well-known, for $F := \nabla f$ with a sufficiently smooth function $f : \mathbb{R}^n \to \mathbb{R}$, the KKT conditions associated to the optimization problem

$$f(x) \to \min_x \quad \text{s.t.} \quad g(x) \leq 0 \tag{2}$$

can be written as system (1). Regardless of this, other applications like KKT conditions associated to variational inequalities lead to such systems [8].

A bunch of methods for computing a solution of the KKT system (1) is based on reformulating (1) as a nonsmooth system of equations by means of a complementarity function (C-function for short). In particular, the C-function [10,11] $\varphi : \mathbb{R}^2 \to \mathbb{R}$ with

$$\varphi(a,b) := a + b - \sqrt{a^2 + b^2}$$

is frequently used and often called Fischer-Burmeister function. Because

$$\varphi(a,b) = 0 \quad \Leftrightarrow \quad a \geq 0,\, b \geq 0,\, ab = 0$$

is satisfied for all $a, b \in \mathbb{R}$ (like for any other C-function), system (1) can be equivalently rewritten as

$$T_\varphi(z) := \begin{pmatrix} \mathcal{L}(x,u) \\ g(x) + s \\ \Phi_\varphi(s,u) \end{pmatrix} = 0 \tag{3}$$

with $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ and $\Phi_\varphi : \mathbb{R}^m \times \mathbb{R}^m \to \mathbb{R}^m$ given by

$$\mathcal{L}(x,u) := F(x) + g'(x)^\top u \quad \text{and} \quad \Phi_\varphi(s,u) := \begin{pmatrix} \varphi(s_1,u_1) \\ \vdots \\ \varphi(s_m,u_m) \end{pmatrix}.$$

For simplicity, we will regard z either as triple $(x,s,u) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$ or as a column vector in $\mathbb{R}^p$ with $p := n + 2m$. This also applies to other vectors in $\mathbb{R}^p$.

By now, conditions which lead to superlinear convergence of Newton-type methods for solving system (3) imply that the method starts in a sufficiently small neighbourhood of an *either isolated or nondegenerate* solution of the KKT system (1), for example see [7,8,10,20–22]. On the other hand, the use of the C-function $\phi_{\min} : \mathbb{R}^2 \to \mathbb{R}$ with

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

and of $T_{\phi_{\min}}$ instead of T_φ, allows (to some extent) local quadratic convergence of Newton-type methods to nonisolated and degenerate solutions of the KKT system (1). Details on these methods for constrained systems of equations and their convergence analysis can be found in [5,6]. Moreover, it is shown in [4,12,18] that this framework is well suited for (constrained) piecewise smooth systems of equations. However, φ and T_φ are *not* piecewise smooth, see [23].

Thus, the question arises whether superlinear convergence can be obtained if T_φ is used instead of a piecewise smooth mapping like $T_{\phi_{\min}}$ under conditions that allow nonisolated and degenerate solutions. Regardless of other possible benefits, a positive answer would enable to exploit the continuous differentiability of the merit function $\|T_\varphi\|^2$ for obtaining global convergence results (see [3], [19, Section 2.4], and [15, Section 3], for example) even if some globalization of the LP-Newton method based on the merit function $\|T_{\phi_{\min}}\|$ has recently been suggested [14], where the occurence of nonstationary accumulation points can be avoided by means of some escape procedure [13].

A discussion in [24, Section 5.5] suggests that the question we just raised cannot be solved within the framework [5], since an assumption therein might be too restrictive. A general answer to this question seems difficult. Therefore, we assume here that, for some solution $z^* = (x^*, s^*, u^*)$ of system (1), the primal part x^* of the solution is isolated, whereas nonisolatedness of the dual solution u^* should be possible so that the polyhedral set

$$\mathcal{U} := \{u \in \mathbb{R}^m \mid (x^*, s^*, u) \in \mathcal{Z}\}$$

may contain more elements than u^* only. Under these circumstances, the weakest assumption for local superlinear convergence of a Newton-type method for solving the KKT system (1) known by now is given in [6, Theorem 5 (c)] for the LP-Newton method and, based on this, in [5] for a constrained Levenberg-Marquardt method, for more details see Subsection 3.1. The subproblems of these methods are linear programs or strongly convex quadratic programs with box constraints, respectively. They arise from the reformulation of the KKT system (1) as a constrained nonsmooth system of equations, namely

$$T_{\phi_{\min}}(z) = 0 \quad \text{s.t.} \quad z \in \Omega,$$

where $\Omega := \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^m_+$.

As explained above, our interest is in using T_φ which is not piecewise smooth. To cope with this, we suggest a Levenberg-Marquardt type method whose subproblems differ from those used in previous works like [1,19,27]. In particular, the subproblems allow to significantly weaken an assumption of the framework in [5] so that longer steps with respect to the u-variables become possible. Based on this and on some new proof techniques, we will show local superlinear convergence with an order of 1.5 under assumptions like in [6, Theorem 5 (c)] .

The paper is organized as follows. In Section 2, the new subproblems are described in detail. Then, in Section 3, we detail the assumptions and discuss their relation to previously used assumptions for related Newton-type methods that use stabilization techniques. In the same section, we present several preliminary results needed for the local convergence analysis that follows in Section 4. Our notation will be fairly standard. Throughout the paper, $\|\cdot\|$ denotes the Euclidean norm in $\mathbb{R}^l$ and $B(w, \delta)$ the closed Euclidean ball around $w \in \mathbb{R}^l$ with radius $\delta > 0$. For some nonempty set $W \subseteq \mathbb{R}^l$, the distance of $\zeta \in \mathbb{R}^l$ to W is defined by $\operatorname{dist}[\zeta, W] := \inf_{w \in W} \|\zeta - w\|$. Moreover, $\operatorname{conv}(W)$ denotes the convex hull of the set W.

2. The novel Levenberg-Marquardt subproblem

Obviously, T_φ is not differentiable at exactly those points $z = (x, s, u)$ with $s_i = u_i = 0$ for at least one index $i \in I$. Then, a usual substitute for T'_φ is given by some mapping

$G : \mathbb{R}^p \to \mathbb{R}^{p\times p}$ with

$$G(z) \in \partial T_\varphi(z) = \operatorname{conv}\left\{ \lim_{\ell\to\infty} T'_\varphi(z^\ell) \,\middle|\, \lim_{\ell\to\infty} z^\ell = z,\ (s_i^\ell, u_i^\ell) \neq (0,0) \text{ for } (i,\ell) \in I \times \mathbb{N} \right\}.$$

Here, $\partial T_\varphi(z)$ denotes Clarke's generalized Jacobian [2, Section 2.6] of T_φ at z. Since T_φ is locally Lipschitz continuous the set $\partial T_\varphi(z)$ is nonempty, closed, convex, and bounded. The definition in [2] applied to the map T_φ shows that

$$G(z) = \begin{pmatrix} F'(x) + \sum_{i=1}^m u_i g_i''(x) & 0_{n\times m} & g'(x)^\top \\ g'(x) & I_{m\times m} & 0_{m\times m} \\ 0_{m\times n} & \operatorname{diag}(\alpha_i) & \operatorname{diag}(\beta_i) \end{pmatrix}$$

with $(\alpha_i, \beta_i) \in \partial\varphi(s_i, u_i)$ for $i \in I$, where

$$\partial\varphi(a,b) = \operatorname{conv}\left\{ \lim_{\ell\to\infty} \varphi'(a_\ell, b_\ell) \,\middle|\, \lim_{\ell\to\infty} \begin{pmatrix} a_\ell \\ b_\ell \end{pmatrix} = \begin{pmatrix} a \\ b \end{pmatrix},\ \begin{pmatrix} a_\ell \\ b_\ell \end{pmatrix} \neq \begin{pmatrix} 0 \\ 0 \end{pmatrix} \text{ for } \ell \in \mathbb{N} \right\} \tag{4}$$

holds for any $(a,b) \in \mathbb{R}\times\mathbb{R}$ according to [2, Theorem 2.5.1]. If all matrices in $\partial T_\varphi(z^*)$ were nonsingular (CD-regularity) at $z^* \in \mathcal{Z}$, the well-known semismooth Newton method applied to $T_\varphi(z) = 0$ would converge locally quadratically to the locally isolated solution z^*. For related regularity assumptions and a convergence analysis of corresponding Newton type methods, in particular for quite general nonsmooth mappings, see [20,22]. However, those assumptions like CD-regularity (or the weaker BD-regularity used in [21]) would imply that z^* is a locally isolated solution. Therefore, we cannot use them here.

In contrast to existing Levenberg-Marquardt methods, we employ not only one but two regularization parameters with a different growth behaviour. More in detail, let $\lambda : \mathbb{R}^p \to (0,\infty)$ and $\mu : \mathbb{R}^p \to (0,\infty)$ be defined by

$$\lambda(z) := \begin{cases} \|T_\varphi(z)\| & \text{if } z \notin \mathcal{Z}, \\ 1 & z \in \mathcal{Z} \end{cases} \quad \text{and} \quad \mu(z) := \begin{cases} \|T_\varphi(z)\|^2 & \text{if } z \notin \mathcal{Z}, \\ 1 & z \in \mathcal{Z}. \end{cases} \tag{5}$$

The obvious difference between $\lambda(z)$ and $\mu(z)$ in the exponent of $\|T_\varphi(z)\|$ is an important ingredient in our analysis below. For any $z \in \mathbb{R}^n$, let $\Psi : \mathbb{R}^p \to [0,\infty)$ be defined by

$$\Psi(\Delta) := \|T_\varphi(z) + G(z)\Delta\|^2 + \lambda(z)\left(\|\Delta_x\|^2 + \|\Delta_s\|^2\right) + \mu(z)\|\Delta_u\|^2,$$

where $\Delta = (\Delta_x, \Delta_s, \Delta_u) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$. For any $z = (x,s,u) \in \Omega = \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^m_+$, let us now consider the subproblem

$$\Psi(\Delta) \to \min_\Delta \quad \text{s.t.} \quad z + \Delta \in \Omega. \tag{6}$$

This is a quadratic minimization problem with box constraints. Thanks to (5), it has a strongly convex objective. Thus, for any $z \in \Omega$, subproblem (6) has a unique solution which we denote by

$$\Delta(z) = (\Delta_x(z), \Delta_s(z), \Delta_u(z)).$$

Given any $z \in \Omega$, the point

$$z^+ := z + \Delta(z) \tag{7}$$

is well-defined and again belongs to Ω. This describes one step of our Newton-type method.

3. Preliminaries

In this section, we first detail the assumptions used later on. This includes a short discussion of assumptions for related existing Newton-type methods. The second subsection provides several basic assertions.

3.1. Assumptions

First recall the blanket assumptions on the smoothness of F and g stated in Section 1. Moreover, as before, let $z^* \in \mathcal{Z}$ be some fixed solution of the KKT system (1). Assumption 1 below guarantees that the x- and the s-part of $z^* = (x^*, s^*, u^*)$ are locally unique, i.e., only the multiplier part u^* is allowed to be nonisolated.

Assumption 1. *There exists* $\delta_* > 0$ *so that*

$$z = (x, s, u) \in \mathcal{Z} \cap B(z^*, \delta_*) \quad \textit{implies} \quad x = x^*.$$

The next assumption requires that a Lipschitzian error bound holds in a neighbourhood of z^*.

Assumption 2. *There exist* $\omega > 0$ *and* $\delta > 0$ *so that*

$$\omega \operatorname{dist}[z, \mathcal{Z}] \leq \|T_\varphi(z)\| \qquad \textit{for all } z \in \Omega \cap B(z^*, \delta).$$

If T_φ in Assumption 2 is replaced by $T_{\min}$ we will obtain an equivalent assumption due to relations between φ and $\phi_{\min}$ stated in [25, Lemma 3.1].

It is well-known that the stabilized SQP method suggested in [26] for nonlinear programs (2) is able to guarantee local superlinear convergence to z^* under conditions that imply the local uniqueness of x^* as requested in Assumption 1. Therefore, let us compare Assumption 2 with those conditions. In [17], a strong second-order sufficient condition is needed, whereas [26] makes use of even stronger assumptions. The weakest condition is provided in [9] and requires that the classic second-order sufficient condition (SOSC) holds at the KKT point of interest. In the same paper, the authors extend this condition to a second-order condition (SOC) at a solution of the more general KKT system (1) not necessarily arising from an optimization problem. Moreover, they extend the stabilized SQP method to a method that sequentially solves stabilized affine variational inequalities. For this method local superlinear convergence is proved if SOC is valid. Finally, it can be seen in [9, Section 4] that even SOC is stronger than requiring Assumption 2. Hence, Assumptions 1 and 2 together are weaker than those used for superlinear convergence of the above stabilized methods.

3.2. Basic assertions

The subsequent lemmas provide several assertions that will be used in the analysis later on. Among these assertions, part (a) of Lemma 3.1, parts (b) – (d) of Lemma 3.2, and Lemma 3.3 will be of particular importance.

Lemma 3.1. *There exists $L > 0$ so that the inequalities*

(a) $\left\| \mathcal{L}(y,v) - \mathcal{L}(x,u) - \mathcal{L}'(x,u) \begin{pmatrix} y-x \\ v-u \end{pmatrix} \right\| \leq L \left(\|y-x\|\|v-u\| + \|y-x\|^2 \right),$

(b) $\|g(y) - g(x) - g'(x)(y-x)\| \leq L\|y-x\|^2$, *and*

(c) $\|T_\varphi(z)\| \leq L \operatorname{dist}[z, \mathcal{Z}]$

are satisfied for all $z = (x,s,u) \in B(z^*,1)$ and all $(y,t,v) \in B(z^*,1)$.

The proof of part (a) exploits not only smoothness properties of F and g but also the structure of $\mathcal{L}$. The proof of parts (b) and (c) are omitted due to their simplicity.

Proof. (a) Let $(x,s,u) \in B(z^*,1)$ and $(y,t,v) \in B(z^*,1)$ be arbitrarily chosen. Then, because of the continuity of $\mathcal{L}'$, Taylor's formula with integral remainder yields

$$\begin{aligned} &\left\| \mathcal{L}(y,v) - \mathcal{L}(x,u) - \mathcal{L}'(x,u) \begin{pmatrix} y-x \\ v-u \end{pmatrix} \right\| \\ &\leq \int_0^1 \left\| (\mathcal{L}'(x+\tau(y-x), u+\tau(v-u)) - \mathcal{L}'(x,u)) \begin{pmatrix} y-x \\ v-u \end{pmatrix} \right\| d\tau. \end{aligned}$$

The norm term within the integral on the right-hand side can be bounded above by means of the triangle inequality and by estimating the resulting parts, i.e., we obtain

$$\begin{aligned} \|(F'(x+\tau(y-x)) - F'(x))(y-x)\| &\leq L_0\|y-x\|^2, \\ \|\sum_{i=1}^m u_i(g_i''(x+\tau(y-x)) - g_i''(x))(y-x)\| &\leq L_0\|y-x\|^2, \\ \|\sum_{i=1}^m \tau(v-u)_i g_i''(x+\tau(y-x))(y-x)\| &\leq L_0\|y-x\|\|v-u\|, \\ \left\|(g'(x+\tau(y-x)) - g'(x))^\top (v-u)\right\| &\leq L_0\|y-x\|\|v-u\| \end{aligned}$$

for any $\tau \in [0,1]$, where $L_0 > 0$ exists due to the local Lipschitz continuity of F', g', and $g_1'', \ldots, g_m''$. Therefore, it follows that

$$\left\| \mathcal{L}(y,v) - \mathcal{L}(x,u) - \mathcal{L}'(x,u) \begin{pmatrix} y-x \\ v-u \end{pmatrix} \right\| \leq 2L_0(\|y-x\|\|v-u\| + \|y-x\|^2),$$

which shows that part (a) is valid with some appropriate $L > 0$. □

Lemma 3.2. *Let $a, b \in \mathbb{R}$ and $(\alpha, \beta) \in \partial\varphi(a, b)$ be arbitrarily chosen. Then, the following assertions are valid:*

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

Proof. Noting that $a^2 + b^2 > 0$ implies differentiability of φ at (a, b), assertions (a) and (b) can be checked directly. To prove (c) note that

$$1 \le \sqrt{1+t} \le 1 + \frac{1}{2}t \qquad \text{for all } t \ge 0. \tag{8}$$

Since $a > 0$ we obtain $(\alpha, \beta) = \varphi'(a, b)$. By part (a), setting $t := \frac{b^2}{a^2}$, and using (8), we get

$$\alpha = 1 - \frac{a}{\sqrt{a^2+b^2}} = 1 - \frac{a}{\sqrt{a^2(1+t)}} = \frac{\sqrt{1+t}-1}{\sqrt{1+t}} \le \sqrt{1+t} - 1 \le \frac{b^2}{2a^2}.$$

The inequality for β in part (d) follows in the same way. Part (e) becomes clear if one takes into account that φ is a continuous C-function and that $\varphi(1, 1) = 2 - \sqrt{2} > 0$. For part (f), see Lemma 3.1 in [25]. By part (a), assertions (g) and (h) can be easily verified for all $(a, b) \neq (0, 0)$. Therefore, due to (4), parts (g) and (h) must also hold for any element (α, β) of $\partial\varphi(0, 0)$. □

Lemma 3.3. *Let $A \in \mathbb{R}^{n\times m}$ and $b \in \mathbb{R}^n$ be given so that the set*

$$U := \{v \in \mathbb{R}^m_+ \mid Av + b = 0\}$$

is nonempty. Then, there exist $C > 0$ and $\rho_0 > 0$ so that, for each $\rho \in (0, \rho_0]$ and each $u \in \mathbb{R}^m_+$ with $\operatorname{dist}[u, U] \le \rho$, there exists $v(\rho, u) \in U$ satisfying

(a) $\|v(\rho, u) - u\| \le C\rho$,
(b) $v_i(\rho, u) \le Cu_i$ *for all* $i \in I$, *and*
(c) $v_i(\rho, u) = 0$ *for all* $i \in I$ *with* $u_i \le \rho$.

Proof. Due to the Weyl-Minkowski theorem, the nonempty polyhedron U can be written as

$$\left\{ \sum_{k=1}^{K} \lambda_k v^k + \sum_{l=1}^{N} \mu_l r^l \,\middle|\, \sum_{k=1}^{K} \lambda_k = 1, \lambda_1 \geq 0, \ldots, \lambda_K \geq 0, \mu_1 \geq 0, \ldots, \mu_N \geq 0 \right\}, \qquad (9)$$

where $\{v^k\}_{k=1}^K$ is the set of extreme points and $\{r^l\}_{l=1}^N$ the set of extreme rays of U. Moreover, since $U \subseteq \mathbb{R}^m_+$, we have $v^k \in \mathbb{R}^m_+$ for $k = 1, \ldots, K$ and $r^l \in \mathbb{R}^m_+$ for $l = 1, \ldots, N$. To proceed, let us define the positive value

$$\rho_0 := \frac{1}{4m} \min\{v_i^k \mid i \in I,\, k = 1, \ldots, K \text{ with } v_i^k > 0\}$$

and let $\rho \in (0, \rho_0]$ and $u \in \mathbb{R}^m_+$ with $\operatorname{dist}[u, U] \leq \rho$ be arbitrarily chosen.

Now, with the index set $I(\rho, u) := \{i \in I \mid u_i \leq \rho\}$, we first show that the polyhedral set

$$\tilde{U} := \{v \in U \mid v_i = 0 \text{ for all } i \in I(\rho, u)\}$$

is not empty (to apply Hoffman's error bound later on). For $u^\perp \in U$ denoting the Euclidean projection of u onto U, we have $\|u^\perp - u\| = \operatorname{dist}[u, U] \leq \rho$. Hence, it follows that

$$u_i^\perp \leq u_i + \rho \qquad \text{for all } i \in I \qquad (10)$$

and, further,

$$u_i^\perp \leq 2\rho \qquad \text{for all } i \in I(\rho, u). \qquad (11)$$

According to (9), there exist $\lambda_k^\perp$ $(k = 1, \ldots, K)$ and $\mu_l^\perp$ $(l = 1, \ldots, N)$ so that

$$u^\perp = \sum_{k=1}^{K} \lambda_k^\perp v^k + \sum_{l=1}^{N} \mu_l^\perp r^l.$$

Then, because of $\mu_l^\perp r^l \in \mathbb{R}^m_+$ for $l = 1, \ldots, N$, we obtain

$$2\rho m \geq \sum_{i \in I(\rho,u)} u_i^\perp = \sum_{i \in I(\rho,u)} \left(\sum_{k=1}^{K} \lambda_k^\perp v_i^k + \sum_{l=1}^{N} \mu_l^\perp r_i^l \right) \geq \sum_{k=1}^{K} \left(\sum_{i \in I(\rho,u)} \lambda_k^\perp v_i^k \right).$$

Since $\sum_{k=1}^K \lambda_k^\perp = 1$ and $\lambda_k^\perp \geq 0$ for $k = 1, \ldots, K$, we further get

$$2\rho m \geq \sum_{k=1}^{K} \lambda_k^\perp \left(\sum_{i \in I(\rho,u)} v_i^k \right) \geq \sum_{k=1}^{K} \lambda_k^\perp \left(\min_{1 \leq k \leq K} \sum_{i \in I(\rho,u)} v_i^k \right) = \min_{1 \leq k \leq K} \sum_{i \in I(\rho,u)} v_i^k.$$

Thus, there exists $k^\diamond \in \{1, \dots, K\}$ with

$$\frac{1}{2m} v_i^{k^\diamond} \le \frac{1}{2m} \sum_{i \in I(\rho,u)} v_i^{k^\diamond} \le \rho \le \rho_0 \qquad \text{for all } i \in I(\rho, u).$$

By the definition of ρ_0, it follows that $v_i^{k^\diamond} = 0$ for all $i \in I(\rho, u)$. This shows that the set $\tilde{U}$ at least contains the extreme point $v^{k^\diamond}$ and, hence, is not empty.

Now, due to Hoffman's error bound, there exists $\tilde{C} > 0$ so that

$$\operatorname{dist}[u^\perp, \tilde{U}] \le \tilde{C} \sum_{i \in I(\rho,u)} u_i^\perp \le 2m\tilde{C}\rho, \tag{12}$$

where (11) implies the last inequality. Note that the constant $\tilde{C}$ depends on the index set $I(\rho, u)$. However, because there exists only a finite number of possible index sets, we assume that $\tilde{C}$ is chosen large enough to guarantee (12). Since $\tilde{U}$ is nonempty and closed, there exists $v(\rho, u) \in \tilde{U}$ with

$$\|v(\rho, u) - u^\perp\| = \operatorname{dist}[u^\perp, \tilde{U}]. \tag{13}$$

Therefore, using (12), we obtain

$$\|v(\rho, u) - u\| \le \|v(\rho, u) - u^\perp\| + \|u^\perp - u\| \le (2m\tilde{C} + 1)\rho. \tag{14}$$

From (13), (12), and (10), we get, for $i \in I \setminus I(\rho, u) = \{i \in I \mid u_i > \rho\}$,

$$v_i(\rho, u) \le u_i^\perp + 2m\tilde{C}\rho \le u_i + \rho + 2m\tilde{C}\rho \le (2 + 2m\tilde{C})u_i.$$

With $C := 2 + 2m\tilde{C}$, this is assertion (b) for $i \in I \setminus I(\rho, u)$. For $i \in I(\rho, u)$, it holds $v_i(\rho, u) = 0$ since $v(\rho, u) \in \tilde{U}$. The latter shows that assertion (c) is valid. Finally, assertion (a) follows from (14). □

4. Local convergence

In this section, we first analyse what a single Levenberg-Marquardt step contributes to the improvement of appropriate quantities. Based on this, local convergence properties of the sequential application of Levenberg-Marquardt steps will be derived.

4.1. Analysis for a single Levenberg-Marquardt step

For simplifying the subsequent analysis within the proofs of this subsection, we use

$$d := \operatorname{dist}[z, \mathcal{Z}] \tag{15}$$

to abbreviate the distance of z to the solution set $\mathcal{Z}$ of the KKT system (1).

Theorem 4.1. *Let Assumption 1 be satisfied. Then, there exist* $L_1 > 0$ *and* $\varrho_1 > 0$ *so that, for any* $z \in B(z^*, \varrho_1) \cap \Omega$*, the unique solution* $\Delta(z)$ *of subproblem* (6) *satisfies*

$$\Psi(\Delta(z)) \leq L_1 \operatorname{dist}[z, \mathcal{Z}]^3.$$

Proof. If $z \in \mathcal{Z}$, the subproblem (6) has the unique solution $\Delta(z) = 0$. Therefore, we assume $z \notin \mathcal{Z}$ throughout this proof.

In order to apply Lemma 3.3, we set $U := \mathcal{U}$. The polyhedron $\mathcal{U}$ was introduced in Section 1 as set of Lagrange multipliers associated to x^* and can be written as

$$\mathcal{U} := \{u \in \mathbb{R}^m_+ \mid F(x^*) + g'(x^*)^\top u = 0, u^\top s^* = 0\}.$$

In what follows, $C > 0$ and $\rho_0 > 0$ are the constants which exist due to Lemma 3.3 for the polyhedron $\mathcal{U}$. Then, with $\sigma := \min\{s_i^* \mid i \in I,\, s_i^* > 0\}$, let ϱ_1 be chosen so that

$$0 < \varrho_1 \leq \min\left\{\frac{1}{(C+1)^2}, \rho_0^2, \frac{1}{2}\sigma, \frac{1}{2}\delta_*\right\}. \tag{16}$$

For an arbitrary but fixed $z = (x, s, u) \in B(z^*, \varrho_1) \cap \Omega$, let $z^\perp = (x^\perp, s^\perp, u^\perp)$ denote the projection of z onto the solution set $\mathcal{Z}$. Note that $x^\perp = x^*$ and $s^\perp = s^*$ follows from Assumption 1 and (16). Keeping in mind the notational convention (15), we have

$$d \leq \|z - z^*\| \leq \varrho_1. \tag{17}$$

Let ρ in Lemma 3.3 be chosen as

$$\rho := \sqrt{d}.$$

Then, because of (17), (16), and $\varrho_1 < 1$, it holds that

$$\rho = \sqrt{d} \leq \sqrt{\varrho_1} \leq \rho_0 \quad \text{and} \quad \operatorname{dist}[u, \mathcal{U}] = \|u - u^\perp\| \leq d \leq \sqrt{d} = \rho.$$

Thus, by Lemma 3.3, a vector $u^\diamond := v\big(\sqrt{d}, u\big) \in \mathcal{U}$ exists so that

$$\|u^\diamond - u\| \leq C\sqrt{d}, \tag{18}$$

$$u_i^\diamond \leq C u_i \qquad \text{for } i \in I, \tag{19}$$

and

$$u_i^\diamond = 0 \qquad \text{for } i \in I \text{ with } u_i \leq \sqrt{d}. \tag{20}$$

Now, let us define

$$z^\diamond := (x^*, s^*, u^\diamond) \quad \text{and} \quad \Delta^\diamond := z^\diamond - z.$$

Due to $u^\diamond \in \mathcal{U}$, $z^\diamond$ is a solution of $T_\varphi(z) = 0$. This implies $z + \Delta^\diamond = z^\diamond \in \Omega$. Hence, $\Delta^\diamond$ is a feasible point for subproblem (6). Moreover, by the definition of d in (15) and by (18),

$$\left\| \begin{pmatrix} \Delta_x^\diamond \\ \Delta_s^\diamond \end{pmatrix} \right\| = \left\| \begin{pmatrix} x^* - x \\ s^* - s \end{pmatrix} \right\| \le d \quad \text{and} \quad \|\Delta_u^\diamond\| = \|u^\diamond - u\| \le C\sqrt{d} \tag{21}$$

follows. We are now in the position to analyse the behaviour of the objective function value $\Psi(\Delta^\diamond)$ of the subproblem (6). To this end, we successively consider all terms in $\Psi(\Delta^\diamond)$, namely

$$\Psi(\Delta^\diamond) = \|t_1(\Delta^\diamond)\|^2 + \|t_2(\Delta^\diamond)\|^2 + \|t_3(\Delta^\diamond)\|^2 + t_4(\Delta^\diamond) \tag{22}$$

with

$$\begin{aligned} t_1(\Delta^\diamond) &:= \mathcal{L}(x,u) + \mathcal{L}'(x,u) \begin{pmatrix} \Delta_x^\diamond \\ \Delta_u^\diamond \end{pmatrix} \\ t_2(\Delta^\diamond) &:= g(x) + s + g'(x)\Delta_x^\diamond + \Delta_s^\diamond, \\ t_3(\Delta^\diamond) &:= \Phi_\varphi(s,u) + \alpha \circ \Delta_s^\diamond + \beta \circ \Delta_u^\diamond, \\ t_4(\Delta^\diamond) &:= \lambda(z)(\|\Delta_x^\diamond\|^2 + \|\Delta_s^\diamond\|^2) + \mu(z)\|\Delta_u^\diamond\|^2, \end{aligned}$$

where $\circ$ in t_3 denotes the Hadamard product and $\alpha, \beta \in \mathbb{R}^m$ are given componentwise by $(\alpha_i, \beta_i) \in \partial\varphi(s_i, u_i)$ for $i \in I$. Note also that

$$\|T_\varphi(z) + G(z)\Delta^\diamond\|^2 = \|t_1(\Delta^\diamond)\|^2 + \|t_2(\Delta^\diamond)\|^2 + \|t_3(\Delta^\diamond)\|^2.$$

To derive an upper bound for $\|t_1(\Delta^\diamond)\|$ let us first note that part (a) of Lemma 3.1 is applicable since

$$(x, s^*, u) \in B(z^*, \varrho_1) \subset B(z^*, 1) \quad \text{and} \quad (x^*, s^*, u^\diamond) \in B(z^*, 1),$$

where we have taken into account (16), and for the second inclusion,

$$\|u^\diamond - u^*\| \le \|u^\diamond - u\| + \|u - u^*\| \le C\sqrt{d} + \varrho_1 \le C\sqrt{\varrho_1} + \varrho_1 \le (C+1)\sqrt{\varrho_1} \le 1,$$

which follows by (18), (17), and (16). Now, since $\mathcal{L}(x^*, u^\diamond) = 0$ due to $u^\diamond \in \mathcal{U}$, part (a) of Lemma 3.1 (with $y := x^*$ and $v := u^\diamond$) and (21) yield

$$\|t_1(\Delta^\diamond)\| \le L(\|\Delta_u^\diamond\|\|\Delta_x^\diamond\| + \|\Delta_x^\diamond\|^2) \le L(Cd^{3/2} + d^2). \tag{23}$$

Similarly, for $t_2(\Delta^\diamond)$, part (b) of Lemma 3.1 with $y := x^*$ and (21) provide

$$\|t_2(\Delta^\diamond)\| \le \|g(x^*) + s^*\| + L\|\Delta_x^\diamond\|^2 = L\|\Delta_x^\diamond\|^2 \le Ld^2. \tag{24}$$

For estimating $\|t_3(\Delta^\diamond)\|$, we consider the i-th entry of $t_3(\Delta^\diamond)$, i.e.,

$$\tau_i := \varphi(s_i, u_i) + \alpha_i(s^* - s)_i + \beta_i(u^\diamond - u)_i$$

for an arbitrary $i \in I$ and distinguish several cases as follows.

1) Suppose that $(s_i, u_i) \neq (0,0)$. Then, $(\alpha_i, \beta_i) = \varphi'(s_i, u_i)$ and part (b) of Lemma 3.2 yields

$$\tau_i = \varphi(s_i, u_i) + \varphi'(s_i, u_i) \begin{pmatrix} s_i^* - s_i \\ u_i^\diamond - u_i \end{pmatrix} = \alpha_i s_i^* + \beta_i u_i^\diamond.$$

1a) If $s_i^* > 0$, then $u_i^\perp = u_i^\diamond = 0$ follows. Moreover, using (15), (17), and (16), we have

$$|s_i^* - s_i| = |s_i^\perp - s_i| \leq \|z^\perp - z\| = d \leq \varrho_1 \leq \frac{1}{2}\sigma \leq \frac{1}{2}s_i^*.$$

This implies

$$s_i \geq s_i^* - d \geq \frac{1}{2}s_i^* \geq \frac{1}{2}\sigma.$$

Similarly, $|u_i - u_i^\perp| \leq d$ implies $u_i \leq d + u_i^\perp = d$. Now, using parts (c) and (h) of Lemma 3.2, we obtain

$$0 \leq \tau_i = \alpha_i s_i^* + \beta_i u_i^\diamond = \alpha_i s_i^* \leq \frac{u_i^2}{2s_i^2} s_i^* \leq \frac{2u_i^2}{s_i^*} \leq \frac{2}{\sigma} d^2.$$

1b) If $s_i^* = 0$ and $u_i \leq \sqrt{d}$, then (20) implies $\beta_i u_i^\diamond = 0$. Hence,

$$\tau_i = \alpha_i s_i^* + \beta_i u_i^\diamond = 0$$

follows.

1c) If $s_i^* = 0$ and $u_i > \sqrt{d}$, then parts (d) and (h) of Lemma 3.2, (19), and $s_i = |s_i - s_i^*| \leq d$, yield

$$0 \leq \tau_i = \alpha_i s_i^* + \beta_i u_i^\diamond = \beta_i u_i^\diamond \leq \frac{s_i^2}{2u_i^2} u_i^\diamond \leq C \frac{s_i^2}{2u_i^2} u_i \leq C \frac{d^2}{2u_i} \leq C d^{3/2}.$$

2) Suppose that $(u_i, s_i) = (0,0)$. It follows from (20) that $u_i^\diamond = 0$ as $u_i = 0 \leq \sqrt{d}$. Moreover, due to (17) and (16), we have

$$|s_i^*| = |s_i - s_i^*| \leq \varrho_1 \leq \frac{1}{2}\sigma.$$

The definition of σ implies $s_i^* = 0$. Thus, we have

$$\tau_i = \alpha_i s_i^* + \beta_i u_i^\diamond = 0.$$

Taking into account the results of the cases 1a), 1b), 1c), and 2), it follows that

$$\tau_i \leq \max\left\{2\sigma^{-1}, C\right\} d^{3/2} \qquad \text{for } i \in I,$$

where $d \leq \varrho_1 \leq 1$ is clear from (16) and (17). Hence, we have

$$\|t_3(\Delta^\diamond)\| \leq \sqrt{m} \max\left\{2\sigma^{-1}, C\right\} d^{3/2}. \tag{25}$$

Finally, let us provide an upper bound for $t_4(\Delta^\diamond)$. Using the definition of $\lambda(z)$ and $\mu(z)$ in (5), the estimates in (21), and part (c) of Lemma 3.1, we obtain

$$t_4(\Delta^\diamond) \le \|T_\varphi(z)\| d^2 + \|T_\varphi(z)\|^2 C^2 d \le L d^3 + L^2 C^2 d^3.$$

This, the estimates (23) – (25), and (22) lead to

$$\Psi(\Delta(z)) \le \Psi(\Delta^\diamond) \le L_1 \operatorname{dist}[z, \mathcal{Z}]^3$$

with some appropriate $L_1 > 0$ (not depending on $z \in B(z^*, \rho_1) \cap \Omega$). □

Theorem 4.2. *Let Assumptions 1 and 2 be satisfied. Then, there exist $L_2 > 0$ and $\varrho_2 \in (0, \varrho_1]$ so that, for all $z \in B(z^*, \varrho_2) \cap \Omega$, the solution $\Delta(z) = (\Delta_x(z), \Delta_s(z), \Delta_u(z))$ of subproblem* (6) *satisfies*

(a) $\|\Delta_x(z)\| \le L_2 \operatorname{dist}[z, \mathcal{Z}], \quad \|\Delta_s(z)\| \le L_2 \operatorname{dist}[z, \mathcal{Z}]$,

(b) $\|\Delta_u(z)\| \le L_2 \operatorname{dist}[z, \mathcal{Z}]^{1/2}$, *and*

(c) $\|T_\varphi(z + \Delta(z))\| \le L_2 \operatorname{dist}[z, \mathcal{Z}]^{3/2}$.

Proof. Let $L_1 > 0, \varrho_1 > 0$ denote the constants from Theorem 4.1 and choose $\varrho_2 \in (0, \varrho_1]$. Then, for an arbitrary but fixed $z \in B(z^*, \varrho_2) \cap \Omega$, the definition of Ψ and Theorem 4.1 provide

$$\|T_\varphi(z) + G(z)\Delta(z)\|^2 + \lambda(z)\left(\|\Delta_x(z)\|^2 + \|\Delta_s(z)\|^2\right) + \mu(z)\|\Delta_u(z)\|^2 \le L_1 d^3. \quad (26)$$

If $z \in \mathcal{Z}$, the assertions of the theorem are obviously true. Thus, we assume $z \notin \mathcal{Z}$. Then, taking into account (26), the definition of the regularization parameters λ and μ in (5), Assumption 2, and recalling (15), we obtain

$$\omega d (\|\Delta_x(z)\|^2 + \|\Delta_s(z)\|^2) \le \|T_\varphi(z)\|(\|\Delta_x(z)\|^2 + \|\Delta_s(z)\|^2) \le L_1 d^3$$

and

$$\omega^2 d^2 \|\Delta_u(z)\|^2 \le \|T_\varphi(z)\|^2 \|\Delta_u(z)\|^2 \le L_1 d^3.$$

This yields

$$\|\Delta_x(z)\| \le \sqrt{\omega^{-1} L_1} d, \quad \|\Delta_s(z)\| \le \sqrt{\omega^{-1} L_1} d, \quad \text{and} \quad \|\Delta_u(z)\| \le \sqrt{\omega^{-2} L_1 d}. \quad (27)$$

Thus, choosing $L_2 > 0$ sufficiently large, assertions (a) and (b) follow for any $\varrho_2 \in (0, \varrho_1]$. Due to this, $\varrho_2 \in (0, \varrho_1]$ can be chosen small enough so that

$$z^+ = z + \Delta(z) \in B(z^*, 1). \quad (28)$$

For the proof of the remaining assertion, we will derive upper bounds for

$$\|T_\varphi(z + \Delta(z))\|^2 = \|T_\varphi(z^+)\|^2 = \|\mathcal{L}(x^+, u^+)\|^2 + \|g(x^+) + s^+\|^2 + \|\Phi_\varphi(s^+, u^+)\|^2. \quad (29)$$

Thanks to (28), part (a) of Lemma 3.1 with $y := x^+$ and $v := u^+$, yields

$$\mathcal{L}(x^+, u^+) = \mathcal{L}(x, u) + \mathcal{L}'(x, u) \begin{pmatrix} x^+ - x \\ u^+ - u \end{pmatrix} + R_1(x^+ - x, u^+ - u)$$

with

$$\|R_1(x^+ - x, u^+ - u)\| \le L(\|u^+ - u\|\|x^+ - x\| + \|x^+ - x\|^2).$$

Hence, using (26) and (27), we have

$$\begin{aligned} \|\mathcal{L}(x^+, u^+)\| &\le \sqrt{L_1} d^{3/2} + L(\|\Delta_u(z)\|\|\Delta_x(z)\| + \|\Delta_x(z)\|^2) \\ &\le \sqrt{L_1} d^{3/2} + L\left(\omega^{-3/2} L_1 d^{3/2} + \omega^{-1} L_1 d^2\right). \end{aligned} \tag{30}$$

For the second term in the right-hand side of (29), setting $y := x^+$, we similarly get with part (b) of Lemma 3.1,

$$\|g(x^+) + s^+\| \le \|g(x) + s + g'(x)(x^+ - x) + (s^+ - s)\| + L\|x^+ - x\|^2.$$

Using (26), and (27), we have

$$\|g(x^+) + s^+\| \le \sqrt{L_1} d^{3/2} + \omega^{-1} L_1 L d^2. \tag{31}$$

The most right term in (29) can be estimated by first considering $\Phi_\varphi(s^+, u^+)$ componentwise. To this end, let $(\alpha_i(z), \beta_i(z)) \in \partial\varphi(s_i, u_i)$ denote the corresponding entries of $G_{n+m+i,n+i}(z)$ and $G_{n+m+i,n+m+i}(z)$. Then, using parts (e) – (h) of Lemma 3.2, we get

$$0 \le \varphi(s_i^+, u_i^+) \le (2 + \sqrt{2}) \min\{s_i^+, u_i^+\} \le \frac{2 + \sqrt{2}}{2 - \sqrt{2}} (\alpha_i(z) s_i^+ + \beta_i(z) u_i^+). \tag{32}$$

Moreover, part (b) of Lemma 3.2 yields (for $(s_i, u_i) = (0, 0)$ this follows directly)

$$\varphi(s_i, u_i) + \alpha_i(z)(s_i^+ - s_i) + \beta_i(z)(u_i^+ - u_i) = \alpha_i(z) s_i^+ + \beta_i(z) u_i^+.$$

Therefore, by means of (32), we have

$$0 \le \varphi(s_i^+, u_i^+) \le (3 + \sqrt{2}) \left(\varphi(s_i, u_i) + \alpha_i(z)(s_i^+ - s_i) + \beta_i(z)(u_i^+ - u_i)\right).$$

Hence, since

$$(T_\varphi(z) + G(z)\Delta(z))_{n+m+i} = \varphi(s_i, u_i) + \alpha_i(z)(s_i^+ - s_i) + \beta_i(z)(u_i^+ - u_i)$$

holds for all $i \in I$, we obtain by (26) that

$$\|\Phi_\varphi(s^+, u^+)\| \le (3 + \sqrt{2}) \sqrt{L_1} d^{3/2}. \tag{33}$$

Summarizing (30), (31), and (33) shows that assertion (c) is valid with some suitable $L_2 > 0$ and $\varrho_2 \in (0, \varrho_1]$. □

Note that the idea of taking an appropriate feasible point of the Levenberg-Marquardt subproblem to derive an upper bound for its optimal value was already used in [19,27]. Now, due to the nonsmoothness of T_φ, constructing and dealing with such a feasible point became significantly more difficult.

4.2. Algorithm and its convergence

At first, we formally describe the Newton-type method whose local convergence behaviour we are going to analyse in the remainder of this section. Let $z^k \in \Omega$ $(k = 0, 1, 2, \ldots)$ be arbitrarily given. Then, z^{k+1} is defined by (7), i.e.,

$$z^{k+1} := z^k + \Delta(z^k), \tag{34}$$

where $\Delta(z^k)$ is obtained as unique solution of subproblem (6) with $z := z^k$. Consequently, for any $z^0 \in \Omega$, the resulting sequence $\{z^k\}$ is well-defined.

Lemma 4.3. *Let $\{z^k\} \subset \mathbb{R}^p$ and $\{r_k\} \subset \mathbb{R}_+$ be sequences, and $r \in [0, 1)$, $R > 0$ numbers so that for $k = 0, 1, 2, \ldots$*

$$\|z^k - z^0\| \le r_0 \frac{R}{1-r} \tag{35}$$

implies

$$r_{k+1} \le r r_k \quad \text{and} \quad \|z^{k+1} - z^k\| \le R r_k. \tag{36}$$

Then, $\{r_k\}$ converges to 0 and $\{z^k\}$ converges to some $\hat{z} \in \mathbb{R}^p$. Moreover, (35) *and* (36) *hold for all $k \ge 0$.*

The previous lemma will be useful to prove convergence of the Newton-type method described above to some solution $\hat{z} \in \mathcal{Z}$. The lemma is a consequence of Lemma 2.9 in [16]. Note that the very last assertion, i.e., that (35) and (36) hold for all $k \in \mathbb{N}$, is shown within the proof of Lemma 2.9 in [16].

Theorem 4.4. *Let Assumptions 1 and 2 be satisfied. Then, there exists $\epsilon > 0$ such that, for any $z^0 \in B(z^*, \epsilon) \cap \Omega$, the sequence $\{z^k\}$ generated according to* (34) *converges to some $\hat{z} \in \mathcal{Z}$. Moreover, the sequences $\{\|T_\varphi(z^k)\|\}$ and $\{\operatorname{dist}[z^k, \mathcal{Z}]\}$ converge to zero with Q-order of at least* 1.5*, whereas the sequence $\{x^k\}$ converges to $\hat{x}$ with R-order of at least* 1.5.

Proof. Let $L_2 > 0$ and $\varrho_2 > 0$ denote the constants from Theorem 4.2. Without loss of generality, we assume that

$$\varrho_2 \le \left(\frac{\omega}{4L_2}\right)^2. \tag{37}$$

By means of Lemma 4.3, we will first show that $\lim_{k\to\infty} z^k = \hat{z}$ for some $\hat{z} \in \mathcal{Z}$. To this end, let

$$r_k := \|T_\varphi(z^k)\|^{1/2}, \quad r := \frac{1}{2}, \quad \text{and} \quad R := 3L_2\omega^{-1/2}$$

be defined. Suppose that $\epsilon > 0$ is small enough such that

$$\epsilon \leq \frac{1}{2}\varrho_2 \quad \text{and} \quad \|T_\varphi(z)\|^{1/2} \leq \frac{\omega^{1/2}}{12L_2}\varrho_2 \quad \text{for all } z \in B(z^*, \epsilon) \cap \Omega. \tag{38}$$

Let us now assume that (35) is valid, i.e.,

$$\|z^k - z^0\| \leq r_0 \frac{R}{1-r} = 6L_2\omega^{-1/2}\|T_\varphi(z^0)\|^{1/2}.$$

Then, by (38), it follows that

$$\|z^k - z^*\| \leq \|z^k - z^0\| + \|z^0 - z^*\| \leq 6L_2\omega^{-1/2}\|T_\varphi(z^0)\|^{1/2} + \epsilon \leq \varrho_2. \tag{39}$$

Thus, $z^k \in B(z^*, \varrho_2) \cap \Omega$. Therefore, parts (a) and (b) of Theorem 4.2, and Assumption 2 imply

$$\|z^{k+1} - z^k\| = \|\Delta(z^k)\| \leq 3L_2 \operatorname{dist}[z^k, \mathcal{Z}]^{1/2} \leq 3L_2\omega^{-1/2}\|T_\varphi(z^k)\|^{1/2} = Rr_k, \tag{40}$$

where $\operatorname{dist}[z^k, \mathcal{Z}] \leq \varrho_2 \leq \varrho_1 \leq 1$ has been taken into account. Furthermore, using part (c) of Theorem 4.2, Assumption 2, and (37), we get

$$\begin{array}{rcl} r_{k+1}^2 &=& \|T_\varphi(z^{k+1})\| \\ &\leq& L_2 \operatorname{dist}[z^k, \mathcal{Z}]^{3/2} \\ &\leq& L_2\omega^{-1} \operatorname{dist}[z^k, \mathcal{Z}]^{1/2}\|T_\varphi(z^k)\| \\ &\leq& L_2\omega^{-1}\varrho_2^{1/2} r_k^2 \leq (rr_k)^2. \end{array}$$

Hence, because of this and (40), Lemma 4.3 implies that $\{r_k^2\} = \{\|T_\varphi(z^k)\|\}$ converges to 0 and $\{z^k\}$ converges to some $\hat{z} \in \mathcal{Z}$. Moreover, from Lemma 4.3, $\|z^k - z^0\| \leq r_0 \frac{R}{1-r}$ follows for all $k \in \mathbb{N}$. This implies, for all $k \in \mathbb{N}$, that (39) and, in turn, $z^k \in B(z^*, \varrho_2) \cap \Omega$ holds. Then, using Assumption 2 (twice) and Theorem 4.2, we obtain

$$\omega \operatorname{dist}[z^{k+1}, \mathcal{Z}] \leq \|T_\varphi(z^{k+1})\| \leq L_2 \operatorname{dist}[z^k, \mathcal{Z}]^{3/2} \leq L_2\omega^{-3/2}\|T_\varphi(z^k)\|^{3/2}$$

for all $k \in \mathbb{N}$. This shows that the sequences $\{\operatorname{dist}[z^k, \mathcal{Z}]\}$ and $\{T_\varphi(z^k)\}$ converge to zero at least with the Q-order of 1.5. Furthermore, since

$$\|x^{k+1} - \hat{x}\| = \|x^{k+1} - x^*\| \leq \operatorname{dist}[z^{k+1}, \mathcal{Z}] \leq \omega^{-1}L_2 \operatorname{dist}[z^k, \mathcal{Z}]^{3/2} \qquad \text{for } k \in \mathbb{N},$$

it follows that the sequence $\{x^k\}$ converges at least with the R-order of 1.5 to $\hat{x}$. □

5. Final remarks

This contribution shall provide a step towards methods with local superlinear convergence for nonsmooth but not piecewise smooth equations with nonisolated solutions. Moreover, there might be chances of improving the results given in Theorem 4.4 with respect to maximal possible convergence rates and to the question whether a one-step

Q-order of $\{x^k\}$ larger than 1 is possible. Let us finally mention that all results remain true if the definition (5) of the regularization parameters λ and μ is modified so that their growth behavior near the solution set $\mathcal{Z}$ remains the same.

Funding

This research was supported by the DFG (Deutsche Forschungsgemeinschaft) under Grant 290762516. The first author's work was also supported by CNPq under Grant 211914/2013-4.

References

[1] R. Behling and A. Fischer. A unified local convergence analysis of inexact constrained Levenberg-Marquardt methods. *Optim Lett*, 6(5):927–940, 2012.

[2] F. H. Clarke. *Optimization and Nonsmooth Analysis.* John Wiley & Sons, New York, 1983.

[3] T. De Luca, F. Facchinei, and C. Kanzow. A semismooth equations approach to the solution of nonlinear complementarity problems. *Math Program*, 75(3):407–439, 1996.

[4] A. Dreves, F. Facchinei, A. Fischer, and M. Herrich. A new error bound result for generalized Nash equilibrium problems and its algorithmic application. *Comput Optim Appl*, 59(1–2):63–84, 2014.

[5] F. Facchinei, A. Fischer, and M. Herrich. A family of Newton methods for nonsmooth constrained systems with nonisolated solutions. *Math Methods Oper Res*, 77(3):433–443, 2013.

[6] F. Facchinei, A. Fischer, and M. Herrich. An LP-Newton method: nonsmooth equations, KKT systems, and nonisolated solutions. *Math Program*, 146(1–2):1–36, 2014.

[7] F. Facchinei, A. Fischer, C. Kanzow, and J.-M. Peng. A simply constrained optimization reformulation of KKT systems arising from variational inequalities. *Appl Math Optim*, 40(1):19–37, 1999.

[8] F. Facchinei and J.-S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems.* Springer Series in Operations Research and Financial Engineering. Springer, New York, 2003.

[9] D. Fernández and M. Solodov. Stabilized sequential quadratic programming for optimization and a stabilized Newton-type method for variational problems. *Math Program*, 125:47–73, 2010.

[10] A. Fischer. A special Newton-type optimization method. *Optimization*, 24(3–4):269–284, 1992.

[11] A. Fischer. An NCP-function and its use for the solution of complementarity problems. In D.-Z. Du, L. Qi, and R. S. Womersley, editors, *Recent Advances in Nonsmooth Optimization*, pages 88–105. World Scientific, Singapore, 1995.

[12] A. Fischer, H. Herrich, A. F. Izmailov, and M. V. Solodov. Convergence conditions for Newton-type methods applied to complementarity systems with nonisolated solutions. *Comput Optim Appl*, 63(2):425–459, 2016.

[13] 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. *Comput Optim Appl*, 2017.

[14] A. Fischer, M. Herrich, A. F. Izmailov, and M. V. Solodov. A globally convergent LP-Newton method. *SIAM J Optim*, 26(4):2012–2033, 2016.

[15] A. Fischer and P. K. Shukla. A Levenberg-Marquardt algorithm for unconstrained multicriteria optimization. *Oper Res Lett*, 36:643–646, 2008.

[16] A. Fischer, P. K. Shukla, and M. Wang. On the inexactness level of robust Levenberg-Marquardt methods. *Optimization*, 59(2):273–287, 2010.

[17] W. W. Hager. Stabilized sequential quadratic programming. *Comput Optim Appl*, 12(1–3):253–273, 1999.

[18] A. F. Izmailov and M. V. Solodov. On error bounds and Newton-type methods for generalized Nash equilibrium problems. *Comput Optim Appl*, 59(1–2):201–218, 2014.

[19] C. Kanzow, N. Yamashita, and M. Fukushima. Levenberg-Marquardt methods with strong local convergence properties for solving equations with convex constraints. *J Comput Appl Math*, 172(2):375–397, 2004.

[20] B. Kummer. Newton's method for non-differentiable functions. In J. Guddat, B. Bank, H. Hollatz, P. Kall, D. Karte, B. Kummer, K. Lommatzsch, L. Tammer, M. Vlach, and K. Zimmermannet, editors, *Advances in Mathematical Optimization*, Mathematical Research, pages 114–125. Akademie Verlag, Berlin, 1988.

[21] L. Qi. Convergence analysis of some algorithms for solving nonsmooth equations. *Math Oper Res*, 18(1):227–244, 1993.

[22] L. Qi and J. Sun. A nonsmooth version of Newton's method. *Math Program*, 58(1–3):353–367, 1993.

[23] L. Qi and P. Tseng. On almost smooth functions and piecewise smooth functions. *Nonlinear Anal Theory Methods Appl*, 67(3):773–794, 2007.

[24] P. K. Shukla. *Levenberg-Marquardt Algorithms for Nonlinear Equations, Multi-objective Optimization, and Complementarity Problems.* PhD thesis, Technische Universität Dresden, Department of Mathematics, 2010.

[25] P. Tseng. Growth behavior of a class of merit functions for the nonlinear complementarity problem. *J Optim Theory Appl*, 89(1):17–37, 1996.

[26] S. Wright. Superlinear convergence of a stabilized SQP method to a degenerate solution. *Comput Optim Appl*, 11(3):253–275, 1998.

[27] N. Yamashita and M. Fukushima. On the rate of convergence of the Levenberg-Marquardt method. *Computing [Suppl.]*, 15:239–249, 2001.