Math Meth Oper Res (2008) 68:235–256
DOI 10.1007/s00186-008-0219-8

ORIGINAL ARTICLE

Globalizing a nonsmooth Newton method via nonmonotone path search

Stephan Bütikofer

Revised: 12 February 2008 / Published online: 22 April 2008

Abstract We give a framework for the globalization of a nonsmooth Newton method. In part one we start with recalling B. Kummer's approach to convergence analysis of a nonsmooth Newton method and state his results for local convergence. In part two we give a globalized version of this method. Our approach uses a path search idea to control the descent. After elaborating the single steps, we analyze and prove the global convergence resp. the local superlinear or quadratic convergence of the algorithm. In the third part we illustrate the method for nonlinear complementarity problems.

Keywords Nonsmooth optimization · Newton's method · Local Lipschitz function · Global convergence

1 Introduction

The local convergence analysis of Newton's method is highly developed both in general and specific framework. We prefer the approach first proposed and analyzed by B. Kummer (see e.g. Kummer 1988, 1992, 2000) for the following reason. He states two conditions for an approximation Gh of local Lipschitz function h between normed vector spaces X and Y (abbreviated $\mathbf{h} \in \mathbf{C^{0,1}(X, Y)}$), which are sufficient for local (superlinear) convergence. So this method can be applied with any kind of generalized derivative Gh. We give the details in Sect. 1.1. For further information on nonsmooth Newton methods we refer e.g. (Pang 1990, 1991; De Luca et al. 1996, 2000; Qi et al. 1993; Facchinei and Pang 2003; Fischer 1997; Griewank 1987; Klatte and Kummer 2002; Kojima and Shindoh 1987; Outrata et al. 1998).

S. Bütikofer (✉)
Institut für Operations Research, Universität Zürich,
Moussonstrasse 15, 8044 Zurich, Switzerland
e-mail: buetikofer@ior.unizh.ch

Locally convergent methods require a starting iterate to be close to the unknown zero of the function h. In order to deal with the situation, where a good starting iterate is not available, we have to introduce globally convergent methods that allow starting iterates far from zero of h. Up to now there are basically three techniques namely path search methods, line search methods and trust region methods (see Facchinei and Pang 2003 for an detailed treatment).

After restating Kummer's approach in Sect. 1.1, we introduce our globalized algorithm in Sect. 2.1. We consider a path search idea. It seems to us natural to work with a (possible) nonlinear path, when using (possible) nonlinear approximations Gh of h. We want to stress at this place that we are not so much interested in the feasibility of the algorithm but rather in the kind of limit points we get, when the algorithm does not stop premature.

Premature termination is discussed in Sect. 2.2. We show there that the algorithm either stops in a stationary point or when the path does not direct along a descent direction of the merit function. In Sect. 2.3 we state a global convergence theorem under the assumption of feasibility of the algorithm. The transition to fast local convergence e.g. the acceptance of the full path length is discussed in Sect. 2.4. We illustrate and explain the application of the algorithm to complementarity problems in Sect. 3. In the last Sect. 4 we compare our method to other known path and line search approaches.

1.1 Newton's method

Following Kummer, a local Newton method to find a zero x^* of a nonsmooth equation

$$h(x) = 0$$

can be given in an abstract framework: given two normed spaces X, Y and $h(x)$ a local Lipschitz function with rank L in a neighborhood of a zero $x^* \in X$ of h, one considers some multifunction

$$Gh : X \times X \rightrightarrows Y,$$

which satisfies at least the following conditions

$$\emptyset \neq Gh(x,u) \text{ and } Gh(x,0) = \{0\} \qquad \forall\, x \in X, \quad \forall\, u \in X. \tag{1}$$

Given an iterate x_k, one has to find $u \in X$ such that

$$\emptyset \neq \alpha \|h(x_k)\| \mathbb{B} \cap [h(x_k) + Gh(x_k, u)], \quad \text{put } \; x_{k+1} := x_k + u, \tag{2}$$

$\mathbb{B}$ is the closed unit ball in X, $\alpha \geq 0$ is an accuracy parameter. $Gh(x,u)$ plays the role of a multivalued generalized (directional) derivative. A linear ansatz is given if $Gh(x,u)$ is the image of a set of linear operators, but a nonlinear ansatz is possible too (see Chap. 10 in Klatte and Kummer 2002). Examples for $Gh(x,u)$ considered in the literature are amongst others the Contingent derivative, the Thibault derivative, the (normal) directional derivative, the Clarke subdifferential or selections out of them (see again Klatte and Kummer 2002; Rockafellar and Wets 1998).

Definition 1 (Klatte and Kummer 2002, Chap. 10) Let h be in $C^{0,1}(X, Y)$, x^* a zero of h and Gh a multifunction, which fulfils (1). We call the triple (h, Gh, x^*) *feasible* if, for each $\epsilon \in (0, 1)$, there are positive r and α such that, whenever $\|x_0 - x^*\| \leq r$, process (2) has solutions and generates iterates satisfying

$$\|x_{k+1} - x^*\| \leq \epsilon \|x_k - x^*\|.$$

In Klatte and Kummer (2002) conditions are given such that superlinear local convergence is ensured. Two types of conditions are essential, namely an injectivity condition (CI) for Gh,

$$\|v\| \geq c\|u\| \;\; \forall\, v \in Gh(x, u), \forall\, u \in X, \forall\, x \in x^* + \delta\mathbb{B}, \;\; (c > 0, \delta > 0 \text{ fixed}) \qquad \text{(CI)}$$

and an approximation condition (CA) for h,

$$h(x) - h(x^*) + Gh(x, u) \subset Gh(x, x + u - x^*) + o(x - x^*)\mathbb{B}, \;\; \forall u \in X, \forall\, x \in x^* + \delta\mathbb{B}, \qquad \text{(CA)}$$

where $o(x - x^*)/\|x - x^*\| \to 0$ as $\|x - x^*\| \to 0$.

Note that (CA) is automatically satisfied if $h \in C^1$ and $Gh(x, u) = Dh(x)u$, but is an essentially restriction in the nonsmooth case, see Example BE.1 in Klatte and Kummer (2002). The existence of an exact ($\alpha = 0$) solution x_{k+1} of (2) is evident if a linear ansatz is given and X, Y are finite dimensional, the injectivity condition (CI) then ensures regularity of all linear operators in $Gh(x, u)$. We quote the main convergence theorem because we will need it in Sect. 2.4.

Theorem 1 (Local convergence I) (Klatte and Kummer 2002, Theorem 10.7) *Let h be in $C^{0,1}(X, Y)$ and Gh a multifunction, which fulfils* (1).

(i) *The triple (h, Gh, x^*) is feasible if there exists $c > 0$, $\delta > 0$ and a function $o(\cdot)$ such that, for all $x \in x^* + \delta\mathbb{B}$, the conditions (CA) and (CI) are satisfied.*

Moreover, having (CA) and (CI), let

$$\epsilon \in (0, 1), \;\; \alpha \in (0, \tfrac{1}{2}c\epsilon L^{-1}], \;\; \textit{and let } r \in (0, \delta] \textit{ be small}$$
$$\textit{enough such that } o(x - x^*) \leq \tfrac{1}{2}c\|x - x^*\|, \quad \forall\, x \in x^* + r\mathbb{B}.$$

Under this condition, the convergence can be quantified as follows:

(ii) *If r even satisfies*

$$o(x - x^*) \leq \frac{1}{2}\alpha c\|x - x^*\| \;\; \forall\, x \in x^* + r\mathbb{B}. \tag{3}$$

then ϵ, α and r fulfil the requirements in the definition of feasibility. In particular (2) *remains solvable if $\|x_0 - x^*\| \leq r$.*

(iii) *If there exists a solution u of* (2) *for every* $x_k \in x^* + r\mathbb{B}$, *then*

$$\|x_{k+1} - x^*\| \le \frac{1}{2}(1+\epsilon)\|x_k - x^*\|, \ provided\ that\ \|x_0 - x^*\| \le r.$$

So (3) *and* $\|x_{k+1} - x^*\| \le \epsilon\|x_k - x^*\|$ *hold for large* k.

(iv) *If all* x_{k+1} *are exact solutions of* (2), *then they fulfil*

$$c\|x_{k+1} - x^*\| \le o(x_k - x^*)\ with\ o(\cdot)\ from\ (CA)\ if\ \|x_0 - x^*\| \le r.$$

Corollary 1 (Local quadratic convergence) *Assume the settings of Theorem* 1*(iv). If additionally* $o(x - x^*) \le q\|x - x^*\|^2$, $q > 0$, *holds for all* x *with* $\|x - x^*\| \le r_1$, $r_1 > 0$, *we get*

$$\|x_{k+1} - x^*\| \le c^{-1}q\|x_k - x^*\|^2\ \ if\ \ \|x_0 - x^*\| \le \min\{r,\ r_1\}.$$

Proof It is a direct implication from Theorem 1 (iv) and the stronger condition on the function $o(\cdot)$. □

2 A path search algorithm

It is the task of this article to find a suitable globalization of this local method by applying and extending approaches to global Newton methods for complementarity problems, finite-dimensional variational inequalities or generalized equations. In our opinion a path search method is particularly suitable here (see the comments in the introduction and in Sect. 4). For other approaches with the help of path search see e.g. Facchinei and Pang (2003) and Ralph (1994).

2.1 The algorithm

Consider a path search algorithm in the spirit of the local method (2) as follows. Let $x_0 \in X$, $\gamma \in (0, 1)$, $\sigma \in (0, 1)$ and $M \in \mathbb{N}_0$ be given:

Step 1 Set $k = 0$.

Step 2 If $h(x_k) = 0$ stop.

Step 3 Construct a path $p_k(\tau) : [0, \bar{\tau}_k] \to X$ with $\bar{\tau}_k \in (0, 1]$, so that

$$\begin{aligned} &p_k(0) = x_k,\ \ p_k \text{ is continuous on } [0, \bar{\tau}_k],\ \ p_k \in C^{0,1}([0, \bar{\tau}_k), X) \text{ and} \\ &\emptyset \ne (h(x_k) + Gh(x_k, p_k(\tau) - x_k)) \cap (1-\tau)\|h(x_k)\| \cdot \mathbb{B},\ \forall \tau \in [0, \bar{\tau}_k] \end{aligned} \qquad (*)$$

Find the smallest nonnegative integer i_k, so that with $i = i_k$

$$\|h(p_k(\sigma^i \bar{\tau}_k))\| \le (1 - \gamma\sigma^i\bar{\tau}_k) \cdot \max_{0 \le j \le m(k)} \|h(x_{k-j})\| \qquad (**)$$

holds, where $m(k)$ is an integer satisfying

$$m(0) = 0 \text{ and } 0 \leq m(k) \leq \min[m(k-1)+1, M], \quad \text{for } k \geq 1. \tag{4}$$

Step 4 Find $j_k \in \{0, ..., i_k\}$ so that

$$\|h(p_k(\sigma^{j_k}\bar{\tau}_k))\| = \min_{0\leq j\leq i_k} \|h(p_k(\sigma^{j}\bar{\tau}_k))\|.$$

Set $\tau_k = \sigma^{j_k}\bar{\tau}_k,\ x_{k+1} = p_k(\tau_k)$ and $k \leftarrow k+1$, go to Step 2.

Let us shortly discuss the single steps.

If the algorithm stops in step 2, one has already found a zero of h. In step 3 one can stop, if the construction of the path p_k is not possible. In this case, it can be shown (see Sect. 2.2) that we are in a stationary point of the *merit function* $\Theta(\mathbf{x}) = \|\mathbf{h}(\mathbf{x})\|$ (for stationarity see Definition 4). The Armijo stepsize in step 3 is necessary because one cannot guarantee a descent on the whole path, when x_k is far away from a zero. We integrated a nonmonotone descent condition, which includes the monotone case by setting $M = 0$. From a theoretical point of view, one cannot prove stronger convergence results, than in the monotone case. But numerical tests show that nonmonotone rules are robust and efficient (Facchinei and Pang 2003; Grippo et al. 1986; Pang et al. 1991; Sun et al. 2002).

2.2 Auxiliary results

In this section we are interested in the question of premature termination of the path search algorithm. Looking at step 3, there are two main questions. Is there a path, which fulfils the intersection (*) and can we descent along this path and fulfil the Armijo-descent condition (**) ?

For the further discussion we have to introduce a new term.

Definition 2 Let h be in $C^{0,1}(X, Y)$, $x \in X$ and $Gh : X \times X \rightrightarrows Y$ a multifunction, which fulfils (1). Gh is called *positive homogeneous* in x, if

$$Gh(x, \lambda u) = \lambda Gh(x, u), \quad \forall \lambda \geq 0 \tag{5}$$

holds for all $u \in X$.

The next proposition shows us, when we are able to construct a path, which fulfils the intersection (*) of step 3 of the algorithm.

Proposition 1 (Existence of a path) *Let h be in $C^{0,1}(X, Y)$ and be $x \in X$ fixed with $h(x) \neq 0$. Assume as well that the multifunction Gh is positive homogeneous in x.*

If there is a $\overline{u} \in X$ and a $\overline{\tau} \in (0, 1]$ with

$$(h(x) + Gh(x, \overline{u} - x)) \cap (1 - \overline{\tau})\|h(x)\| \cdot \mathbb{B} \neq \emptyset, \tag{6}$$

then there exists a path $p(\tau) : [0, \overline{\tau}] \to X$, which fulfils the conditions (*) from the algorithm's step 3.

Proof Consider the function $F : (-\infty, 1] \to \mathbb{R}$ defined by

$$F(s) = \|h(x) + sv\|,$$

where $v \in Gh(x, \overline{u} - x))$ fulfils the intersection (6).

The function F has the following properties:

- $F(1) = (1 - \overline{\tau})\|h(x)\|, \ F(0) = \|h(x)\|$
- F is convex on $(-\infty, 1]$.
- F is continuous on $(-\infty, 1]$.

Let us define $s_{\min}$ by

$$s_{\min} = \min_{s \in (-\infty, 1]} \{s \mid F(s) = (1 - \overline{\tau})\|h(x)\|\}$$

(see example (ii) at the end of the Sect. 2.2).

$F|_{(-\infty, s_{\min}]}$ is injective, otherwise there would be $s_1, s_2 \in (-\infty, s_{\min}]$ with $F(s_1) = F(s_2)$ and $s_1 \neq s_2$ and hence $s_3 \in (s_1, s_2)$, which minimizes F on $[s_1, s_2]$. Then s_3 minimizes F on $(-\infty, s_{\min}]$, which gives a contradiction to the definition of $s_{\min}$.

Therefore $F^{-1} : [\ \|(1 - \overline{\tau})h(x)\|, +\infty\ ) \to (-\infty, s_{\min}]$ exists and is continuous on $[\ \|(1 - \overline{\tau})h(x)\|, +\infty\ )$.

F^{-1} is even a local Lipschitz function on $(\|(1 - \overline{\tau})h(x)\|, \|h(x)\|]$. This follows from a inverse function Theorem Clarke (1983), since $0 \notin \partial F(s), \ \forall s \in (-\infty, s_{\min})$, where $\partial F(s)$ denotes the convex subdifferential . Otherwise there would be a contradiction to the definition of $s_{\min}$ like above.

Now we can define a continuous function $s(\tau) : [0, \overline{\tau}] \to [0, s_{\min}]$ by

$$s(\tau) = F^{-1}((1 - \tau)\|h(x)\|), \text{ and it holds } s \in C^{0,1}([0, \overline{\tau}), \mathbb{R}).$$

The desired continuous path on $[0, \overline{\tau}]$ is

$$p(\tau) = s(\tau)(\overline{u} - x) + x, \quad \text{with } p(\tau) \in C^{0,1}([0, \overline{\tau}), X).$$

It holds

$$\|h(x) + s(\tau)v\| = (1 - \tau)\|h(x)\|, \ \forall \tau \in [0, \overline{\tau}] \text{ and } s(\tau)v \in Gh(x, p(\tau) - x)$$

by construction of $s(\tau)$ and the positive homogeneity of the multifunction $Gh(x, \cdot)$. □

Corollary 2 (Extension of the path) *Assume the settings of Proposition 1. The path $p(\tau) : [0, \overline{\tau}] \to X$ constructed there can be extended, i.e there exists a τ' in $[\overline{\tau}, 1]$ and a path $p_1(\tau) : [0, \tau'] \to X$, which fulfils the conditions* (*) *from the algorithm's step 3. The paths $p_1(\tau)$ and $p(\tau)$ coincide on* $[0, \overline{\tau}]$.

Proof Consider again the function $F(s) : \mathbb{R} \to \mathbb{R}$ with

$$F(s) = \|h(x) + sv\|,$$

but this time defined on $\mathbb{R}$. F is still a convex function on $\mathbb{R}$. We distinguish two cases:

1. $F(s)$ attains the global minimum on $\mathbb{R}$.
   We get an interval $[\underline{s}, \overline{s}]$ of global minimizers, where $s_{\min} \leq \underline{s} \leq \overline{s} \leq \infty$ and $s_{\min}$ is defined as in the proof of Proposition 1.
   We can use $\underline{s}$ to calculate the maximal path length $\tau_{\max}$

$$F(\underline{s}) = (1 - \tau_{\max})\|h(x)\| \iff \tau_{\max} = \frac{F(\underline{s}) - \|h(x)\|}{-\|h(x)\|}$$

   By the same arguments as in Proposition 1, we conclude that the path $p(\tau)$ constructed there can be extended to a path $p_1(\tau)$ defined on an interval $[0, \tau']$ with any $\tau' \in [\overline{\tau}, \tau_{\max}]$.
2. $F(s)$ does not attain the global minimum on $\mathbb{R}$.
   In this case we conclude in the same way as in Proposition 1 the existence of a path $p_1(\tau)$ defined on $[0, \tau']$ with a $\tau' \in (\overline{\tau}, 1)$. □

In the examples at the end of the section we calculate a path for a norm induced by a scalar product. We have an example there too that shows that it was necessary to work with $s_{\min}$ in the proof of Proposition 1.

It is clear that in the absence of convexity, we have to take a closer look at some kind of "stationary" points. For a comprehensible notation, we will denote in the rest of the article *the norm function* by $\mathbf{n}(\mathbf{x}) = \|\mathbf{x}\|$.
In this article we work with the following stationarity term.

Definition 3 (Approximation of the merit function $\Theta(x)$) Let h be in $C^{0,1}(X, Y)$ and let the multifunction $Gh : X \times X \rightrightarrows Y$ fulfil (1). Then we define *the function* $\mathbf{S}\Theta : X \times X \rightrightarrows \mathbb{R}$ by

$$S\Theta(x, u) = \bigcup_{v \in Gh(x,u)} n'(h(x); v), \quad \forall\, u \in X,$$

where $n'(h(x); v)$ denotes the (standard) directional derivative of $n(y) = \|y\|$ at the point $y = h(x)$ in direction v.

Definition 4 (S-stationarity) A point x is called *S-stationary* for the merit function Θ, if

$$S\Theta(x, u) \geq 0, \quad \forall\, u \in X$$

holds, where $S\Theta$ is the multifunction from Definition 3.

So far we did not claim any quality property of the approximation Gh of h. We did not want to introduce a general approximation condition for Gh but rather introduce them at the suited place, i.e. when they are needed for the proofs.

Anyhow we want to discuss shortly an approximation condition, which has a kind of minimal quality. It is met by many generalized derivatives addressed in the literature and by the ones we want to use. We will see below that it holds also for $S\Theta$.

Definition 5 Let h be in $C^{0,1}(X, Y)$, $x \in X$ and $Gh : X \times X \rightrightarrows Y$ a multifunction, which fulfils (1). The multifunction Gh fulfils a (weak approximation) *condition* (NA) in x for h, if

$$h(x+u) - h(x) \subseteq Gh(x,u) + o(u) \cdot \mathbb{B}, \quad \forall\, u \in X \tag{NA}$$

holds.

It is well known that "chain rules" do not hold in general for generalized derivatives (see e.g. Fusek 1994; Klatte and Kummer 2002). In the light of the conditions (CA) and (CI) we can show that the full composition of the generalized derivatives (as in Definition 3) still fulfils the two conditions (unpublished until now).

The next lemma and the following corollary show that for the multifunction $S\Theta$ the (weak approximation) condition (NA) still holds, if (NA) holds for h.

Lemma 1 (Conservation of the condition (NA) under composition) *Let g be in $C^{0,1}(Y, Z)$, h in $C^{0,1}(X, Y)$ and $x \in X$. Assume that the multifunction $Gh : X \times X \rightrightarrows Y$ resp. the multifunction $Gg : Y \times Y \rightrightarrows Z$ fulfils the condition (NA) in $x \in X$ for h resp. in $h(x) \in Y$ for g.*

Then the multifunction $Gf : X \times X \rightrightarrows Z$ defined by

$$Gf(x,u) = \bigcup_{v \in Gh(x,u)} Gg(h(x), v), \quad \forall\, u \in X$$

fulfils the condition (NA) in x for $f = g \circ h$, if the multifunction $Gg(h(x), \cdot)$ is Lipschitz, i.e. for every pair u_1, u_2 in X and every point $v_1 \in Gg(h(x), u_1)$ there exists a point v_2 in $Gg(h(x), u_2)$ and a constant L with

$$\|v_1 - v_2\| \leq L\|u_1 - u_2\|.$$

Proof The proof is similar to the proof of the chain rule for differentiable functions. We start with

$$\begin{aligned} f(x+u) - f(x) &= g(h(x+u)) - g(h(x)) \\ &\subseteq Gg(h(x), h(x+u) - h(x)) + o_g(h(x+u) - h(x))t \\ &= Gg(h(x), v + o_h(u)s) + o_g(h(x+u) - h(x))t, \end{aligned}$$

where $v \in Gh(x,u)$, $s \in \mathbb{B}_Y$ and $t \in \mathbb{B}_Z$ are suitable chosen. Therefore we find $w \in Gg(h(x), v + o_h(u)s)$ so that we can write

$$\begin{aligned} f(x+u) - f(x) &= w + o_g(h(x+u) - h(x))t \\ &= z + (w - z) + o_g(h(x+u) - h(x))t \end{aligned}$$

with $z \in Gg(h(x), v)$ and by the Lipschitz property of $Gg(h(x), \cdot)$, we can find z with

$$\|z - w\| \leq L\|v - (v + o_h(u)s)\| = Lo_h(u).$$

It remains to show that

$$\lim_{\|u\|\to 0} \frac{o_g(h(x+u)-h(x))}{\|u\|} = \lim_{\|u\|\to 0} \left(\frac{\frac{o_g(h(x+u)-h(x))}{\|h(x+u)-h(x)\|}}{\frac{\|u\|}{\|h(x+u)-h(x)\|}} \right) = 0$$

holds.

But this is true because of the Lipschitz property of $h(x)$ and the definition of o_g. □

Corollary 3 (The condition NA for $S\Theta$) *Let h be in $C^{0,1}(X, Y)$ and the multifunction Gh fulfils the condition (NA) in x for h.*

Then the multifunction $S\Theta$ from Definition 3 fulfils the condition (NA) in x for Θ.

Proof The assertion follows easily from Lemma 1 and the properties of the norm function $n(x)$. □

It is now easy to see that the upper Dini derivative of $\Theta(x)$

$$\Theta^D(x; u) = \limsup_{t\downarrow 0} \frac{\Theta(x+tu) - \Theta(x)}{t}$$

lies in $S\Theta(x; u)$, if (NA) and positive homogeneity hold in $x \in X$ for Gh and the set $Gh(x, u)$ is closed. Therefore every S-stationary point x is also a Dini stationary point (Facchinei and Pang 2003) and so in this spirit we did not invent "new" stationary points.

After this short insert about approximation and stationarity of the merit function $\Theta(x)$ we return to the question of premature termination. The next Proposition gives another confirmation that it was reasonable to work with $S\Theta$ and S-stationarity.

Proposition 2 (S-stationarity) *Let $x \in X$ be fixed and h be in $C^{0,1}(X, Y)$ and let the multifunction Gh be positive homogeneous.*

Then there is no path $p(\tau)$, which fulfils the conditions (*) from the algorithm's step 3 if and only if x is S-stationary.

Proof "⇒" by Proposition 1 we get

$$\|h(x) + v\| \geq \|h(x)\|, \quad \forall v \in \bigcup_{u\in X} Gh(x, u)$$

Let be $u \in X$ and $w \in S\Theta(x, u)$ two arbitrary elements. We have to show that $w \geq 0$ holds.

By definition of the $S\Theta(x, u)$, we have a sequence $\{t_k\}_{k\in\mathbb{N}}$, $t_k \downarrow 0$ and $v \in Gh(x, u)$ with

$$w = \lim_{k\to\infty} \frac{\|h(x) + t_k v\| - \|h(x)\|}{t_k}.$$

The positive homogeneity of $Gh(x, \cdot)$ implies that $t_k v \in Gh(x, t_k u)$ and so the fraction in the above limit is positive by assumption.

"⇐" From the convexity of the norm function $n(x)$ and S-stationarity we get

$$0 \le n'(h(x); v) \le \|h(x) + v\| - \|h(x)\|, \quad \forall v \in \bigcup_{u \in X} Gh(x, u).$$

Again with Proposition 1 we deduce that there is no path $p(\tau)$, which fulfils the conditions (*) from the algorithm's step 3. □

Proposition 3 (Descent condition) *Let h be in $C^{0,1}(X, Y)$, γ in $(0, 1)$ and Gh fulfils the following approximation condition in x namely*

$$h(x + u) - h(x) - Gh(x, u) \subseteq o(u) \cdot \mathbb{B}. \tag{7}$$

Suppose as well that there exists a path $p(\tau) : [0, \overline{\tau}] \to X$, which fulfils the conditions (*) from the algorithm's step 3. Then there exists $\tau' \le \overline{\tau}$ with

$$\|h(p(\tau))\| \le (1 - \gamma\tau)\|h(x)\|, \quad \forall \tau \in [0, \tau'].$$

Proof Let γ be given in $(0, 1)$ and assume that τ' does not exist.

Then there exists a sequence $\{\tau_k\}_{k \in \mathbb{N}}$, $\tau_k \downarrow 0$ with

$$\|h(p(\tau_k))\| > (1 - \gamma\tau_k)\|h(x)\|, \quad \forall k \in \mathbb{N}.$$

Using condition (7) and $p(0) = x$, we get

$$\begin{aligned} (1 - \gamma\tau_k)\|h(x)\| &< \|h(x) + v(\tau_k) + o(p(\tau_k) - x)\| \\ &\overset{(*)}{\le} \|(1 - \tau_k)h(x)\| + \|o(p(\tau_k) - x)\|, \end{aligned}$$

where $v(\tau_k) \in Gh(x, p(\tau_k) - x)$ fulfils the intersection (*).

$\Rightarrow \quad \tau_k(\gamma - 1)\|h(x)\| + \|o(p(\tau_k) - x)\| > 0$

$\Rightarrow \quad (\gamma - 1)\|h(x)\| > 0$ because the path $p(\tau)$ is Lipschitz in 0.

This gives a contradiction to the choice of $\gamma \in (0, 1)$. □

Remark 1 Under the conditions of Proposition 3 it holds:

(a) the Armijo-stepsize rule in (**) of step 3 is realisable.

(b) If $\overline{u}$, $v \in Gh(x, \overline{u} - x)$ and $\overline{\tau}$ fulfil the intersection (6), then $\overline{u} - x$ is a descent direction of $\Theta(x)$:

$$\begin{aligned} S\Theta(x)(\overline{u} - x) &\ni n'(h(x); v) = n'(h(x); (1 - \overline{\tau})\|h(x)\|s - h(x)) \\ &\le n(h(x) + (1 - \overline{\tau})\|h(x)\|s - h(x)) - n(h(x)) \\ &= -\overline{\tau}\|h(x)\| < 0 \end{aligned}$$

with $s \in \mathbb{B}$, where the second last inequality follows from the convexity of $n(x)$.

Examples

(i) (Condition (CA) and (CI))
Unlike the smooth case (see Proposition 1) the conditions (CA) and (CI) in the iteration point do not assure the existence of a path $p(\tau)$ in step 3 of the algorithm for a nonsmooth function h. Consider for this the following setting

$$h : \mathbb{R} \to \mathbb{R},\ h(x) = |x| + 1,\ Gh(x, u) = h'(x; u)$$

The conditions (CA) and (CI) are fulfilled in $x = 0$, but $x = 0$ is a S-stationary point.

(ii) (Illustration of Proposition 1)
Consider the function $h : \mathbb{R}^2 \to \mathbb{R}^2$ defined by

$$h(x, y) = \begin{pmatrix} -1.5x + 3.5 \\ \exp(y)(1 - y) \end{pmatrix}.$$

The data given in Proposition 1 be $(x, y) = (1, 0)$ and $\overline{u} - (x, y) = (1, 0)$. We work with the Fréchet derivative $Dh(x)u$ instead the multifunction $Gh(x, u)$ and the maximum norm. First we calculate $\overline{\tau}$:

$$\begin{aligned} \left\| h(1,0) + Dh(1,0) \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\|_\infty &= \left\| \begin{pmatrix} 2 \\ 1 \end{pmatrix} + \begin{pmatrix} -1.5 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\|_\infty \\ &= 1 = (1 - 0.5)\|h(1,0)\|_\infty \end{aligned}$$

We get $\overline{\tau} = 0.5$.
The Function $F(s)$ can be determined explicitly

$$F(s) = \left\| \begin{pmatrix} 2 \\ 1 \end{pmatrix} + s \begin{pmatrix} -1.5 \\ 0 \end{pmatrix} \right\|_\infty = \begin{cases} 2 - 1.5s, & \text{if } s \le \frac{2}{3}; \\ 1, & \text{if } \frac{2}{3} \le s \le 2; \\ -2 + 1.5s, & \text{if } 2 \le s. \end{cases}$$

It follows that $s_{\min} = \frac{2}{3}$ and since $0 \in \partial F(\frac{2}{3})$ we can not further extend the path.

(iii) (Calculation of the path for a norm induced by a scalar product)
We are again in the setting of Proposition 1 and we assume that the norm is induced by a scalar product, i.e. $\|x\|^2 = \langle x, x \rangle$.
We can write the function $F(s)$ from the proof of Proposition 1 as

$$F(s) = (\langle h(x) + sv, h(x) + sv \rangle)^{\frac{1}{2}} = (\|h(x)\|^2 + 2s\langle h(x), v \rangle + s^2\|v\|^2)^{\frac{1}{2}}.$$

By quadratic extension we get

$$F(s) = \left(\|v\|^2 \left(s + \frac{\langle h(x), v \rangle}{\|v\|^2} \right)^2 + \|h(x)\|^2 - \frac{\langle h(x), v \rangle^2}{\|v\|^2} \right)^{\frac{1}{2}}$$

and therefore

$$\underline{s} = -\frac{\langle h(x), v\rangle}{\|v\|^2}, \quad F(\underline{s}) = \left(\|h(x)\|^2 - \frac{\langle h(x), v\rangle^2}{\|v\|^2}\right)^{\frac{1}{2}},$$

where $\underline{s}$ is the unique global minimalpoint of $F(s)$.
We compute the continuous function $s(\tau) : [0, \tau_{\max}] \to [0, \underline{s}],\ s(\tau) \in C^{0,1}([0, \tau_{\max}), \mathbb{R})$ from the proof of Proposition 1, with $\tau_{\max}$ defined as in the proof of corollary 2,
i.e. $\tau_{\max} = \frac{F(\underline{s})-\|h(x)\|}{-\|h(x)\|}$ and

$$\begin{aligned} s(\tau) &= F^{-1}((1-\tau)\|h(x)\|) \\ &= \frac{-2\langle h(x), v\rangle - (4\langle h(x), v\rangle^2 - 4\|v\|^2(\|h(x)\|^2 - (1-\tau)^2\|h(x)\|^2))^{\frac{1}{2}}}{2\|v\|^2}. \end{aligned}$$

(We used the solution formula for quadratic equations).

2.3 Global convergence

After the discussion above about premature termination we can now state our main (global) convergence theorem. As mentioned in the introduction, we are not so much interested in conditions assuring the feasibility of the algorithm. We rather ask us what kind of points we calculate, when we do not have premature termination.

Theorem 2 (Global convergence) *Let the sequence $\{x_k\}_{k\in\mathbb{N}}$ be generated by the nonmonotone path-search algorithm in Sect. 2.1. We define $l(k)$ as an integer, such that*

$$k - m(k) \le l(k) \le k \text{ and } \|h(x_{l(k)})\| = \max_{0\le j\le m(k)} \|h(x_{k-j})\|$$

holds for every $k \in \mathbb{N}$, where $m(k)$ is defined in (4).

Then it holds that every accumulation point x^* of $\{x_{l(k)}\}_{k\in\mathbb{N}}$ is a zero of $h \in C^{0,1}(X, Y)$, if the following (technical) condition (8) is fulfilled.

$$\lim_{k\in K,\ k\to\infty} \frac{\|h(p_{l(k)-1}(\sigma^{i_{l(k)-1}-1}\overline{\tau}_{l(k)-1})) - h(x_{l(k)-1}) - v_{l(k)-1}\|}{\sigma^{i_{l(k)-1}-1}\overline{\tau}_{l(k)-1}} = 0 \qquad (8)$$

holds for all convergent subsequences $\{x_{l(k)}\}_{k\in K}$ of $\{x_{l(k)}\}_{k\in\mathbb{N}}$ with $\lim_{k\in K,\ k\to\infty} \sigma^{i_{l(k)-1}-1}\overline{\tau}_{l(k)-1} = 0$, where

$$v_{l(k)-1} \in Gh(x_{l(k)-1}, p_{l(k)-1}(\sigma^{i_{l(k)-1}-1}\overline{\tau}_{l(k)-1}) - x_{l(k)-1})$$

is a solution of the intersection in (*) from the algorithm's step 3.

Remark 2 Condition (8) implies that the path $p_{l(k)-1}(\cdot)$ is well defined at the point $\sigma^{i_{l(k)-1}-1}$. Therefore $i_{l(k)-1}$ has to be bigger than one at least for large k's.

Proof Since $m(k)$ is bounded, it follows that $l(k)$ is unbounded and by definition it holds $m(k+1) \leq m(k)+1$ for every $k \in \mathbb{N}$.
In step 3 of the algorithm we get

$$\begin{aligned}\|h(x_{l(k+1)})\| &= \max_{0\leq j\leq m(k+1)} \|h(x_{k+1-j})\| \\ &\leq \max_{0\leq j\leq m(k)+1} \|h(x_{k+1-j})\| \\ &= \max[\|h(x_{l(k)})\|, \|h(x_{k+1})\|] = \|h(x_{l(k)})\| \quad \forall k \in \mathbb{N},\end{aligned}$$

i.e. $\{\|h(x_{l(k)})\|\}$ is monotone decreasing and hence convergent.
We consider two cases:

1. $\lim_{k\to\infty} \|h(x_{l(k)})\| = 0$, then the assertion follows.
2. $\lim_{k\to\infty} \|h(x_{l(k)})\| = \eta > 0$

Then it holds $\nu_k = \sigma^{i_{l(k)-1}}\overline{\tau}_{l(k)-1} \to 0$, otherwise there would be $\epsilon \in (0,1)$ and a subsequence $\{\nu_{k_i}\}_{i\in\mathbb{N}}$ of $\{\nu_k\}_{k\in\mathbb{N}}$ with $\nu_{k_i} \geq \epsilon > 0,\ \forall\, i \in \mathbb{N}$.
$\Rightarrow \quad (1-\gamma\nu_{k_i}) \leq (1-\gamma\epsilon)\ \forall\, i \in \mathbb{N}$
$\Rightarrow \quad \|h(x_{l(k_i)})\| \leq (1-\gamma\nu_{k_i})\|h(x_{l(l(k_i)-1)})\| \leq (1-\gamma\epsilon)\|h(x_{l(l(k_i)-1)})\|\ \forall\, i \in \mathbb{N}$,
Since $\{l(k_i)\}_{i\in\mathbb{N}}$ is unbounded, it follows $\|h(x_{l(k)})\| \to 0$ as k goes to infinity and we get a contradiction.
Therefore it holds $\lim_{k\to\infty} \nu_k = 0$ and consequently

$$\tilde{\nu}_k = \frac{\nu_k}{\sigma} = \sigma^{i_{l(k)-1}-1}\overline{\tau}_{l(k)-1} \underset{k\to\infty}{\longrightarrow} 0.$$

Accordant the algorithm's step 3 we can make the following estimates:

$$\begin{aligned}&(1-\gamma\tilde{\nu}_k)\|h(x_{l(l(k)-1)})\| \overset{(**)}{<} \|h(p_{l(k)-1}(\tilde{\nu}_k))\| \\ &\quad = \|h(x_{l(k)-1}) + v_{l(k)-1} + h(p_{l(k)-1}(\tilde{\nu}_k)) - h(x_{l(k)-1}) - v_{l(k)-1}\| \\ &\quad \overset{(*)}{\leq} (1-\tilde{\nu}_k)\|h(x_{l(k)-1})\| + \|h(p_{l(k)-1}(\tilde{\nu}_k)) - h(x_{l(k)-1}) - v_{l(k)-1}\| \\ &\quad \leq (1-\tilde{\nu}_k)\|h(x_{l(l(k)-1)})\| + \|h(p_{l(k)-1}(\tilde{\nu}_k)) - h(x_{l(k)-1}) - v_{l(k)-1}\|,\end{aligned}$$

where $v_{l(k)-1} \in Gh(x_{l(k)-1}, p_{l(k)-1}(\tilde{\nu}_k) - x_{l(k)-1})$ is a solution of the intersection in (*) of the algorithm's step 3.
Hence it holds

$$\begin{aligned}&\sigma^{i_{l(k)-1}-1}\overline{\tau}_{l(k)-1}(1-\gamma)\|h(x_{l(l(k)-1)})\| \\ &\quad \leq \|h(p_{l(k)-1}(\sigma^{i_{l(k)-1}-1}\overline{\tau}_{l(k)-1})) - h(x_{l(k)-1}) - v_{l(k)-1}\|.\end{aligned}$$

After dividing both sides by $\sigma^{i_{l(k)-1}-1}\overline{\tau}_{l(k)-1}$ and passing to the limit we get by assumption (8) that $(1-\gamma)\eta \leq 0$ holds. The choice of $\gamma \in (0,1)$ implies $\eta \leq 0$. This contradicts the assumption of case 2 and the assertion follows. □

For a illustration and remarks on assumption (8) we refer to the Sects. 3 and 4.

Corollary 4 *Under the assumptions of Theorem 2, it holds that every accumulation point x^* of $\{x_k\}_{k\in\mathbb{N}}$ is a zero of $h \in C^{0,1}(X, Y)$.*

Proof We already know from the proof of Theorem 2 that $\{\|h(x_{l(k)})\|\}_{k\in\mathbb{N}}$ converges to zero. By definition we get

$$\|h(x_{l(k)})\| \geq \|h(x_k)\| \quad \forall k \in \mathbb{N}.$$

Therefore we have for any convergent subsequence $\{x_{k_i}\}_{i\in\mathbb{N}}$ of $\{x_k\}_{k\in\mathbb{N}}$ with limit x^*

$$0 \leq \|h(x^*)\| = \lim_{i\to\infty} \|h(x_{k_i})\| \leq \lim_{i\to\infty} \|h(x_{l(k_i)})\| = 0$$

□

2.4 Superlinear and quadratic convergence

We are interested in the fast local convergence. As usual in this field we expect a transition to full step length, when we are close enough to zero.

We prove this part in two steps. First we show that there exists a path with minimal path length, when the iteration point is close enough to a feasible zero x^* of h. Then we show that the sequence converges at least at linear rate, when all accumulation points are feasible.

Lemma 2 (Minimal path length) *Let x^* be a zero of $h \in C^{0,1}(X, Y)$ and let the triple (h, Gh, x^*) fulfil the conditions (CI) and (CA).*

Then there exist for every γ in $(0, 1)$ a triple (ϵ, α, r) with $\epsilon \in (0, 1)$, $\alpha > 0$ and $r > 0$, so that whenever $x \in x^* + r\mathbb{B}$ holds and $Gh(x, \cdot)$ is positive homogeneous, we find a path

$$p(\tau) : [0, \overline{\tau}] \to X, \ with\ p(0) = x\ and\ 1 \geq \overline{\tau} \geq (1 - \alpha),$$

which fulfils the conditions (*) from the algorithm's step 3. Moreover it holds

$$\|h(p(\overline{\tau}))\| \leq (1 - \gamma\overline{\tau})\|h(x)\|, \tag{9}$$

i.e. $\overline{\tau}$ is accepted by the descent condition (**) from the algorithm's step 3.

Proof We first prove the existence of a path.

From Theorem 1 (i) we know that there exists a triple (ϵ, α, r), such that we find $u \in X$ with

$$\emptyset \neq \alpha\|h(x)\|\mathbb{B} \cap [h(x) + Gh(x, u)],$$

whenever $\|x - x^*\| \leq r$ holds.

By Proposition 1 we deduce the existence of a path $p(\tau) : [0, \overline{\tau}] \to X$, $p(0) = x$ and $\alpha \geq (1 - \overline{\tau}) \geq 0$, which fulfils the conditions (*) from the algorithm's step 3. This proves the first part.

For the second part we choose (ϵ, α, r) such that

$$\begin{aligned} &\epsilon \in (0, 1),\ \alpha \in (0, \tfrac{1}{2}c\epsilon L^{-1}],\ \text{and let } r \in (0, \delta] \text{ be small enough} \\ &\text{such that } o(x - x^*) \leq \min\{\tfrac{1}{2}c, \tfrac{1}{2}\alpha c\} \cdot \|x - x^*\|,\ \forall\, x \in x^* + r\mathbb{B}, \end{aligned} \tag{10}$$

where δ and $o(\cdot)$ stem from the definition of the conditions (CA) and (CI). Note that with this choice all the assertions of Theorem 1 hold true.
The conditions (1) and (CA) show that for $x \in x^* + r\mathbb{B}$ there exists $v \in Gh(x, x^* - x)$ and $s \in \mathbb{B}$ with

$$v = h(x^*) - h(x) + o(x - x^*)s.$$

Together with the condition (CI) and (10) we deduce

$$c\|x - x^*\| \leq \|v\| \leq \|h(x^*) - h(x)\| + \frac{1}{2}c\|x - x^*\|$$

and therefore

$$\frac{c}{2}\|x - x^*\| \leq \|h(x^*) - h(x)\|. \tag{11}$$

From (11) and the Lipschitz continuity of h we get also $\frac{c}{2} \leq L$ and by the choice of α in (10) $\alpha \leq \epsilon$.

Finally using Theorem 1 (ii), (iii) and (10), it follows

$$\|h(p(\overline{\tau}))\| \leq L\|p(\overline{\tau}) - x^*\| \leq L\epsilon\|x - x^*\| \leq 2L\epsilon c^{-1}\|h(x)\|.$$

So it suffices for showing (9) that there exists a triple (ϵ, α, r) fulfilling besides (10) also

$$(1 - \gamma\overline{\tau}) \geq 2L\epsilon c^{-1} \quad \text{or equivalently,} \tag{12}$$

$$\gamma \leq \frac{1 - 2L\epsilon c^{-1}}{\overline{\tau}} \leq \frac{1 - 2L\epsilon c^{-1}}{1 - \epsilon}, \tag{13}$$

where the last inequality follows from $\epsilon \geq \alpha \geq (1 - \overline{\tau})$. Since

$$\lim_{\epsilon \downarrow 0} \left(\frac{1 - 2L\epsilon c^{-1}}{1 - \epsilon} \right) = 1 \text{ and } \gamma < 1$$

holds, the existence of a triple (ϵ, α, r) fulfilling (10) and (12) is shown. □

Theorem 3 (Local convergence II) *Let $\{x_k\}_{k\in\mathbb{N}}$ be generated by the path search algorithm in Sect. 2.1 and assume that one accumulation point $\overline{x}$ of $\{x_k\}_{k\in\mathbb{N}}$ is a zero of h and the triple $(h, Gh, \overline{x})$ fulfils the conditions* (CI) *and* (CA).

Consider for the point $\overline{x}$ a triple (ϵ, α, r) chosen according to Lemma 2.

If the algorithm chooses in step 3 for any iteration point $x_k \in \overline{x} + r\mathbb{B}$ a path $p_k(\tau) : [0, \overline{\tau}_k] \to X$, $p_k(0) = x_k$ with $1 \geq \overline{\tau}_k \geq (1 - \alpha)$, then the following statements hold.

1. *The sequence $\{x_k\}_{k\in\mathbb{N}}$ converges to $\overline{x}$, i.e. $\overline{x}$ is the only accumulation point.*
2. *The sequence $\{x_k\}_{k\in\mathbb{N}}$ converges with linear rate*

$$\|x_{k+1} - \overline{x}\| \leq \epsilon \|x_k - \overline{x}\|.$$

3. *If there exists $k' \in \mathbb{N}$ so that $\overline{\tau}_k = 1$ for all $k \geq k'$, we get superlinear convergence*

$$\|x_{k+1} - \overline{x}\| \leq c^{-1} o(x_k - \overline{x})$$

with the function $o(\cdot)$ from the condition (CA) and the constant c from (CI).
If additionally $o(x - x^*) \leq q\|x - x^*\|^2$, $q > 0$, holds for all x with $\|x - x^*\| \leq r_1$, $r_1 > 0$, we get quadratic convergence

$$\|x_{k+1} - x^*\| \leq c^{-1} q \|x_k - x^*\|^2$$

for all k sufficient large.

Proof Let x_k be the first iterate in $\overline{x} + r\mathbb{B}$. Since $(1 - \overline{\tau}_k) \leq \alpha$ holds, we get

$$\emptyset \neq \alpha\|h(x_k)\|\mathbb{B} \cap [h(x_k) + Gh(x_k, p_k(\overline{\tau}_k) - x_k)],$$

i.e. $(p_k(\overline{\tau}_k) - x_k)$ is a solution of the intersection (2).

In the foregoing Lemma 2 we showed that $\overline{\tau}_k$ is accepted by the nonmonotone descent condition (**) in step 3 of the algorithm. The next iterate x_{k+1} is therefore given by

$$x_{k+1} = p_k(\overline{\tau}_k) = x_k + (p_k(\overline{\tau}_k) - x_k).$$

Comparing with the local process (2) we see that the path search algorithm produces the same iterates like the local Newton method. The assertions 1, 2, and 3 from Theorem 3 follow from Theorem 1 and from Corollary 1. □

Remark 3 One situation, where we can guarantee the minimal path length $\overline{\tau}_k \geq (1-\alpha)$ as claimed in Theorem 3, is the following.

From Lemma 2 we know the existence of a path with path length $\overline{\tau}_k \geq (1 - \alpha)$ whenever $x_k \in x^* + r\mathbb{B}$ holds. If we assume that in every iteration point $x_k \in x^* + r\mathbb{B}$ we can find the global solution of the minimization problem

$$\min\{\|h(x_k) + v\| \mid v \in Gh(x_k, u), u \in X\} \tag{14}$$

then due to Lemma 2 the optimal value of (14) has to be smaller than $\alpha\|h(x_k)\|$. Using Proposition 1 we get the existence of a path with the desired minimal path length.

In Sect. 3 we give an application, where we are able to solve (14).

3 Application to nonlinear complementarity problems

We understand about a nonlinear complementarity problem the following.

Given local Lipschitz functions $a,\ b\colon \mathbb{R}^n \to \mathbb{R}^n$, one has to find x such that

$$a(x) \geq 0,\ b(x) \geq 0 \quad \text{and} \quad \langle a(x), b(x)\rangle = 0.$$

With $y \in \mathbb{R}^n$, we rewrite the conditions as $a(x) = y^+$, $b(x) = -y^-$, which yields the equation

$$F(x, y) = 0, \quad \text{where } F_1(x, y) = a(x) - y^+, \quad F_2(x, y) = -b(x) - y^-, \tag{15}$$

where $y_i^+ = \max\{0, y_i\}$, $y_i^- = \min\{0, y_i\}$.

$F(x, y)$ has the form of a so-called *generalized Kojima function*, for more details see Klatte and Kummer (2002). In the rest of the article $a(x)$, $b(x)$ will always be one time continuously differentiable. We take $F(x, y)\colon \mathbb{R}^{2n} \to \mathbb{R}^{2n}$ as our model local Lipschitz function. Notice that $F(x, y)$ is nonsmooth even if $a(x)$ and $b(x)$ are smooth.

The function $F(x, y)$ has a special structure, which allows a product representation $F(x, y) = M(x)N(y)$, where

$$M(x) = \begin{pmatrix} a(x) & -E & 0 \\ -b(x) & 0 & -E \end{pmatrix} \text{ is a } (2n \times (2n + 1))\text{-matrix,}$$

E is the $(n \times n)$ identity matrix and $N(y) = (1, y^+, y^-)^T \in \mathbb{R}^{1+2n}$.

The product structure of $F(x, y)$ admits to compute *the (standard) directional derivative* $\mathbf{F}'((\mathbf{x}, \mathbf{y}); (\mathbf{u}, \mathbf{v}))$ at the point (x, y) in direction (u, v) defined by

$$F'((x, y); (u, v)) = \lim_{t \downarrow 0} \frac{F((x, y) + t(u, v)) - F(x, y)}{t}.$$

We apply the following product rule to $F(x, y)$.

Proposition 4 (Product rule) (Klatte and Kummer 2002, Corollary 6.10 *Let* $F(x, y) = M(x)N(y)$ *be the Kojima function* (15). *Then the product rule of differentiation holds for* $F'((x, y); (u, v))$, *i.e.*

$$F'((x, y); (u, v)) = [DM(x)(u)]N(y) + M(x)[N'(y; v)], \tag{16}$$

where $DM(x)$ *is the Jacobian of* $M(x)$.

Using the product rule (16) we get the following representation of $F'((x,y);(u,v))$,

$$F'((x,y);(u,v)) = \left\{ \left(\begin{array}{cccc} Da_1(x)u & -r_1 v_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ Da_n(x)u & 0 & \cdots & -r_n v_n \\ -Db_1(x)u & -(1-r_1)v_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -Db_n(x)u & 0 & \cdots & -(1-r_n)v_n \end{array} \right) \,\middle|\, r \in R(y,v) \right\}$$

where the set $R(y,v)$

$$R(y,v) = \left\{ r \in \{0,1\}^n \,\middle|\, \begin{array}{l} r_i = 1, \text{ if } y_i > 0 \text{ or if } y_i = 0,\ v_i \geq 0; \\ r_i = 0, \text{ if } y_i < 0 \text{ or if } y_i = 0,\ v_i < 0 \end{array} \right\}$$

stems from the directional derivative of the functions y_i^+ and y_i^-.

For practical reason we transform this representation. We set

$$\alpha_i = r_i v_i \quad \text{and} \quad \beta_i = (1-r_i)v_i$$

and arrive at

$$F'((x,y);(u,\alpha+\beta)) = \left\{ \left(\begin{array}{cccc} Da_1(x)u & -\alpha_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ Da_n(x)u & 0 & \cdots & -\alpha_n \\ -Db_1(x)u & -\beta_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -Db_n(x)u & 0 & \cdots & -\beta_n \end{array} \right) \,\middle|\, \begin{array}{l} \beta_i = 0, \text{ if } y_i > 0, \\ \alpha_i = 0, \text{ if } y_i < 0, \\ \alpha_i \geq 0 \geq \beta_i, \text{ if } y_i = 0, \\ \alpha_i \beta_i \geq 0, \text{ if } y_i = 0 \end{array} \right\}$$

For further calculation details we refer to Klatte and Kummer (2002) and Ponomarenko (2003).

In the light of the convergence Theorem 3 we are interested in a path with maximal path length. Proposition 1 together with the formulas for $F'((x,y);(u,v))$ above show that we can reduce this problem to the following optimization problem (see also (14)). At the current iteration point (x,y) we are looking in the algorithm's step 3 for the global solution of

$$\begin{array}{rll} & \min_{u,\alpha,\beta} & \left\| F(x,y) + \begin{pmatrix} Da(x) & -E & 0 \\ -Db(x) & 0 & -E \end{pmatrix} \begin{pmatrix} u \\ \alpha \\ \beta \end{pmatrix} \right\|_2^2 \\ s.t. & \beta_i = 0, & \text{if } y_i > 0 \\ & \alpha_i = 0, & \text{if } y_i < 0 \\ & \alpha_i \geq 0 \geq \beta_i, & \text{if } y_i = 0 \\ & \alpha_i \beta_i \geq 0, & \text{if } y_i = 0, \end{array} \tag{17}$$

where $\|\cdot\|_2$ denotes the Euclidean norm.

Therefore we have to solve an optimization problem with convex quadratic objective and linear complementarity constraints in every Newton step. If the number of zero components of the variable y is small, the problem (17) could be solved effectively by either complete enumeration Pang et al. (1991) (see references there too) or a branch and bound scheme (Liu and Zhang 2002; Pang et al. 1991; Zhang and Liu 2001). In general, the Newton step problem (17) is potentially much simpler than the original problem (15) and seems to be a reasonable subproblem.

By using similar techniques as in Pang (1991) and Xiao and Harker (1994) we are able to prove that Theorem 2 holds without the technical condition (8) if we assume that the Lipschitz constants L_k of the paths $p_k(\tau) : [0, \sigma^{i_k-1}\overline{\tau}_k] \to X$ stay bounded. In contrast to Pang (1991) and Xiao and Harker (1994) we do not need the existence of exact solutions of (2) for omitting (8). We do not give the proof here because it goes beyond the scope of this article.

4 Summary and conclusions

In this paper, we have presented a globalizing framework for a nonsmooth Newton method introduced by Kummer (1988, 1992). It was possible to extend the abstract framework of the local method with the help of a path search algorithm. We have established global and local superlinear respectively quadratic convergence under adjusted assumptions. We discussed the causes for premature termination, worked out the calculation of a path for specific situations and gave an interesting application of the framework to nonlinear complementarity problems.

The future research will deal with applications to specific optimization problems presentable as nonsmooth equations (as in Sect. 3) and numerical experiments.

We want to close the paper with a short comparison of our method to two known approaches from the literature.

Our method is similar to the work of Ralph (1994) (see also Facchinei and Pang 2003). He also uses the idea of searching along a path instead of a line segment as a natural way to handle the difficulties of nonsmooth equations. The main difference lies in (*) from the algorithm's step 3. Ralph asks a much stronger condition on the path, namely

$$h(x_k) + Gh(x_k, p_k(\tau) - x_k) = (1 - \tau)h(x_k) \quad \forall \tau \in [0, \bar{\tau}_k], \tag{18}$$

where we allow any descent of our local model of h into the ball with radius $\|(1 - \tau)h(x_k)\|$. This strong condition (18) on the path $p(\tau)$ reflects in the assumptions on the multifunction $Gh(x, u)$. In order to fulfill the Eq. (18) Ralph assumes that $Gh(x, u)$ is a nonsingular uniform Newton approximation on X for h (see again Ralph 1994; Facchinei and Pang 2003). It is easy to see that in this case our technical assumption (8) is fulfilled too. Therefore our convergence theory is less restrictive in its assumptions, which do not guarantee the existence of a solution a priori.

In addition we provide in Proposition 1 and example (iii) in Sect. 2.2 a method to compute a path in a general setting. Ralph uses a modification of Lemke's Algorithm for this. Our method still have to prove its numerical robustness against what Ralph's approach is implemented in the PATH-Solver (Dirkse and Ferris 1995).

Line search damping of general Newton methods for nonsmooth equations is another approach and considered by many authors e.g. Han et al. (1992), Pang (1991) and Pang et al. (1991). A detailed treatment of the collected work in this field is given in Facchinei and Pang (2003).

Line search methods are based on applying a routine to minimize a nonnegative merit function $\Theta(x)$ that satisfies:

$$\Theta(x) = 0 \Leftrightarrow h(x) = 0$$

$\Theta(x)$ is assumed to be in $C^{0,1}(X, \mathbb{R})$. In general these methods are looking for any descent direction d, where the Dini derivative $\Theta^D(x; u)$ is negative and reduce the merit function on the linear path $x + \tau u$ with a Armijo search. In Facchinei and Pang (2003) and Pang et al. (1991) the descent direction is computed with the help of the minimizing problem

$$\min \left\{ \Theta^D(x; u) + \frac{1}{2} u^T H u \mid u \in X \right\}, \tag{19}$$

where H is a symmetric, positive definite matrix. Under a technical assumption similar to our condition (8) it can be shown that every accumulation point is a Dini stationary point of $\Theta(x)$.

In Han et al. (1992) and Pang (1991) they apply this general line search procedure to a nonsmooth Newton method. $\Theta(x)$ is there the square of the Euclidean norm. $Gh(x, u) : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ is a function, which stands for a generalized derivative. They assume that they can solve the Newton equation

$$h(x) + Gh(x, u) = 0 \tag{20}$$

in every step and receive the descent direction u out of it. Again under technical assumptions comparable to our condition (8) they now can show that every accumulation point is a zero of h. In Pang (1991) and Xiao and Harker (1994) they have been able to omit the technical condition in the special case of $Gh(x, u) = h'(x; u)$, where h is a nonsmooth reformulation of the stationary points condition of a nonlinear program or a variational inequality with the help of the min-operator. We could omit the technical condition (8) for the nonlinear complementarity problem too if we make use of the directional derivative (see also the remark at the end of Sect. 3). Furthermore we believe that we can omit the technical condition (8) for any generalized Kojima function $F(x, y) = M(x)N(y)$ (Klatte and Kummer 2002), when $M(x)$ is continuously differentiable.

Let us point out the differences to our approach. Comparing (17) and (*) in the algorithm's step 3 to (19) and (20) we work in general with a different descent direction u, e.g. we do not assume the solvability of (20). Additionally we try in (**) to minimize the merit function $\Theta(x)$ along a possible nonlinear path $p(\tau)$, which we think is appropriate in this setting. The clear transition to the local method and the local convergence properties shown in Theorem 3 sustains this belief.

Acknowledgments First I would like to thank Diethard Klatte for his advice and his support. The comments and suggestions of the referees are gratefully acknowledged.

References

Clarke FH (1983) Optimization and nonsmooth analysis. Wiley, New York

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

De Luca T, Facchinei F, Kanzow C (2000) Theoretical and numerical comparison of some semismooth algorithm for complementarity problems. Comput Optim Appl 16:173–205

Dirkse SP, Ferris MC (1995) The PATH solver: a non-monotone stabilization scheme for mixed complementarity problems. Optim Methods Softw 5:123–156

Facchinei F, Pang J-S (2003) Finite-dimensional variational inequalities and complementarity problems, vol II. Springer, Berlin

Fischer A (1997) Solutions of monotone complementarity problems with locally Lipschitzian functions. Math Program B 76:513–532

Fusek P (1994) Über Kettenregeln in Gleichungsform für Ableitungen nichtglatter Funktionen. Diplomarbeit, Institut für angewandte Mathematik Humboldt-Universität zu Berlin, Berlin

Griewank A (1987) The local convergence of Broyden-like methods on Lipschitzian problems in Hilbert spaces. SIAM J Numer Anal 24:684–705

Grippo L, Lampariello F, Lucidi S (1986) A nonmonotone line search technique for Newton's method. SIAM J Numer Anal 23:707–716

Han S, Pang J, Rangaraj N (1992) Globally convergent Newton methods for nonsmooth equations. Math Oper Res 17:586–607

Klatte D, Kummer B (2002) Nonsmooth equations in optimization. Kluwer, Dordrecht

Kojima M (1980) Strongly stable stationary solutions in nonlinear programs. In: Robinson SM (ed) Analysis and computation of fixed points. Academic Press, New York, pp 93–138

Kojima M, Shindoh S (1987) Extensions of Newton and quasi-Newton methods to systems of PC^1 equations. J Oper Res Soc Jpn 29:352–372

Kummer B (1988) Newton's method for non-differentiable functions. In: Guddat J et al. (eds) Advances in Mathematical Optimization. Akademie Verlag Berlin, Ser. Math. Res. vol 45, pp 114–125

Kummer B (1992) Newton's method based on generalized derivatives for nonsmooth functions: convergence analysis. In: Oettli W, Pallaschke D (eds) Advances in optimization. Springer, Berlin, pp 171–194

Kummer B (2000) Generalized Newton and NCP-methods: Convergence, regularity, actions. Dis Math Diff Incl 20:209–244

Liu GS, Zhang JZ (2002) A new branch and bound algorithm for solving quadratic programs with linear complementarity constraints. J Comput Appl Math 146:77–87

Outrata J, Kocvara M, Zowe J (1998) Nonsmooth approach to optimization problems with equilibrium constraints. Kluwer, Dordrecht

Pang J (1990) Newton's method for B-differentiable equations. Math Oper Res 15(2):311–341

Pang J (1991) A B-differentiable equation-based, globally and locally quadratically convergent algorithm for nonlinear programs, complementarity and variational inequality problems. Math Program 51:101–131

Pang J, Han S, Rangaraj N (1991) Minimization of locally Lipschitz functions. Siam J Optim 1:57–82

Ponomarenko A (2003) Lösungsmethoden für Variationsungleichungen. Dissertation, Fachbereich Mathematik Humboldt-Universität zu Berlin, Berlin

Qi L, Sun L (1993) A nonsmooth version of Newton's method. Math Program 58:353–367

Ralph D (1994) Global convergence of damped Newton's method for nonsmooth equations via the path search. Math Oper Res 19(2):352–389

Rockafellar RT, Wets RJ-B (1998) Variational analysis. Springer, Berlin

Sun W, Han J, Sun J (2002) Global convergence of nonmonotone descent methods for unconstrained optimization problems. J Comput Appl Math 146:89–98
Xiao B, Harker PT (1994) A nonsmooth Newton method for variational inequalities, I: Theory, II: Numerical results. Math Program 65:151–216
Zhang JZ, Liu GS (2001) A new extreme point algorithm and its application in PSQP algorithms for solving mathematical programs with linear complementarity constraints. J Global Optim 19:345–361