Hybrid Newton-type method for a class of semismooth equations

Terms of use:

Publisher copyright

(Article begins on next page)

13 May 2024

Hybrid Newton-Type Method for a Class of Semismooth Equations *

S. Pieraccini[†]

Abstract

In this paper, we present a hybrid method for the solution of a class of composite semismooth equations encountered frequently in applications. The method is obtained by combining a generalized finite difference Newton method to a inexpensive direct search method. We prove that, under standard assumptions, the method is globally convergent with a local rate of convergence which is superlinear or quadratic. We also report several numerical results obtained applying the method to suitable reformulations of well known nonlinear complementarity problems.

Keywords: Nonsmooth systems, generalized Newton methods, hybrid methods, global convergence.

1 Introduction

Let us consider the problem of solving the system of nonlinear equations

$$H(x) = 0 \tag{1}$$

where $H : \mathbb{R}^n \mapsto \mathbb{R}^n$ is assumed to be a composite semismooth function of the form $H(x) = (\Phi \circ G)(x)$, being $G : \mathbb{R}^n \mapsto \mathbb{R}^m$ a smooth function and $\Phi : \mathbb{R}^m \mapsto \mathbb{R}^n$ a semismooth one with a simpler structure with respect to H. Alternatively, the case $H(x) = (E \circ \Psi)(x)$ could be considered, with $\Psi : \mathbb{R}^n \mapsto \mathbb{R}^m$ semismooth and $E : \mathbb{R}^m \mapsto \mathbb{R}^n$ smooth. The analysis of the latter case is almost the same of the former, with a few minor modifications, so it is omitted here.

Problems of the considered class are frequent in applications, for example they arise from proper reformulations of nonlinear complementarity problems and of the Karush-Kuhn-Tucker conditions for variational inequalities problems, or from the discretization of some nondifferentiable Dirichlet problems (Refs. 1-4).

A general approach to the solution of semismooth systems, introduced in Refs. 5-6, consists in using generalized Newton methods, in which the scheme is the same as the classical Newton method, but the Jacobian matrix is replaced by the Clarke generalized Jacobian or other proper generalizations. While such generalized Jacobians can be easily computed on simple academic test problems, this is not true for problems coming from real applications. For composite semismooth functions some recent papers suggested to use mixed strategies easier to apply to the latter problems. For example, the idea of applying different techniques to the "smooth part" and to the "nonsmooth part" was recently considered by some authors (see Ref. 2 and references therein); in particular, in Ref. 4 a secant method is proposed using difference approximations of the Jacobian of the smooth component G and generalized Jacobians of the nonsmooth component Φ.

Most of the previous methods, under proper assumptions, were proved to be locally convergent with superlinear or quadratic rate of convergence. In recent years, several authors have proposed globally convergent modifications of generalized Newton methods for some classes of nonsmooth problems. The global convergence is usually obtained by using some linesearch strategy (Refs. 6-7). More robust global algorithms are obtained combining some generalizations of Newton method with the steepest descent method for a smooth merit function (Refs. 8-9). We remark that the availability of a smooth merit function is an essential factor in the analysis and design of most global methods. In fact, they are applied to nonsmooth equations, arising from nonlinear complementarity and related problems, in which $H(x)$ is semismooth and the natural merit function

$$\theta(x) = \frac{1}{2} H(x)^T H(x)$$

is continuously differentiable.

Numerical results in literature show that, as in the smooth case, the hybrid approach is very promising also for this class of nonsmooth systems. Here we propose a hybrid method for composite semismooth equations with smooth

natural merit function: it combines a fast local Newton-type method with line-search (the basic method) with a cheap global direct search method for the minimization of $\theta(x)$ (the auxiliary method). The idea is the following: if the basic method fails to find a new iterate within a reasonable computational effort, we suppose that the used search direction or the local model is not adequate and we turn to the auxiliary method. The basic method exploits the suggestion in Ref. 4, using finite difference approximations of the Jacobian matrices of G; the auxiliary method is a derivative-free method which only needs to evaluate the objective function at suitable points around the iterates. The discretization steplengths in the basic method are chosen in such a way that all the computed values of G are reused by the auxiliary method if a switch occurs. In this way, turning to the auxiliary method results to be an inexpensive tool to avoid breakdown or stalling. The combination of finite difference Newton methods and direct search methods appeared to give rise to very robust schemes in the smooth case (Refs. 10-11) and this suggested to use a similar approach also in the nonsmooth case.

The hybrid method is described in Section 2 and it is proved to be globally convergent under standard assumptions, with superlinear or quadratic rate of convergence. The application of the proposed method to nonlinear complementarity problems is shown in Section 3, where several numerical experiments on well known test problems from literature are reported. From the results we obtained, the method seems to be robust and reliable, while some suggestions are deduced to further improve its performance. Section 4 contains some final comments and perspectives.

Throughout the paper $\|\cdot\|$ will represent the 2-norm of a vector or matrix. Further, a continuously differentiable function F will be said to be a C^1-function and $F'(x)$ will denote its Jacobian evaluated at x; if, moreover, F' is locally Lipschitzian, F will be said to be a LC^1-function. Finally, the gradient of a smooth real function $f : \mathbb{R}^n \mapsto \mathbb{R}$ evaluated at x will be denoted by $\nabla f(x)$.

Below we report some basic definitions and results from nonsmooth analysis (Refs. 5-6, 12).

If $H : \mathbb{R}^n \mapsto \mathbb{R}^n$ is a locally Lipschitzian function, according to Rademacher's theorem H is differentiable almost everywhere. Let D_H be the set of points at which H is differentiable. The B-subdifferential of H at x is defined as

$$\partial_B H(x) = \{\lim_{x_k \to x} H'(x_k) : x_k \in D_H\}$$

and the Clarke generalized Jacobian of H at x is

$$\partial H(x) = \mathrm{co}\partial_B H(x).$$

Definition 1.1 *$H : \mathbb{R}^n \mapsto \mathbb{R}^n$ is said to be semismooth at x if it is locally Lipschitzian and the limit*

3

$$\lim_{V \in \partial H(x+td'), d' \to d, t \downarrow 0} \{Vd'\}$$

exists for any $d \in \mathbb{R}^n$.

Proposition 1.1 *If* $H : \mathbb{R}^n \mapsto \mathbb{R}^n$ *is semismooth at* x *then it is directionally differentiable at* x *and for any* $d \to 0$ *and* $W \in \partial H(x+d)$ *we have*

$$Wd - H'(x;d) = o(\|d\|),$$

where $H'(x;d)$ *denotes the directional derivative of* H *at* x *in the direction* d, *and*

$$H(x+d) - H(x) - Wd = o(\|d\|).$$

Definition 1.2 $H : \mathbb{R}^n \mapsto \mathbb{R}^n$ *is said to be strongly semismooth at* x *if it is semismooth at* x *and if*

$$Wd - H'(x;d) = O(\|d\|^2)$$

for any $d \to 0$ *and* $W \in \partial H(x+d)$.

Proposition 1.2 *If* $H : \mathbb{R}^n \mapsto \mathbb{R}^n$ *is strongly semismooth at* x *then*

$$H(x+d) - H(x) - Wd = O(\|d\|^2)$$

for any $d \to 0$ *and* $W \in \partial H(x+d)$.

2 Hybrid Method

The generalized Newton method for solving (1) can be described as follows: given $x_k \in \mathbb{R}^n$, compute x_{k+1} by

$$x_{k+1} = x_k + d_k, \quad d_k = -W_k^{-1} H(x_k)$$

where $W_k \in \partial H(x_k)$, as in [5], or $W_k \in \partial_B H(x_k)$, as in [6].

It is known that the computation of iteration matrices in $\partial H(x_k)$ or $\partial_B H(x_k)$ is, in general, a not easy task for problems coming from real applications and moreover it can be very expensive, even for composite functions. As suggested in [4], for $x \in \mathbb{R}^n$, we consider the pseudo Jacobian $J(x)$ of H given by

$$J(x) = \{W \in \mathbb{R}^{n \times n} : W = UG'(x), U \in \partial_B \Phi(y)|_{y=G(x)}\}.$$

We suppose that some elements of $\partial_B \Phi(y)$ are actually computable at a low cost and we approximate $G'(x)$ by finite difference. Given a discretization steplength $h \in \mathbb{R}$, $h \neq 0$, we use the finite difference approximation $G_h(x)$ of $G'(x)$ defined as follows:

4

$$G_h(x)e_j = (G(x + he_j) - G(x))/h, \quad j = 1, ..., n, \tag{2}$$

where e_j is the j-th coordinate vector in $\mathbb{R}^n$. We remark that (2) corresponds to forward or backward differences according to the sign of h. By using (2), a finite difference pseudo Jacobian can be defined in a natural way as the set

$$J_h(x) = \{W \in \mathbb{R}^{n \times n} : W = UG_h(x), U \in \partial_B \Phi(y)|_{y=G(x)}\}. \tag{3}$$

The following Lemma states conditions ensuring the nonsingularity of the matrices in $J_h(x)$. This Lemma slightly generalizes Lemma 5.2 in [4], where h is assumed equal to $\varepsilon \|H(x)\|$ for some positive constant ε; since in [4] the proof is omitted, here we give the complete proof to the reader's convenience.

Lemma 2.1 *Let $H = \Phi \circ G$, with Φ locally Lipschitzian and G a LC^1-function. Let $x^* \in \mathbb{R}^n$ be such that all the matrices $W_* \in J(x^*)$ are nonsingular. Then there exist a neighborhood N^* of x^* and two constants $\hat{h}, C > 0$ such that all the matrices $W \in J_h(x)$, with $x \in N^*$ and $|h| < \hat{h}$, are nonsingular and $\|W^{-1}\| \leq C$.*

Proof. Since $\partial_B \Phi(\cdot)$ is upper semicontinuous, i.e. for each $\varepsilon > 0$ there exists $\delta_1 > 0$ such that if $\|y_1 - y_2\| < \delta_1$ then $\partial_B \Phi(y_1) \subset \partial_B \Phi(y_2) + \varepsilon B$, where B is the open unit ball in $\mathbb{R}^{m \times n}$ ([12]), it follows that for every $U_1 \in \partial_B \Phi(y_1)$ there exists $U_2 \in \partial_B \Phi(y_2)$ such that $\|U_1 - U_2\| < \varepsilon$. Besides, due to the continuity of G, there exists $\hat{\delta}_1$ such that if $\|x - x^*\| < \hat{\delta}_1$ then $\|G(x) - G(x^*)\| < \delta_1$. Finally, since G' is Lipschitzian, it is easily proved that for each $\varepsilon > 0$ there exist $\delta_2 > 0$ and $\hat{h} > 0$ such that if $\|x - x^*\| < \delta_2$ and $|h| < \hat{h}$, then $\|G'(x^*) - G_h(x)\| < \varepsilon$.

Let $\varepsilon > 0$ be given and x such that $\|x - x^*\| < \delta$, with $\delta = \min(\hat{\delta}_1, \delta_2)$. Let us take any $W = UG_h(x)$, with $|h| < \hat{h}$ and $U \in \partial_B \Phi(y)|_{y=G(x)}$. Corresponding to U there exists $U_* \in \partial_B \Phi(y)|_{y=G(x^*)}$ such that $\|U - U_*\| < \varepsilon$; let us pose $W_* = U_* G'(x^*)$ and consider the quantity $\|W_* - W\|$. We have

$$
\begin{aligned}
\|W_* - W\| &= \|U_* G'(x^*) - UG'(x^*) + UG'(x^*) - UG_h(x)\| \leq \\
&\leq \|U_* - U\| \|G'(x^*)\| + (\|U - U_*\| + \|U_*\|) \|G'(x^*) - G_h(x)\| \leq \\
&\leq \varepsilon(\|G'(x^*)\| + \varepsilon + \|U_*\|) =: C_*.
\end{aligned}
$$

By assumption, the matrix W_* is nonsingular; then,

$$\|I - W_*^{-1}W\| \leq \|W_*^{-1}\| \|W_* - W\| \leq \|W_*^{-1}\| C_*$$

and, by fixing ε in such a way that $\|I - W_*^{-1}W\| < 1$, the perturbation lemma can be applied. This yields

$$\|W^{-1}W_*\| \leq 1/(1 - \|I - W_*^{-1}W\|),$$

from which it follows

$$
\begin{aligned}
\|W^{-1}\| \quad &\leq \quad \|W^{-1}W_*\|\|W_*^{-1}\| \leq \|W_*^{-1}\|/(1 - \|I - W_*^{-1}W\|) \leq \\
&\leq \quad \|W_*^{-1}\|/(1 - \|W_*^{-1}\|C_*) =: C. \qquad \square
\end{aligned}
$$

By using (3), a finite difference generalized Newton method can be defined as follows:

$$
x_{k+1} = x_k + d_k, \quad d_k = -W_k^{-1}H(x_k), \quad W_k \in J_{h_k}(x_k) \qquad (4)
$$

where the steplength h_k has to be properly chosen. The convergence properties of method (4) are stated in the following theorem: the proof can be easily obtained by the one given in [4] for Theorem 5.3 and therefore it is omitted.

Theorem 2.1 *Let x^* be a solution of (1) and let $H = \Phi \circ G$, with Φ semismooth at $G(x^*)$ and G a LC^1-function. Suppose that all the matrices $W_* \in J(x^*)$ are nonsingular. Then there exist a neighborhood N^* of x^* and $\hat h > 0$ such that, if $|h_k| < \hat h$ for every k, the iterative procedure (4) is well defined in N^* and the sequence generated by it converges to x^*. If $h_k \to 0$, the rate of convergence is superlinear. If $h_k = O(\|x_k - x^*\|)$ and Φ is strongly semismooth at $G(x^*)$, then the rate of convergence is quadratic.*

Method (4) has therefore nice local properties. In order to construct a global method based upon it, we may first of all endow it with a linesearch strategy: given $\beta \in (0,1)$, $\lambda \in (0,1)$ and $M > 0$, we look for the smallest index $j \in \{0, ..., M\}$ such that

$$
\|H(x_k + \lambda^j d_k)\| < (1 - \lambda^j \beta)\|H(x_k)\|.
$$

If such an index is found, the iteration is successful: we set $\lambda_k = \lambda^j$ and $x_{k+1} = x_k + \lambda_k d_k$. We remark that only a fixed number of trials are allowed in this strategy. This is reasonable taking into account that the direction d_k given by the finite difference method (4) can fail to be a descent direction; further, from a practical point of view, too small steplengths λ_k have to be avoided. Method (4) with the linesearch strategy now described will be our basic method, named b-method in the sequel.

If, for a given x_k, the computed matrix W_k is singular or the linesearch is not successful, we turn to the auxiliary method, named a-method, which is a simple direct search method for the unconstrained minimization problem

$$
\min_{x \in \mathbb{R}^n} \theta(x).
$$

As in all direct search methods ([13]), the search of x_{k+1} in the a-method is performed by comparing the values of $\theta(x)$ at suitable points around x_k; neither derivatives nor explicit approximations of derivatives are computed. More precisely, the search of x_{k+1} is carried on in two phases. In the first phase we compare $\theta(x_k)$ with the values of $\theta(x)$ at n trial points $x_k + \varepsilon_k e_j$, $j = 1, ..., n$, where $\varepsilon_k > 0$ is a given steplength; if we find a trial point ξ such that $\theta(\xi) < \theta(x_k)$, we take ξ as the new iterate x_{k+1}, else, in the second phase, we check the points $x_k - \varepsilon_k e_j$, $j = 1, ..., n$. If neither this set contains a point ξ such that $\theta(\xi) < \theta(x_k)$, then we half ε_k and go back to the first phase. From this short description it results that, given a starting point $x_0 \in \mathbb{R}^n$ and an initial steplength $\varepsilon_0 > 0$, the a-method generates a sequence $\{x_k\}$ such that $\theta(x_{k+1}) < \theta(x_k)$ for $k \geq 0$.

Under the assumptions that θ is a C^1-function, the level set

$$\mathcal{L}(x_0) = \{x \in \mathbb{R}^n : \theta(x) \leq \theta(x_0)\}$$

is bounded and the set

$$\mathcal{S} = \{x \in \mathcal{L}(x_0) : \nabla\theta(x) = 0\}$$

contains an isolated point, it is possible to prove the global convergence of the a-method to a stationary point for θ (see Refs. 10-11 and Proposition 1.2.38 in [14]).

In the hybrid method, the b-method and the a-method are combined as follows: given the current iterate x_k and a steplength ε_k, first of all the b-method is tried with $h_k = \varepsilon_k$; if a failure occurs in the computation of d_k or in the linesearch, the first phase of the iteration of the a-method is activated; if even it fails to compute x_{k+1}, then a new trial is made with the b-method, which is invoked with $h_k = -\varepsilon_k$; if this also fails, the second phase of the iteration of the a-method is activated. If, for a given ε_k, both methods fail both trials, then we half ε_k and start again. A sketch of the iteration is given below:

Step 0. Given $x_k \in \mathbb{R}^n$, $\varepsilon_k > 0$, $\beta \in (0, 1)$, $\lambda \in (0, 1)$, $M > 0$.

Step 1. Compute a matrix $U_k \in \partial_B \Phi(y)|_{y=G(x_k)}$.

Step 2. Set $h_k = \varepsilon_k$; compute $\xi_j = x_k + \varepsilon_k e_j$, for $j = 1, ..., n$, and $G_{h_k}(x_k)$ by (2); set $W_k = U_k G_{h_k}(x_k)$.

Step 3. Try an iteration of the b-method. If it is successful, then update ε_k and go to Step 8.

Step 4. Set $\xi = \arg\min\{\theta(\xi_j), j = 1, ..., n\}$; if $\theta(\xi) < \theta(x_k)$, then set $x_{k+1} = \xi$, $\varepsilon_{k+1} = \varepsilon_k$ and go to Step 8.

Step 5. Set $h_k = -\varepsilon_k$; compute $\xi_j = x_k - \varepsilon_k e_j$, for $j = 1, ..., n$, and $G_{h_k}(x_k)$ by (2); set $W_k = U_k G_{h_k}(x_k)$.

Step 6. Try an iteration of the b-method. If it is successful, then update ε_k and go to Step 8.

Step 7. Set $\xi = \operatorname{argmin}\{\theta(\xi_j), j = 1, ..., n\}$; if $\theta(\xi) < \theta(x_k)$, then set $x_{k+1} = \xi$, $\varepsilon_{k+1} = \varepsilon_k$ and go to Step 8; else, set $\varepsilon_k = \varepsilon_k/2$ and go to Step 2.

Step 8. End of iteration.

After a successful iteration of the b-method (steps 3 and 6), ε_k has to be properly updated. Here we'll use the following rule:

$$\varepsilon_{k+1} = \min(\varepsilon_k, \|x_{k+1} - x_k\|, \|H(x_k)\|) \qquad (5)$$

which, as shown in the sequel, allows to obtain good convergence properties; these properties essentially follow from the monotonicity of $\{\varepsilon_k\}$ and from the presence of the term $\|H(x_k)\|$ in (5). Moreover, (5) takes into account the twofold meaning of the parameter ε_k; in fact, ε_k gives the steplength used in the forward or backward finite difference formula, but it represents at the same time the steplength used by the a-method: so the term $\|x_{k+1} - x_k\|$ is useful, by a practical point of view, in order to relate ε_k to the scaling of the problem with respect to the independent variable.

From the algorithm it is clear that when the a-method is invoked (steps 4 and 7) all the values of G already computed by the b-method (steps 2 and 5) are fully exploited. Hence, the extra computational cost involved depends only on the cost of evaluating $\Phi(y)$, which is very low in many problems coming from real applications.

The convergence properties of the hybrid method are stated in the theorems which follow. These are standard results and we give only a sketch of the proofs. In the theorems, θ is assumed to be a C^1-function: as we already remarked, this assumption, exploited in several works concerning semismooth equations (see, e.g., Refs. 7-9, 15), is satisfied in a wide and interesting class of problems arising from complementarity and variational inequalities problems.

Theorem 2.2 *Let H be such that $H = \Phi \circ G$, with Φ a locally Lipschitzian function and G a LC^1-function. Let $x_0 \in \mathbb{R}^n$ and $\varepsilon_0 > 0$ be given. Assume that θ is a C^1-function, the level set $\mathcal{L}(x_0)$ is compact and the set $\mathcal{S}$ contains an isolated point. Then every accumulation point for $\{x_k\}$ is a stationary point for θ.*

Proof. Sketch of Proof. First of all, the compactness of $\mathcal{L}(x_0)$ and the monotonicity of the sequence $\{\theta(x_k)\}$ imply that some accumulation points for $\{x_k\}$ exist. If there exists $\hat{k}$ such that x_k is computed by the a-method for every $k \geq \hat{k}$, the result directly follows by the properties of the a-method. Otherwise,

for an infinite number of indices k the iterate x_k is computed by the b-method and $\|H(x_k)\| < (1 - \lambda^M \beta)\|H(x_{k-1})\|$; since the whole sequence $\{\|H(x_k)\|\}$ is decreasing, from $(1 - \lambda^M \beta) < 1$ it follows that $\|H(x_k)\| \to 0$. By known calculus rules ([12]) we have

$$\nabla \theta(x) = W H(x) \qquad (6)$$

for any matrix $W \in \partial_B H(x)$. Taking into account the continuity of $\nabla \theta$ and the boundedness of the matrices in $\partial_B H(x)$, from (6) the desired result follows. $\square$

Remark 2.1 *In order to prove a stronger convergence result some assumptions have to be added. For example, if x^* is an accumulation point for $\{x_k\}$ isolated in $\mathcal{S}$ and there exists $C > 0$ such that $\|W_k^{-1}\| \le C$ for every k such that x_{k+1} is computed by the b-method, then it can be proved that $x_k \to x^*$; the proof uses arguments which are standard for similar results (see, e.g., Refs. 14, 16).*

Remark 2.2 *Theorem 2.2 and Remark 2.1 state conditions under which the sequence generated by the hybrid method converges to a stationary point for θ. Stronger assumptions are needed in order to obtain the convergence to a solution of (1). For example, (6) implies that if x^* is a stationary point for θ such that at least one matrix $W \in \partial_B H(x^*)$ is nonsingular, then x^* solves (1).*

We can now state the following theorem about the rate of convergence of the hybrid method.

Theorem 2.3 *Suppose that the assumptions of Theorem 2.2 hold. Let x^* be a solution of (1) such that $x_k \to x^*$ and all the matrices $W_* \in J(x^*)$ are nonsingular. If Φ is semismooth at $G(x^*)$, the rate of convergence is superlinear. If Φ is strongly semismooth at $G(x^*)$, then the rate of convergence is quadratic.*

Proof. Sketch of Proof. By Lemma 2.1, for k large enough all the iteration matrices W_k are nonsingular and uniformly bounded; then, $d_k \to 0$ due to $H(x_k) \to 0$. Therefore, we can use classical arguments (cfr. Theorem 4.4 in [9]) to prove that

$$\|H(x_k + d_k)\| = o(\|H(x_k)\|)$$

and from this it follows, for k large enough,

$$\|H(x_k + d_k)\| < (1 - \beta)\|H(x_k)\|$$

so that the iteration of the b-method is successful with $\lambda_k = 1$. Since, under our assumptions, it is proved ([17]) that $\|H(x_k)\| = O(\|x_k - x^*\|)$, due to the choice (5) for ε_k, the results about the rate of convergence directly follow from the properties of the local method (4) stated in Theorem 2.1. $\square$

3 Numerical Results

The potential of the proposed hybrid method was tested by several numerical experiments on test problems coming from nonlinear complementarity problems (NCP for short) and box constrained variational inequalities widely used in the literature. While referring to [17] for the complete experimentation, we report in this section results about some semismooth systems with smooth natural merit function coming from NCP.

We recall that the NCP consists in finding $x \in \mathbb{R}^n$ such that

$$x \geq 0, \qquad F(x) \geq 0, \qquad x^T F(x) = 0 \tag{7}$$

where $F : \mathbb{R}^n \mapsto \mathbb{R}^n$ is a given smooth function.

Let $\phi(a, b) = \sqrt{a^2 + b^2} - a - b$ be the well known Fischer-Burmeister NCP-function ([18]); since $\phi(a, b) = 0$ if and only if $a \geq 0$, $b \geq 0$ and $ab = 0$, x^* solves (7) if and only if it solves the nonlinear system

$$H(x) \equiv \begin{pmatrix} \phi(x_1, F_1(x)) \\ \vdots \\ \phi(x_n, F_n(x)) \end{pmatrix} = 0. \tag{8}$$

It is evident that $H(x) = (\Phi \circ G)(x)$, being $G : \mathbb{R}^n \mapsto \mathbb{R}^{2n}$ given by

$$G(x) = \begin{pmatrix} x \\ F(x) \end{pmatrix}$$

and $\Phi : \mathbb{R}^{2n} \mapsto \mathbb{R}^n$ by

$$\Phi(y) = \begin{pmatrix} \phi(y_1, y_{n+1}) \\ \vdots \\ \phi(y_n, y_{2n}) \end{pmatrix}.$$

The function G has the same degree of smoothness as F and Φ is a strongly semismooth function. In addition, for a given $y \in \mathbb{R}^{2n}$, the matrices $U \in \partial_B \Phi(y) \subset \mathbb{R}^{n \times 2n}$ are easily computed by the following explicit formulas ([4]):

$$U = \begin{pmatrix} A(y) \ B(y) \end{pmatrix}$$

where $A(y)$ and $B(y)$ are $n \times n$ diagonal matrices defined by, respectively,

$$a_{ii}(y) = \begin{cases} y_i / \sqrt{y_i^2 + y_{i+n}^2} - 1 & \text{if } y_i^2 + y_{i+n}^2 \neq 0 \\ \alpha_i - 1 & \text{otherwise} \end{cases}$$

and

$$b_{ii}(y) = \begin{cases} y_{i+n}/\sqrt{y_i^2 + y_{i+n}^2} - 1 & \text{if } y_i^2 + y_{i+n}^2 \neq 0 \\ \beta_i - 1 & \text{otherwise} \end{cases}$$

for $i = 1, ..., n$, with α_i and β_i arbitrary non negative constants such that $\alpha_i^2 + \beta_i^2 = 1$. In our experiments we obtained substantially similar results using several different choices for α_i and β_i; in this section we report results obtained by using $\alpha_i = \beta_i = \sqrt{2}/2$ for $i = 1, ..., n$.

The properties of reformulation (8) of the NCP have been widely studied (Refs. 8-9, 15, 19). In particular, H is semismooth if F is a C^1-function, and strongly semismooth if F is a LC^1-function. A key property is that the natural merit function θ is a C^1-function even if H is nonsmooth. Moreover, under some suitable assumptions on F (for example if F is an uniform P-function), the level sets of θ are compact.

The hybrid method was implemented in a double precision FORTRAN77 code and run on a HP J5000 work station (machine precision $\approx 10^{-16}$). The linear algebraic systems were solved by LU factorization, using the double precision version of the routines F07ADF and F07AEF of the NAG Fortran Library Mk 17 (corresponding to the routines DGETRF and DGETRS of LAPACK). The stopping criterion was

$$\|H(x_k)\| \leq 10^{-6},$$

while failure was declared when 300 iterations were performed without satisfying the stopping criterion or when ε_k was reduced below 10^{-11}.

Throughout the computational experiments, we used several values for the parameters β, λ and M in the linesearch strategy; the results given in this section were obtained with $\beta = 0.025$, $\lambda = 0.5$, $M = 4$.

Below some numerical results are reported about a few significant problems, chosen to highlight the typical behaviour of the method. For each problem, several starting points from literature and several values for ε_0 were used. Further, a comparison was made with the results obtained by the local method (4), which was implemented with two different choices for the discretization steplengths:

$$h_k = \sqrt{\varepsilon_m}\|H(x_k)\| \tag{9}$$

as in the local secant method studied in [4], and

$$h_k = \begin{cases} \sqrt{\varepsilon_m}\|x_k\| & \text{if } \|x_k\| \neq 0 \\ \sqrt{\varepsilon_m} & \text{otherwise} \end{cases} \tag{10}$$

which is a standard choice in the context of discrete Newton methods for smooth systems of nonlinear equations ([20]). For all but one (see Example 3.1) the problems here reported the two choices produced identical results.

In the tables we summarize the results by the following data:

SP is the starting point;

I_l is the total number of iterations performed by the local method (4);

ε_0 is the initial steplength;

I_h is the total number of iterations performed by the hybrid method;

I_a is the number of iterations performed by the a-method;

LU is the number of LU factorizations;

FE is the number of evaluations of the function G, i.e. of the function F;

$\|H(x_f)\|$ is the norm of H in the final iterate x_f.

A "-" sign in the tables denotes a failure. We remark that the number of LU factorizations is reported since the number of linear systems to be solved at each iteration of the hybrid method is not fixed, so that the iterations may have different computational cost depending on how many LU factorizations are performed. We also remark that FE includes the number of function evaluations needed to compute the matrices $G_{h_k}(x_k)$.

Example 3.1. Watson Fourth Problem (Refs. 8-9, 21-22). This is a NCP of dimension $n = 5$ proposed in [22]; the function F is the following:

$$F_i(x) = 2(x_i - i + 2) \exp\{\sum_{j=1}^{5}(x_j - j + 2)^2\}, \text{ for } i = 1, ..., 5.$$

The problem admits the degenerate solution $x^* = (0, 0, 1, 2, 3)$. We recall that a degenerate solution for a NCP is a solution x^* such that, for at least one index i, $x_i^* = F_i(x^*) = 0$. It is worth mentioning that the function H defined by (8) is non differentiable at a degenerate solution. The starting points used were the following: $\pi_1 = (0, ..., 0)$, $\pi_2 = (1, ..., 1)$, $\pi_3 = (2, ..., 2)$, $\pi_4 = (3, ..., 3)$, $\pi_5 = (-1, ..., -1)$, $\pi_6 = (-2, ..., -2)$ and $\pi_7 = (-3, ..., -3)$; all these points, excluded π_4 and π_7, were suggested in [9]. Due to the presence of the exponential, the problem is prone to overflow and therefore starting points too large in norm have to be avoided.

In Table 1 some results for this problem are given. We see that the problem resulted to be quite difficult for the local method (4), which failed in several cases with both the choices (9) and (10) for h_k; the successes reported in Table 1 were obtained with the second choice for h_k. In fact, using the first one, the method required a larger number of iterations in order to satisfy the stopping criterion (typically 19-20 iterations instead of 3-4); this is due to the fact that the norm of H became very large at the first iterations ($\|H\| \simeq 10^7$ at the second iteration), and the steplengths given by (9) were therefore too large.

Concerning the hybrid method, we give in Table 1 results obtained with two values for M ($M = 4$ and $M = 8$) and two values for ε_0 ($\varepsilon_0 = 0.1$ and $\varepsilon_0 = 1$). For each considered value of M, the method converged from all the starting points, using one or both the ε_0 values considered; the failures for $\varepsilon_0 = 0.1$ with the starting point π_7 were caused by overflow problems.

For $\varepsilon_0 = 0.1$, the numbers of iterations I_h for all the starting points are favourably compared with those obtained by other hybrid algorithms for semismooth reformulations of NCP (see Refs. 8-9). Using $\varepsilon_0 = 1$ the number of iterations is substantially reduced.

Table 1 show that, for some starting points, several iterations were performed by the a-method, due to failures of the linesearch strategy; concerning this, let us note that increasing the number M of allowed trials in the linesearch, in fact, didn't avoid to invoke the a-method: using $M = 8$ instead of $M = 4$ increased the cost of the method in term of function evaluations but didn't reduce the number of iterations performed by the a-method.

Finally, the results show that when the local method (4) was successful, the number of iterations performed by the hybrid method was larger. This is probably due to the fact that the request of monotone decrease for $\|H\|$ at each iteration of the hybrid method was hardly satisfied in this problem, for which the local method exhibited an highly nonmonotone behaviour. We presume that a nonmonotone linesearch strategy could behave very well in this case.

Example 3.2. Problem HS66 (Refs. 8, 21, 23). This is the 66th nonlinear programming problem from the Hock-Schittkowski collection ([24]) reformulated as a nonlinear complementarity problem. The NCP has dimension $n = 8$ and is defined by

$$
\begin{aligned}
F_1(x) &= -0.8 + x_4 \exp\{x_1\} + x_6 \\
F_2(x) &= -x_4 + x_5 \exp\{x_2\} + x_7 \\
F_3(x) &= 0.2 - x_5 + x_8 \\
F_4(x) &= x_2 - \exp\{x_1\} \\
F_5(x) &= x_3 - \exp\{x_2\} \\
F_6(x) &= 100 - x_1 \\
F_7(x) &= 100 - x_2 \\
F_8(x) &= 10 - x_3
\end{aligned}
$$

The starting points we used for this problem are the following: the points $\pi_1 = (1, ..., 1)$, $\pi_2 = (2, ..., 2)$, $\pi_3 = (1, 1, 1, 0, ..., 0)$, $\pi_4 = (-1, -1, -1, 1, ..., 1)$, $\pi_5 = (1, 1, 1, -10, ..., -10)$, $\pi_6 = (1, 1, 1, -1, ..., -1)$, $\pi_7 = (-1, -1, -1, 0, 1, 2, 3, 4)$ and $\pi_8 = (0, 0, 0, 1, ..., 1)$, used in [23]; the point $\pi_s = (0, 1.05, 2.9, 0, ..., 0)$, defined in [21] from the standard point in [24], and some multiples of π_s listed in Table 2, where the results obtained with $\varepsilon_0 = 0.1$ are reported.

For this problem both the local and the hybrid method succeeded for almost all the starting points, converging to $x^* \simeq (0.184126, 1.20217, 3.32732, 0.665464, 0.200000, 0, 0, 0)$. The numbers of iterations performed by the two methods were in general comparable, except for a few cases (starting points $2 \cdot \pi_s$, $3 \cdot \pi_s$ and $5 \cdot \pi_s$), where the larger computational effort for the hybrid method was probably due to the required monotonicity, as for Example 3.1. We also notice that, due to the contribution of the a-method, the hybrid method was able to recover the failure of the local method starting from π_s.

Example 3.3. Problem HS34 (Refs. 8, 21, 23). As the previous, this NCP comes from the reformulation of a nonlinear programming problem, the 34th, from the Hock-Schittkowski collection ([24]). The dimension of the NCP is still $n = 8$ and the function F is given by

$$
\begin{aligned}
F_1(x) &= -1 + x_4 \exp\{x_1\} + x_6 \\
F_2(x) &= -x_4 + x_5 \exp\{x_2\} + x_7 \\
F_3(x) &= -x_5 + x_8 \\
F_4(x) &= x_2 - \exp\{x_1\} \\
F_5(x) &= x_3 - \exp\{x_2\} \\
F_6(x) &= 100 - x_1 \\
F_7(x) &= 100 - x_2 \\
F_8(x) &= 10 - x_3
\end{aligned}
$$

The starting points are the same used for the Problem HS66. In Table 3 we report the results obtained with $\varepsilon_0 = 0.1$ starting from the points π_1 to π_8, for which method (4) succeeded. The hybrid method was also successful, in general with a larger computational effort and with several iterations performed by the a-method.

In Table 4 we show the results obtained with the starting point π_s and its multiples, using $\varepsilon_0 = 1$. Here, the situation is completely different: for almost all such points the local method (4) failed, while the hybrid method was able to satisfy the stopping criterion from all but one the starting points; in many cases the a-method was invoked several times. We remark that, for these starting points, the hybrid method was successful also using values for ε_0 smaller than 1, even if with a larger number of iterations.

The computed solution for this problem, whatever starting point was used, was $x^* \simeq (0.834032, 2.30259, 10.0000, 0.434294, 0.043429, 0, 0, 0.043429)$.

Example 3.4. Josephy Problem (Refs. 8, 19, 21, 23, 25-26). This is a classical NCP included in the MCPLIB library ([26]); it has dimension $n = 4$ and is described by the following function:

$$
\begin{aligned}
F_1(x) &= 3x_1^2 + 2x_1x_2 + 2x_2^2 + x_3 + 3x_4 - 6 \\
F_2(x) &= 2x_1^2 + x_2^2 + x_1 + 3x_3 + 2x_4 - 2 \\
F_3(x) &= 3x_1^2 + x_1x_2 + 2x_2^2 + 2x_3 + 3x_4 - 1 \\
F_4(x) &= x_1^2 + 3x_2^2 + 2x_3 + 3x_4 - 3
\end{aligned}
$$

The problem has the unique solution $x^* = (\sqrt{6}/2, 0, 0, 1/2)$. We used all the starting points suggested in the `cpstart.m` file of the `MCPLIB` library: $\pi_1 = (0, ..., 0)$, $\pi_2 = (1, ..., 1)$, $\pi_3 = (100, ..., 100)$, $\pi_4 = (1, 0, 1, 0)$, $\pi_5 = (1, 0, 0, 0)$, $\pi_6 = (0, 1, 1, 0)$, $\pi_7 = (0, 1, 0, 1)$ and $\pi_8 = (1.25, 0, 0, 0.5)$. The results, obtained with $\varepsilon_0 = 0.01$, are reported in Table 5. We remark that almost all the failures of the local method (4) were recovered by the hybrid method, without large computational efforts, thanks to the linesearch strategy.

Example 3.5. Kojima Problem (Refs. 8-9, 19, 21, 23, 25-26). As the previous, this is a classical NCP included in the `MCPLIB` library; the dimension is $n = 4$ and the problem is defined by:

$$
\begin{aligned}
F_1(x) &= 3x_1^2 + 2x_1x_2 + 2x_2^2 + x_3 + 3x_4 - 6 \\
F_2(x) &= 2x_1^2 + x_2^2 + x_1 + 10x_3 + 2x_4 - 2 \\
F_3(x) &= 3x_1^2 + x_1x_2 + 2x_2^2 + 2x_3 + 9x_4 - 9 \\
F_4(x) &= x_1^2 + 3x_2^2 + 2x_3 + 3x_4 - 3
\end{aligned}
$$

The problem is widely used in literature since it has two solutions, $x_1^* = (\sqrt{6}/2, 0, 0, 1/2)$ and $x_2^* = (1, 0, 3, 0)$, the first of them being degenerate. The starting points suggested in the `MCPLIB` library are the same as for the Josephy Problem (π_1 to π_8). The results, obtained with $\varepsilon_0 = 0.01$, are reported in Table 6, where a superscript d denotes that the method reached the degenerate solution. Also for this problem the hybrid method had a regular behaviour and converged from almost all the starting points, without excessive computational effort; moreover, for most of the points, it satisfied the stopping criterion in less iteration than method (4).

4   Conclusions and Perspectives

The computational results presented in Section 3 show that the hybrid approach proposed in this paper may yield robust and effective algorithms for the solution of composite semismooth equations. In the meantime, the comparison with the local method (4) indicates that even better performance could be attained by relaxing the requirement that the value of the merit function decreases at each

iteration; this may be accomplished, for example, by using some nonmonotone linesearch strategy (see, e.g., Refs. 11, 27).

The numerical examples presented in this paper are still just academic examples. It is our intent consider in future work testing the hybrid approach on problems coming from real applications. This goal implies two aspects. On one hand, semismooth nonlinear equations that model real applications are usually of large or very large scale. Therefore the linear algebraic systems cannot be solved by direct methods and considering inexact methods is of crucial importance. On the other hand, some recent papers (Refs. 19, 28) show that NCP-functions different from the Fischer-Burmeister function lead to semismooth reformulations of nonlinear complementarity and variational inequality problems with appealing theoretical and computational properties. The practical advantages of such reformulations should be particularly interesting for large problems. A cost to be paid is in general the lack of smoothness for the natural merit function. Hence, we intend to investigate the possibility of making our algorithms more flexible and extending our approach to semismooth systems with nondifferentiable merit functions.

References

[1] CHEN, X., MATSUNAGA, N., and YAMAMOTO, T., *Smoothing Newton Methods for Nonsmooth Dirichlet Problems*, Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, Edited by M. Fukushima and L. Qi, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 65-79, 1999.

[2] JIANG, H., QI, L., CHEN, X., and SUN, D., *Semismoothness and Superlinear Convergence in Nonsmooth Optimization and Nonsmooth Equations*, Nonlinear Optimization and Applications, Edited by G. Di Pillo and F. Giannessi, Plenum Press, New York, NY, pp. 197-212, 1996.

[3] PANG, J.-S., and QI, L., *Nonsmooth Equations: Motivation and Algorithms*, SIAM Journal on Optimization, Vol. 3, pp. 443-465, 1993.

[4] POTRA, F. A., QI, L., and SUN, D., *Secant Methods for Semismooth Equations*, Numerische Mathematik, Vol. 80, pp. 305-324, 1998.

[5] QI, L., and SUN, J., *A Nonsmooth Version of Newton's Method*, Mathematical Programming, Vol. 58, pp. 353-367, 1993.

[6] QI, L., *Convergence Analysis of Some Algorithms for Solving Nonsmooth Equations*, Mathematics of Operations Research, Vol. 18, pp. 227-244, 1993.

[7] JIANG, H., and RALPH, D., *Global and Local Superlinear Convergence Analysis of Newton-Type Methods for Semismooth Equations with Smooth*

Least Squares, Reformulation: Nonsmooth, Piecewise Smooth, Semismooth and Smoothing Methods, Edited by M. Fukushima and L. Qi, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 181-209, 1999.

[8] DE LUCA, T., FACCHINEI, F., and KANZOW, C., *A Semismooth Equation Approach to the Solution of Nonlinear Complementarity Problems*, Mathematical Programming, Vol. 75, pp. 407-439, 1996.

[9] JIANG, H., and QI, L., *A New Nonsmooth Equation Approach to Nonlinear Complementarity Problems*, SIAM Journal on Control and Optimization, Vol. 35, pp. 178-193, 1997.

[10] BELLAVIA, S., GASPARO, M. G., and MACCONI, M., *A Switching-Method for Nonlinear Systems*, Journal of Computational and Applied Mathematics, Vol. 71, pp. 83-93, 1996.

[11] GASPARO, M. G., *A Nonmonotone Hybrid Method for Nonlinear Systems*, Optimization Methods and Software, Vol. 13, pp. 79-92, 2000.

[12] CLARKE, F. H., *Optimization and Nonsmooth Analysis*, John Wiley and Sons, New York, NY, 1983.

[13] LEWIS, R. M., TORCZON, V., and TROSSET, M. W., *Direct Search Methods: Then and Now*, Journal of Computational and Applied Mathematics, Vol. 124, pp. 191-207, 2000.

[14] POLAK, E., *Optimization: Algorithms and Consistent Approximations*, Springer Verlag, New York, NY, 1997.

[15] FACCHINEI, F., and SOARES, J., *A New Merit Function for Nonlinear Complementarity Problems and a Related Algorithm*, SIAM Journal on Optimization, Vol. 7, pp. 225-247, 1997.

[16] MORÉ, J. J., and SORENSEN, D. C., *Computing a Trust Region Step*, SIAM Journal on Scientific and Statistical Computing, Vol. 4, pp. 553-572, 1983.

[17] PIERACCINI, S., *Metodi Ibridi per Equazioni Nonlineari Nondifferenziabili di Tipo "Semismooth"*, PhD Thesis, Department of Mathematics, University of Milan, Milan, Italy, 2000.

[18] FISCHER, A., *A Special Newton-Type Optimization Method*, Optimization, Vol. 24, pp. 269-284, 1992.

[19] KANZOW, C., and KLEINMICHEL, H., *A New Class of Semismooth Newton-Type Methods for Nonlinear Complementarity Problems*, Computational Optimization and Applications, Vol. 11, pp. 227-251, 1998.

[20] KELLEY, C. T., *Iterative Methods for Linear and Nonlinear Equations*, Frontiers in Applied Mathematics, SIAM, Philadelphia, PA, 1999.

[21] FACCHINEI, F., and SOARES, J., *Testing a New Class of Algorithms for Nonlinear Complementarity Problems*, Variational Inequalities and Network Equilibrium Problems, Edited by F. Giannessi and A. Maugeri, Plenum Press, New York, NY, pp. 69-83, 1995.

[22] WATSON, L. T., *Solving the Nonlinear Complementarity Problem by a Homotopy Method*, SIAM Journal on Control and Optimization, Vol. 17, pp. 36-46, 1979.

[23] GEIGER, C., and KANZOW, C., *On the Resolution of Monotone Complementarity Problems*, Computational Optimization and Applications, Vol. 5, pp. 155-173, 1996.

[24] HOCK, W., and SCHITTKOWSKI, K., *Text Examples for Nonlinear Programming Codes*, Lecture Notes in Economics and Mathematical Systems, Springer Verlag, Berlin, Germany, Vol. 187, 1981.

[25] CHEN, B., CHEN, X., and KANZOW, C., *A Penalized Fischer-Burmeister NCP-Function: Theoretical Investigation and Numerical Results*, Preprint 126, Institute of Applied Mathematics, University of Hamburg, Hamburg, Germany, 1997.

[26] DIRSKE, S. P., and FERRIS, M. C., *MCPLIB: A Collection of Nonlinear Mixed Complementarity Problems*, Optimization Methods and Software, Vol. 5, pp. 319-345, 1995.

[27] FERRIS, M. C., and LUCIDI, S., *Nonmonotone Stabilization Methods for Nonlinear Equations*, Journal of Optimization Theory and Applications, Vol. 81, pp. 53-71, 1994.

[28] LOPES, V. L. R., MARTÍNEZ, J. M., and PÉREZ, R., *On the Local Convergence of Quasi-Newton Methods for Nonlinear Complementarity Problems*, Applied Numerical Mathematics, Vol. 30, pp. 3-22, 1999.