

ON THE SECANT METHOD AND NONDISCRETE
MATHEMATICAL INDUCTION

IOANNIS K. ARGYROS

(Las Cruces)

Abstract. The method of nondiscrete mathematical induction is used to find error bounds for the Secant method. We assume only that the operator has Hölder continuous derivatives. In case the Fréchet-derivative of the operator satisfies a Lipschitz condition our results reduce to the ones obtained by F. Potra (Num. Math. 1982).

Introduction. Consider the equation

$$(1) \quad f(x) = 0$$

where f is a nonlinear operator mapping a subset E_f of a Banach space E_1 into another Banach space E_2 .

Here we are concerned with finding solutions of (1) using the secant iterations

$$(2) \quad x_{n+1} = x_n - \delta f(x_{n-1}, x_n)^{-1} f(x_n)$$

$$(3) \quad x_{n+1} = x_n - \delta f(x_{-1}, x_0)^{-1} f(x_n)$$

where x_{-1} and x_0 are two points in the domain of f , and δf is a consistent approximation of f' .

This work is based upon the elegant work of F. Potra included in [4] concerning the error analysis of the Secant method. One of Potra's basic assumptions is the fact that essentially the linear operator f' is Lipschitz continuous. However in the presence of some interesting examples (see part (III)), where f' is only Hölder continuous we extend most of the results contained in [4] for the iteration (3). We leave the extension of the results for (2) to the motivated reader.

We furnish two examples in part (III) to show that our results can be applied whereas the equivalent results in [4] cannot.

Since our results are drawn almost in the same lines with the ones in [4], we will need to restate some here.

I. Preliminaries. Consider a class \mathcal{C} of pairs (f, v_0) where f is as above and $v_0 = (x_{-k+1}, \dots, x_0)$ is a system of k points from E_f . We want to attach to each pair $(f, v_0) \in \mathcal{C}$ a sequence $\{x_n\}$, $n = 0, 1, 2, \dots$ of points of E_f converging to a root x^* of (1). To achieve this we associate with

the pair (f, v_0) an operator $F: E \subset E_f^p \rightarrow E_1$, where $k \geq p$ and try to obtain a sequence $\{x_n\}$, $n = 0, 1, 2, \dots$ by the scheme:

$$(4) \quad x_{n+1} = F(x_{n-p+1}, \dots, x_n), \quad n = 0, 1, 2, \dots$$

The above scheme will yield a sequence $\{x_n\}$, $n = 0, 1, 2, \dots$, if $u_0 = (x_{-p+1}, \dots, x_0)$ is an admissible system of starting points in the sense given by the following definition:

DEFINITION 1. Consider an operator $F: E \subset E_1^p \rightarrow E_1$ and define recursively

$$\tilde{E}_0 = E, \quad \tilde{E}_{n+1} = \{u = (y_1, \dots, y_p) \in \tilde{E}_n; (y_2, \dots, y_p, F(u)) \in \tilde{E}_n\}, \quad n = 0, 1, 2.$$

Any $u_0 \in E_\infty = \bigcap_{n \geq 0} \tilde{E}_n$ will be called an admissible system of starting points for the scheme (4).

If u_0 is an admissible system of starting points for the scheme (4) we shall say that (4) is well defined.

DEFINITION 2. Let C be a class of pairs (f, v_0) where f is a nonlinear operator defined on a subset E_f of a Banach space E_1 with values in a Banach space E_2 , and $v_0 = (x_{-k+1}, \dots, x_0) \in E$. Let $p \leq k$. By an iterative procedure of type $(p; 1)$ for the class C , we mean an application which associates with any $(f, v_0) \in C$ an operator $F: E \subset E_f^p \subset E_1$ having the following two properties:

- (i) $u_0 = (x_{-p+1}, \dots, x_0)$ is an admissible system of starting points for the scheme (4);
- (ii) the sequence $\{x_n\}$, $n = 0, 1, 2, \dots$ given by (4) converges to a root x^* of (1).

Having an iterative procedure of type $(p; 1)$ for the class C it is important to find a function $\alpha: \mathbf{Z}_+ \rightarrow \mathbf{R}_+$ and a function $\beta: \mathbf{R}_+^p \rightarrow \mathbf{R}_+$ such that the following inequalities are satisfied

$$(5) \quad d(x_n, x^*) \leq \alpha(n)$$

$$(6) \quad d(x_n - x^*) \geq \beta(d(x_{n-p+1}, x_{n-p}), \dots, d(x_p, x_{p-1}))$$

for every pair $(f, x_0) \in C$ and every positive integer n .

The inequalities (5) are called a priori estimates because the right hand side can be computed before obtaining the points x_1, \dots, x_n via (4), while the inequalities (6) are called a posteriori estimates because their right hand side can be computed only after obtaining these points.

The estimates (5) and/or (6) will be called sharp if there exists a pair $(f, u_0) \in C$ for which these estimates are attained for all $n = 1, 2, 3, \dots$.

In the study of (4) we use the nondiscrete mathematical induction. The method was initiated by V. Pták by refining the closed graph theorem [3], [8]. V. Pták used this method to investigate iterative algorithms of type (4) with $p = 1$. In [3] the method was extended for any p .

Here we restate the results obtained in the above mentioned paper. Let T denote either the set of all positive numbers, or an interval of the form $(0, b] = \{x \in \mathbf{R}; 0 < x \leq b\}$. Let ω be a mapping of the cartesian product T^p into T and let us consider the "iterates" $\omega^{(n)}$ of ω given for each $t = (t_1, \dots, t_p) \in T^p$ by the following scheme:

(7) $\omega^{(0)}(t) = t_p, \omega^{(n+1)}(t) = \omega^{(n)}(t_2, \dots, t_p, \omega(t)), \quad n = 0, 1, 2, \dots$

DEFINITION 3. A mapping $\omega: T^p \rightarrow T$, with the above iteration law, is called a rate of convergence of type $(p; 1)$ on T , if the series

$$(8) \quad \sigma(t) = \sum_{n=0}^{\infty} \omega^{(n)}(t)$$

is convergent for all $t \in T^p$.

From now on F will be a mapping of E into E_3 , where E_3 is a complete metric space, and E a subset of the cartesian product E_3^p . We attach to F the mapping $\bar{F}: E \rightarrow E_3^p$, defined for every $u = (y_1, \dots, y_p) \in E$ by

$$(9) \quad \bar{F}(u) = (y_2, \dots, y_p, F(u)).$$

Denoting $u_n = (x_{n-p+1}, \dots, x_n)$ we have

$$(10) \quad u_{n+1} = \bar{F}(u_n), \quad n = 0, 1, 2, \dots$$

Similarly we attach to ω the mapping $\bar{\omega}: T^p \rightarrow T^p$ defined by

$$(11) \quad \bar{\omega}(t) = (t_2, \dots, t_p, \omega(t)), \quad t = (t_1, \dots, t_p) \in T^p.$$

Denote by $\bar{\omega}^{(n)}$ the iterates of $\bar{\omega}$ in the sense of the usual composition of functions, that is

$$\bar{\omega}^{(0)}(t) = t, \quad \bar{\omega}^{(n+1)}(t) = \bar{\omega}(\bar{\omega}^{(n)}(t)).$$

Then (7) becomes

$$(12) \quad \omega^{(0)}(t) = t_p, \quad \omega^{(n+1)}(t) = \omega(\bar{\omega}^{(n)}(t)).$$

Finally, we introduce the notation

$$\beta(t) = \sigma(t) - t_p.$$

From (8) and (11) it follows that

$$\beta(t) = \sigma(\bar{\omega}(t)).$$

With the above notation we can state the following proposition whose proof can be found in [3] or [4].

PROPOSITION 1. Let E_3 be a complete metric space and let E be a subset of E_3^p . Let us consider the operators $F: E \rightarrow E_3$ and $Z: T^p \rightarrow \exp(E)$, where $\exp(E)$ denotes the class of all subsets of E . Let ω be a rate of convergence of type $(p; 1)$ on T .

If there exists $u_0 = (x_{-p+1}, \dots, x_0) \in E$ and $t_0 \in T^p$ such that

$$(13) \quad u_0 \in Z(t_0)$$

and if the relations

$$(14) \quad F(u) \in Z(\bar{\omega}(t)),$$

(15) $d(T^p(u), y_p) \leq t_p$
 are satisfied for all $t = (t_1, \dots, t_p) \in T^p$ and $u = (y_1, \dots, y_p) \in Z(t)$,
 then:

- (i) the iteration (4) is well defined.
 (ii) There exists an $x^* \in E_2$ such that $x^* = \lim_{n \rightarrow \infty} x_n$.
 (iii) The following relations are satisfied for all $n = 0, 1, 2, \dots$

$$(16) \quad u_n \in Z(\bar{\omega}^{(n)}(t_0)),$$

$$(17) \quad d(x_{n+1}, x_n) \geq \omega^{(n)}(t_0),$$

$$(18) \quad d(x_n, x_0) \leq \sigma(t_0) - \sigma(\bar{\omega}^{(n)}(t_0)),$$

$$(19) \quad d(x_n, x^*) \leq \sigma(\bar{\omega}^{(n)}(t_0)).$$

(iv) Let n be a positive integer and let $d_n \in T^p$; if $u_{n-1} \in Z(d_n)$,
 then

$$(20) \quad d(x_n, x^*) \leq \beta(d_n).$$

Since we are only going to consider iteration (3) and only indicate what will follow for (2) we assume from now on that $p = 1$. We will need the definition:

DEFINITION 4. Let E_1 and E_2 be two Banach spaces and let E_1 be a subset of E_2 . Let $f: E_1 \rightarrow E_2$ be a nonlinear operator which is Fréchet differentiable on E_1 . We say that the Fréchet-derivative $f'(x)$ is Hölder continuous over E_1 if for some $c > 0$ and $q \in [0, 1]$, and all $x, y \in E_1$

$$(21) \quad \|f'(x) - f'(y)\| \leq c \|x - y\|^q.$$

In this case we say $f'(\cdot) \in H_{E_1}(c, q)$.

DEFINITION 5. Let E_1 and E_2 be two Banach spaces and let E_1 be a convex subset of E_2 . Let $f: E_1 \rightarrow E_2$ be a nonlinear operator which is Fréchet-differentiable on E_1 . A mapping $\delta f: E_1 \times E_1 \rightarrow L(E_1, E_2)$, (the space of bounded linear operators from E_1 to E_2) will be called a consistent generalized approximation of f' , if there exists a constant $d > 0$ such that

$$(22) \quad \|\delta f(x, y) - f'(z)\| \leq d(\|x - z\|^q + \|y - z\|^q), \quad q \in [0, 1],$$

and for all $x, y, z \in E_1$.

The above condition implies the Hölder continuity of f' . Since,

$$(23) \quad \begin{aligned} \|f'(x) - f'(y)\| &= \|(f'(x) - \delta f(x, y)) + (\delta f(x, y) - f'(y))\| \\ &\leq d(\|x - x\|^q + \|y - x\|^q) + d(\|x - y\|^q + \|y - y\|^q) \\ &\leq 2d\|x - y\|^q. \end{aligned}$$

That is

$$(23) \quad \|f'(x) - f'(y)\| \leq c\|x - y\|^q, \quad c = 2d \text{ and for all } x, y \in E_1.$$

Also, as in [2] we can easily show that

$$(24) \quad \|f(x) - f(y) - f'(x)(x - y)\| \leq \frac{e}{1+q} \|x - y\|^{1+q}$$

for all $x, y \in E_1$.

Finally, for all $x, y, u, v \in E_1$ we have

$$(25) \quad \begin{aligned} \|f(u) - f(v) - \delta f(x, y)(u - v)\| &= \\ &= \|(f(u) - f(v) - f'(v)(u - v)) + (f'(v) - \delta f(x, y))(u - v)\| \\ &\leq \frac{2d}{1+q} \|u - v\|^{1+q} + d(\|x - v\|^q + \|y - v\|^q) \|u - v\| \\ &\leq d \left(\frac{2}{1+q} \|u - v\|^q + \|x - v\|^q + \|y - v\|^q \right) \|u - v\|. \end{aligned}$$

Let $C(h_0, q_0, r_0)$ be the class of all triplets (f, x_0, x_{-1}) satisfying the following properties:

(P₁) f is a nonlinear operator having the domain of definition E_f included into a Banach space E_1 and taking values in a Banach space E_2 .

(P₂) x_0 and x_{-1} are two points of E_f such that

$$\|x_0 - x_{-1}\| \leq q_0, \quad \|x_0 - x_{-1}\| < \mu.$$

(P₃) f is Fréchet-differentiable in the open ball $U = U(x_0, \mu) = \{x \in E_f / \|x - x_0\| < \mu\}$ and continuous on its closure \bar{U} .

(P₄) there exists a consistent generalized approximation δf of f' such that $D_0 := \delta f(x_{-1}, x_0)$ is invertible and

$$(26) \quad \|D_0^{-1}(\delta f(x, y) - f'(z))\| \leq h_0(\|x - z\|^q + \|y - z\|^q)$$

for all $x, y, z \in U$ and some $h_0 \geq d \cdot \|D_0^{-1}\|$.

(P₅) the following inequality is satisfied:

$$(27) \quad \|D_0^{-1}f(x_0)\| \leq r_0.$$

(P₆) Assume that for $r \in (0, r_0]$, $q_0 > 0$ and for fixed $q \in [0, 1]$, the following estimate holds:

$$(28) \quad h_0 \left[\frac{2}{q+1} r^q + (\sigma(r_0) - \sigma(r) + q_0)^2 + (\sigma(r_0) - \sigma(r))^q \right] r \leq \omega(r)$$

where

$$(29) \quad \sigma(r) = x_0 - a,$$

$$(30) \quad \omega(r) = h_0 \left\{ \left[\sqrt[1+q]{\frac{r}{h_0} + a^{1+q}} - r \right]^{1+q} - a^{1+q} \right\},$$

$$(31) \quad x_0(r) = x_0 = \sqrt[1+q]{\frac{r}{h_0} + a^{1+q}},$$

and a is the minimum positive solution of (if it exists)

$$(32) \quad (x_0(r_0) + q_0)^{1+q} - (x_0(r_0))^{1+q} = \frac{q_0}{h_0}.$$

We will use the estimate

$$\begin{aligned} \|D_0^{-1}(f(u) - f(v) - \delta f(x, y)(u - v))\| &= \|D_0^{-1}(f(u) - f(v) - f'(v)(u - v) \\ &\quad + D_0^{-1}(f'(v) - \delta f(x, y))(u - v))\| \\ &\leq \frac{2d \|D_0^{-1}\|}{1+q} \|u - v\|^{1+q} + h_0(\|x - v\|^q + \|y - v\|^q) \|u - v\| \end{aligned}$$

(by (24) and (26))

$$(33) \quad \leq h_0 \left[\frac{2}{1+q} \|u - v\|^q + \|x - v\|^q + \|y - v\|^q \right] \|u - v\|.$$

II. Main results. Using (3) we shall show that if $(f, x_0, x_{-1}) \in \mathcal{C}(h_0, q_0, r_0)$ then (1) has a solution x^* which is unique in a certain neighborhood of x_0 .

We will need the following lemma whose proof as similar to Lemma 1 in [4] is omitted.

LEMMA 1. *If $h_0 > 0$, $q_0 \geq 0$, $r_0 \geq 0$ are such that the equation (32) has a minimum positive solution a . Then the function ω given by (30) is a rate of convergence of type (1,1) on the interval $T = (0, r_0]$ and the corresponding σ -function is given by (29).*

We will now prove the main result.

THEOREM 1. *If $(f, x_0, x_{-1}) \in \mathcal{C}(h_0, q_0, r_0)$, then*

(a) *the sequence $\{x_n\}$, $n = 0, 1, 2, \dots$ is well defined on $U = U(x_0, \mu_0)$, where $\mu_0 = \sigma(r_0)$ remains in U and converges to a solution x^* of (1) such that:*

$$(34) \quad \|x_{n+1} - x^*\| \leq \sigma(\omega^{(n)}(r_0)), \quad n = 0, 1, 2, \dots$$

$$(35) \quad \|x_n - x^*\| \leq \sigma(\|x_n - x_{n-1}\|) - \|x_n - x_{n-1}\|, \quad n = 0, 1, 2, \dots$$

where ω , σ are given by (30) and (29) respectively.

Proof. Define the mappings $F: U \rightarrow E_1$ and $Z: T = (0, r_0] \rightarrow \exp(E_1)$ by

$$(36) \quad F(x) = x - D_0^{-1}f(x)$$

$$(37) \quad Z(r) = \{x \in E_1 / \|x - x_0\| \leq \sigma(r_0) - \sigma(r), \|D_0^{-1}f(x)\| \leq r\}.$$

By, $\sigma(r_0) = \mu_0$, it follows that $Z(r) \subset U$. If $r \in (0, r_0]$, $x \in Z(r)$ and $w = F(x)$, then we have

$$(38) \quad \|w - x_0\| \leq \|w - x\| + \|x - x_0\| \leq r + \sigma(r_0) - \sigma(r) - \sigma(\omega(r_0)) - \sigma(\omega(r)).$$

Since $w = F(x)$ implies $f(x) + D_0(w - x) = 0$, using (33) we have

$$\begin{aligned} \|D_0^{-1}f(w)\| &= \|D_0^{-1}(f(w) - f(x) - D_0(w - x))\| \\ &\leq h_0 \left[\frac{2}{1+q} \|w - x\|^q + (\|x - x_0\| + \|x_0 - y_0\|^q + \|x_0 - x\|^q) \|w - x\| \right] \\ &\leq h_0 \left[\frac{2}{1+q} r^q + (\sigma(r_0) - \sigma(r) + q_0)^q + (\sigma(r_0) - \sigma(r))^q \right] r \\ (39) \quad &\leq \omega(r), \text{ (by (28)).} \end{aligned}$$

By (36), (37), (38) and (39) it follows that the hypotheses (13), (14) and (15) of proposition 1 are satisfied. The estimates (34) follow then from (19), while, corresponding to (16) and (17), we have

$$(40) \quad x_{n-1} \in Z(\omega^{(n-1)}(r_0)), \|x_n - x_{n-1}\| \leq \omega^{(n-1)}(r_0), \quad n = 1, 2, 3, \dots$$

Using (40) and the fact that σ increases on $(0, r_0]$ we have $x_{n-1} \in Z(\|x_n - x_{n-1}\|)$, so that according to (iv) of proposition 1 it follows that (35) hold for $n = 1, 2, \dots$

Let now $n \rightarrow \infty$ in (3) to get $f(x^*) = 0$. This completes part (a) of the theorem.

Part (b) and (c) follow identically as proposition 2 in [4].

That completes the proof of the theorem.

We can talk about the uniqueness of the solution x^* in a certain neighborhood of x_0 but the motivated reader can easily produce the analog of theorem 2 in [4] that describes the uniqueness of x^* .

At this point we prefer not to pursue the goal of investigating iteration (2) with our new hypotheses but instead refer to a couple of interesting examples where our results can be applied and the corresponding ones in [4] cannot.

III. Applications. *Example 1.* Consider the function G defined on $[0, b]$ by

$$G(t) = \frac{2}{3} t^{\frac{3}{2}} + t - 3$$

for some $b > 0$.

Let $\|\cdot\|$ denote the max norm on \mathbb{R} , then

$$\|G''(t)\| = \max_{t \in [0, b]} \left| \frac{1}{2} t^{-\frac{1}{2}} \right| = \infty,$$

which implies that the basic hypothesis in [2] (the Lipschitz continuity of f' for $q \neq 1$ in [4]) for the application of Newton's method is not satisfied

for finding a solution of the equation

$$(41) \quad G(t) = 0.$$

However, it can easily be seen that $G'(t)$ is Hölder continuous on $[0, b]$ with

$$c = 1 \text{ and } q = \frac{1}{2}.$$

Therefore under the assumptions of theorem 1, iteration (3) will converge to a solution t^* of (41).

A more interesting nontrivial application for theorem 1 is given by the following example.

Example 2. Consider the differential equation

$$(42) \quad \begin{aligned} x'' + x^{1+q} &= 0, \quad q \in [0, 1] \\ x(0) &= x(1) = 0. \end{aligned}$$

We divide the interval $[0, 1]$ into n subintervals and we set $h = \frac{1}{n}$.

Let $\{v_k\}$ be the points of subdivision with

$$0 = v_0 < v_1 < \dots < v_n = 1.$$

A standard approximation for the second derivative is given by

$$x_i'' = \frac{x_{i-1} - 2x_i + x_{i+1}}{h^2}, \quad x_i = x(v_i), \quad i = 1, 2, \dots, n-1.$$

Take $x_0 = x_n = 0$ and define the operator $\tilde{F}: \mathbb{R}^{n-1} \rightarrow \mathbb{R}^{n-1}$ by

$$(43) \quad \tilde{F}(x) = H(x) + h^2\varphi(x)$$

$$H = \begin{bmatrix} 2 & -1 & & & 0 \\ -1 & 2 & & & \\ & & \ddots & & \\ & & & \ddots & \\ 0 & & & -1 & 2 \end{bmatrix},$$

$$\varphi(x) = \begin{bmatrix} x_1^{1+q} \\ x_2^{1+q} \\ \vdots \\ x_{n-1}^{1+q} \end{bmatrix},$$

and

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \end{bmatrix}.$$

Then

$$\tilde{F}'(x) = H + h^2(q+1) \begin{bmatrix} x_1^q & & & 0 \\ & x_2^q & & \\ & & \ddots & \\ 0 & & & x_{n-1}^q \end{bmatrix}$$

Newton's method cannot be applied to the equation

$$(44) \quad \tilde{F}(x) = 0.$$

We may not be able to evaluate the second Fréchet-derivative since it would involve the evaluation of quantities of the form x_i^{2q} and they may not exist.

We will face the same difficulty in verifying the Lipschitz continuity of \tilde{F}' .

Let $x \in \mathbb{R}^{n-1}$, $H \in \mathbb{R}^{n-1} \times \mathbb{R}^{n-1}$ and define the norms of x and H by

$$\|x\| = \max_{1 \leq j \leq n-1} |x_j|$$

$$\|H\| = \max_{1 \leq j \leq n-1} \sum_{k=1}^{n-1} |h_{jk}|.$$

For all $x, z \in \mathbb{R}^{n-1}$ for which $|x_i| > 0$, $|z_i| > 0$, $i = 1, 2, \dots, n-1$ we obtain, for $q = \frac{1}{2}$ say,

$$\begin{aligned} \|F''(x) - F''(z)\| &= \left\| \text{diag} \left\{ \left(1 + \frac{1}{2}\right) h^2 \left(x_j^{\frac{1}{2}} - z_j^{\frac{1}{2}}\right) \right\} \right\| \\ &= \frac{3}{2} h^2 \max_{1 \leq j \leq n-1} |x_j^{\frac{1}{2}} - z_j^{\frac{1}{2}}| \leq \frac{3}{2} h^2 [\max |x_j - z_j|]^{\frac{1}{2}} \\ &= \frac{3}{2} h^2 \|x - z\|^{\frac{1}{2}}. \end{aligned}$$

Therefore, under the assumptions of theorem 1, iteration (3) will converge to the solution x^* of (44).

Remarks. (a) Note that for $q = 1$ our results reduce to the ones in [4] and condition (28) is then immediately satisfied for the particular choice of ω and σ given by (30) and (29) respectively.

(b) Using Rolle's theorem one can give sufficient conditions in terms of r_0 , q_0 , q and h_0 that guarantee the existence of a minimum positive solution a of (32).

REFERENCES

1. Davis, H. T., *Introduction to nonlinear differential and integral equations*, Dover Publ., New York, 1962.
2. Kantorovich, L. V., Akilov, G. P., *Functional analysis in normed spaces*, Oxford, Pergamon Press, 1964.
3. Potra, F. A., Pták, V., *Nondiscrete induction and iterative processes*, Pitman Publ., 1984.
4. Potra, F. A., Pták, V., *An error analysis of the Secant method*, Numer. Math., **38** (1982), 427-445.
5. Rheinboldt, W. C., *Numerical Analysis of Parametrized Nonlinear Equations*. John Wiley Publ., 1986.
6. Rheinboldt, W. C., *A unified convergence theory for a class of iterative processes*, SIAM J. Numer. Anal., **5** (1968), 42-63.
7. Rheinboldt, W. C., *Iterative solution of nonlinear equations in several variables*, Academic Press, New York, 1970.
8. Pták, V., *Nondiscrete mathematical induction*. In: *General topology and its relations to modern analysis and algebra IV*. Lecture notes in Mathematics 609, pp. 166-178. Berlin, Heidelberg, New York, Springer, 1977.

Received 20 VII 1988

Department of Mathematics
New Mexico State Univ.
Las Cruces, NM 88003