PROJECTED FILTER TRUST REGION METHODS FOR A SEMISMOOTH LEAST SQUARES FORMULATION OF MIXED COMPLEMENTARITY PROBLEMS

Christian Kanzow and Stefania Petra

University of Würzburg
Institute of Mathematics
Am Hubland
97074 Würzburg
Germany

e-mail: kanzow@mathematik.uni-wuerzburg.de
petra@mathematik.uni-wuerzburg.de

September 28, 2005 (revised June 19, 2006)

Abstract: A reformulation of the mixed complementarity problem as a box constrained overdetermined system of semismooth equations or, equivalently, a box constrained nonlinear least squares problem with zero residual is presented. Based on this reformulation, a trust region method for the solution of mixed complementarity problems is considered. This trust region method contains elements from different areas: A projected Levenberg-Marquardt step in order to guarantee local fast convergence under suitable assumptions, affine scaling matrices which are used to improve the global convergence properties, and a multidimensional filter technique to accept a full step more frequently. Global convergence results as well as local superlinear/quadratic convergence is shown under appropriate assumptions. Moreover, numerical results for the MCPLIB indicate that the overall method is quite robust.

Keywords: Complementarity problems, nonlinear least squares reformulation, semismooth functions, trust region methods, filter method, Cauchy step, global convergence, quadratic convergence.

1 Introduction

Let $m \geq n$ and a semismooth function $\Phi : \mathcal{O} \to \mathbb{R}^m$ be given which is defined on an open neighbourhood $\mathcal{O} \subseteq \mathbb{R}^n$ of a box $\mathcal{B} := [l, u]$, where the lower bounds $l = (l_1, \ldots, l_n)^T$ and upper bounds $u = (u_1, \ldots, u_n)^T$ satisfy $-\infty \leq l_i < u_i \leq +\infty$ for all $i = 1, \ldots, n$. We consider the box constrained overdetermined system

$$\Phi(x) = 0, \quad x \in \mathcal{B}, \tag{1}$$

which is equivalent to the bound constrained least squares problem

$$\min \frac{1}{2}\|\Phi(x)\|^2 \quad \text{s.t.} \quad x \in \mathcal{B} \tag{2}$$

provided that this problem has a zero residual at the solution. If all bounds are infinite, this includes the unconstrained least squares problem.

Our motivation for considering this kind of problem comes from a recent reformulation of the mixed complementarity problem as an (unconstrained and semismooth) least squares problem, see [28]. This reformulation is different from the many existing formulations of the (mixed) complementarity problem as a square system of equations (both unconstrained and constrained), see, for example, [1, 3, 7, 13, 14, 23, 24, 25, 37, 38]. While all these reformulations as a square system of equations typically lead to some efficient algorithms, it is our experience, however, that the nonsquare reformulation from [28] is significantly more robust than the corresponding counterpart based on a reformulation with $m = n$.

In this paper, we therefore try to improve the method from [28] further. To this end, we incorporate two additional features: First, we add the bound constraints explicitly into the reformulation so that all iterates stay feasible with respect to these constraints. This should improve the robustness of the method from [28] since this avoids spurious stationary points outside the box $\mathcal{B}$, for example. Second, we add a filter in our trust region method in order to improve the efficiency of the method. In fact, preliminary numerical experiments showed that the local method behaves very good, so we try to accept the full (Levenberg-Marquardt-type) step as often as possible.

Compared with standard methods for constrained or unconstrained least squares problems (see [10, 2] and references therein), we note that our problem (1), (2) has two special properties that have to be taken into account and that will be exploited in our approach: On the one hand, the mapping Φ from (1) is only semismooth and not continuously differentiable. This point is important when we apply our approach to mixed complementarity problems where Φ is not differentiable everywhere. On the other hand, a solution x^* of (1) gives a zero residual for the least squares formulation in (2), i.e., we have $\|\Phi(x^*)\| = 0$. This property is quite unusual for general least squares problems.

The organization of this paper is as follows: We begin by considering the abstract problem (1), (2) without referring to any specific reformulation of the mixed complementarity problem. The local projected Levenberg-Marquardt-type method for the

solution of (1), (2) is discussed in Section 2. An affine-scaling trust region globalization of this local method is presented in Section 3. The global and local convergence properties of this method are investigated in Sections 4 and 5, respectively. We then incorporate the filter technique into the globalized method in Section 6 and show that all global and local convergence results still hold for this filter trust region method. The application to mixed complementarity problems is presented in Section 7 where we give the details of a reformulation of such a mixed complementarity problem as a problem of the form (1), (2). Numerical results are then presented in Section 8, and we conclude with some final remarks in Section 9.

A word on notation: $\mathbb{R}^n$ denotes the n-dimensional real vector space. The Euclidean vector norm or its associated matrix norm are denoted by $\|\cdot\|$, while $\|\cdot\|_\infty$ is used for the maximum norm in $\mathbb{R}^n$. The open ball with radius $\varepsilon > 0$ around a point x^* is denoted by $B_\varepsilon(x^*)$, where the radius is taken with respect to the Euclidean norm. If $G : \mathbb{R}^n \to \mathbb{R}^m$ is a vector-valued mapping, G_i is used for its ith component function. If G is continuously differentiable, we write $G'(x)$ for the Jacobian of G at a point $x \in \mathbb{R}^n$, whereas $\nabla G(x)$ denotes the transposed Jacobian. In particular, if $m = 1$, $\nabla G(x)$ is the gradient, which is considered to be a column vector. On the other hand, if $G : \mathbb{R}^n \to \mathbb{R}^m$ is locally Lipschitz only, then $\partial G(x)$ is the generalized Jacobian of G at x in the sense of Clarke [4]. The Euclidean projection of a vector $x \in \mathbb{R}^n$ onto the feasible set $\mathcal{B}$ will be denoted by $P_{\mathcal{B}}$. We make frequent use of the Landau symbols $O(\cdot)$ and $o(\cdot)$ which are defined as follows: Given two sequences $\{\alpha_k\}$ and $\{\beta_k\}$ converging to zero, we write $\alpha_k = O(\beta_k)$ if $\limsup_{k\to\infty} \alpha_k/\beta_k < \infty$, whereas $\alpha_k = o(\beta_k)$ means that $\lim_{k\to\infty} \alpha_k/\beta_k = 0$. We also assume that the reader is familiar with the notion of a (strongly) semismooth function. On of the central properties of a semismooth function G is that, for any sequence $\{x^k\}$ converging to some point x^* and any sequence $\{H_k\}$ with $H_k \in \partial G(x^k)$, we have

$$\big\|G(x^k) - G(x^*) - H_k(x^k - x^*)\big\| = o(\|x^k - x^*\|). \tag{3}$$

Similarly, a strongly semismooth function has the related property

$$\big\|G(x^k) - G(x^*) - H_k(x^k - x^*)\big\| = O(\|x^k - x^*\|^2). \tag{4}$$

For more details on (strongly) semismooth functions, the reader is referred to [35, 34, 33, 12].

2 Local Projected Levenberg-Marquardt Method

In this section, we present a local method for the box constrained overdetermined system of equations (1) and the related nonlinear least squares problem (2). To motivate our method, recall that (2) is an unconstrained least squares problem if all bounds are infinite. The standard (nonsmooth) Levenberg-Marquardt method for this class of problems is an iterative procedure of the form

$$x^{k+1} := x^k + p^k_{LM}, \quad k = 0, 1, \ldots,$$

where p_{LM}^k is the solution of the linear system

$$\big(H_k^T H_k + \nu_k I\big) p_{LM} = -H_k^T \Phi(x^k), \qquad H_k \in \partial\Phi(x^k). \tag{5}$$

Now, if all or some of the bounds are finite, we consider the following projected Levenberg-Marquardt method:

$$x^{k+1} := x^k + p_{PLM}^k, \quad k = 0, 1, \ldots, \tag{6}$$

where

$$p_{PLM}^k := P_{\mathcal{B}}(x^k + p_{LM}^k) - x^k \tag{7}$$

and p_{LM}^k denotes the unconstrained Levenberg-Marquardt direction from (5). Formally, we therefore obtain the following method, see also [29].

Algorithm 2.1 (Projected Levenberg-Marquardt Method)

(S.0) Choose $x^0 \in \mathcal{B}$, and set $k := 0$.

(S.1) If x^k satisfies a suitable termination criterion: STOP.

(S.2) Choose $H_k \in \partial\Phi(x^k), \nu_k > 0$, and compute p_{LM}^k from (5).

(S.3) Compute p_{PLM}^k from (7).

(S.4) Set $x^{k+1} = x^k + p_{PLM}^k, k \leftarrow k+1$, and go to (S.1).

Note that Algorithm 2.1 is well-defined, and that all iterates x^k stay in the box $\mathcal{B}$. Of course, Algorithm 2.1 is a local method only. In order to state the local convergence properties of this method, we need the following result which follows from the upper semicontinuity of the generalized Jacobian.

Lemma 2.2 *Let Φ be semismooth and x^* be a solution of problem (1) such that all elements from $\partial\Phi(x^*)$ have full rank. Then there exist constants $\varepsilon > 0$ and $c > 0$ such that $\|(H^T H)^{-1}\| \le c$ for all $H \in \partial\Phi(x)$ and all $x \in \mathbb{R}^n$ with $x \in B_\varepsilon(x^*)$.*

Proof. The proof is similar to the one of [28, Lemma 2.5], so we skip the details here. □

We are now in the position to state the main local convergence properties of Algorithm 2.1. Later, this result will facilitate the local convergence analysis of a trust-region globalization of this method.

Theorem 2.3 *Let Φ be semismooth and x^* be a solution of (1) such that all matrices $H_* \in \partial\Phi(x^*)$ have full rank. Then the following statements hold:*

(a) There exist constants $\varepsilon > 0$ and $\hat{\nu} > 0$ such that, for every $x^0 \in \mathcal{B} \cap B_\varepsilon(x^*)$ and all $\nu_k \in (0, \hat{\nu}]$, Algorithm 2.1 generates a sequence $\{x^k\}$ converging at least Q-linearly to x^*.

(b) The rate of convergence is Q-superlinear if, in addition, $\nu_k \to 0$.

(c) The rate of convergence is Q-quadratic if, in addition, $\nu_k = O(\|\Phi(x^k)\|)$ *and* Φ *is strongly semismooth.*

Proof. Lemma 2.2 implies that there are constants $\varepsilon_1 > 0$ and $c > 0$ such that

$$\big\|(H^T H + \nu I)^{-1}\big\| \le c \quad \forall x \in B_{\varepsilon_1}(x^*),\ \forall H \in \partial\Phi(x),\ \forall \nu > 0.$$

Furthermore, from the upper semicontinuity of the generalized Jacobian, we obtain the existence of constants $\varepsilon_2 > 0$ and $\alpha > 0$ with

$$\|H^T\| \le \alpha \quad \forall x \in B_{\varepsilon_2}(x^*),\ \forall H \in \partial\Phi(x).$$

Moreover, the semismoothness of Φ implies that there is a constant $\varepsilon_3 > 0$ with

$$\big\|\Phi(x) - \Phi(x^*) - H(x - x^*)\big\| \le \frac{1}{4\alpha c}\|x - x^*\| \quad \forall x \in B_{\varepsilon_3}(x^*),\ \forall H \in \partial\Phi(x),$$

see (3). Now take $\varepsilon := \min\{\varepsilon_1, \varepsilon_2, \varepsilon_3\}$ and $\hat{\nu} := \frac{1}{4c}$. Suppose that the kth iterate $x^k \in \mathcal{B}$ belongs to the ball $B_\varepsilon(x^*)$ (in the beginning, this is true since we choose $x^0 \in B_\varepsilon(x^*)$). Then we have the following identity:

$$\begin{aligned}
\big(H_k^T H_k + \nu_k I\big)(x^k + p_{LM}^k - x^*) &= \big(H_k^T H_k + \nu_k I\big)(x^k - x^*) + \big(H_k^T H_k + \nu_k I\big)p_{LM}^k \\
&= \big(H_k^T H_k + \nu_k I\big)(x^k - x^*) - H_k^T \Phi(x^k) \\
&= -H_k^T\big(\Phi(x^k) - \Phi(x^*) - H_k(x^k - x^*)\big) + \nu_k(x^k - x^*).
\end{aligned}$$

Premultiplying with $(H_k^T H_k + \nu_k I)^{-1}$ and taking norms, we therefore obtain

$$\begin{aligned}
&\|x^k + p_{LM}^k - x^*\| \\
&\le \big\|(H_k^T H_k + \nu_k I)^{-1}\big\|\big[\|H_k^T\|\,\|\Phi(x^k) - \Phi(x^*) - H_k(x^k - x^*)\| + \nu_k\|x^k - x^*\|\big] \\
&\le c\big[\alpha\|\Phi(x^k) - \Phi(x^*) - H_k(x^k - x^*)\| + \hat{\nu}\|x^k - x^*\|\big] \\
&\le c\big(\alpha\frac{1}{4\alpha c} + \frac{1}{4c}\big)\|x^k - x^*\| \;=\; \frac{1}{2}\|x^k - x^*\|.
\end{aligned}$$

This implies

$$\begin{aligned}
\|x^{k+1} - x^*\| &= \|x^k + p_{PLM}^k - x^*\| \;=\; \|P_{\mathcal{B}}(x^k + p_{LM}^k) - x^*\| \\
&= \|P_{\mathcal{B}}(x^k + p_{LM}^k) - P_{\mathcal{B}}(x^*)\| \;\le\; \|x^k + p_{LM}^k - x^*\| \;\le\; \frac{1}{2}\|x^k - x^*\|
\end{aligned}$$

since the projection operator is nonexpansive. In particular, this shows that x^{k+1} also belongs to the ε-ball around x^*, and using an induction argument, it follows that $\|x^{k+1} - x^*\| \le \frac{1}{2}\|x^k - x^*\|$ for all $k \in \mathbb{N}$. This shows that the sequence $\{x^k\}$ converges to x^* at least Q-linearly.

The proof of statements (b) and (c) is now standard by exploiting the assumptions on the sequence $\{\nu_k\}$ together with the (strong) semismoothness of Φ. We therefore skip the details. □

We close this section with the following interesting note.

Remark 2.4 *Suppose Φ is continuously differentiable with a locally Lipschitzian Jacobian Φ', and suppose that we choose $\nu_k := \nu\|\Phi(x^k)\|^2$ for some constant $\nu > 0$ in Algorithm 2.1. Then it was shown in [29] that the sequence $\{x^k\}$ generated by Algorithm 2.1 is locally quadratically convergent under an error bound condition. This condition is weaker than the full rank assumption made in Theorem 2.3 and might also hold in some situations where the solution set is not (locally) unique. However, in the present paper we are mainly interested in a nonsmooth mapping Φ, and whether the result from [29] holds in this situation is not clear.*

3 Globalized Projected Levenberg-Marquardt Method

We now want to develop a globalized version of Algorithm 2.1 for the solution of problem (1). To this end, we exploit the relation to the nonlinear least squares formulation from (2), and make the following assumption.

(A) The mapping $\Psi(x) := \frac{1}{2}\|\Phi(x)\|^2$ is continuously differentiable with gradient $\nabla\Psi(x) = H^T\Phi(x)$ for an arbitrary $H \in \partial\Phi(x)$.

This assumption certainly holds if Φ itself is continuously differentiable, but we will see that it also holds in some situations where Φ is a nonsmooth mapping, cf. Section 7. In view of Assumption (A), the Levenberg-Marquardt direction p_{LM}^k may equivalently be computed by solving the linear system $\big(H_k^T H_k + \nu_k I\big)p_{LM} = -\nabla\Psi(x^k)$, see (5).

Our globalized method is a trust-region algorithm that contains elements from affine-scaling methods. In fact, we exploit an observation by Coleman and Li [5] who noted that the first-order optimality conditions of the least squares problem (2) are equivalent to the nonlinear system of equations

$$G(x) = 0 \quad \text{with} \quad G(x) := D(x)^r \nabla\Psi(x),$$

where $r > 0$ and

$$D(x) := \operatorname{diag}\big(d_1(x), \ldots, d_n(x)\big)$$

is a suitable *scaling matrix* satisfying the following conditions:

$$d_i(x) \begin{cases} = 0, & \text{if } x_i = l_i \text{ and } [\nabla\Psi(x)]_i > 0, \\ = 0, & \text{if } x_i = u_i \text{ and } [\nabla\Psi(x)]_i < 0, \\ \geq 0, & \text{if } x_i \in \{l_i, u_i\} \text{ and } [\nabla\Psi(x)]_i = 0, \\ > 0, & \text{else.} \end{cases} \qquad i = 1, \ldots, n. \tag{8}$$

Several different scaling matrices may be found in the literature, see, e.g., [5, 27, 40]. For the sake of simplicity, we will always use the following scaling matrix in this paper:

$$d_i(x) := \begin{cases} \min\{1, x_i - l_i\}, & \text{if } [\nabla\Psi(x)]_i > 0, \\ \min\{1, u_i - x_i\}, & \text{if } [\nabla\Psi(x)]_i < 0, \\ \min\{1, x_i - l_i, u_i - x_i\}, & \text{if } [\nabla\Psi(x)]_i = 0. \end{cases} \quad i = 1, \ldots, n. \tag{9}$$

This scaling matrix was suggested by Ulbrich [38] and satisfies some additional properties that will be used in our convergence analysis, see Lemmas 4.2 and 4.7. We note, however, that other choices are possible, too, like the one from [27].

For notational convenience, we write

$$D_k := D(x^k) \quad \text{and} \quad g^k := \nabla\Psi(x^k).$$

Let

$$q_k(p) := p^T g^k + \frac{1}{2} p^T \big(H_k^T H_k + \nu_k I\big) p$$

be a quadratic approximation of $\Psi(x^k+p)-\Psi(x^k)$, where x^k denotes the current iterate. We then compute a search direction p^k as an approximate solution of the trust region subproblem

$$\min\, q_k(p) \quad \text{s.t.} \quad x^k + p \in \mathcal{B}, \;\; \|p\|_\infty \le \Delta_k \tag{10}$$

for some trust region radius $\Delta_k > 0$. Note that this is a box constrained quadratic program with feasible set

$$X_k := [l - x^k, u - x^k] \cap [-\Delta_k, +\Delta_k]^n. \tag{11}$$

Whether $x^k + p^k$ can be accepted as the new iterate x^{k+1} then depends on the ratio

$$r_k := \frac{\text{ared}_k(p^k)}{\text{pred}_k(p^k)} \tag{12}$$

of the actual and predicted reductions

$$\text{ared}_k(p) := \Psi(x^k) - \Psi(x^k + p) \quad \text{and} \quad \text{pred}_k(p) := -q_k(p),$$

respectively. In order to guarantee nice global convergence results, the approximate solution p^k of the trust region subproblem (10) has to satisfy at least the *feasibility condition*

$$p^k \in X_k \iff x^k + p^k \in \mathcal{B},\; \|p^k\|_\infty \le \Delta_k, \tag{13}$$

and the *fraction of Cauchy decrease* condition

$$q_k(p^k) \le \alpha q_k(p_C^k), \tag{14}$$

where $\alpha \in (0, 1]$ is a given constant, and $p_C^k = p(t_k)$ denotes the scaled *Cauchy step*, where t_k is defined as the solution of the one-dimensional subproblem

$$\begin{array}{ll} \min_t q_k\big(p(t)\big) \quad \text{s.t.} & p(t) = -tD_k^2 g^k, \;\; t \ge 0, \\ & \|p(t)\|_\infty \le \Delta_k, \;\; x^k + p(t) \in \mathcal{B}. \end{array} \tag{15}$$

Note that, since $q_k(p_C^k) \le q_k(0) = 0$, we may take the Cauchy step $p^k = p_C^k$ itself in order to get a suitable approximate solution of the trust region subproblem (10) satisfying (14). The overall method is as follows.

Algorithm 3.1 (Scaled Trust Region Method)

(S.0) Choose $x^0 \in \mathcal{B},\ \Delta_0 > 0, 0 < \rho_1 < \rho_2 < 1,\ 0 < \sigma_1 < 1 < \sigma_2,\ \varepsilon \geq 0,\ \eta \in (0,1),\ \Delta_{\min} > 0,\ \alpha \in (0,1]$, and set $k := 0$.

(S.1) If $\|D_k g^k\| \leq \varepsilon$: STOP.

(S.2) Choose $H_k \in \partial\Phi(x^k), \nu_k > 0$, and compute p^k_{LM} using (5).

(S.3) Compute p^k_{PLM} from (7). If $\|\Phi(x^k + p^k_{PLM})\| \leq \eta\|\Phi(x^k)\|$ holds, set $x^{k+1} := x^k + p^k_{PLM},\ \Delta_{k+1} := \max\{\Delta_{\min}, \sigma_2\Delta_k\}$, and go to step (S.6); otherwise go to step (S.4).

(S.4) Compute an approximate solution p^k of the trust region subproblem (10) satisfying (13) and (14), and define r_k by (12). If $r_k \geq \rho_1$, we call the iteration k successful and set $x^{k+1} := x^k + p^k$; otherwise we set $x^{k+1} := x^k$.

(S.5) Update the trust region radius as follows:

$$\Delta_{k+1} := \begin{cases} \sigma_1\Delta_k, & \text{if } r_k < \rho_1, \\ \max\{\Delta_{\min}, \Delta_k\}, & \text{if } r_k \in [\rho_1, \rho_2), \\ \max\{\Delta_{\min}, \sigma_2\Delta_k\}, & \text{if } r_k \geq \rho_2. \end{cases}$$

(S.6) Set $k \leftarrow k + 1$, and go to (S.1).

Throughout the rest of this paper, we denote by

$$\hat{g}^k := D_k g^k$$

the scaled gradient. Using this notation, we can state the following result which is standard for trust region methods and provides a lower bound for the predicted reduction pred_k, cf. [5, Lemma 3.1]. We therefore skip the proof.

Lemma 3.2 *Let p^k be an approximate solution of the subproblem* (10) *satisfying the fraction of Cauchy decrease condition* (14)*. Then*

$$\text{pred}_k(p^k) \geq \frac{1}{2}\alpha\|\hat{g}^k\| \min\Big\{\frac{\|\hat{g}^k\|}{\|H_k^T H_k + \nu_k I\|}, \Delta_k, 1\Big\}.$$

As a direct consequence of Lemma 3.2, we have that Algorithm 3.1 is well-defined since the denominator $\text{pred}_k(p^k) = -q_k(p^k)$ in the definition of r_k is nonzero for all $k \in \mathbb{N}$ because otherwise the algorithm would have stopped at step (S.1). More precisely, Lemma 3.2 shows that the denominator $\text{pred}_k(p^k)$ is always positive. This, in turn, implies that $\Psi(x^{k+1}) \leq \Psi(x^k)$ for all iterations $k \in \mathbb{N}$ for which the test in step (S.3) does not hold. On the other hand, if this test is satisfied, we also have $\Psi(x^{k+1}) \leq \Psi(x^k)$. Consequently, the entire sequence $\{\Psi(x^k)\}$ is monotonically decreasing. We will use this fact several times in our subsequent convergence analysis.

Moreover, the proof of Lemma 3.2 clearly shows how the scaled Cauchy step p^k_C can be computed in practice.

4 Global Convergence

The aim of this section is to prove some global convergence results for Algorithm 3.1. To this end, we assume that Algorithm 3.1 does not terminate after a finite number of iterations. Furthermore, we recall that Algorithm 3.1 uses two different search directions, namely the projected Levenberg-Marquardt step p^k_{PLM} and the Cauchy-like step p^k. The former will be used in order to prove fast local convergence, whereas the latter is the main tool for proving global convergence results. Our first result basically shows that the global convergence properties are not destroyed by using the projected Levenberg-Marquardt direction.

Theorem 4.1 *If the direction p^k_{PLM} is accepted an infinite number of times in step (S.3) of Algorithm 3.1, we have* $\lim_{k\to\infty} \|\Phi(x^k)\| = 0$.

Proof. We already observed that the entire sequence $\{\Psi(x^k)\}$ is monotonically decreasing. Obviously, this implies that the whole sequence $\{\|\Phi(x^k)\|\}$ is also monotonically decreasing. Since the test $\|\Phi(x^k + p^k_{PLM})\| \leq \eta\|\Phi(x^k)\|$ is accepted an infinite number of times in view of our assumptions, we therefore get $\|\Phi(x^k)\| \to 0$ for $k \to \infty$ since $\eta \in (0, 1)$. □

For a complete convergence analysis of Algorithm 3.1, it remains to consider the case where the direction p^k_{PLM} from step (S.3) is accepted only a finite number of times. In the following global convergence analysis, we therefore assume without loss of generality that the direction p^k_{PLM} is never accepted in step (S.3). Hence, in all iterations $k \in \mathbb{N}$, we take the approximate solution p^k from step (S.4).

The technique of proof is similar to the one in [24, 30] for a square system of equations. In addition, we present another convergence result which is based on a stronger smoothness property of the mapping $x \mapsto D(x)\nabla\Psi(x)$. To this end, we first note that $\nabla\Psi$ is continuous on $\mathcal{O}$ since Ψ is continuously differentiable on this set by Assumption (A). However, the scaling $x \mapsto D(x)$ is discontinuous at certain points x. Nevertheless, our first result states that the mapping $x \mapsto D(x)\nabla\Psi(x)$ is continuous, see also [38, Lemma 6.1].

Lemma 4.2 *The mapping $x \mapsto D(x)\nabla\Psi(x)$ is continuous on $\mathcal{O}$.*

Proof. The proof is straightforward and therefore omitted here. □

From now on, we always assume that, in addition to Assumption (A), the following condition is satisfied.

(B) The sequence $\{\nu_k\}$ is bounded.

The following result will be used in order to show that every accumulation point of a sequence generated by Algorithm 3.1 is a KKT-point of (2).

Lemma 4.3 *Let $\{x^k\}$ be a sequence generated by Algorithm 3.1, and let $\{x^k\}_K$ be a subsequence converging to a point $x^* \in \mathcal{B}$. If x^* is not a KKT-point of (2), then $\liminf_{k\to\infty, k\in K} \Delta_k > 0$.*

Proof. The proof is similar to those of corresponding results given in [24, 30], so we do not give the details here. □

As a direct consequence of Lemma 4.3, we now show that there are infinitely many successful iterations (provided that Algorithm 3.1 does not terminate at a stationary point of (2) after a finite number of iterations).

Lemma 4.4 *Let $\{x^k\}$ be a sequence generated by Algorithm 3.1. Then there are infinitely many successful iterations.*

Proof. Assume that the number of successful iterations in (S.4) is finite. Then there exists an index $k_0 \in \mathbb{N}$ such that $r_k < \rho_1$ and $x^k = x^{k_0}$ for all $k \geq k_0$. Hence $\{\Delta_k\} \to 0$ and $\{x^k\} \to x^{k_0}$. This contradicts Lemma 4.3, since $\hat{g}^{k_0} = D(x^{k_0})\nabla\Psi(x^{k_0}) \neq 0$. □

We are now in the position to state our first global convergence result.

Theorem 4.5 *Let $\{x^k\}$ be any sequence generated by Algorithm 3.1. Then every accumulation point of $\{x^k\}$ is a KKT-point of (2).*

Proof. We first recall that, as a consequence of Theorem 4.1, we may assume without loss of generality that the search direction is always computed by step (S.4) of Algorithm 3.1.

Let x^* be an accumulation point of $\{x^k\}$, and let $\{x^k\}_K$ be a subsequence converging to x^*. In view of Lemma 4.4, we can assume, without loss of generality, that all $k \in K$ are successful iterations, since $x^{k+1} = x^k$ for all nonsuccessful iterations k. Suppose that x^* is not a KKT-point of (2). Then it follows from Lemma 4.2, the upper semicontinuity of the generalized Jacobian, and Assumption (B) that there exist suitable constants $\beta_1 > 0$ and $\beta_2 > 0$ such that

$$\|\hat{g}^k\| \geq \beta_1 \quad \text{and} \quad \|H_k^T H_k + \nu_k I\| \leq \beta_2 \tag{16}$$

for all $k \in K$. Since the iterations $k \in K$ are successful, we have $r_k \geq \rho_1$ for all $k \in K$. By Lemma 3.2 and the fact that the entire sequence $\{\Psi(x^k)\}$ is decreasing and bounded from below, we have

$$\begin{aligned}
\Psi(x^0) &\geq \sum_{k=0}^{\infty} \big(\Psi(x^k) - \Psi(x^{k+1})\big) \geq \sum_{k=0}^{\infty} \rho_1 \operatorname{pred}_k(p^k) \geq \rho_1 \sum_{k\in K} \operatorname{pred}_k(p^k) \\
&\geq \frac{\alpha\rho_1}{2} \sum_{k\in K} \|\hat{g}^k\| \min\Big\{\frac{\|\hat{g}^k\|}{\|H_k^T H_k + \nu_k I\|}, \Delta_k, 1\Big\} \geq \frac{\alpha\rho_1\beta_1}{2} \sum_{k\in K} \min\Big\{\frac{\beta_1}{\beta_2}, \Delta_k, 1\Big\}.
\end{aligned}$$

This implies $\{\Delta_k\}_K \to 0$, a contradiction to Lemma 4.3. $\square$

We want to give two additional global convergence results which are more traditional in the context of trust region methods, see, e.g., [6]. To this end, we first introduce the following assumption.

(C) The sequence $\{H_k\}$ is bounded.

Then we can state the following result which is weaker than Theorem 4.5 in the sense that it does not guarantee that every accumulation point is a KKT-point. However, it will be used in the subsequent result in order to state a stronger convergence theorem.

Theorem 4.6 *Suppose that Assumptions (A), (B), and (C) hold, and let $\{x^k\}$ be any sequence generated by Algorithm 3.1. Then*

$$\liminf_{k\to\infty} \|\hat{g}^k\| = 0. \tag{17}$$

Proof. Suppose there exists a constant $\beta_1 > 0$ such that $\|\hat{g}^k\| \geq \beta_1$ for all $k \in \mathbb{N}$. Assumptions (B) and (C) imply the existence of a constant $\beta_2 > 0$ such that $\|H_k^T H_k + \nu_k I\| \leq \beta_2$ for all $k \in \mathbb{N}$. We denote the set of all successful iterates by S and note that it has infinite cardinality by Lemma 4.4. Since the entire sequence $\{\Psi(x^k)\}$ is monotonically decreasing, we get from Lemma 3.2 that

$$\Psi(x^0) \geq \sum_{k\in S} \big(\Psi(x^k) - \Psi(x^{k+1})\big) \geq \sum_{k\in S} \rho_1 \operatorname{pred}_k(p^k) \geq \frac{\alpha\rho_1\beta_1}{2} \sum_{k\in S} \min\Big\{\frac{\beta_1}{\beta_2}, \Delta_k, 1\Big\}.$$

Hence $\sum_{k\in S} \Delta_k < \infty$. Using $\|x^{k+1} - x^k\|_\infty = \|p^k\|_\infty \leq \Delta_k$ for all $k \in S$, we get $\sum_{k\in S} \|x^{k+1} - x^k\|_\infty < \infty$. Since $\|x^{k+1} - x^k\|_\infty = 0$ for all $k \notin S$, we obtain

$$\sum_{k=0}^{\infty} \|x^{k+1} - x^k\|_\infty < \infty.$$

Hence $\{x^k\}$ is a Cauchy sequence and therefore convergent to a point x^*. Theorem 4.5 then implies that x^* is a KKT-point of (2). Consequently, we have $\hat{g}^* := D(x^*)\nabla\Psi(x^*) = 0$. However, Lemma 4.2 implies that $\hat{g}^k \to \hat{g}^* = 0$, a contradiction to our assumption. Hence (17) holds. $\square$

In order to state our final convergence result, we need to introduce another assumption.

(D) The function $\nabla\Psi$ is uniformly continuous and bounded on the box $\mathcal{B}$.

Note that Assumption (D) automatically holds under Assumption (A) if the box $\mathcal{B}$ is a compact set. As a consequence of Assumption (D), we get the following preliminary result, see also [38].

Lemma 4.7 *Under Assumption (D), the mapping $x \mapsto D(x)\nabla\Psi(x)$ is uniformly continuous on $\mathcal{B}$.*

Proof. Let $\varepsilon > 0$ and $i \in \{1, \dots, n\}$ be arbitrary. Since, by assumption, $\nabla\Psi$ is bounded on $\mathcal{B}$, there is a constant $b > 0$ such that $\|\nabla\Psi(x)\|_\infty \leq b$ for all $x \in \mathcal{B}$. Since $\nabla\Psi$ is uniformly continuous, we have the existence of a constant $\bar{\delta} > 0$ such that

$$-\frac{\varepsilon}{4} \leq [\nabla\Psi(y)]_i - [\nabla\Psi(x)]_i \leq \frac{\varepsilon}{4} \tag{18}$$

holds for all $x, y \in \mathcal{B}$ with $|y_i - x_i| \leq \bar{\delta}$. Now define $\delta := \min\{\bar{\delta}, \frac{3\varepsilon}{4b}\}$, let $x, y \in \mathcal{B}$ with $\|y - x\|_\infty \leq \delta$ be arbitrary, and set $T(x, y) := D(y)\nabla\Psi(y) - D(x)\nabla\Psi(x)$. Then

$$\begin{aligned} \big|[T(x,y)]_i\big| &= \big|d_i(y)[\nabla\Psi(y)]_i - d_i(x)[\nabla\Psi(x)]_i\big| \\ &\leq d_i(y)\big|[\nabla\Psi(y)]_i - [\nabla\Psi(x)]_i\big| + \big|d_i(y) - d_i(x)\big|\,\big|[\nabla\Psi(x)]_i\big| \\ &\overset{(9)}{\leq} \big|[\nabla\Psi(y)]_i - [\nabla\Psi(x)]_i\big| + \big|d_i(y) - d_i(x)\big|\,\big|[\nabla\Psi(x)]_i\big|. \end{aligned} \tag{19}$$

We now distinguish several cases.
Case 1. If $[\nabla\Psi(x)]_i = 0$, we immediately obtain $\big|[T(x,y)]_i\big| \leq \varepsilon$ from (19) and (18).
Case 2. If $[\nabla\Psi(x)]_i > 0$, we consider two subcases:
Case 2.1. If $[\nabla\Psi(x)]_i \leq \frac{3\varepsilon}{8}$ we obtain $\big|[T(x,y)]_i\big| \leq \frac{\varepsilon}{4} + 2\frac{3\varepsilon}{8} = \varepsilon$ from (19).
Case 2.2. If $[\nabla\Psi(x)]_i > \frac{3\varepsilon}{8}$, we obtain

$$[\nabla\Psi(y)]_i = \big([\nabla\Psi(y)]_i - [\nabla\Psi(x)]_i\big) + [\nabla\Psi(x)]_i \geq -\frac{\varepsilon}{4} + \frac{3\varepsilon}{8} = \frac{\varepsilon}{8} > 0$$

from the first inequality in (18). Hence both $[\nabla\Psi(x)]_i$ and $[\nabla\Psi(y)]_i$ are positive, and (19) together with the definition of the scaling matrix $D(x)$ becomes

$$\begin{aligned} \big|[T(x,y)]_i\big| &\leq \big|[\nabla\Psi(y)]_i - \nabla\Psi(x)]_i\big| + \big|\min\{1, y_i - l_i\} - \min\{1, x_i - l_i\}\big|[\nabla\Psi(x)]_i \\ &\leq \big|[\nabla\Psi(y)]_i - \nabla\Psi(x)]_i\big| + \big|y_i - x_i\big|\,[\nabla\Psi(x)]_i \leq \frac{\varepsilon}{4} + \frac{3\varepsilon}{4b}b = \varepsilon. \end{aligned}$$

Case 3. Analogously, for $[\nabla\Psi(x)]_i < 0$, we consider the following two subcases:
Case 3.1. If $[\nabla\Psi(x)]_i \geq -\frac{3\varepsilon}{8}$, we obtain $\big|[T(x,y)]_i\big| \leq \frac{\varepsilon}{4} + 2\frac{3\varepsilon}{8} = \varepsilon$ from (19).
Case 3.2. If $[\nabla\Psi(x)]_i < -\frac{3\varepsilon}{8}$, the second inequality in (18) implies

$$[\nabla\Psi(y)]_i = \big([\nabla\Psi(y)]_i - [\nabla\Psi(x)]_i\big) + [\nabla\Psi(x)]_i \leq \frac{\varepsilon}{4} - \frac{3\varepsilon}{8} = -\frac{\varepsilon}{8} < 0.$$

Hence $[\nabla\Psi(x)]_i$ and $[\nabla\Psi(y)]_i$ are both negative, and we obtain

$$\begin{aligned} \big|[T(x,y)]_i\big| &\leq \big|[\nabla\Psi(y)]_i - \nabla\Psi(x)]_i\big| + \big|\min\{1, u_i - y_i\} - \min\{1, u_i - x_i\}\big|\,\big|[\nabla\Psi(x)]_i\big| \\ &\leq \big|[\nabla\Psi(y)]_i - \nabla\Psi(x)]_i\big| + \big|y_i - x_i\big|\,\big|[\nabla\Psi(x)]_i\big| \leq \frac{\varepsilon}{4} + \frac{3\varepsilon}{4b}b = \varepsilon \end{aligned}$$

from (19).

Summarizing all three cases, we obtain $\|T(x,y)\|_\infty \leq \varepsilon$ for all $\|y-x\|_\infty \leq \delta$. The assertion of the lemma therefore holds. $\square$

We are now in the position to state our final global convergence result.

Theorem 4.8 *Suppose that Assumptions (A), (B), (C), and (D) hold, and let $\{x^k\}$ be any sequence generated by Algorithm 3.1. Then*

$$\lim_{k\to\infty} \|\hat{g}^k\| = 0. \tag{20}$$

Proof. Suppose that (20) does not hold. Then there is a constant $\varepsilon > 0$ and a subsequence $\{x^k\}_K, K \subseteq \mathbb{N}$, such that

$$\|\hat{g}^k\| \geq 2\varepsilon \quad \forall k \in K. \tag{21}$$

In view of Theorem 4.6, we can find, for each $k \in K$, an index $\ell(k) > k$ such that

$$\|\hat{g}^\ell\| \geq \varepsilon \ \ \forall k \leq \ell < \ell(k) \quad \text{and} \quad \|\hat{g}^{\ell(k)}\| < \varepsilon. \tag{22}$$

Let $\beta_2 > 0$ be a constant such that $\|H_k^T H_k + \nu_k I\| \leq \beta_2$ for all $k \in \mathbb{N}$, cf. Assumptions (B) and (C). Given $k \in K$, take an arbitrary index ℓ with $k \leq \ell < \ell(k)$ and suppose, for the moment, that iteration ℓ is successful. Then Lemma 3.2 implies

$$\begin{aligned}
\Psi(x^\ell) - \Psi(x^{\ell+1}) &\geq \rho_1 \operatorname{pred}_\ell(p^\ell) \geq \frac{1}{2}\alpha\rho_1\|\hat{g}^\ell\| \min\Big\{\frac{\|\hat{g}^\ell\|}{\|H_\ell^T H_\ell + \nu_\ell I\|}, \Delta_\ell, 1\Big\} \\
&\geq \frac{1}{2}\alpha\rho_1\varepsilon \min\Big\{\frac{\varepsilon}{\beta_2}, \Delta_\ell, 1\Big\} \geq \frac{1}{2}\alpha\rho_1\varepsilon \min\Big\{\frac{\varepsilon}{\beta_2}, \|x^{\ell+1}-x^\ell\|_\infty, 1\Big\}.
\end{aligned}$$

Since $\{\Psi(x^k)\}$ converges, we therefore get $\Psi(x^\ell) - \Psi(x^{\ell+1}) \geq \frac{1}{2}\alpha\rho_1\varepsilon\|x^{\ell+1}-x^\ell\|_\infty$ for all these ℓ sufficiently large. Trivially, this inequality also holds for all nonsuccessful iterations. Consequently, we get

$$\begin{aligned}
\frac{1}{2}\alpha\rho_1\varepsilon\|x^{\ell(k)}-x^k\|_\infty &\leq \frac{1}{2}\alpha\rho_1\varepsilon \sum_{\ell=k}^{\ell(k)-1} \|x^{\ell+1}-x^\ell\|_\infty \\
&\leq \sum_{\ell=k}^{\ell(k)-1} \big(\Psi(x^\ell) - \Psi(x^{\ell+1})\big) = \Psi(x^k) - \Psi(x^{\ell(k)})
\end{aligned}$$

for all $k \in K$. The convergence of the entire sequence $\{\Psi(x^k)\}$ therefore implies $\big\{\|x^{\ell(k)}-x^k\|\big\}_K \to 0$. In view of Lemma 4.7, we then get $\big\{\|\hat{g}^{\ell(k)}-\hat{g}^k\|\big\}_K \to 0$. On the other hand, it follows from (21) and (22) that

$$\|\hat{g}^{\ell(k)}-\hat{g}^k\| \geq \|\hat{g}^k\| - \|\hat{g}^{\ell(k)}\| \geq 2\varepsilon - \varepsilon = \varepsilon.$$

This contradiction completes the proof. $\square$

5 Local Convergence

In this section, we consider the local behavior of Algorithm 3.1. Taking into account Theorem 2.3, it follows that we only have to show that the projected Levenberg-Marquardt direction p^k_{PLM} from step (S.3) is automatically accepted in a neighbourhood of a solution of (1). To this end, we begin with the following preliminary result.

Lemma 5.1 *Let $x^* \in \mathbb{R}^n$ be a solution of (1) such that all elements from $\partial\Phi(x^*)$ have full rank, and let $\bar{\nu} > 0$. Then there exist constants $\varepsilon > 0$ and $\kappa > 0$ such that*

$$p^T(H^TH + \nu I)p \geq \kappa\|p\|^2$$

for all $H \in \partial\Phi(x)$ and all $x \in \mathbb{R}^n$ with $\|x - x^*\| \leq \varepsilon$ and all $\nu \in [0, \bar{\nu}]$, i.e., the matrices $H^TH + \nu I$ are uniformly positive definite.

Proof. It follows from Lemma 2.2 that there exist constants $\varepsilon > 0$ and $c > 0$ such that

$$\|(H^TH + \nu I)^{-1}\| \leq c \quad \forall x \in B_\varepsilon(x^*),\ \forall H \in \partial\Phi(x),\ \forall \nu \in [0, \bar{\nu}]. \tag{23}$$

Since

$$\|(H^TH + \nu I)^{-1}\| = \frac{1}{\lambda_{\min}(H^TH + \nu I)} =: \frac{1}{\lambda^x_{\min}}$$

for all $x \in B_\varepsilon(x^*)$, all $H \in \partial\Phi(x)$ and all $\nu \in [0, \bar{\nu}]$, we obtain from (23)

$$(p)^T(H^TH + \nu I)p \geq \lambda^x_{\min}\|p\|^2 \geq \frac{1}{c}\|p\|^2 \quad \forall x \in B_\varepsilon(x^*),\ \forall H \in \partial\Phi(x),\ \forall \nu \in [0, \bar{\nu}].$$

Hence the assertion holds with $\kappa := \frac{1}{c}$. □

We also need the following result in order to establish our main local convergence theorem.

Lemma 5.2 *Let $x^* \in \mathbb{R}^n$ be a solution of problem (1) such that all elements from $\partial\Phi(x^*)$ have full rank. Then there exist constants $\varepsilon > 0$ and $\gamma > 0$ such that*

$$\|\Phi(x)\| \geq \gamma\|x - x^*\| \quad \textit{for all } x \in B_\varepsilon(x^*).$$

Proof. Lemma 5.1 implies that there are constants $\varepsilon_1 > 0$ and $\kappa > 0$ such that

$$\|H(x - x^*)\|^2 = (x - x^*)^T H^T H(x - x^*) \geq \kappa\|x - x^*\|^2 \tag{24}$$

holds for all $x \in B_{\varepsilon_1}(x^*)$ and all $H \in \partial\Phi(x)$. Furthermore, the semismoothness of Φ implies that there is a constant $\varepsilon_2 > 0$ such that

$$\|\Phi(x) - \Phi(x^*) - H(x - x^*)\| \leq \frac{\sqrt{\kappa}}{2}\|x - x^*\| \tag{25}$$

holds for all $x \in B_{\varepsilon_2}(x^*)$ and all $H \in \partial\Phi(x)$, cf. (3). Setting $\varepsilon := \min\{\varepsilon_1, \varepsilon_2\}$, we obtain from (24) and (25) that, for all $x \in B_\varepsilon(x^*)$ and all $H \in \partial\Phi(x)$, we have

$$\begin{aligned}\|\Phi(x)\| &= \big\|H(x-x^*) + \big(\Phi(x) - \Phi(x^*) - H(x-x^*)\big)\big\| \\ &\geq \|H(x-x^*)\| - \|\Phi(x) - \Phi(x^*) - H(x-x^*)\| \\ &\geq \sqrt{\kappa}\|x-x^*\| - \frac{\sqrt{\kappa}}{2}\|x-x^*\| = \frac{\sqrt{\kappa}}{2}\|x-x^*\|.\end{aligned}$$

The statement therefore holds with $\gamma := \frac{\sqrt{\kappa}}{2}$. □

We are now in the position to state the main convergence result of this section.

Theorem 5.3 *Suppose that Assumptions (A), (B) hold, and let $\{x^k\}$ be a sequence generated by Algorithm 3.1. Assume that $x^* \in \mathcal{B}$ is an accumulation point of $\{x^k\}$ such that x^* is a solution of problem (1) and such that all elements from $\partial\Phi(x^*)$ have full rank. Then the following statements hold:*

(a) The entire sequence $\{x^k\}$ converges to x^*.

(b) The direction p^k_{PLM} in (S.3) is always accepted for k sufficiently large so that the next iterate is given by $x^{k+1} = x^k + p^k_{PLM}$, provided that $\nu_k \to 0$.

(c) The rate of convergence is Q-superlinear if $\nu_k \to 0$.

(d) The rate of convergence is Q-quadratic if $\nu_k = O(\|\Phi(x^k)\|)$ and, in addition, Φ is strongly semismooth.

Proof. (a) To establish that the entire sequence $\{x^k\}$ converges to x^*, we first note that x^* is an isolated solution of (1). This follows immediately from Lemma 5.2. Since Algorithm 3.1 generates a decreasing sequence $\{\Psi(x^k)\}$ and x^* is a zero of Φ (and Ψ), it follows that the entire sequence $\{\Psi(x^k)\}$ converges to zero. Hence every accumulation point of the sequence $\{x^k\}$ is a solution of (1). Consequently, x^* is an isolated accumulation point of the sequence $\{x^k\}$.

Now let $\{x^k\}_K$ denote any subsequence converging to x^*, and note that $\Phi(x^*) = 0$ and, therefore, $\nabla\Psi(x^*) = 0$.

For all $k \in \mathbb{N}$ with the search direction p^k_{PLM} coming from (S.3), we have

$$\begin{aligned}\|x^{k+1} - x^k\| = \|p^k_{PLM}\| &= \|P_{\mathcal{B}}(x^k + p^k_{LM}) - x^k\| = \|P_{\mathcal{B}}(x^k + p^k_{LM}) - P_{\mathcal{B}}(x^k)\| \\ &\leq \|x^k + p^k_{LM} - x^k\| = \|-(H_k^T H_k + \nu_k I)^{-1}\nabla\Psi(x^k)\| \qquad (26) \\ &\leq \|(H_k^T H_k + \nu_k I)^{-1}\|\,\|\nabla\Psi(x^k)\|.\end{aligned}$$

We now consider the iterates $k \in \mathbb{N}$ where the search direction is the Cauchy-like step p^k coming from (S.4). Using Lemma 5.1, it follows that there is a constant $\kappa > 0$ such that

$$\kappa\|p^k\|^2 \leq (p^k)^T(H_k^T H_k + \nu_k I)p^k \qquad (27)$$

for all $k \in K$ sufficiently large. On the other hand, since $q_k(p^k) \leq 0$, we get

$$\frac{1}{2}(p^k)^T(H_k^T H_k + \nu_k I)p^k \leq -\nabla\Psi(x^k)^T p^k. \tag{28}$$

From (27), (28), and the Cauchy-Schwarz inequality, we obtain

$$\kappa\|p^k\|^2 \leq 2\|\nabla\Psi(x^k)\|\,\|p^k\|.$$

This implies

$$\|x^{k+1} - x^k\| \leq \|p^k\| \leq \frac{2}{\kappa}\|\nabla\Psi(x^k)\|. \tag{29}$$

Since $\{\nabla\Psi(x^k)\}_K \to \nabla\Psi(x^*) = 0$, we obtain from (26), (29), and the boundedness of the sequence $\{\|(H_k^T H_k + \nu_k I)^{-1}\|\}$ (see Lemma 2.2) that $\{\|x^{k+1} - x^k\|\}_K \to 0$. Hence statement (a) follows from [31, Lemma 4.10].

(b), (c), (d) We only have to prove statement (b), since (c) and (d) then follow directly from Theorem 2.3.

To this end, we first recall from the proof of Theorem 2.3 that

$$\|x^k + p_{LM}^k - x^*\| = o(\|x^k - x^*\|), \tag{30}$$

provided that $\nu_k \to 0$. Furthermore, Lemma 5.2 implies that there is a constant $\gamma > 0$ such that

$$\|\Phi(x^k)\| \geq \gamma\|x^k - x^*\| \tag{31}$$

for all $k \in \mathbb{N}$ sufficiently large. Using (31) and (30), we obtain

$$\begin{aligned}\frac{\|\Phi(x^k + p_{LM}^k)\|}{\|\Phi(x^k)\|} &\leq \frac{\|\Phi(x^k + p_{LM}^k)\|}{\gamma\|x^k - x^*\|} = \frac{\|\Phi(x^k + p_{LM}^k) - \Phi(x^*)\|}{\gamma\|x^k - x^*\|} \\ &\leq \frac{L\|x^k + p_{LM}^k - x^*\|}{\gamma\|x^k - x^*\|} \to 0,\end{aligned}$$

where $L > 0$ denotes the local Lipschitz constant of Φ (note that Φ is semismooth and therefore, in particular, locally Lipschitzian). Hence we have

$$\|\Phi(x^k + p_{LM}^k)\| = o(\|\Phi(x^k)\|). \tag{32}$$

Using the definition of p_{PLM}^k and exploiting the definition of the projection $P_{\mathcal{B}}$, we obtain

$$\begin{aligned}\|p_{PLM}^k - p_{LM}^k\| &= \|P_{\mathcal{B}}(x^k + p_{LM}^k) - (x^k + p_{LM}^k)\| \\ &\overset{x^* \in \mathcal{B}}{\leq} \|x^* - (x^k + p_{LM}^k)\| \overset{(30)}{=} o(\|x^k - x^*\|) \overset{(31)}{=} o(\|\Phi(x^k)\|).\end{aligned} \tag{33}$$

From (33) and (32), we now get

$$\begin{aligned}\|\Phi(x^k + p_{PLM}^k)\| &\leq \|\Phi(x^k + p_{PLM}^k) - \Phi(x^k + p_{LM}^k)\| + \|\Phi(x^k + p_{LM}^k)\| \\ &\leq L\|p_{PLM}^k - p_{LM}^k\| + \|\Phi(x^k + p_{LM}^k)\| = o(\|\Phi(x^k)\|),\end{aligned}$$

and this shows that the test in (S.3) is passed by the direction p_{PLM}^k for all $k \in \mathbb{N}$ sufficiently large. □

6 A Projected Filter Trust Region Method

In this section, we present a variant of Algorithm 3.1 by adding a filter technique into our projected Levenberg-Marquardt trust region method. We will show that this method has essentially the same global and local convergence properties as Algorithm 3.1 itself. However, the filter allows a nonmonotone behaviour of the sequence $\{\Psi(x^k)\}$ by accepting the full projected Levenberg-Marquardt step even in some situations where we get no decrease of the merit function Ψ.

Originally, filter methods were proposed in the year 2002 by Fletcher and Leyffer [19] for the solution of constrained optimization problems, see also [18, 22, 39, 41] for some further developments in this direction. Extensions of the filter idea to the solution of nonlinear systems of equations can be found in Fletcher and Leyffer [20] as well as in Gould et al. [22]. Here we adapt the multidimensional filter approach from [22] and incorporate that idea into our method for the solution of problem (1).

More precisely, we simplify the approach from [22] to some extent and present a special case of that filter approach only. This version is tailored to the case where we apply our method to mixed complementarity problems in the following sections.

To describe the filter idea, let $r \in \{1, \ldots, m-1\}$ be any given number, and partition the mapping $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ into

$$\Phi(x) := \begin{pmatrix} \Phi_A(x) \\ \Phi_B(x) \end{pmatrix} \quad \text{with} \quad \Phi_A : \mathbb{R}^n \to \mathbb{R}^r, \;\; \Phi_B : \mathbb{R}^n \to \mathbb{R}^{m-r}.$$

Then define a mapping $\theta : \mathbb{R}^n \to \mathbb{R}^2$ by

$$\theta(x) := \big(\theta_1(x), \theta_2(x)\big)^T := \big(\|\Phi_A(x)\|, \|\Phi_B(x)\|\big)^T.$$

We say that a vector $x \in \mathbb{R}^n$ *dominates* another vector $y \in \mathbb{R}^n$ if $\theta_i(x) \le \theta_i(y)$ for both $i = 1$ and $i = 2$. Now suppose that we are at the kth iteration of a suitable method and that we have generated certain iterates $x^0, x^1, \ldots, x^k$. Then a *filter* $\mathcal{F}_k$ at the kth iteration is a subset

$$\mathcal{F}_k \subseteq \big\{\theta(x^0), \theta(x^1), \ldots, \theta(x^k)\big\}$$

such that none of the elements $\theta(x^l) \in \mathcal{F}_k$ dominates another element from the set $\mathcal{F}_k$.

Assume that a filter $\mathcal{F}_k$ is given, and that we have computed a new vector y (which we hope to become x^{k+1}). The question is when $\theta(y)$ becomes an element of the new filter $\mathcal{F}_{k+1}$. A straightforward idea would be to add $\theta(y)$ to the old filter $\mathcal{F}_k$ if $\theta(y)$ is not dominated by any element from $\mathcal{F}_k$. However, this notion is not strong enough in order to prove suitable convergence results.

Following [22], we therefore call y *acceptable* for the filter $\mathcal{F}_k$ if there is a constant $\gamma_\theta > 0$ such that, for each element $\theta(x^l) \in \mathcal{F}_k$, there is an index $j \in \{1, 2\}$ with

$$\theta_j(y) \le \theta_j(x^l) - \gamma_\theta \|\theta(y)\|.$$

Loosely speaking, this means that $\theta(y)$ is acceptable if, for each element $\theta(x^l) \in \mathcal{F}_k$, the new candidate $\theta(y)$ is sufficiently smaller than $\theta(x^l)$ in at least one of the two

components. In this case we define the new filter by

$$\mathcal{F}_{k+1} := \mathcal{F}_k \cup \{\theta(y)\}$$

and remove all elements from $\mathcal{F}_k$ that are dominated by $\theta(y)$. Moreover, we accept $x^{k+1} := y$ as our new iterate. On the other hand, if y is not acceptable for the filter $\mathcal{F}_k$, we simply set $\mathcal{F}_{k+1} := \mathcal{F}_k$.

Incorporating this filter idea into Algorithm 3.1, we obtain the following method.

Algorithm 6.1 (Scaled Filter Trust Region Method)

(S.0) Choose $x^0 \in \mathcal{B},\ \Delta_0 > 0, 0 < \rho_1 < \rho_2 < 1,\ 0 < \sigma_1 < 1 < \sigma_2,\ \varepsilon \geq 0,\ \eta \in (0,1),\ \Delta_{\min} > 0,\ \alpha \in (0,1], M > 0$, and set $\mathcal{F}_0 := \{\theta(x^0)\}, k := 0$.

(S.1) If $\|D_k g^k\| \leq \varepsilon$: STOP.

(S.2) Choose $H_k \in \partial\Phi(x^k), \nu_k > 0$, and compute p^k_{LM} using (5).

(S.3) Compute p^k_{PLM} from (7). If $x^k + p^k_{PLM}$ is acceptable for the filter $\mathcal{F}_k$ and $\|\Phi(x^k + p^k_{PLM})\| \leq M$, set $x^{k+1} := x^k + p^k_{PLM},\ \Delta_{k+1} := \max\{\Delta_{\min}, \sigma_2\Delta_k\}, \mathcal{F}_{k+1} := \mathcal{F}_k \cup \{\theta(x^{k+1})\}$ (and remove all entries from $\mathcal{F}_k$ that are dominated by $\theta(x^{k+1})$), and go to step (S.7); otherwise set $\mathcal{F}_{k+1} := \mathcal{F}_k$, and go to (S.4).

(S.4) If $\|\Phi(x^k + p^k_{PLM})\| \leq \eta\|\Phi(x^k)\|$ holds, set $x^{k+1} := x^k + p^k_{PLM},\ \Delta_{k+1} := \max\{\Delta_{\min}, \sigma_2\Delta_k\}$, and go to (S.7); otherwise, go to step (S.5).

(S.5) Compute an approximate solution p^k of the trust region subproblem (10) satisfying (13) and (14), and define r_k by (12). If $r_k \geq \rho_1$, we call the iteration k successful and set $x^{k+1} := x^k + p^k$; otherwise we set $x^{k+1} := x^k$.

(S.6) Update the trust region radius as follows:

$$\Delta_{k+1} := \begin{cases} \sigma_1\Delta_k, & \text{if } r_k < \rho_1, \\ \max\{\Delta_{\min}, \Delta_k\}, & \text{if } r_k \in [\rho_1, \rho_2), \\ \max\{\Delta_{\min}, \sigma_2\Delta_k\}, & \text{if } r_k \geq \rho_2. \end{cases}$$

(S.7) Set $k \leftarrow k + 1$, and go to (S.1).

Note that Algorithm 6.1 differs from Algorithm 3.1 only in (S.3) where we added the filter strategy. Furthermore, note that we have a constant M which we assume to be sufficiently large in practice such that the test $\|\Phi(x^k + p^k_{PLM})\| \leq M$ is always satisfied. From a theoretical point of view, however, this constant M is needed and plays the role of a safeguard in order to prevent the sequence $\{\|\Phi(x^k)\|\}$ to become too large. In fact, we have the following simple note.

Remark 6.2 *The sequence $\{x^k\}$ generated by Algorithm 6.1 has the property that $\|\Phi(x^k)\| \leq \max\{\|\Phi(x^0)\|, M\}$ for all $k \in \mathbb{N}$. This can be seen by induction. For $k = 0$, this inequality holds trivially. Hence suppose that it holds for some $k \geq 0$, and consider the iterate x^{k+1}. If this iterate is computed in step (S.3), we have $\|\Phi(x^{k+1})\| \leq M$. Otherwise, we have $\|\Phi(x^{k+1})\| \leq \|\Phi(x^k)\|$, and the statement then follows from the induction hypothesis.*

The following result shows what happens if the new vector $x^k + p^k_{PLM}$ is accepted an infinite number of times by our filter step in (S.3).

Theorem 6.3 *Assume there are infinitely many iterations k such that $x^{k+1} = x^k + p^k_{PLM}$ is accepted in the filter step (S.3) of Algorithm 6.1. Then $\lim_{k\to\infty} \|\Phi(x^k)\| = 0$.*

Proof. Let $K \subseteq \mathbb{N}$ denote the infinite subset such that $x^{k+1} = x^k + p^k_{PLM}$ is accepted for all $k \in K$ in step (S.3) of Algorithm 6.1. In the first part of the proof, we show that

$$\lim_{k\in K} \|\theta(x^{k+1})\| = 0. \tag{34}$$

Suppose this is not true. Then we may assume that there is a infinite subset $\bar{K} \subseteq K$ and a constant $\varepsilon > 0$ such that

$$\|\theta(x^{k+1})\| \geq \varepsilon \quad \forall k \in \bar{K}. \tag{35}$$

In view of step (S.3) of Algorithm 6.1, the sequence $\{\theta(x^{k+1})\}_{k\in\bar{K}}$ is bounded. Hence there is another subset $\hat{K} \subseteq \bar{K}$ with

$$\lim_{k\in\hat{K}} \theta(x^{k+1}) = \theta^* \tag{36}$$

for some number $\theta^* \in \mathbb{R}^2$ satisfying $\|\theta^*\| \geq \varepsilon$.

For the moment, consider a fixed index $k \in \hat{K}$. Furthermore, let $l_k \in \hat{K}$ denote the index in $\hat{K}$ previous to k. Since x^{k+1} was acceptable for the filter $\mathcal{F}_k$, we have

$$\theta_j(x^{k+1}) \leq \theta_j(x^{l_k+1}) - \gamma_\theta \|\theta(x^{k+1})\| \tag{37}$$

for at least one index $j \in \{1, 2\}$. This statement does not depend on $\theta(x^{l_k+1})$ still being in the filter $\mathcal{F}_k$. Indeed, if $\theta(x^{l_k+1}) \notin \mathcal{F}_k$, it must be dominated by an entry $\theta(x^l)$ in the filter $\mathcal{F}_k$. Since $x^{k+1}, k \in \hat{K}$, is acceptable for $\theta(x^l) \in \mathcal{F}_k$, there is an index $j \in \{1, 2\}$ such that $\theta_j(x^{k+1}) \leq \theta_j(x^l) - \gamma_\theta\|\theta(x^{k+1})\| \leq \theta_j(x^{l_k+1}) - \gamma_\theta\|\theta(x^{k+1})\|$, where the last inequality holds since $\theta(x^l)$ dominates $\theta(x^{l_k+1})$. Hence (37) holds. Together with (35), we obtain $\theta_j(x^{k+1}) - \theta_j(x^{l_k+1}) \leq -\gamma_\theta\varepsilon$ for at least one index $j \in \{1, 2\}$. However, the left-hand side converges to zero for at least one index $j \in \{1, 2\}$ because of (36). This contradiction shows that (34) holds.

As an immediate consequence of (34), we also obtain $\lim_{k\in K} \|\Phi(x^{k+1})\| = 0$. However, for all $k \notin K$, we have $\|\Phi(x^{k+1})\| \leq \|\Phi(x^k)\|$. Hence we obtain $\lim_{k\to\infty} \|\Phi(x^k)\| = 0$, and this completes the proof. □

Theorem 6.3 shows that every accumulation point of a sequence generated by Algorithm 6.1 is actually a solution of (1) and not just a stationary point of (2), provided the filter is accepted an infinite number of times. Hence we get a very strong global convergence result in this case. Moreover, it is easy to see that all statements of the local convergence result from Theorem 5.3 remain true in this case.

On the other hand, if the filter is accepted only a finite number of times, then Algorithm 6.1 eventually reduces to Algorithm 3.1, and in this case Algorithm 6.1 has precisely the same convergence properties of Algorithm 3.1 as described in the previous sections.

7 Application to Mixed Complementarity Problems

In this section, we present the reformulation of the mixed complementarity problem that fits into the framework (1), (2) discussed in the previous sections. We also note that the assumptions for local or global convergence hold under appropriate conditions. The reformulation we use here is taken from [28] except that we add the bound constraints explicitly in the present paper.

In order to introduce the mixed complementarity problem, it is convenient to consider the variational inequality problem first, see [11] for more details. Given a function $F : \mathbb{R}^n \to \mathbb{R}^n$ and a nonempty, closed and convex set $X \subseteq \mathbb{R}^n$, the variational inequality problem consists in finding a point $x^* \in X$ such that $F(x^*)^T(x - x^*) \geq 0$ for all $x \in X$. If the feasible set X is a box of the form $X = [l, u]$ with lower bounds $l = (l_1, \ldots, l_n)^T$ and upper bounds $u = (u_1, \ldots, u_n)^T$ satisfying $-\infty \leq l_i < u_i \leq +\infty$ for all $i \in \{1, \ldots, n\}$, we obtain the mixed complementarity problem.

In order to present a reformulation of this mixed complementarity problem, let us introduce the following partition of the index set $I := \{1, \ldots, n\}$:

$$\begin{aligned} I_l &:= \{i \in I \mid -\infty < l_i < u_i = \infty\}, \\ I_u &:= \{i \in I \mid -\infty = l_i < u_i < \infty\}, \\ I_{lu} &:= \{i \in I \mid -\infty < l_i < u_i < \infty\}, \\ I_f &:= \{i \in I \mid -\infty = l_i < u_i = \infty\}. \end{aligned}$$

Furthermore, let $\phi : \mathbb{R}^2 \to \mathbb{R}$ denote the Fischer-Burmeister function

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

which has the interesting property that

$$\phi(a, b) = 0 \iff a \geq 0, b \geq 0, ab = 0,$$

see [16]. We now define the operator $\Phi : \mathbb{R}^n \to \mathbb{R}^{2n}$ componentwise as follows ($i =$

$1, \ldots, n$):

$$\Phi_i(x) := \begin{cases} \lambda\phi(x_i - l_i, F_i(x)) & \text{if } i \in I_l, \\ -\lambda\phi(u_i - x_i, -F_i(x)) & \text{if } i \in I_u, \\ \lambda\phi(x_i - l_i, \phi(u_i - x_i, -F_i(x))) & \text{if } i \in I_{lu}, \\ -\lambda F_i(x) & \text{if } i \in I_f, \end{cases}$$

$$\Phi_{n+i}(x) := \begin{cases} (1-\lambda)\phi_+(x_i - l_i, F_i(x)) & \text{if } i \in I_l, \\ (1-\lambda)\phi_+(u_i - x_i, -F_i(x)) & \text{if } i \in I_u, \\ (1-\lambda)(\phi_+(x_i - l_i, F_i(x)) + \phi_+(u_i - x_i, -F_i(x))) & \text{if } i \in I_{lu}, \\ -(1-\lambda)F_i(x) & \text{if } i \in I_f. \end{cases}$$

Here, $\lambda \in (0, 1]$ is a given parameter, and ϕ_+ denotes the mapping $\phi_+(a, b) := \max\{0, a\}\max\{0, b\}$. It was noted in [28] that the overdetermined system of equations $\Phi(x) = 0$ is equivalent to the mixed complementarity problem. Obviously, the same holds for the box constrained reformulation $\Phi(x) = 0, x \in [l, u]$, which we prefer here because this avoids some problems like mappings F which are not defined outside the feasible region $X = [l, u]$ or possible stationary points outside this set. Hence we arrive at a problem of the form (1). Moreover, it was noted in [28] that the corresponding merit function $\Psi(x) := \frac{1}{2}\|\Phi(x)\|^2$ is continuously differentiable.

Furthermore, Φ is semismooth if F is continuously differentiable, and strongly semismooth if, in addition, F' is locally Lipschitzian. The full rank assumption for all elements in the generalized Jacobian $\partial\Phi(x^*)$ at a solution x^* of the mixed complementarity is satisfied under a condition that is called R-regularity in [28]. Finally, we note that, in a very natural way, the definition of Φ leads to the partition $\Phi = (\Phi_A, \Phi_B)$, where Φ_A denotes the first n components and Φ_B the last n components of Φ. Hence we can apply our filter technique from the previous section with $r := n$ to this reformulation of mixed complementarity problems.

We stress, however, that there are other reformulations of the (mixed) complementarity problem that also satisfy the assumptions of this paper. For example, ϕ does not have to be the Fischer-Burmeister function. Other mappings are also possible, see, e.g., [3, 14, 26, 37]. Moreover, it is also possible to reformulate general variational inequality problems as a nonsmooth system of equations (via the corresponding KKT-conditions, see [11]) in such a way that all assumptions of this paper hold.

8 Numerical Results

We implemented Algorithm 6.1 in MATLAB and tested the algorithm on a number of mixed complementarity problems from the MCPLIB collection (see [9]) using the reformulation from the previous section.

Preliminary numerical experiments showed that the local method from Algorithm 2.1 behaves extremely good. We therefore decided to use the local method as a preprocessor before starting the main algorithm. More precisely, we first allow at most

20 iterations of the local method, and then switch to the globalized trust-region filter method from Algorithm 6.1. Moreover, we follow an idea by Ulbrich [38] and compute x^{k+1} in a slightly different way, replacing the formula

$$x^{k+1} = x^k + p^k_{PLM} = P_{\mathcal{B}}(x^k + p^k_{LM}) = x^k + P_{\mathcal{B}-x^k}(p^k_{LM})$$

from (6) by $x^{k+1} = x^k + P_{X_k}(p^k_{LM})$, where X_k denotes the set from (11). Note that, locally, this does not change anything since $p^k_{LM} \to 0$ whereas the trust-region radius $\Delta_k \geq \Delta_{\min}$ is bounded away from zero. In particular, neither the global nor the local convergence theory is affected by this modification.

We next describe the initialization of our method: The starting point x^0 is the one from the MCPLIB collection. It always belongs to the box $\mathcal{B}$, so there is no need to project it onto the feasible set. The Levenberg-Marquardt parameter ν_k is chosen as follows: For smaller problems with $n < 100$, we first estimate the condition number of the matrix $H_k^T H_k$. If this estimated condition number is larger than 10^{25}, we set $\nu_k := 10^{-6}/(k+1)$, otherwise we set $\nu_k := 10^{-16}$. In all other cases, we take $\nu_k := 0$. (Note the the condition estimator becomes expensive for larger problems, so we do not use it for problems with $n \geq 100$.) We terminate our iteration if one of the following conditions hold:

$$\Psi(x^k) \leq 10^{-10} \quad \text{or} \quad \|\hat{g}(x^k)\| \leq 10^{-6} \quad \text{or} \quad k > 500 \quad \text{or} \quad \Delta_k \leq 10^{-12}.$$

The remaining parameters used by our method are $\lambda = 0.1, \alpha = 10^{-4}, \rho_1 = 10^{-4}, \rho_2 = 0.75, \sigma_1 = 0.5, \sigma_2 = 2, \Delta_0 = 10$, and $\Delta_{\min} = 10^{-6}$.

If the preprocessor is not able to solve a problem or if $\|p^k_{PLM}\| \leq 10^{-12}$, we switch to Algorithm 6.1 starting with the best point computed so far. We then test whether our projected step p^k_{PLM} is acceptable for the current filter or satisfies the descent condition in (S.4). If this is not the case and p^k_{PLM} also fails to satisfy the fraction of Cauchy decrease condition, we compute p^k in (S.5) by solving the trust-region subproblem (10) exactly. Here, the QP-solver MINQ from Neumaier [32] is used. This is a MATLAB program for bound constrained quadratic programs, and we allow at most n inner iterations for each call of this QP-solver. Unfortunately, we sometimes do not succeed in solving the trust-region subproblem even with a higher number of inner QP-iterations. The number of errors produced by MINQ grows with the dimension n of the mixed complementarity problem. For this reason, we exclude from our test all problems of the MCPLIB with size $n > 160$. Alternatively, we could compute an approximate solution of the trust-region subproblem like a Cauchy step or a simple dogleg step, however, according to our numerical tests, it is better to solve (or try to solve) the QP-subproblem exactly.

Our numerical results are summarized in Table I. In this table the first column gives the name of the problem; i_{tot} gives the total number of outer iterations (adding the iteration numbers from the preprocessor and the main algorithm). The entry '–' is used to indicate that the algorithm terminated unsuccessfully; $\Psi(x^f)$ and $\|\hat{g}(x^f)\|$ denote the values of $\Psi(x)$ and $\|\hat{g}(x)\|$ at the final iterate $x = x^f$; i_{fil} gives the number

of filter steps taken and i_{des} the number of descent steps satisfying the criterion in (S.4) of Algorithm 6.1. The remaining two columns contain nonzero numbers only if we solve our QP-subproblem using MINQ. The entries of column i_{TR} report the number of successful (left) and the number of unsuccessful trust-region steps (right); moreover, column i_{QP} gives the number of QP-problems that were solved successfully (left) and unsuccessfully (right). In the latter case, we do not stop our iteration if the final approximate solution provided by MINQ satisfies the fraction of Cauchy decrease condition.

Table I: Numerical results for MCPLIB test problems

Problem	i_{tot}	$\Psi(x^f)$	$\|\hat{g}(x^f)\|$	i_{fil}	i_{des}	i_{TR}		i_{QP}	
						succ	unsucc	solv	unsolv
badfree	4	1.589642e-14	3.088351e-08	0	0	0	0	0	0
bertsekas	11	1.276482e-11	3.156252e-05	0	0	0	0	0	0
billups	–	2.000000e-06	0.000000e+00	0	0	0	0	0	0
choi	5	2.649619e-16	4.405966e-10	0	0	0	0	0	0
colvdual	14	9.073964e-11	9.450997e-05	0	0	0	0	0	0
colvnlp	6	4.885855e-16	4.901420e-08	0	0	0	0	0	0
cycle	4	8.921959e-12	4.224224e-07	0	0	0	0	0	0
degen	4	3.151895e-17	1.122835e-09	0	0	0	0	0	0
duopoly	–	5.163723e+00	4.977950e-07	14	0	22	5	14	13
ehl_k40	12	1.511446e-13	1.842138e-04	0	0	0	0	0	0
ehl_k60	15	3.103870e-11	6.406644e-04	0	0	0	0	0	0
ehl_k80	14	9.474365e-13	3.872440e-03	0	0	0	0	0	0
ehl_kost	17	5.651902e-12	1.512516e-02	0	0	0	0	0	0
electric	52	1.745588e-11	1.238040e-02	23	0	4	5	4	0
explcp	4	7.407629e-14	3.849074e-08	0	0	0	0	0	0
freebert	11	4.545773e-11	5.974924e-05	0	0	0	0	0	0
gafni	10	6.420657e-13	9.795657e-06	0	0	0	0	0	0
games	13	4.384397e-13	1.618921e-05	0	0	0	0	0	0
hanskoop	14	1.231700e-11	3.573738e-06	0	0	0	0	0	0
hydroc06	7	5.792347e-19	1.566474e-09	0	0	0	0	0	0
hydroc20	10	2.322843e-16	5.090809e-05	0	0	0	0	0	0
jel	8	3.651083e-18	1.601171e-08	0	0	0	0	0	0
josephy	2	2.989144e-11	6.436729e-05	0	0	0	0	0	0
kojshin	2	3.004186e-11	6.452607e-05	0	0	0	0	0	0
mathinum	4	3.024771e-12	6.673325e-07	0	0	0	0	0	0
mathisum	8	2.199503e-16	1.559264e-08	0	0	0	0	0	0
methan08	4	6.252855e-13	2.274457e-02	0	0	0	0	0	0
nash	4	2.354633e-19	7.457045e-09	0	0	0	0	0	0

Table I: Numerical results for MCPLIB test problems (continued)

Problem	i_{tot}	$\Psi(x^f)$	$\|\hat{g}(x^f)\|$	i_{fil}	i_{des}	$i_{\text{ TR}}$		$i_{\text{ QP}}$	
						succ	unsucc	solv	unsolv
ne-hard	20	5.337625e-11	1.323158e-04	0	0	0	0	0	0
pgvon106	69	1.387117e-12	1.325755e-01	4	0	11	34	42	3
pies	29	1.739675e-13	1.328305e-03	8	0	0	1	0	0
powell	4	5.659114e-11	7.740843e-06	0	0	0	0	0	0
powell_mcp	2	2.728284e-13	6.048681e-06	0	0	0	0	0	0
qp	2	1.603357e-31	1.025472e-15	0	0	0	0	0	0
scarfanum	3	3.605079e-12	1.796300e-06	0	0	0	0	0	0
scarfasum	3	3.604872e-12	2.255827e-06	0	0	0	0	0	0
scarfbsum	120	1.507297e-11	9.669907e-04	7	1	57	35	89	2
shubik	–	1.407769e-07	4.303409e-03	74	8	161	237	290	91
simple-ex	25	1.079537e-13	4.142756e-07	5	0	0	0	0	0
simple-red	10	2.173645e-11	5.866415e-06	0	0	0	0	0	0
sppe	3	4.251553e-11	1.849914e-04	0	0	0	0	0	0
tinloi	6	7.639515e-12	7.292264e-04	0	0	0	0	0	0
tobin	2	1.474605e-14	1.352957e-05	0	0	0	0	0	0

Table I shows that our method was able to solve the majority of all test examples. Most of them were solved in less than 20 iterations and, therefore, by our preprocessor which turns out to be very effective. We have failures only on three problems, namely `billups`, `duopoly`, and `shubik`. For `billups` and `duopoly`, we terminate with a stationary point, whereas the function value in the final iteration of `shubik` is very small, but does not satisfy our termination criterion. We also stress that we have a relatively high number of unsolved QP-subproblems for the two examples `duopoly` and `shubik` which might be the reason for the failure of the overall algorithm. However, we also tried some other QP-solvers, but, in general, MINQ seems to be a good choice for bound constrained quadratic programs. We should mention that we changed example `pgvon106` slightly by adding a small number to the lower bounds. This prevents difficulties in computing the function value $F(x)$ when x is close to the lower bounds.

We finally stress that both the preprocessor and the filter technique improved the overall behaviour of our method significantly. To illustrate this point, we present in Table II the number of failures on the examples from Table I for all possible combinations (with or without preprocessor, and with or without filter). The first line in Table II corresponds to the results for the method presented in Table I. Despite the fact that this method has the smallest number of failures, we also note that the filter technique was sometimes very helpful in decreasing the number of iterations. For example, the method without preprocessor and without filter takes 20 iterations to solve problem

Table II: Number of failures on MCPLIB problems using different options

method	# of failures
with preprocessor, with filter	3
with preprocessor, no filter	5
no preprocessor, with filter	9
no preprocessor, no filter	13

`colvdual`, whereas the same method, but with the filter technique switched on, only needs 13 iterations.

9 Final Remarks

We have presented a filter trust-region method for the solution of semismooth least squares problems with box constraints and applied this method to a corresponding reformulation of mixed complementarity problems that was recently introduced in [28]. The method has similar convergence properties as the one from [28], but it generated feasible iterates in contrast to the method from [28]. In particular, the projected local method used in this paper seems to be very efficient and surprisingly robust, and many of the test examples from the MCPLIB can be solved successfully in this way. Some problems arise as soon as we have to find a solution of our QP-subproblems. In particular, solving these QP-subproblems is time-consuming, so we do not suggest to apply this approach to large-scale problems.

However, motivated by the success of our preprocessor, i.e., the local projected Levenberg-Marquardt method from Algorithm 2.1, we also tested a combination of the two approaches from [28] and this paper in order to see whether it is possible to get a method that is both efficient and robust for small- and large-dimensional problems. In fact, using this preprocessor for the first 20 iterations and then switching to the unconstrained method from [28], we are able to solve all test problems (including the larger ones) from the MCPLIB collection with the only exception of problems `duopoly` and `bishop`. This is even better than the results presented in Section 8, and certainly a significant improvement over other existing methods, possibly with the exception of the PATH solver (see [8]) whose newer versions, however, use restarts and several other techniques so that a direct comparison is not possible. An implementation of this method is available from the homepages of the authors.

Acknowledgement. The authors would like to thank the two referees for their very careful reading and useful comments on the first version of this paper.

References

[1] Billups, S.C., 1995, Algorithms for Complementarity Problems and Generalized Equations. PhD thesis, Computer Sciences Department, University of Wisconsin, Madison.

[2] Björck, Å., 1996, *Numerical Methods for Least Squares Problems.* SIAM, Philadelphia, PA.

[3] Chen, B., Chen, X., and Kanzow, C., 2000, A penalized Fischer-Burmeister NCP-function. *Math. Program.*, **88**, 211–216.

[4] Clarke, F.H., 1983, *Optimization and Nonsmooth Analysis.* John Wiley & Sons, New York (reprinted by SIAM, Philadelphia, 1990).

[5] Coleman, T.F. and Li, Y., 1996, An interior trust region approach for nonlinear minimization subject to bounds. *SIAM J. Optim.*, **6**, 418–445.

[6] Conn, A.R., Gould, N.I.M. and Toint, Ph.L., 2000, *Trust-Region Methods.* SIAM, Philadelphia, PA.

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

[8] Dirkse, S.P. and Ferris, M.C., 1995: The PATH solver: A non-monotone stabilization scheme for mixed complementarity problems. *Optim. Meth. Soft.*, **5**, 123–156.

[9] Dirkse, S.P. and Ferris, M.C., 1995: MCPLIB: A collection of nonlinear mixed complementarity problems. *Optim. Meth. Soft.*, **5**, 319–345.

[10] Dennis, J.E. and Schnabel, R.B., 1983, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations.* Prentice Hall, Englewood Cliffs, NJ.

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

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

[13] Facchinei, F. and Soares, J., 1997, A new merit function for nonlinear complementarity problems and a related algorithm. *SIAM J. Optim.*, **7**, 225–247.

[14] Ferris, M.C., Kanzow, C. and Munson, T.S., 1999, Feasible descent algorithms for mixed complementarity problems. *Math. Program.*, **86**, 475–497.

[15] Ferris, M.C. and Pang, J.-S., 1997, Engineering and economic applications of complementarity problems. *SIAM Rev.*, **39**, 669–713.

[16] Fischer, A., 1992, A special Newton-type optimization method. *Optim.*, **24**, 269–284.

[17] Fischer, A., 1997, Solution of monotone complementarity problems with locally Lipschitzian functions. *Math. Program.*, **76**, 513–532.

[18] Fletcher, R., Gould, N.I.M., Leyffer, S., Toint, Ph.L. and Wächter, A., 2002, Global convergence of trust-region SQP-filter algorithms for nonlinear programming. *SIAM J. Optim.*, **13**, 635–659.

[19] Fletcher, R. and Leyffer, S., 2002, Nonlinear programming without a penalty function. *Math. Program.*, **91**, 239–269.

[20] Fletcher, R. and Leyffer, S., 2003, Filter-type algorithms for solving systems of algebraic equations and inequalities. In: G. Di Pillo and A. Murli (Eds) *High-Performance Algorithms and Software in Nonlinear Optimization*. (Kluwer, Dordrecht, The Netherlands), pp. 259–278.

[21] Fletcher, R., Leyffer, S. and Toint, Ph.L., 2002, On the global convergence of a filter-SQP algorithm. *SIAM J. Optim.*, **13**, 44–59.

[22] Gould, N.I.M., Leyffer, S. and Toint, Ph.L., 2004, A multidimensional filter algorithm for nonlinear equations and nonlinear least squares. *SIAM J. Optim.*, **15**, 17–38.

[23] Jiang, H., 1999, Global convergence analysis of the generalized Newton and Gauss-Newton methods of the Fischer-Burmeister equation for the complementarity problem. *Math. Oper. Res.*, **24**, 529–543.

[24] Jiang, H., Fukushima, M., Qi, L. and Sun, D., 1998: A trust region method for solving generalized complementarity problems. *SIAM J. Optim.*, **8**, 140–157.

[25] Kanzow, C., 2001, Strictly feasible equation-based methods for mixed complementarity problems. *Numer. Math.*, **89**, 135–160.

[26] Kanzow, C. and Kleinmichel, H., 1998, A new class of semismooth Newton-type methods for nonlinear complementarity problems. *Comput. Optim. Appl.*, **11**, 227–251.

[27] Kanzow, C. and Klug, A., 2006, On affine-scaling interior-point Newton methods for nonlinear minimization with bound constraints. *Comput. Optim. Appl.*, **35**, 177–197.

[28] Kanzow, C. and Petra, S., 2004, On a semismooth least squares formulation of complementarity problems with gap reduction. *Optim. Meth. Soft.*, **19**, 507–525.

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

[30] Kanzow, C. and Zupke, M., 1999, Inexact trust-region methods for nonlinear complementarity problems. In: M. Fukushima and L. Qi (Eds), *Reformulation - Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods* (Kluwer Academic Press, Dordrecht, The Netherlands), pp. 211–233.

[31] Moré, J. and Sorensen, D.C., 1983, Computing a trust region step. *SIAM J. Sci. Stat. Comput.*, **4**, 553–572.

[32] Neumaier, A., 1998, MINQ – General definite and bound constrained indefinite quadratic programming. Availabe online at: www.mat.univie.ac.at/˜neum/software/minq/

[33] Pang, J.-S. and Qi, L., 1993, Nonsmooth equations: motivation and algorithms. *SIAM J. Optim.*, **3**, 443–465.

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

[35] Qi, L. and Sun, J., 1993, A nonsmooth version of Newton's method. *Math. Program.*, **58**, 353–367.

[36] Robinson, S.M., 1980, Strongly regular generalized equations. *Math. Oper. Res.*, **5**, 43–62.

[37] Sun, D. and Qi, L., 1999, On NCP-functions. *Comput. Optim. Appl.*, **13**, 201–230.

[38] Ulbrich, M., 2001, Nonmonotone trust-region methods for bound-constrained semismooth equations with applications to nonlinear mixed complementarity problems. *SIAM J. Optim.*, **11**, 889–916.

[39] Ulbrich, S., 2004, On the superlinear local convergence of a filter-SQP method. *Math. Program.*, **100**, 217–245.

[40] Ulbrich, M., Ulbrich, S. and Heinkenschloss, M., 1999, Global convergence of trust-region interior-point algorithms for infinite-dimensional nonconvex minimization subject to pointwise bounds. *SIAM J. Control Optim.*, **37**, 731–764.

[41] Ulbrich, M., Ulbrich, S. and Vicente, L.N., 2004, A globally convergent primal-dual interior-point filter method for nonlinear programming. *Math. Program.*, **100**, 379–410.