NEWTON's METHOD FOR NON-DIFFERENTIABLE FUNCTIONS

Bernd Kummer[1)]

Abstract

The paper discusses extensions of Newton's method for solving equations with non-differentiable functions. Sufficient conditions of convergence, estimations for the speed as well as the range of convergence, and partial classes of functions allowing such extension are given. An example shows the difficulties which may arise in relation to regular zeros of Lipschitz equations.

Keywords

Newton's method, multifunction, Lipschitz functions, regularity, Clarke Jacobian, conditions of convergence

0. Introduction

The main objective of this paper is the study of Newton's method for solving the equation

$$f(x) = 0 \qquad (2)$$

where $f: X \to Y$ is a (not necessarily differentiable) function, and X as well as Y are real Banach spaces. In order to extend Newton's idea to this case let G be some upper semicontinuous multifunction which assigns to each $x \in X$ some non-empty bounded subset $G(x)$ of the space $L(X,Y)$ of all linear, continuous operators A from X into Y. Let, further, $\bar{x}$ be a solution of (1) such that for all $A \in G(\bar{x})$, $A^{-1} \in L(Y,X)$ exists and that $\sup_{A \in G(\bar{x})} \| A^{-1} \|$ is finite.

Then, the following algorithm may be considered

ALG 1

step 1 $\quad x^0 \in X$ (near $\bar{x}$)

step k $\quad$ select any $A \in G(x^k)$ and solve $f(x^k) + A(x-x^k) = 0$ to obtain x^{k+1}, put $k = k+1$, go to step k.

Obviously, ALG 1 coincides with Newton's method if both f is continuously (Frechét-) differentiable and $G(x) = \{Df(x)\}$. The aim of our paper is twofuld. At the one hand some sufficient condition for ALG 1 being (locally, superlinearly) converging will be derived. This condition shows, particularly, the convergence of ALG 1 in the following cases:

1) Humboldt-Universität zu Berlin, Sekt. Math. PSF 1297, GDR Berlin 1086

I $f=f(x)$ is a continuous selection of finitely many functions $f^{\nu}: X \to Y$ being continuously differentiable (Frechét-), briefly $f=CS(f^1,\dots,f^m)$, $f^{\nu} \in C^1$. The mapping G is then defined as

$$G(x) = \{D\, f^{\nu}(x) \,/\, f^{\nu}(x) = f(x)\}.$$

II f: $R^n \to R^n$ is locally Lipschitz, $G(x) = \partial f(x)$, the Clarke Jacobian of f at x, and card $\partial f(\bar{x}) = 1$.

At the other hand we present some example of a Lipschitz function f: $R \to R$ with the following properties:

$\partial f(x) \subset [a,b]$, $a > 0$

$Df(\bar{x})$ exists

For $G = \partial f$, ALG 1 do not converge with almost all x^0 (though f is continuously differentiable at all x^k attained).

Concerning case I some remarks are appropriate. First, that case has been discussed in the paper M. Kojima and S. Susumu (1986).
There, the following assumptions are imposed

(I.1) $X = Y = R^n$

(I.2) $X = cl(int\, X_{\nu})$ with $X_{\nu} = \{x \,/\, f^{\nu}(x) = f(x)\}$.

I.1 implies (by W.W. Hager's famous theorem 1979) that f is locally Lipschitz and that, consequently, the Clarke-Jacobians $\partial f(x)$ always exist. By I.2, then the inclusion $G(x) \subset \partial f(x)$ is ensured.
The main advantage of the present approach consists less in considering Banach spaces instead of R^n than in the fact that mappings G of a large class may be used in ALG 1 not supposing the existence of $\partial f(x)$. At the same time, the main difficulty is to find, for given f, such a mapping G which makes ALG 1 converging and allows to determine $A \in G(x^k)$ in an efficient way.
The paper is organized as follows.

In section 1 the basic results will be formulated. They allow to study the equation (1) as well as the complementory problem

$$0 \in f(x) + N(x) \qquad (2)$$

where N is a given multifunction.

In section 2 we specify these results for (1) and (2) and discuss sufficient conditions of convergence.

Finally, in section 3, some numerical topics in view of concrete realizations are added.

Notations: For subsets A_1, A_2 of a Banach space X, for elements $x \in X$ and $r \in R$, we use the (Minkovski-) notations.

$$A_1 + rA_2 = \{a_1 + ra_2 \,/\, a_1 \in A_1,\ a_2 \in A_2\}$$
$$x + rA_2 = \{x + ra_2 \,/\, a_2 \in A_2\}$$
$$-A_1 = (-1)A_1.$$

Further, B_X denotes the closed unit ball in X.

1. Basic results

Let X and Y be (real) Banach spaces and suppose $F: X \times X \longrightarrow Y$ to be a multifunction assigning to each pair $(x,t) \in X \times X$ some subset $F(x,t)$ of Y. We consider t as being some parameter and study the sets S(t) of all x satisfying

$$0_Y \in F(x,t). \tag{1.1}$$

If we fix any $x^0 \in X$ and solve, for given x^k, the inclusion

$$0_Y \in F(x,x^k)$$

to obtain x^{k+1} (if such solution exists) we will call this procedure ALG 2. Thus, ALG 2 is defined by

$$x^{k+1} \in S(x^k),\ x^0 \text{ given}$$

and covers ALG 1 where F is specified as

$$\begin{aligned} F(x,t) &= f(t) + G(t)(x-t) := \\ &:= \{f(t)+A(x-t) \ / \ A \in G(t)\}, \\ &G(t) \subset L(X,Y). \end{aligned} \tag{1.2}$$

The convergence of ALG 2 obviously depends on the behaviour of the solution sets S(t) for t near to some parameter $\bar{t}$ of interest. We are going now to investigate this crucial fact in a more detailed manner.

Let $\bar{t} \in X$ be arbitrarily fixed and suppose M to be some non-empty subset of $S(\bar{t})$. Depending on $\bar{t}$ and M, we define a function $\tau = \tau(\varepsilon,t)$ as follows

$$\tau(\varepsilon,t) = \inf \left\{ \delta > 0 \ / \ F(x,\bar{t}+t) \subset F(x,\bar{t}) + \delta B_Y \ \text{ for all } x \in M + \varepsilon B_X \right\}$$

$(\varepsilon > 0)$.

For given $y \in Y$, let L(y) denote the set of all x satisfying $y \in F(x,\bar{t})$.

We shall say that M is regular (related to $\bar{t}$) if there are positive reals $\bar{\varepsilon}$ and $\bar{l}$ such that

$$(M+\bar{\varepsilon} B_X) \cap L(y) \subset M+\bar{l}\cdot \|y\| \cdot B_X \quad \forall y \in \bar{\varepsilon} B_Y \tag{1.3}$$

and

$$\lim_{t \to 0} \tau(\bar{\varepsilon},t) = 0 \tag{1.4}$$

hold. In the case $M = \{\bar{t}\}$ we will simply say that $\bar{t}$ is regular.

Proposition 1. If M is regular (related to $\bar{t}$) then there is some $\alpha > 0$ such that both

$$x \in S(t+\bar{t}),\ \|t\| < \alpha \quad \text{and} \quad r \overset{Df}{=} \operatorname{dist}(x,M) < \bar{\varepsilon} \tag{1.5}$$

imply

$$r \leq \bar{l} \cdot \inf_{r' > r} \tau(r',t) \tag{1.6}$$

where $\bar{\varepsilon}$ and $\bar{l}$ are from (1.3) and (1.4). Moreover, such α is defined by the following conditions

$$0 < \alpha < \bar{\varepsilon}$$
$$\tau(\bar{\varepsilon},t) < \bar{\varepsilon} \quad \forall t \in \alpha B_X \tag{1.7}$$

Proof. Let (1.5) and (1.7) be fulfilled, and let β belong to the open interval $(0, \bar{\varepsilon} - \tau(\bar{\varepsilon},t))$. Since $0 \in F(x,t+\bar{t})$ and $\mathrm{dist}(x,M) = r$, for each $r' > r$, we have $x \in M+r'B_X$. By the definition of τ one finds some y such that $\|y\| < \tau(r',t)+\beta$ and $0 \in F(x,\bar{t}) + y$.
Obviously, we may assume that $r' < \bar{\varepsilon}$ what implies

$$\tau(r',t) + \beta < \bar{\varepsilon} .$$

In view of (1.3) we may then estimate

$$x \in M + \bar{l}\cdot(\tau(r',t) + \beta)\cdot B_X \qquad (\forall \beta > 0).$$

This establishes the inequality (1.6).
As a particular result of the proposition we obtain that the solution sets $S(\bar{t}+t)$, restricted to the neighbourhood $M + \frac{\bar{\varepsilon}}{2} B_X$ of M, form an upper semicontinuous (at t = 0) mappings (in Hausdorff's sense), briefly u.s.c. (H).
It is very clear that several simple assumptions allow to replace the estimation (1.6) by

$$r \leq \bar{l}\cdot\tau(r,t). \tag{1.6)'$$

Such assumptions are M compact or $\tau(.,t)$ continuous.
Our main hypothesis for what follows is the estimation

$$\tau(\varepsilon,t) \leq c\cdot\|t\| + a_0(t)\cdot\varepsilon + a_1(t) \tag{1.8}$$

where we suppose $c \geq 0$ and

$a_k = a_k(t)$ to be non-negative functions satisfying

$$\lim_{t\to 0,\,\|t\|\neq 0} a_k(t)\cdot\|t\|^{-k} = 0 \qquad (k=0,k=1).$$

Since we aim at local statements only we shall seen (1.8) to be true . this inequality holds for sufficiently small ε and $\|t\|$.
The motivation of the hypothesis (1.8) is given by

Proposition 2. Let F be defined via (1.2) where $G: X \to L(X,Y)$ is u.s.c. (H) at $\bar{t}$ and $G(\bar{t})$ is bounded. Let $M = \{\bar{t}\}$. Then, the estimation (1.8) holds (locally) whenever f is locally Lipschitz near $\bar{t}$.

Proof. Since G is u.s.c. (H) at $\bar{t}$ there is (by definition) some positive functions $\delta = \delta(t)$ satisfying $G(\bar{t}+t) \subset G(\bar{t}) + \frac{1}{2}\delta(t)B_{L(X,Y)}$;
$\delta(t) \to 0$ as $t \to 0$. Therefore, to each $A_t \in G(\bar{t}+t)$, corresponds some $A_t^0 \in G(\bar{t})$ such that

$\|A_t - A_t^o\| < \delta(t).$

Consider any

$z \in F(x, \bar{t}+t)$ where $x \in \bar{t} + \varepsilon B_X$.

In view of (1.2) we may write

$$z = f(\bar{t}+t) + A_t(x-(\bar{t}+t)).$$

Setting $y = f(\bar{t}) + A_t^o(x-\bar{t})$ we obtain

$$y \in F(x, \bar{t})$$

and

$$\|z-y\| \le \|f(\bar{t}+t)-f(t)-A_t t\| + \delta(t)\ \|x-\bar{t}\|.$$

Consequently (with $a_o(t) = \delta(t)$), the estimation

$$\tau(\varepsilon, t) \le \|f(\bar{t}+t)-f(\bar{t})-A_t t\| + a_o(t)\cdot\varepsilon$$

is true.

Finally, since $\|A_t\|$ is bounded for sufficiently small $\|t\|$ and since f is Lipschitzian near $\bar{t}$, there is some c such that (1.8) holds true (with $a_1 \equiv 0$).

Note that the assumption $M = \{\bar{t}\}$ is needed anyway in order to ensure the convergence of the sequence $\{x^k\}$ in ALG 2 to $\bar{t}$.

At the same time one sees that, if F is given by (1.2) and $\bar{t}$ is regular, the mappings $A \in G(\bar{t})$ are injective and $f(\bar{t}) = 0$. Let us, additionally, assume that $f(\bar{t} + \varepsilon B)$ contains some neighbourhood of 0_Y in Y (for each $\varepsilon > 0$). If nothing more is known about f and if we want to guarantee the existence of x^{k+1} in ALG 1 for x^k near $\bar{t}$ we have to impose that all $A \in G(t)$ (for t near $\bar{t}$) are also surjective.

Therefore, our assumptions

$$G(x) \text{ bounded, } G \text{ u.s.c. (H) at } \bar{x},\ A^{-1} \in L(Y,X) \text{ exists for all } A \in G(\bar{x}) \text{ and } \sup_{A \in G(\bar{x})} \|A^{-1}\| < \infty \qquad (*)$$

which have been imposed in relation to ALG 1 (see introduction) are natural and ensure that the assumptions of the Proposition 1 and 2 are satisfied.

We proceed now to study the convergence of ALG 2 under the following hypothesis.

$\bar{t}$ is a regular solution of $0_Y \in F(x, \bar{t})$ (see 1.3, 1.4) (1.9)

There exists $\bar{\delta} > 0$ such that

$\emptyset \ne S(\bar{t}+t) \subset \bar{t} + \delta(t) \cdot B_X$ for $\|t\| < \bar{\delta}$ where (1.10)

$\delta(t) \to 0$ as $t \to 0_X$.

<u>Theorem 1.</u> Let (1.9) and (1.10) be fulfilled, and let (1.8) be true for $\varepsilon \le \bar{\delta}$, $\|t\| \le \bar{\delta}$. Suppose, additionally that the numbers $\bar{\varepsilon}$ and $\bar{I}$ in (1.3) and (1.4) satisfy

$\varepsilon < \bar{\delta}$ and $c\bar{l} < 1$.

Then, there exists $\alpha > 0$ such that ALG 2 generates an infinite sequence x^k satisfying

$$\|x^{k+1}-\bar{t}\| \le 2\frac{\bar{l}c}{1+\bar{l}c}\|x^k-\bar{t}\| + \frac{2\bar{l}}{1+\bar{l}c}a_1(x^k-\bar{t}) \qquad (**)$$

whenever $\|x^0-\bar{t}\| < \alpha$.

The number α shows the property above if

$$0 < \alpha < \bar{\varepsilon},$$

and if $\|t\| < \alpha$ implies

$$\delta(t) < \bar{\varepsilon} \qquad (1.11)$$

$$c\|t\| + a_o(t)\cdot\bar{\varepsilon} + a_1(t) < \bar{\varepsilon} \qquad (1.12)$$

$$\bar{l}\, a_o(t) < \frac{1}{2}(1-\bar{l}c). \qquad (1.13)$$

and

$$a_1(t) \le \frac{1}{4}\cdot\frac{1-\bar{l}c}{\bar{l}}\cdot\|t\|. \qquad (1.14)$$

Proof. Since $c\bar{l} < 1$, the existence of $\alpha > 0$ such that $\|t\| < \alpha$ implies (1.11),...,(1.14) is ensured.

Let α be choosen in the given way, and let, for some fixed t, $\|t\| < \alpha$. We consider any solution x of $0 \in F(x,\bar{t}+t)$, i.e. $x \in S(\bar{t}+t)$, and put $r = \|x-\bar{t}\|$. Because of (1.10) and (1.11) such solution x exists, and $r \le \delta(t) < \bar{\varepsilon}$.

By (1.8) and (1.12) we obtain $\tau(\bar{\varepsilon},t) < \bar{\varepsilon}$. Therefore, proposition 1 may be applied with $M = \{\bar{t}\}$. This leads to the inequality

$$r \le \bar{l}\cdot\inf_{r'>r}\tau(r',t) \le \bar{l}(c\|t\| + a_o(t)r + a_1(t)). \qquad (1.15)$$

Viewing (1.13) we may continue

$$r \le \bar{l}c\|t\| + \frac{1}{2}(1-\bar{l}c)r + \bar{l}\, a_1(t)$$

$$r\cdot\left(\frac{1+\bar{l}c}{2}\right) \le \bar{l}c\|t\| + \bar{l}a_1(t)$$

$$r \le 2\frac{\bar{l}c}{1+\bar{l}c}\|t\| + \frac{2\bar{l}}{1+\bar{l}c}a_1(t). \qquad (1.16)$$

Finally, recalling (1.14), the right-hand side in (1.16) is bounded above by

$$\frac{2\bar{l}c + \frac{1}{2} - \frac{1}{2}\bar{l}c}{1+\bar{l}c}\cdot\|t\| =$$

$$\frac{1}{2}\cdot\frac{1+3\bar{l}c}{1+\bar{l}c}\cdot\|t\| = \gamma\cdot\|t\|$$

with $\gamma < 1$. Hence

$$r \le \gamma\|t\|, \quad \gamma < 1. \qquad (1.17)$$

Setting $t = x^k-\bar{t}$ $(k=0,1,2,...)$ and $x = x^{k+1}$ the inequality (1.17) shows both that ALG 2 is converging for $\|x^0-\bar{t}\| < \alpha$ and that $S(x^k)$ is non-empty. The inequality (1.16) gives the estimation (**).

To discuss the theorem let us assume (1.9) and (1.10) to be satisfied (These are, of course, no trivial assumptions). According to the theorem we have then to estimate

$$\tau(\varepsilon,t) = \inf\left\{\delta>0 \;/\; F(x,\bar{t}+t)\subset F(x,\bar{t})+\delta B_Y \ \text{ for all } x\in \bar{t}+\varepsilon B_X\right\}$$

in the form (1.8) where the constant c is as small as possible. If $c < \frac{1}{l}$ the local convergence of ALG 2 is ensured.
In the case $0 < c < \frac{1}{l}$ this convergence is a linear one (or better).
In the case $c = 0$ it is of the order a_1 and, consequently, better than linear.
Finally, if such estimation for τ is impossible or is only possible with $c \geq \frac{1}{l}$, we cannot say anything.

As a simple consequence of the inequality (1.15) one observes, for the case $0 \leq c < \frac{1}{l}$, that

$$\limsup_{\substack{k\to\infty \\ x^k\neq\bar{t}}} \frac{\|x^{k+1}-\bar{t}\|}{\|x^k-\bar{t}\|} \leq c\,l \qquad \text{whenever } \|x^0-\bar{t}\| < \alpha .$$

2. Special types of multifunctions F

2.1. F is given by (1.2)

Let us suppose the multifunction G in (1.2) to satisfy the conditions (*) with $\bar{x} = \bar{t}$ and $f(\bar{x}) = 0$. Let f be locally Lipschitzian.
Then, we already know that (1.9) is true and that (1.8) holds with some c, a_0 and a_1. To verify (1.10) is an elementary task of functional analysis which we can omit here.
Thus, the number c in (1.8) becomes the key for applying Theorem 1.

Proposition 3: Let G satisfy the condition (*) with $\bar{x} = \bar{t}$, and let $f(\bar{x}) = 0$.
Then, the estimation (1.8) holds with $c = 0$ if

$$\limsup_{\substack{t\to 0,\, t\neq 0 \\ A_t\in G(\bar{t}+t)}} \frac{1}{\|t\|}\cdot \|f(\bar{t}+t)-f(\bar{t})-A_t t\| = 0. \tag{2.1}$$

Proof. The statement immediately follows from the estimation

$$\tau(\varepsilon,t) \leq \|f(\bar{t}+t)-f(t)-A_t t\| + a_0(t)\cdot\varepsilon$$

derived in the proof of Proposition 2.

Let us discuss the condition (2.1) for the cases I and II in the introduction.

I $\quad f = CS(f^1,\ldots,f^m)$, $f^\nu \in C^1(X,Y)$

I.1 $\quad G(x) = \{Df^\nu(x) \ / \ f^\nu(x) = f(x)\}$.

Since, for each ν, f^ν is continuously Frechét differentiable there is some function $a_1^\nu = a_1^\nu(t)$ satisfying

$$\lim_{t\to 0, t\neq 0} a_1^\nu(t)\cdot \|t\|^{-1} = 0$$

and

$$\|f^\nu(\bar{t}+t)-f^\nu(\bar{t})-Df^\nu(\bar{t})t\| \leq a_1^\nu(t)$$

Defining the finite sets

$$I(X) = \{\nu / \ f^\nu(x) = f(x)\}$$

and taking into account that f is continuous we obtain the existence of some $\delta > 0$ such that

$$I(t+\bar{t}) \subset I(\bar{t}) \qquad \text{for } \|t\|<\delta.$$

Thus, if $\|t\|<\delta$, for each ν such that

$$Df^\nu(\bar{t}+t) \in G(\bar{t}+t)$$

we have $f^\nu(\bar{t}) = f(\bar{t})$. This yields the estimation

$$\|f(\bar{t}+t)-f(\bar{t})-A_t t\| = \|f^\nu(\bar{t}+t)-f^\nu(\bar{t})-Df^\nu(\bar{t}+t)t\|$$

$$\leq a_1^\nu(t) + \|-Df^\nu(\bar{t}+t) + Df^\nu(\bar{t})\| . \|t\|$$

and shows (2.1).

In order to fulfil the remaining hypotheses of Proposition 3 we need:

$f(\bar{t}) = 0$ and

$[Df^\nu(\bar{t})]^{-1} : Y \longrightarrow X$ exists and is bounded for each $\nu \in I(0)$.

We consider now the case I under the assumptions

I.2 $\quad X = Y = R^n$, $G(x) = \partial f(x)$

where $\partial f(x)$ denotes the Clarke-Jacobian of f at x (see F.H. Clarke 1983)

$$\partial f(x) = \text{conv}\{A \ / \ A = \lim Df(x^k),\ f \text{ is Frechét-differentiable at } x^k,\ x^k \to x\}.$$

Since f is locally Lipschitz (see W.W. Hager, 1979) the mapping $G = \partial f$ fulfils (*) if (at $\bar{x} = \bar{t}$) all matrices in $\partial f(\bar{t})$ are regular.
For proving (2.1) we need

<u>Proposition 4.</u> Let $f = CS(f^1,\ldots,f^m)$, $f^\nu \in C^1(R^n,R^n)$. Then it holds

$$\partial f(x) = \text{conv}\{Df^\nu(x) \ / \ \nu \in J(x)\}$$

where $J(x)$ is the set of all ν such that some sequence $x^k \to x$ exists satisfying both

$$f^\nu(x^k) = f(x^k) \text{ and } Df^\nu(x^k) = Df(x^k).$$

Proof. Let C be the convex hull of these $Df^\nu(x)$. Since $f^\nu \in C^1$ the inclusion $C \subset \partial f(x)$ is trivial. To verify $\partial f(x) \subset C$ we firstly consider any point x where $Df(x)$ exists. Without loss of generality, let $Df(x)=0$. We fix any $u \in R^n$, $u \neq 0$. Since $f = CS(f^1,\dots,f^m)$ there are some sequence of positive $\lambda_k \to 0$ and some ν such that

$f(x+\lambda_k u) = f^\nu(x+\lambda_k u)$ for all k.

One then obtains

$$0 = \lim \frac{1}{\lambda_k} (f(x+\lambda_k u)-f(x)) = \lim \frac{1}{\lambda_k} (f^\nu(x+\lambda_k u)-f^\nu(x)).$$

Hence $Df^\nu(x)u = 0$.

Setting $U^\nu = \{u \,/\, Df^\nu(x)u = 0\}$ this means

$$R^n \subset \bigcup_{\nu \in V(x)} U^\nu \tag{2.2}$$

with $V(x) = \{\nu /\ f^\nu(x^k) = f(x^k)$ for some sequence $x^k \to x\}$.

If $Df^\nu(x) \neq 0$ the subspace U^ν is a proper one. Since $V(x)$ is finite the inclusion (2.2) yields

$\dim U^\nu = n$ for at least one $\nu \in V(x)$

and, consequently, $Df^\nu(x) = 0 = Df(x)$.

Now, let $A \in \partial f(x)$ where x is arbitrarily fixed and, A is any extremal point of the compact convex set $\partial f(x)$. Then, some sequence $x^k \to x$ exists such that f is differentiable at each x^k and

$A = \lim Df(x^k)$.

Recalling the first part of the proof one can assign to each k some $\nu(k)$ such that

$f^\nu(x^k) = f(x^k)$ and $Df^\nu(x^k) = Df(x^k)$

for $\nu = \nu(k)$. After selecting some subsequence satisfying $\nu(k) \equiv \nu$ (for all k) this shows $A \in C$ and leads to $\partial f(x) \subset C$.

Knowing proposition 4 the proof of (2.1), in the case I.2, can be organized as in the case I.1. before. We have only to note that it is enough to consider extremal points $A_t \in G(\bar t+t)$ in (2.1) and we must replace the set $I(x)$ by $J(x)$.

Finally, let us look at the <u>case II</u> in the introduction, i.e.

$f: R^n \to R^n$ locally Lipschitz, $G(x) = \partial f(x)$

$\operatorname{card} \partial f(\bar x) = 1$.

The verification of (2.1), for $\bar t = \bar x$, then follows via

$\|f(\bar t+t) - f(\bar t) - Df(\bar t)t\| \leq a_1(t)$

and

$$\limsup_{A_t \in \partial f(\bar t+t),\, t \to 0} \|A_t - Df(\bar t)\| = 0 \tag{2.3}$$

by the standard estimation

$$\|f(\bar{t}+t)-f(\bar{t})-A_t t\| \le a_1(t)+\|A_t-Df(\bar{t})\|\cdot\|t\|.$$

The condition (*) means nothing else than that $Df(\bar{x})$ is a regular matrix.

It should be mentioned that, in each of these cases, ALG 1 locally superlinear converges where the estimations are given by the theorem.

2.2. Complementarity problems

Let F be given by the formula

$$F(x,t) = f(t)+G(t)(x-t) + N(x). \qquad (2.4)$$

This mapping differs from (1.2) by the fixed multifunction N not depending on t, and it is related to the complementarity problem (2)

$$0 \in f(x) + N(x).$$

The mapping N, most often, is a normal cone mapping associated with some closed convex set. In this case we have $Y = X^*$.

If we modify ALG 1 by solving, in step k,

$$0 \in f(x^k) + A(x-x^k) + N(x), \qquad (2.5)$$

then the results above may be applied. Particularly, the function τ will not depend on N. However, it becomes now more difficult to ensure regularity (see 1.3) as well as solvability of each inclusion (2.5) for x^k near to some solution $\bar{x}$ of (2) and for any $A \in G(x^k)$. In this context, we refer to B. Kummer 1987, 1984 and to the basic paper S.M. Robinson 1979. For the case that (2) describes a Kuhn-Tucker-system, the papers D. Klatte/K. Tammer (1987) and D. Klatte(1988; in the present volume) are very interesting.

2.3. A pathological example

In this section, we define a real Lipschitz function f having the following properties.

$f(0) = 0$, $Df(0) = 1$, $\partial f(0) = [\frac{1}{2}, 2]$

f^{-1} exists

ALG 1 (with $G = \partial f$) fails to converge for each starting point $x^0 \neq 0$ provided that $Df(x^0)$ exists.

For the equation $f(x) - \lambda = 0$, ALG 1 finds the solution $x(\lambda)$ by one step whenever $\lambda \neq 0$ and $\|x^0-x(\lambda)\| < \varepsilon(\lambda)$.

In order to construct f we fix any natural number $n > 1$ and consider the interval $I_n = [\frac{1}{n}, \frac{1}{n-1}]$. Let m and m' be the middle points of I_n and I_{2n}, respectively. Setting

$$a = \frac{2n}{4n-1}, \quad b = \frac{8n-4}{4n-3}$$

we define two linear functions by

$$f_n^1(x) = a(x+m),\ f_n^2(x) = b(x-m').$$

They fulfil

$$f_n^1(\tfrac{1}{n-1}) = \tfrac{1}{n-1},\ f_n^2(\tfrac{1}{n}) = \tfrac{1}{n},$$

$$f_n^1(m) < f_n^2(m),\quad a < b.$$

The point z defined by $f_n^1(z) = f_n^2(z)$ then belongs to the open interval $(\frac{1}{n},m)$. Note that a,b,m,m' and z depend on n. Now, we define f:

(i) $f(0) = 0$

$$(ii)_n\ f(x) = \begin{cases} f_n^1(x) & z \le x \le \frac{1}{n-1} \\ f_n^2(x) & \frac{1}{n} \le x \le z \end{cases}$$

(iii) $f(x) = f_2^1(x) \qquad x \ge 1$

(iv) $f(x) = -f(-x) \qquad x < 0$

For $x \in I_n$, we have

$$\frac{1}{n} \le f(x) \le \frac{1}{n-1}$$

and

$$\frac{n-1}{n} \le \frac{f(x)}{x} \le \frac{n}{n-1},$$

viewing (iv) that leads to $Df(0) = 1$.
Since, for $n \to \infty$, one obtains $\lim a = \frac{1}{2}$, $\lim b = 2$ the equation $\partial f(0) = [\frac{1}{2},2]$ is evident.
Let us apply ALG 1 ($G = \partial f$) with $x^0 \in (z,\frac{1}{n-1})$.
Then $x^1 = -m,\ x^2 = m,\ x^3 = -m, \ldots$
If $x^0 \in (\frac{1}{n},z)$, then the sequence $x^2 = m',\ x^3 = -m',\ x^4 = m', \ldots$ will be generated.
If x^0 belongs to the corresponding negative intervals or $|x^0| > 1$, then the same situation (alternating sequences) will appear.
Finally, the equation $f(x) = \lambda$ ($\lambda \neq 0$) will be solved by ALG 1 after one step since f is linear between $x(\lambda)$ and x^0 (near to $x(\lambda)$).
It should be noted that, at $\bar{t} = 0$, f does not satisfy the condition (2.3).

3. Some numerical topics

3.1. Approximation of G(x)

It is well-known from the ordinary Newton method that, in the case of regularity, the matrices $A = Df(x^k)$ may approximatively be determined up to some error ε_k. If $\varepsilon_k \to 0$ ($k \to \infty$) then the order of convergence remains better than linear; if $\varepsilon_k \le \delta$ and δ is sufficiently small, then this order is linear and the better as the upper bound δ is smaller. The same situation appears for ALG 1 provided that $A_k \in G(x^k) + \varepsilon_k \cdot B_{L(X,Y)}$ (in step k) and that the hypotheses of Proposition 3 including (2.1) are fulfilled. Indeed, after assigning to A_k some $A \in G(x^k)$ such that $\|A_k - A\| \le 2\varepsilon_k$ one easily may estimate the distance d between the solutions y^{k+1} and x^{k+1} of

$$f(x^k) + A_k(x-x^k) = 0$$

and

$$f(x^k) + A(x-x^k) = 0,$$

respectively, i.e.

$$d \le \|A_k^{-1} - A^{-1}\| \, \|f(x^k)\|.$$

Since (*) and (2.1) imply the existence of $L < \infty$ such that $\|f(\bar{t}+t)-f(\bar{t})\| \le L \, \|t\|$ for $\|t\| < \frac{1}{L}$ an estimation of the type

$$d \le \delta(\varepsilon_k) \cdot L \cdot \|x^k - \bar{t}\| \; ; \; \delta(\varepsilon) \to 0 \text{ as } \varepsilon \to 0$$

is true. Together with (**) this gives the desired (local) convergence of the sequence y^k.

3.2. Difference quotients

Let $X = Y = R^n$. In the cases I.1, I.2 and II being considered in 2.1 one can try to approximate some elements of $G(x^k)$ by difference quotients. As case II is concerned the only question remaining is that of the reasonable step-size. For case I, however, an additional problem arises which we discuss now in relation to I.1.

Let $f = CS(f^1, f^2)$ where $f^\nu: R^2 \to R^2$, $f^\nu \in C^1$. Assume $f = f^1$ for $x_1 \ge x_2$ and $f = f^2$ for $x_1 \le x_2$. Finally, let $x^k = 0$.

If A is some approximation of $Df^1(0)$ or $Df^2(0)$ then it is an approximation of G(0), too.

But a direct difference approximation of f via

$$\frac{\partial f}{\partial x_i} \approx \frac{1}{\varepsilon} \; (f(\varepsilon e_i) - f(0)),$$

where e_i denotes the i-th unit vector, leads to a matrix D being near to

$$\left(\frac{\partial f^1}{\partial x_1} \, , \, \frac{\partial f^2}{\partial x_2}\right)$$

and, possible, far from G(0).

References:

Clarke, F.H. Optimization and non-smooth analysis, Wiley Interscience, New York (1983)

Hager, W.W. Lipschitz continuity for constrained processes. SIAM. J. Control and Optimiz. 17(1979) 321-338

Kojima, M.; Susumu, S. Extensions of Newton and Quasinewton Methods to Systems of PC^1 Equations. Journ. of Oper. Research, Society of Japan Vol. 29, No. 4, Dez. 1986

Kummer, B. /1/ Generalized equations: solvability and regularity, Math. Progr. Study 21(1984), 199-212

/2/ Linearly and non-linearly perturbed optimization problems. In Guddat/Jongen/Kummer/Nožička (eds.): Parametric optimization and related topics. Akademie-Verlag Berlin (1987)

Robinson, S.M. Generalized equations and their solutions: Part I: Basic theory. Math. Progr. Study 10(1979), 128-141

Klatte, D.; Tammer, K. On second order sufficient optimality conditions for $C^{1,1}$-optimization problems. Optimization (1987, to appear).