INSTITUTUL
DE
MATEMATICA

INSTITUTUL NATIONAL
PENTRU CREATIE
STIINTIFICA SI TEHNICA

ISSN 0250 3638

ON THE APOSTERIORI ERROR ESTIMATES FOR NEWTON'S METHOD

by

F.-A. POTRA

PREPRINT SERIES IN MATHEMATICS
No.19/1981

BUCURESTI

ON THE APOSTERIORI ERROR ESTIMATES FOR NEWTON'S METHOD

by

F.-A. POTRA*)

March 1981

*) Department of Mathematics, National Institute for Scientific and Technical Creation, Bd.Pacii 220, 79622 Bucharest, Romania.

ON THE APOSTERIORI ERROR ESTIMATES FOR NEWTON'S METHOD

by

Florian-Alexandru POTRA

A b s t r a c t. New sharp aposteriori error estimates for Newton's method are given under the hypotheses of Kantorovich's theorem. One shows that they are more advantageous than the known ones from the point of view of the accuracy and of the cost of the information used.

1. Introduction

Let us consider the equation:

$$f(x)=0, \tag{1}$$

where f is a nonlinear operator defined on a subset $\mathcal{D}_f$ of a Banach space X and taking values into a Banach space Y. We are given an initial point $x_o \in \mathcal{D}_f$ and we want to produce a sequence $(x_n)_{n \geqslant 0}$ of points of $\mathcal{D}_f$ converging to a root x^* of the equation (1). To this effect we attach to the pair (f,x_o) a mapping $F:\mathcal{D}_F \subset X \longrightarrow X$ and consider the following recurrent scheme:

$$x_{n+1}=F(x_n) \qquad n=0,1,2,\ldots \tag{2}$$

Generally, $\mathcal{D}_F$ is included into $\mathcal{D}_f$. For example if one takes

(3) $$F(x)=x-f'(x)^{-1}f(x)$$

then $\mathcal{D}_F$ is composed of those points x of $\mathcal{D}_f$ for which the Fréchet derivative f'(x) exists and is invertible.

If x_o belongs to $\mathcal{D}_F$ then we may obtain by (2) a point $x_1 \in X$; if x_1 also belongs to $\mathcal{D}_F$ then we may obtain a new point $x_2 \in X$. We are interested in the case where $x_n \in \mathcal{D}_F$ for n=0,1,2,...

Definition 1. Consider a mapping $F : \mathcal{D}_F \subset X \to X$ and define recurrently

$$\mathcal{D}_o = \mathcal{D}_F\ ,\quad \mathcal{D}_{n+1} = \{x \in \mathcal{D}_n\ ;\ F(x) \in \mathcal{D}_n\}\ ,\qquad n=0,1,2,\ldots$$

The set $\mathcal{D} = \bigcap_{n \geq 0} \mathcal{D}_n$ is called the set of admissible starting points for F. If $x_o \in \mathcal{D}$, then the pair (F,x_o) is called a well defined iterative algorithm.■

Thus, if (F,x_o) is a well defined iterative algorithm, then we may obtain by (2) a sequence (x_n) of points of $\mathcal{D}_f$. Under adequate hypotheses this sequence will converge to a root x^* of the equation (1). This way of solving equations is usually called iterative procedure. In what follows we give this notion a precise meaning.

Let us consider a class $\mathcal{B}$ of (Banach) spaces and let us denote by $\mathcal{M}$ the class of all mappings $\varphi : \mathcal{D}_\varphi \subset X \longrightarrow Y$ where $X,Y \in \mathcal{B}$. Let $\mathcal{A}$ be the class of all pairs (F,x_o), where

$F: \mathcal{D}_F \subset X \longrightarrow X$ is an element of $\mathcal{M}$ and x_o belongs to $\mathcal{D}_F$ such that (F, x_o) is a well defined iterative algorithm.

Definition 2. Let $\mathcal{C}$ be a class of pairs (f, x_o) where $f: \mathcal{D}_f \subset X \rightarrow Y$ is an element of $\mathcal{M}$ and $x_o \in \mathcal{D}_f$. A mapping

$$\mathcal{F}: \mathcal{C} \rightarrow \mathcal{A} \quad , \quad \mathcal{F}(f, x_o) = (F, x_o)$$

will be called a convergent iterative procedure for the class $\mathcal{C}$ if for each $(f, x_o) \in \mathcal{C}$ the iterative algorithm $(F, x_o) = \mathcal{F}(f, x_o)$ produces a sequence (x_n) having the properties

(4) $$x_n \in \mathcal{D}_f \ , \ n=0,1,2,\ldots; \ \lim_{n\to\infty} x_n = x^*; \ f(x^*) = 0 \ . \blacksquare$$

Let k_o and r_o be two positive numbers and let us denote by $\mathcal{C}(k_o, r_o)$ the class of all pairs (f, x_o) satisfying the following conditions:

(c_1) f is a (nonlinear) operator defined on a subset $\mathcal{D}_f$ of a Banach space X and with values in a Banach space Y, and x_o is a point of $\mathcal{D}_f$.

(c_2) The operator f is Fréchet differentiable in the open ball $U = \{x \in X; \|x - x_o\| < s\}$ and continuous on the closure $\bar{U}$ of this ball.

(c_3) The linear operator $D_o = f'(x_o)$ is invertible and for all $x, y \in U$ we have

(5) $$\| D_o^{-1}(f'(x) - f'(y)) \| \leq k_o \| x - y \| .$$

(c_4) The following inequalities hold:

(6) $\|D_o^{-1} f(x_o)\| \leq r_o$,

(7) $2k_o r_o \leq 1$,

(8) $s \geq k_o^{-1}(1-\sqrt{1-2k_o r_o}) =: s_o$.

We note that in the above definition of $\mathcal{C}(k_o, r_o)$ the spaces X and Y are not fixed.

In fact the conditions defining the class $\mathcal{C}(k_o, r_o)$ represent an "affine invariant" version of the hypothesis of Kantorovich's theorem (see [1]). We may define a convergent iterative procedure for the class $\mathcal{C}(k_o, r_o)$ by associating with each $(f, x_o) \in \mathcal{C}(k_o, r_o)$ the iterative algorithm (F, x_o), with F given by (3). The recurrent scheme (2) becomes in this case

(9) $$x_{n+1} = x_n - f'(x_n)^{-1} f(x_n) \qquad n=0,1,2,\ldots$$

This iterative procedure will be called Newton's method for the class $\mathcal{C}(k_o, r_o)$ and will be denoted by $\mathcal{N}$.

Let us return now to our general discussion and suppose that we are given a convergent iterative procedure $\mathcal{F} : \mathcal{C} \to \mathcal{A}$. In this case we can associate with each pair $(f, x_o) \in \mathcal{C}$, a sequence (x_n) which converges to a root x^* of the equation $f(x)=0$. In order to emphasize the fact that this sequence is attached by the iterative procedure $\mathcal{F}$ to the pair (f, x_o), we shall write

(10) $$x_k = [\mathcal{F}(f,x_o)]^k \qquad k=0,1,2,\ldots$$

We want to find estimates of the distances $\|x_n - x^*\|$, $n=0,1,2,\ldots$ which should be valid for all pairs $(f,x_o)\in\mathcal{C}$. One way of doing this, is to determine a function $\alpha : \mathbb{Z}_+ \to \mathbb{R}_+$ such that

(11) $$\|x_n - x^*\| \leq \alpha(n)$$

for all $n\in\mathbb{Z}_+$, and for all $(f,x_o)\in\mathcal{C}$.

If

(12) $$\lim_{n\to\infty} \alpha(n) = 0$$

then (11) gives us the possibility of computing in advance the number of steps required to obtain any desired precision. However the error bounds given by (11) are in most cases very pessimistic. One could obtain better error bounds trying to find a sequence of functions $\gamma_n : \mathbb{R}_+^{n(n+1)/2} \longrightarrow \mathbb{R}_+$ such that

(13) $$\|x_n - x^*\| \leq \gamma_n(\|x_1 - x_o\|, \|x_2 - x_o\|, \ldots, \|x_n - x_{n-1}\|)$$

for all positive integers n and all $(f,x_o)\in\mathcal{C}$. Corresponding to (12) we require that

(14) $$\lim_{n\to\infty} \gamma_n(\|x_1 - x_o\|, \|x_2 - x_o\|, \ldots, \|x_n - x_{n-1}\|) = 0$$

for all $(f,x_o)\in\mathcal{C}$.

Let us note that the right-hand side of (13) can be computed only _after_ obtaining the points $x_1,\ldots,x_n$ via the iterative procedure. That's why (13) are called <u>aposteriori</u> error bounds in contrast with (11) which are called <u>apriori</u> error bounds. It is conceivable to obtain aposteriori error bounds using other "information" than the relative distances $\|x_i - x_j\|$, $0 \leq i < j \leq n$.

In what follows we shall give an example where it is impossible to find a function α verifying (11) and (12), but where we can obtain aposteriori error bounds using an adequate information. We consider the class

$$\mathscr{C}_o = \{(f,x_o); f:[0,1] \to \mathbb{R},\ x_o \in]0,1],\ f(0) \leq 0,\ f(x_o) > 0,\ f \text{ is twice differentiable on } [0,1],\ f' \text{ is increasing and } f'' \text{ is non-decreasing on }]0,1[\ \}.$$

It is easy to see that Newton's method represents a convergent iterative procedure for this class. If m is an integer greater than one, then the pair $(f_m, 1)$, where $f_m(t) = t^m$, obviously belongs to $\mathscr{C}_o$. For this pair the recurrent scheme (9) reduces to

$$x_{n+1}^{(m)} = \frac{m-1}{m} x_n^{(m)}, \qquad n=0,1,2,\ldots$$

It follows that $x_n^{(m)} = \left(\frac{m-1}{m}\right)^n$ for $n=0,1,2,\ldots$.

Let $\alpha : \mathbb{Z}_+ \to \mathbb{R}_+$ be a function satisfying (11) for all $(f,x_o) \in \mathscr{C}_o$. For each $n \in \mathbb{Z}_+$ we have

$$\alpha(n) \geq \sup_{(f,x_o)\in\mathscr{C}_o} \|x_n - x^*\| \geq \sup_{m \geq 2} \left(\frac{m-1}{m}\right)^n = 1$$

and so (12) cannot be satisfied.

Now, let (f,x_o) be a pair of $\mathscr{C}_o$. The recurrent scheme (9) yields a sequence (x_n) decreasingly convergent to a root x^* of the equation $f(x)=0$. Applying the mean theorem we have

$f(x_n)=f'(\xi)(x_n-x^*)$, $x^*<\xi<x_n$. Writing

$$\int_{x^*}^{\xi}(\xi-t)f''(\xi)dt \geq \int_{x^*}^{\xi}(f'(\xi)-f'(t))dt = \int_{\xi}^{x_n}(f'(t)-f'(\xi))dt \geq \int_{\xi}^{x_n}(t-\xi)f''(\xi)dt$$

we deduce that $\xi \geq (x_n+x^*)/2 \geq x_n/2$. Hence we deduce the following aposteriori error bounds for Newton's method in the class $\mathcal{C}_o$.

(15) $$|x_n-x^*| \leq \frac{f(x_n)}{f'(x_n/2)} , \qquad n=0,1,2,\ldots$$

It is easy to prove that

(15') $$\lim_{n\to\infty} \frac{f(x_n)}{f'(x_n/2)} = 0$$

for all $(f,x_o)\in \mathcal{C}_o$.

The estimate (15) uses the "information" represented by the values $f(x_n)$ and $f'(x_n/2)$. In what follows we shall define in a general way the notion of information attached to an iterative procedure and the notion of error estimate for a given convergent iterative procedure and a given information. Let c_o be the set of all sequences of real numbers converging to zero and let us denote

$$\mathcal{P}_m=\{(a_n)_{n\geq 0} \in c_o \; ; \; a_n=0 \text{ for all } n > m\}, \quad \mathcal{P}=\bigcup_{m\geq 0}\mathcal{P}_m$$

<u>Definition 3</u>. Let $\mathcal{F}:\mathcal{C}\to\mathcal{A}$ be a convergent iterative procedure for the class $\mathcal{C}$. A mapping $I:\mathbb{Z}_+ \times \mathcal{C}\to\mathcal{P}$ is called an information attached to the iterative procedure $\mathcal{F}$ if for each $n\in\mathbb{Z}_+$ there exists a $m_n\in\mathbb{Z}_+$ such that

$$\mathcal{G}_n := \{I(n,(f,x_0));\ (f,x_0)\in\mathcal{C}\}\subset\mathcal{P}_{m_n}.$$

For every n the mapping

$$I_n:\mathcal{C}\to\mathcal{G}_n,\ I_n(f,x_0)=I(n,\ (f,x_0))$$

is called the information at the n'th step.■

<u>Definition 4</u>. Let $\mathcal{F}:\mathcal{C}\to\mathcal{A}$ be a convergent iterative procedure and let $I:\mathbb{Z}_+\times\mathcal{C}\to\mathcal{P}$ be an information attached to it. Let $\mathcal{G}$ be the range of the information I (i.e. $\mathcal{G}=\bigcup_{n\geq 0}\mathcal{G}_n$). A mapping $\beta:\mathcal{G}\longrightarrow\mathbb{R}_+$ is called an error estimate for the iterative procedure $\mathcal{F}$ which uses the information I, if we have

(16) $$\|x_n-x^*\|\leq\beta(I_n(f,x_0)),\qquad n=1,2,\ldots$$

and

(17) $$\lim_{n\to\infty}\beta(I_n(f,x_0))=0$$

for all $(f,x_0)\in\mathcal{C}$.■

The set of all error estimates for the iterative procedure $\mathcal{F}$ which use the same information I, will be denoted by $\mathcal{E}(\mathcal{F},I)$. According to the above introduced terminology, a function $\alpha:\mathbb{Z}_+\to\mathbb{R}_+$ satisfying (11) and (12) is an error estimate for the iterative procedure $\mathcal{F}$ which uses the information $I^{(1)}$ given by

(18) $$I_n^{(1)}(f,x_0)=n\ .$$

We have seen that such an error estimate does not exist in case of Newton's method for the class $\mathcal{C}_o$. On the other hand (15) and (15') show that the function $\beta_o(s,t)=s/t$ is an error estimate for Newton's method in the class $\mathcal{C}_o$ which uses the informtion $I^{(0)}$, where :

$$I_n^{(0)}(f,x_o)=(f(x_n)\ ,\ f'(x_n/2))\ .$$

Now let us give some examples of information which will be used to obtain sharp error estimates for Newton's method in the class $\mathcal{C}(k_o,r_o)$. At the n'th step these informations can be written as follows:

(19) $$I_n^{(2)}(f,x_o)=\| x_n-x_{n-1} \| \ ,$$

(20) $$I_n^{(3)}(f,x_o)=(n,\ \|x_n-x_{n-1}\|)\ ,$$

(21) $$I_n^{(4)}(f,x_o)=(\ \|x_n-x_o\|\ ,\ \|x_n-x_{n-1}\|)\ ,$$

(22) $$I_n^{(5)}(f,x_o)=(\ \|x_n-x_o\|\ ,\ \|D_0^{-1}f(x_n)\|)\ ,$$

(23) $$I_n^{(6)}(f,x_o)=(k_{n-1}\ ,\|x_n-x_{n-1}\|),\ k_{n-1}=\sup_{x,y\in U}\frac{\|f'(x_{n-1})^{-1}(f'(x)-f'(y))\|}{\|x-y\|}\ .$$

In the above formulae (x_n) represents, of course, the sequence obtained from (f,x_o) via the recurrent scheme (9). (i.e. $x_k=[\mathcal{N}(f,x_o)]^k$). We see that the first three informations are formulated only in terms of the relative distances $\|x_i-x_j\|$. The general form of this type of information is

(24) $$I_n(f,x_o)=(n, \|x_1-x_o\|, \|x_2-x_o\|, \ldots, \|x_n-x_{n-1}\|).$$

$I_n^{(5)}(f,x_o)$ depends also upon the quantity $\|D_o^{-1}f(x_n)\|$. If we use Newton's method to solve a given equation then at the n'th step, that is at the moment when we have computed the points $x_1, x_2, \ldots, x_n$, the value $f(x_n)$ does not need to be computed yet. If we don't want to go further (i.e. to compute x_{n+1}) we don't have to compute $f(x_n)$ at all. In this case for obtaining the information (22) we have to do some little extra work. Concerning the quantity k_{n-1} appearing in (23), let us remark that at the n'th step $f'(x_{n-1})$ has already been computed. However with the exception of the scalar case, where we can take $k_{n-1}=$ $=k_o|f'(x_o)|/|f'(x_{n-1})|$, the cost of obtaining k_{n-1} might be very high.

After these comments let us present some error estimates for Newton's method in the class $\mathscr{C}(k_o,r_o)$, which use the informations described above.

We denote with $\mathscr{J}^{(k)}$ the range of the information $I^{(k)}$. To simplify the formulae we introduce the constants

(25) $$a=k_o^{-1}(1-2k_or_o)^{1/2},$$

(26) $$b=\frac{1-k_o(a+r_o)}{k_or_o},$$

and the sequence

(27) $$s_n=\begin{cases} 2ab^{2^n}/(1-b^{2^n}) & \text{if } a>0 \\ k_o^{-1}2^{-n} & \text{if } a=0 \end{cases}.$$

With the above notation we define the functions :

(28) $\beta_1 : \mathcal{G}^{(1)} \to \mathbb{R}_+ , \quad \beta_1(n) = s_n$

(29) $\beta_2 : \mathcal{G}^{(2)} \to \mathbb{R}_+ , \quad \beta_2(r) = (a^2+r^2)^{1/2} - a$

(30) $\beta_3 : \mathcal{G}^{(3)} \to \mathbb{R}_+ , \quad \beta_3(n,r) = r^2/(s_n+2a)$

(31) $\beta_3^* : \mathcal{G}^{(3)} \to \mathbb{R}_+ , \quad \beta_3^*(n,r) = rs_n/(s_{n-1}-s_n)$

(32) $\beta_4 : \mathcal{G}^{(4)} \to \mathbb{R}_+ , \quad \beta_4(q,r) = k_o^{-1} - q - [(k_o^{-1}-q)^2 - r^2]^{1/2}$

(33) $\beta_5 : \mathcal{G}^{(5)} \to \mathbb{R}_+ , \quad \beta_5(q,t) = k_o^{-1} - q - [(k_o^{-1}-q)^2 - 2k_o^{-1}t]^{1/2}$

(34) $\beta_6 : \mathcal{G}^{(6)} \to \mathbb{R}_+ , \quad \beta_6(k,r) = k^{-1}[1-kr-(1-2kr)^{1/2}]$,

In Section 3 we shall prove that the above defined functions are error estimates for Newton's method in the class $\mathcal{C}(k_o,r_o)$ in the sense of Definition 4, i.e.

(35) $\beta_j \in \mathcal{E}(\mathcal{N}, I^{(j)})$, $j=1,2,\ldots,6$; $\beta_3^* \in \mathcal{E}(\mathcal{N}, I^{(3)})$.

With the exception of β_4 and β_5 all the other error estimates are known (eventually under a different formulation). Thus β_1 and β_3^* have been obtained by Gragg and Tapia [2] , β_3 is due to Miel ([4], [5]), β_2 to Potra and Pták [10] and β_6 to Kornstaedt [3]. We note that another form of β_1 had been found, under somewhat different conditions, by Ostrowski([8],[9]). All these error estimates are sharp in the sense of the following definition:

Definition 5. Let $\mathcal{F}$ be a convergent iterative procedure for a class $\mathcal{C}$ and I an information attached to it. An error estimate $\beta \in \mathcal{E}(\mathcal{F}, I)$ is called sharp if there exists a pair $(f, x_o) \in \mathcal{C}$ such that $\beta(I_n(f, x_o)) = \| x_n - x^* \|$ for all n. ∎

In the next definition we give a natural criterium for comparing two different estimates of a given iterative procedure.

Definition 6. Let $\mathcal{F}$ be a convergent iterative procedure for a class $\mathcal{C}$ and let I, I^* be two informations attached to it. Consider two error estimates $\beta \in \mathcal{E}(\mathcal{F}, I)$ and $\beta^* \in \mathcal{E}(\tilde{\mathcal{F}}, I^*)$. We say that β is more accurate than β^* and denote this by writing $\beta \prec \beta^*$ if

$$\beta(I_n(f, x_o)) \leq \beta^*(I_n^*(f, x_o)) \tag{36}$$

for all positive integers n and all (f, x_o) from $\mathcal{C}$. ∎

In Section 3 we shall prove that between the error estimates for Newton's method in the class $\mathcal{C}(k_o, r_o)$ there exist the following relations

$$\beta_1 \succ \beta_3^* \succ \beta_2 \succ \beta_3 \succ \begin{matrix} \beta_5 \\ \curlywedge \\ \beta_4 \\ \curlyvee \\ \beta_6 \end{matrix} . \tag{37}$$

We see that β_4 is the most accurate among the error estimates using an information of type (24). This fact strongly recommends its usefulness in numerical applications.

2. Nondiscrete induction

In the proof of the main theorem from the next section we shall use the method of nondiscrete mathematical induction of V.Pták. For the motivation and the general principles of this method see Pták [11] and [13]. The method of nondiscrete induction is based on the notion of rate of convergence, or small function.

Definition 7. Let T denote either the set of all positive real numbers or a half-open interval of the form $]0,t_o]$. A function $w:T \rightarrow T$ is called a rate of convergence (or small function) on T if the series

$$t+w(t)+w(w(t))+ \dots \tag{38}$$

is convergent for all $t \in T$.∎

For the sake of simplicity it will be convenient to denote by w^n the n'th iterate of w in the sense of usual function composition (i.e. $w^o(t)=t$; $w^{n+1}(t)=w(w^n(t))$, $n=0,1,2,\dots$). Let us denote by s(t) the sum of the series (38). The functions w and s are obviously connected by the following functional relation

$$s(t)=t+s(w(t)) \tag{39}$$

It turns out that this relation caracterizes in some sense the rates of convergence.

<u>Proposition 1</u>. If $w:T \longrightarrow T$ and $h:T \longrightarrow \mathbb{R}_+$ are two functions satisfying the relation

$$h(t)=t+h(w(t))$$

for all $t \in T$, then w is a rate of convergence on T.

Moreover, if the limit $h_o=\lim\limits_{t\downarrow 0} h(t)$ exists then we have $s(t)=\sum\limits_{n\geqslant 0} w^n(t)=h(t)-h_o$.

<u>Proof</u>. For each $t\in T$ and each positive integer n we have

$$t+w(t)+\dots+w^n(t)=h(t)-h\ w^{n+1}(t))\leqslant h(t) \quad .■$$

Using the above proposition it is easy to check that, for any given constant a, the function

$$w(t)=\frac{t^2}{2(t^2+a^2)^{1/2}} \tag{40}$$

is a rate of convergence on the semi-axis $T=\{t;t>0\}$ and the sum of its iterates is

$$s(t)=t-a+(t^2+a^2)^{1/2} \quad . \tag{41}$$

Let us state now the induction theorem.

<u>Proposition 2</u>. Let (X,d) be a complete metric space and let $\mathcal{D}_F$ be a subset of X. Consider a mapping $F: \mathcal{D}_F \to X$ and a point $x_o \in \mathcal{D}_F$. If we can attach to the pair (F,x_o) a rate of

convergence w on an interval T and a family of subsets $Z(t) \subset \mathcal{D}_F$, $t \in T$, such that

(i) $x_o \in Z(r_o)$ for some $r_o \in T$,

(ii) $F(x) \in Z(w(t))$ and $d(F(x),x) \leqslant t$ for all $t \in T$ and $x \in Z(t)$,

then:

1° (F,x_o) is a well defined iterative algorithm and the sequence (x_n) produced by it converges to a point $x^* \in \mathcal{D}_F$;

2° the following relations are satisfied for all $n \in \mathbb{Z}_+$:

(42) $$x_n \in Z(w^n(r_o)),$$

(43) $$d(x_n, x_{n+1}) \leqslant w^n(r_o),$$

(44) $$d(x_n, x^*) \leqslant s(w^n(r_o)) \quad . \blacksquare$$

The proof of the above proposition is very simple and it will be left as an exercise. The interested reader could consult, for example, [10].

The inequalities (44) give the possibility of obtaining apriori error estimates for iterative procedures. These estimates have the following property: if they are attained at a certain step then they will be attained for all subsequent steps. More precisely we have

Proposition 3. Under the hypotheses of Proposition 2, suppose that equality is attained in (44) for a certain n_o. Then equality will be attained in (43) and (44) for all $n \geqslant n_o$.

Proof. Because $d(x_{k+1},x_k)\leq w^k(r_o)$, for all $k\geq 0$, we can write

$$\sum_{k\geq n_o} w^k(r_o)=s(w^{n_o}(r_o))=d(x_{n_o},x^*)\leq \sum_{k\geq n_o} d(x_{k+1},x_k)\leq \sum_{k\geq n_o} w^k(r_o) .$$

It follows that $d(x_{k+1},x_k)=w^k(r_o)$ for all $k\geq n_o$. The proof is complete.■

3. Error bounds for Newton's method

In this section we shall study the convergence of Newton's method in the class $\mathcal{C}(k_o,r_o)$. We shall use the following two lemmas:

Lemma 1. If $(f,x_o)\in\mathcal{C}(k_o,r_o)$ then the following inequality holds for all $x,y\in\bar{U}$:

$$\| D_o^{-1}[f(x)-f(y)-f'(y)(x-y)]\| \leq \tfrac{1}{2}k_o\| x-y\|^2 . \tag{45}$$

Proof. Use the integral representation

$$f(x)-f(y)=\int_0^1 f'(y+t(x-y))(x-y)dt \tag{46}$$

and then apply condition (5). (see also [7, 3.2.12]). ■

Lemma 2. Let k_o and r_o be two positive numbers satisfying inequality (7) and consider the constant a from (25). If (s_n) is the sequence given in (27) and w,s the functions defined in (40), (41), then the following relations hold for n=0,1,2,...

(47) $s_o=k_o^{-1}-a, \quad s_{n+1}=s_n^2/(2s_n+2a)$

(48) $w^n(r_o)=s_n-s_{n+1}$

(49) $s(w^n(r_o))=s_n$

Proof. (47) can be verified directly while (48) and (49) can easily be proved by induction. ∎

Now we may state

Theorem 1. If $(f,x_o)\in \mathcal{E}(k_o,r_o)$ then the iterative algorithm (9) is well defined, the sequence (x_n) produced by it converges to a root x^* of the equation $f(x)=0$ and the following inequalities are satisfied for n=0,1,2,... .

(50) $$\|x_n-x_{n+1}\| \leqslant s_n-s_{n+1},$$

(51) $$\|x_n-x^*\| \leqslant s_n,$$

where (s_n) is the sequence given by (27).

Proof. We shall use Proposition 2 with w and s given by (40) and (41). First let us observe that with the constant a given by (25) we have $s(r_o)=s_o$. Hence the closed ball with center x_o and radius $s(r_o)$ is included in U (see (8)). Let $\mathcal{D}_F$ be the set of those $x\in U$ for which the linear operator $f'(x)$ is invertible and let us consider the mapping $F: \mathcal{D}_F \to Y$ given by (3). For any $r > 0$ let us define

(52) $$Z(r)=\{x\in X;\ \|x-x_o\|\leqslant s(r_o)-s(r),\ f'(x) \text{ is invertible},\ \|f'(x)^{-1}f(x)\|\leqslant r\}.$$

We have obviously $Z(r_o)=\{x_o\}$ so that hypothesis (i) of Proposition 2 is satisfied. Suppose now $x\in Z(r)$ and denote $y=F(x)$. It follows immediately that

(53) $$\|y-x\| = \|f'(x)^{-1}f(x)\| \leqslant r .$$

Using (39) we obtain

(54) $$\|y-x_o\| \leqslant \|y-x\|+\|x-x_o\|\leqslant r+s(r_o)-s(r)=s(r_o)-s(w(r)).$$

In particular this shows that $y\in U$. Applying (5) we have

$$\|D_o^{-1}(D_o-f'(y))\|\leqslant k_o\|y-x_o\|\leqslant k_o[k_o^{-1}-a-s(w(r))]=1-k_o(a^2+r^2)^{1/2} .$$

According to Banach's Lemma it follows that the linear operator $f'(y)$ is invertible and that

$$\|(D_o^{-1}f'(y))^{-1}\|\leqslant k_o^{-1}(r^2+a^2)^{-1/2} .$$

$y=F(x)$ implies $f(x)+f'(x)(y-x)=0$; hence, using Lemma 1, we obtain

$$\|D_o^{-1}f(y)\| = \|D_o^{-1}(f(y)-f(x)-f'(x)(y-x))\|\leqslant 2^{-1}k_o\|y-x\|^2\leqslant 2^{-1}k_o r^2 .$$

From the last inequalities it follows that

$$(55)\quad \|f'(y)^{-1}f(y)\| \leq \|(D_o^{-1}f'(y))^{-1}D_o^{-1}f(y)\| \leq 2^{-1}r^2(r^2+a^2)^{-1/2}=w(r)$$

Now (53)-(55) show that hypothesis (ii) of Proposition 2 is also satisfied, and by virtue of Lemma 2 the proof of the theorem is complete. ∎

From the above theorem it follows, using the terminology introduced in Section 1, that Newton's method ($\mathcal{N}$) is a convergent iterative procedure for the class $\mathcal{C}(k_o,r_o)$. It also follows that the function β_1 defined by (28) is an error estimate for this process ($\beta_1 \in \mathcal{E}(\mathcal{N},I^{(1)})$). In the following proposition we shall prove that this estimate is sharp in the sense of Definition 5.

<u>Proposition 4</u>. For any pair of positive numbers k_o,r_o satisfying the inequality (7) there exists a function $f:\mathbb{R}\longrightarrow\mathbb{R}$ and a point $x_o\in\mathbb{R}$ having the property that $(f,x_o)\in\mathcal{C}(k_o,r_o)$ and for which the estimates (50) and (51) are attained for all n.

<u>Proof</u>. Take $f(x)=2^{-1}k_o(x^2-a^2)$ and $x_o=k_o^{-1}$, where the constant a is given by (25). We have obviously $(f,x_o)\in\mathcal{C}(k_o,r_o)$ and $x_o-a=s_o$. The rest follows from Proposition 3 with $n_o=0$. ∎

Using Theorem 1 we can prove that the function β_6 given by (34) is also an error estimate for Newton's method in the class $\mathcal{C}(k_o,r_o)$. ($\beta_6\in\mathcal{E}(\mathcal{N},I^{(6)})$).

Proposition 5. Under the hypotheses of Theorem 1 the following inequality holds for n=1,2,3,...

$$\|x_n - x^*\| \leq k_{n-1}^{-1}\left[1 - k_{n-1}\|x_n - x_{n-1}\| - (1 - 2k_{n-1}\|x_n - x_{n-1}\|\ )^{1/2}\ \right] \tag{56}$$

where

$$k_{n-1} = \sup_{x,y\in U} \frac{\|f'(x_{n-1})^{-1}(f'(x) - f'(y))\|}{\|x-y\|} .$$

Proof. Consider a pair $(f, x_o) \in \mathcal{C}(k_o, r_o)$ and denote $r_{n-1} = \|f'(x_{n-1})^{-1} f(x_{n-1})\|$. We want to prove that $(f, x_{n-1}) \in \mathcal{C}(k_{n-1}, r_{n-1})$. This reduces to the demonstration of the relation

$$2\,k_{n-1} r_{n-1} \leq 1 . \tag{57}$$

From Theorem 1 it follows $r_{n-1} = \|x_n - x_{n-1}\| \leq s_{n-1} - s_n$, so that we have:

$$\|D_o^{-1}(f'(x_o) - f'(x_{n-1}))\| \leq k_o\|x_{n-1} - x_o\| \leq k_o \sum_{j=0}^{n-2} \|x_{j+1} - x_j\| \leq k_o(s_o - s_{n-1}) .$$

According to Banach's lemma this implies that

$$\|(D_o^{-1} f'(x_{n-1}))^{-1}\| \leq k_o^{-1}(a + s_{n-1})^{-1} .$$

Hence

$$k_{n-1} = \sup_{x,y\in U} \frac{1}{\|x-y\|} \|(D_o^{-1} f'(x_{n-1}))^{-1} D_o^{-1}(f'(x) - f'(y))\|$$

$$\leq \|(D_o^{-1} f'(x_{n-1}))^{-1}\| \sup_{x,y\in U} \frac{1}{\|x-y\|} \| D_o^{-1}(f'(x) - f'(y))\| \leq (a + s_{n-1})^{-1} .$$

Using Lemma 2 we obtain

$$2k_{n-1}r_{n-1} \leqslant 2\frac{s_{n-1}-s_n}{a+s_{n-1}} = 2\frac{w^{n-1}(r_o)}{a+s(w^{(n-1)}(r_o))} =$$

$$=2\frac{w^{n-1}(r_o)}{w^{n-1}(r_o)+[(w^{n-1}(r_o))^2+a^2]^{1/2}} \leqslant 1 .$$

Denote now $a_{n-1}=k_{n-1}^{-1}(1-2k_{n-1}r_{n-1})^{1/2}$. Applying Theorem 1 to the pair $(f,x_{n-1})\in\mathcal{C}(k_{n-1},r_{n-1})$ we deduce the inequality

$$\| x_n-x^* \| \leqslant k_{n-1}^{-1}-a_{n-1}-r_{n-1} ,$$

which is exactly the error bound (56) .■

In the next proposition we shall prove that $\beta_5 \in \mathcal{E}(\mathcal{N}, I^{(5)})$.

Proposition 6. Under the hypotheses of Theorem 1 the following inequality holds for n=0,1,2,...

(58) $$\|x_n-x^*\| \leqslant k_o^{-1}-\|x_n-x_o\| - [(k_o^{-1}-\|x_n-x_o\|)^2-2k_o^{-1}\|D_o^{-1}f(x_n)\|]^{1/2} .$$

Proof. First let us remark that

(59) $$f(x_n)=f(x_n)-f(x^*)=(\int_0^1 f'(x^*+t(x_n-x^*))dt)(x_n-x^*) .$$

We want to prove that the linear operator $A=\int_0^1 f'(x^*+t(x_n-x^*))dt$ is invertible. To this effect we note that according to (5) we have:

$$\|D_o^{-1}(f'(x_o)-A)\| \leqslant \tfrac{1}{2}k_o(\|x_n-x_o\|+\|x^*-x_o\|) \leqslant \tfrac{1}{2}k_o(2\|x_n-x_o\|+\|x_n-x^*\|) .$$

From Theorem 1 it follows that

$$2\cdot\|x_n-x_o\|+\|x_n-x^*\| < 2(\|x_n-x_o\|+\|x_n-x^*\|) \leqslant 2s_o \leqslant 2k_o^{-1} .$$

By virtue of Banach's lemma, A is invertible and the following norm estimation holds:

(60) $$\|(D_o^{-1}A)^{-1}\| \leqslant \left[1-\tfrac{1}{2}k_o(2\|x_n-x_o\|+\|x_n-x^*\|)\right]^{-1}$$

Finally from (59) and (60) we deduce that

$$\|x_n-x^*\| = \|A^{-1}f(x_n)\| = \|(D_o^{-1}A)^{-1}D_o^{-1}f(x_n)\| \leqslant \frac{2\|D_o^{-1}f(x_n)\|}{2-2k_o\|x_n-x_o\|-k_o\|x_n-x^*\|}$$

and it is easy to see that this relation implies (58) . ∎

Now we are able to prove that all the functions defined in (28)-(34) are sharp error estimates for Newton's procedure in the class $\mathcal{C}(k_o,r_o)$. We recall the fact that for each k=1,2,...,6 β_k uses the information $I^{(k)}$ and β_3^* uses the information $I^{(3)}$ (these informations are defined by the formulae (18)-(23)).

<u>Theorem 2</u>. The functions β_1, β_2, β_3, β_3^*, β_4, β_5 and β_6 defined by the relations (28)-(34) are sharp error estimates for Newton's method in the class $\mathcal{C}(k_o,r_o)$ and the relation "$\prec$" introduced in Definition 6 orders them as shown in diagram (37).

<u>Proof</u>. We have already proved that $\beta_j \in \mathcal{E}(\mathcal{N},I^{(j)})$ for

$j=1,5,6$ (see Theorem 1, Proposition 5 and Proposition 6). Moreover we have shown that the error estimate β_1 is sharp (see Proposition 4). It follows that the proof of our theorem would be complete if we could demonstrate the validity of the relations indicated in diagram (37).

Let us consider a pair $(f,x_o)\in \mathcal{C}(k_o,r_o)$. We denote by (x_n) the sequence generated by Newton's method applied to this pair and by x^* its limit. For proving the relation $\beta_4 \succ \beta_6$ we have to evaluate the quantity

$$k_{n-1}=\sup_{x,y\in U}\frac{\| f'(x_{n-1})^{-1}(f'(x)-f'(y))\|}{\| x-y \|}\,.$$

Using the identity

$$f'(x_{n-1})^{-1}(f'(x)-f'(y))=(f'(x_o)^{-1}f'(x_{n-1}))^{-1}f'(x_o)^{-1}(f'(x)-f'(y))$$

and condition (5) it is easy to prove that

$$k_{n-1}\leq k_o/(1-k_o\| x_{n-1}-x_o\|)\,.$$

Denoting by $\tilde{k}_{n-1}$ the right hand side of the above inequality we have

$$\beta_6(k_{n-1},\|x_n-x_{n-1}\|)\leq \beta_6(\tilde{k}_{n-1},\|x_n-x_{n-1}\|)=k_o^{-1}-\|x_n-x_{n-1}\|$$

$$-\|x_{n-1}-x_o\|-\left[(k_o^{-1}-\|x_n-x_{n-1}\|-\|x_{n-1}-x_o\|)^2-\|x_n-x_{n-1}\|^2\right]^{1/2}$$

$$\leq k_o^{-1}-\|x_n-x_o\|-\left[(k_o^{-1}-\|x_n-x_o\|)^2-\|x_n-x_{n-1}\|^2\right]^{1/2}=\beta_4(\|x_n-x_o\|,\|x_n-x_{n-1}\|)\,.$$

The relation $\beta_4 \succ \beta_5$ reduces to the inequality

$$2k_o^{-1}\|D_o^{-1}f(x_n)\| \leqslant \|x_n-x_{n-1}\|^2$$

which can be obtained noticing that $f(x_n)=f(x_n)-f(x_{n-1})-$ $-f'(x_{n-1})(x_n-x_{n-1})$ and applying Lemma 1.

In order to prove the relation $\beta_3 \succ \beta_4$ we have to show that

$$(61) \qquad k_o^{-1}-\|x_n-x_o\|+\left[(k_o^{-1}-\|x_n-x_o\|^2)-\|x_n-x_{n-1}\|^2\right]^{1/2} \geqslant s_n+2a$$

Using (50) we get

$$k_o^{-1}-\|x_n-x_o\|+\left[(k_o^{-1}-\|x_n-x_o\|)^2-\|x_n-x_{n-1}\|^2\right]^{1/2}$$

$$\geqslant k_o^{-1}-(s_o-s_n)+\left[(k_o^{-1}-s_o+s_n)^2-(s_{n-1}-s_n)^2\right]^{1/2} .$$

The right hand side of the above inequality equals s_n+2a because, according to Lemma 2, we have $s_o=k_o^{-1}-a$ and $(s_n+a)^2-(s_n-s_{n-1})^2=a^2$. The relation $\beta_2 \succ \beta_3$ can be proved immediately observing that

$$(\|x_n-x_{n-1}\|^2+a^2)^{1/2}-a=\frac{\|x_n-x_{n-1}\|^2}{a+(\|x_n-x_{n-1}\|^2+a^2)^{1/2}}$$

$$\geqslant \frac{\|x_n-x_{n-1}\|^2}{a+\left[(s_{n-1}-s_n)^2+a^2\right]^{1/2}} = \frac{\|x_n-x_{n-1}\|^2}{s_n+2a} .$$

Using the fact that the function $\varphi(t)=t/\left[a+(t^2+a^2)^{1/2}\right]$ is increasing in t we can prove in a simple way the relation $\beta_3^* \succ \beta_2$. Indeed we have

$$\frac{\|x_n-x_{n-1}\|^2}{a+(\|x_n-x_{n-1}\|^2+a^2)^{1/2}} \leqslant \frac{s_{n-1}-s_n}{a+\left[(s_{n-1}-s_n)^2+a^2\right]^{1/2}}\|x_n-x_{n-1}\| =$$

$$=\frac{s_n}{s_{n-1}-s_n}\|x_n-x_{n-1}\| .$$

The relation $\beta_1 \succ \beta_3^*$ is obvious because $\|x_n-x_{n-1}\| \leqslant s_{n-1}-s_n$. The proof is complete.■

Let us test now the error estimates presented above on a very simple example proposed in 10 and used also in 6 . Namely we consider the scalar cubic $f(x)=\frac{1}{3}(x^3-1)$. Taking $x_o=1.3$ we have $f(x_o)/f'(x_o)=r_o=0.2360946745$. With $s=2r_o$ we obtain $k_o=$ $=2.0972655019$. It can easily be checked that (f,x_o) (k_o,r_o).

We have performed four steps of the iterative algorithm (9) in double precision on a CDC-3500.

In the following table we give the results within a precision of $\frac{1}{2}10^{-10}$.

n	1	2	3	4
x_n	1.0639053254	1.0037617275	1.0000140800	1.0000000002
β_1	0.1937717784	0.0779910691	0.0243428971	0.0041562278
β_2	0.1937717784	0.0293510766	0.0001493512	0.0000000021
β_3	0.1937717784	0.0210451135	0.0001187900	0.0000000020
β_3^*	0.1937717784	0.0405133423	0.0017004978	0.0000028989
β_4	0.1937717784	0.0103103864	0.0000397184	0.0000000006
β_5	0.1009636891	0.0059993187	0.0000224673	0.0000000003
β_6	0.1937717784	0.0070741048	0.0000250351	0.0000000004

REFERENCES

[1] DEUFLHARD, P. and HEINDL, G., Affine invariant convergence theorems for Newton's method and extensions for related methods. SIAM.J.Numer.Anal., 16, 1 (1979), pp.1-10.

[2] GRAGG, W.B. and TAPIA, R.A., Optimal error bound for the Newton-Kantorovich theorem, SIAM.J.Numer.Anal., 11, 1 (1974), pp.10-13.

[3] KORNSTAEDT, H-J., Funktionalungleichungen und Iterationsverfahren, Aequationes Math. 13 (1975), pp.21-45.

[4] MIEL, G.J., The Kantorovich theorem with optimal error bounds, Amer.Math.Monthly, 86 (1979), pp.212-215.

[5] MIEL, G.J., Majorizing sequences and error bounds for iterative methods, Math.Comp., 34 (1980), pp.185-202.

[6] MIEL, G.J., An updated version of the Kantorovich theorem for Newton's methos. MRC Technical Summary Report, Univ. of Wisconsin-Madison, 1980.

[7] ORTEGA, J.M. and RHEINBOLDT, W.C., Iterative solution of nonlinear equations in several variables, Academic Press, N.Y., 1970.

[8] OSTROWSKI, A.M., La méthod de Newton dans les espaces de Banach, C.R.Acad.Sci Paris 272 (A), (1971), pp.1251-1253.

[9] OSTROWSKI, A.M., Solution of equations in Euclidian and Banach spaces, Academic Press, N.Y. 1973.

[10] POTRA, F-A. and PTÁK, V., Sharp error bounds for Newton's process. Numer.Math., 34 (1980), 63-72.

[11] PTÁK, V., Nondiscrete mathematical induction and iterative existence proofs, Linear algebra and its applications, 13 (1976), 223-236.

[12] PTÁK, V., The rate of convergence of Newton's process. Numer. Math., 25 (1976), 279-285.

[13] PTÁK, V., Nondiscrete mathematical induction in: General Topology and its Relations to Modern Analysis and Algebra IV, pp.166-178, Lecture Notes in Mathematics 609, Springer, Berlin 1977.