Electronic Journal: Southwest Journal of Pure and Applied Mathematics
Internet: http://rattler.cameron.edu/swjpam.html
ISSN 1083-0464
Issue 1, July 2003, pp. 56–65.
Submitted: June 20, 2002. Published: July 1 2003.

A UNIFYING THEOREM ON NEWTON'S METHOD IN A BANACH SPACE WITH A CONVERGENCE STRUCTURE UNDER WEAK ASSUMPTIONS

Ioannis K. Argyros

Abstract. We provide a semilocal analysis for Newton's method on a Banach space with a convergence structure. Our main result contains as special cases earlier results on semilocal theorems of Newton–Kantorovich-type as well as global results based on monotonicity. It turns out that under the same hypotheses our sufficient convergence conditions are weaker and the error bounds on the distances involved finer than before. Some numerical examples are provided to show that our results apply where others fail.

A.M.S. (MOS) Subject Classification Codes. 65B05, 65G99, 47H17, 49M15, CR:1.5.

Key Words and Phrases. Newton's method, Banach space with a convergence structure, order convex operator, Newton–Kantorovich hypothesis, monotone convergence, multilinear/bounded/symmetric operator.

1. Introduction

In this study we are concerned with the problem of approximating a zero x^* of an operator G on a Banach space X. Our results will be in terms of the operator

$$F(x) := AG(x_0 + x), \tag{1}$$

where x_0 is the initial guess for Newton's method

$$x_0 = 0, \quad x_{n+1} = x_n + F'(x_n)^*(-F(x_n)) \quad (n \geq 0), \tag{2}$$

A is an approximation to $G'(x_0)^{-1}$ and $F'(x_n)^*$ is an operator to be precised later.

Semilocal convergence theorems were given in [1], [2], [4], [5] under various assumptions on a Banach space with a convergence structure. Here we introduce finer "majorizing" sequences and show that under the same hypotheses as in [5] we can obtain semilocal results under weaker assumptions and with finer error bounds.

We complete our study with some examples to justify the claims made above.

Department of Mathematics, Cameron University, Lawton, OK 73505, U.S.A.
E-mail Address: ioannisa@cameron.edu

2. Semilocal Convergence Analysis

To make the study as self-contained as possible we need to state some definitions and lemmas that can also be found in [5] (see also [1], [2]).

Definition 1. A Banach space with a convergence structure is a triplet (X, V, E) such that:

(3) $(X, \|\cdot\|)$ is a real Banach space;
(4) $(V, C, \|\cdot\|_V)$ is a real Banach space which is partially ordered by the closed convex cone C with the norm $\|\cdot\|_V$ assumed to be monotone on C;

E is a closed convex cone in $X \times V$ such that

$$\{0\} \times C \subseteq E \subseteq X \times C; \tag{5}$$

the operator $|\cdot|\colon D \to C$ is well defined:

$$|x| := \inf\{r \in C \mid (x, r) \in E\}$$

where

(6) $x \in D = \{x \in X \mid \exists r \in C\colon (x, r) \in E\}$;
(7) $\|X\| \leq \| \, |x| \|_V$ for all $x \in D$.

The set

$$U(p) = \{x \in X \mid (x, p) \in E\} \tag{8}$$

is a generalized neighborhood of zero.

The convergence analysis is based on monotonicity considerations in the space $X \times V$.

We have

$$(x_n, r_n) \leq (x_{n+m}, r_{n+m}) \Rightarrow 0 \leq (x_{n+m} - x_n, r_{n+m} - r_n) \tag{9}$$

for all increasing sequences $(x_n, r_n) \in E^\pi$. If

$$r_n \to r \tag{10}$$

we get

$$0 \leq (x_{n+m} - x_n, r - r_n) \tag{11}$$

and by (5) of Definition 1

$$\|x_{n+m} - x_n\| \leq \|r - r_n\|_V \to 0 \quad \text{as } n \to \infty. \tag{12}$$

Hence $\{x_n\}$ $(n \geq 0)$ is a Cauchy sequence in a Banach space and as such it converges to some point $x^* \in X$. Let $\{v_n\} \in C^\pi$ be an increasing sequence and set $r_n = v_0 - v_n$ to obtain error estimates:

$$0 \leq (x_{n+m} - x_n, v_n - v_{n+m}) \leq (x_{n+m} - x_n, v_n). \tag{13}$$

Note that if $x_n \to x^*$ as $n \to \infty$ this implies

$$|x^* - x_n| \leq v_n \quad (n \geq 0). \tag{14}$$

Definition 2. An operator $L \in C^1(V_0 \to V)$ on an open subset V_0 of an ordered Banach space V is called order convex on $[a,b] \subseteq V_0$ if

$$c, d \in [a,b],\ c \le d \Rightarrow L'(d) - L'(c) \in L_+(V), \tag{15}$$

where,

$$L_+(V^n) = \{L \in L(V^n) \mid 0 \le x_i \Rightarrow 0 \le L(x_1, \dots, x_n)\}. \tag{16}$$

Definition 3. The set

$$\begin{aligned} B(A) = \big\{ L \in L_+(V^n) \mid (x_i, r_i) \in E \\ \Rightarrow [A(x_1, \dots, x_n), L(r_1, \dots, r_n)] \in E \big\} \end{aligned} \tag{17}$$

is the set of bounds for $A \in L(X^n)$.

Lemma 1. *Let $A\colon [0,1] \to L(X^n)$ and $L\colon [0,1] \to L_+(V^n)$ be continuous operators. Then the following implication holds*:

$$\forall t \in [0,1]\colon L(t) \in B(A(t)) \Rightarrow \int_0^1 L(t)dt \in B\Big(\int_0^1 A(t)dt\Big). \tag{18}$$

Definition 4. Let $Q\colon Y \to Y$ be an operator on a subset Y of a normed space. By $Q^n(x)$ we mean $Q(Q^{n-1}(x))$ and in case of convergence we write

$$Q^\infty(x) = \lim_{n\to\infty} Q^n(x). \tag{19}$$

Definition 5. Let $A \in L(X)$ and $y \in X$ be given. Then we have

$$\begin{aligned} A^* y = z \Leftrightarrow z \in Q^\infty(0),\ Q(x) = (I - A)(x) + y \\ \Leftrightarrow z = \sum_{i=0}^{\infty} (I - A)^i (y). \end{aligned} \tag{20}$$

Lemma 2. *Let $L \in L_+(V)$ and $p, r \in C$ be given such that*

$$L(r) + p \le r \quad \text{and} \quad L^n(r) \to 0 \quad \text{as } n \to \infty. \tag{21}$$

Then the operator

$$(I - L)^*\colon [0,p] \to [0,p]$$

is well defined and continuous.

Lemma 3. *Let $A \in L(X)$, $L \in B(A)$, $y \in D$ and $r \in C$ be given such that*:

$$L(r) + |y| \le r \quad \text{and} \quad L^n(r) \to 0 \quad \text{as } n \to \infty. \tag{22}$$

Then the following hold:

$$x = (I - A)^*(y) \tag{23}$$

is well defined, $x \in D$, and

$$|x| \le (I - L)^* |y| \le r. \tag{24}$$

We can now provide the main semilocal convergence theorem for Newton's method (2).

Theorem 1. *Assume: given a Banach space X with a convergence structure (X, V, E) with $V = (V, C, \|\cdot\|_V)$, an operator $F \in C^1(X_0 \subseteq X \to X)$, operators M, L_0, $L \in L^1(V_0 \subseteq V \to V)$, and a point $p \in C$ such that the following conditions hold*:

$$U(p) \subseteq X_0, \quad [0,p] \subseteq V_0; \tag{25}$$

M, L_0, L are order convex on $[0,p]$, and such that for $x, y, z \in U(a)$ with $|x| \le p$, $|y| + |z| \le p$

$$L_0'(|x|) - L_0'(0) \in B(F'(0) - F'(x)), \tag{26}$$
$$L'(|y| + |z|) - L'(|y|) \in B(F'(y) - F'(y+z)), \tag{27}$$
$$M(p) \le p, \tag{28}$$
$$L_0(p_0) \le M(p_0) \quad \text{for all } p_0 \in [0,p], \tag{29}$$
$$L_0'(p_0) \le M'(p_0) \quad \text{for all } p_0 \in [0,p], \tag{30}$$
$$L_0'(0) \in B(I - F'(0)), \quad (-F(0), L_0(0)) \in E, \tag{31}$$
$$M'(p)^n(p) \to 0 \quad \text{as } n \to \infty, \tag{32}$$
$$M'(d_n)(b - d_n) + L(d_n) \le M(b) \quad \text{for all } n \ge 0, \tag{33}$$

where,

$$d_0 = 0,\ d_{n+1} = L(d_n) + L_0'((x_n))(c_n),\ c_n = |x_{n+1} - x_n| \quad (n \ge 0) \tag{34}$$

and

$$b = M^\infty(0). \tag{35}$$

Then sequence $\{x_n\}$ $(n \ge 0)$ generated by Newton's method (2) is well defined, remains in E^π, is monotone and converges to a unique zero x^* in $U(b)$, where b is the smallest fixed point of M in $[0,p]$. Moreover the following bounds hold for all $n \ge 0$

$$d_n \le b, \tag{36}$$
$$|x^* - x_n| \le b - d_n, \tag{37}$$
$$|x^* - x_n| \le M^n(p) - M^n(0), \tag{38}$$
$$c_n + d_n \le d_{n+1}, \tag{39}$$

and

$$c_n = |x_{n+1} - x_n|. \tag{40}$$

Proof. We first show that (x_n, d_n) is well defined and monotone for all $n \ge 0$ using induction on n.

By (5), (22) and the order convexity of L_0 we get

$$L_0'(0)(b) + |-F(0)| \le L_0(b - 0) + L_0(0) = L_0(b) \le M(b) = b. \tag{41}$$

Hence equation

$$r = (I - F'(0))(r) - F(0) \tag{42}$$

has a solution x_1, which is well defined, and $(x_1, b) \in E$. We also have

$$x_1 = (I - F'(0))(x_1) + (-F(0)), \tag{43}$$

$$|x_1| \leq L_0'(0)|x_1| + L(0) = d_1, \tag{44}$$

and

$$d_1 = L_0'(0)|x_1| + L(0) \leq M'(0)(b) + L(0) \leq M(b) = b. \tag{45}$$

Assume (x_k, d_k) is well defined and monotone with

$$0 \leq (x_{n-1}, d_{k-1}) \leq (x_k, d_k), \quad k = 1, 2, \ldots, n \tag{46}$$

and

$$d_k \leq b. \tag{47}$$

We must solve

$$r = (I - F'(x_n))(r) + (-F(x_n)). \tag{48}$$

Using (26) and (31) we can have

$$\begin{aligned} |I - F'(x_n)| &= |I - F'(x_n) - F'(0) + F'(0)| \\ &= |(I - F'(0)) + (F'(0) - F'(x_n))| \\ &\leq L_0'(0) + L_0'(|0| + |x_n|) - L_0'(|0|) = L_0'(|x_n|). \end{aligned} \tag{49}$$

That is

$$L_0'(|x_n|) \in B(I - F'(x_n)). \tag{50}$$

By the approximation

$$-F(x_n) = -F(x_n) + F(x_{n-1}) + F'(x_{n-1})(x_1 - x_{n-1}) \tag{51}$$

we can have in turn using (27), (33) and (34)

$$\begin{aligned} |-F(x_n)| &\leq \int_0^1 [L'(|x_{n-1}| + tc_{n-1}) - L'(|x_{n-1}|)]c_{n-1}dt \\ &= L(|x_{n-1}| + c_{n-1}) - L(|x_{n-1}|) - L'(|x_{n-1}|)c_{n-1} \\ &\leq L(d_{n-1} + d_n - d_{n-1}) - L(d_{n-1}) - L_0'(|x_{n-1}|)c_{n-1} \\ &= L(d_n) - d_n, \end{aligned} \tag{52}$$

and

$$L_0'(|x_n|)(b - d_n) + |-F(x_n)| + d_n \leq M'(d_n)(b - d_n) + L(d_n) \leq M(b) = b. \tag{53}$$

Hence we showed

$$M'(|x_n|)(b-d_n)+|-F(x_n)|\leq b-d_n. \tag{54}$$

By Lemma 3, x_{n+1} is defined and

$$c_n\leq b-d_n. \tag{55}$$

Therefore d_{n+1} is well defined too, and we obtain:

$$d_{n+1}\leq L(d_n)+L_0'(d_n)(b-d_n)\leq L(d_n)+M'(d_n)(b-d_n)\leq M(b)=b. \tag{56}$$

Moreover we can have:

$$c_n+d_n\leq L_0'(|x_n|(c_n)+|-F(x_n)|+d_n\leq L_0'(|x_n|)(c_n)+L(d_n)=d_{n+1}, \tag{57}$$

which shows the monotonicity $(x_n,d_n)\leq(x_{n+1},d_{n+1})$ $(n\geq 0)$.

Furthermore we get

$$M^n(0)\leq d_n\leq b, \tag{58}$$

and since

$$M^n(0)\to b \quad \text{as } n\to\infty \tag{59}$$

we get

$$d_n\to b \quad \text{as } n\to\infty. \tag{60}$$

That is, $x_n\to x^*\in U(b)$, and $F(x^*)=0$. Clearly $(y_n,M^n(0))$ is monotone in $X\times V$, and if $x^{**}\in U(p)$ is a solution of F, then as above we get

$$|x^{**}-y_n|\leq M^n(p)-M^n(0)\leq M'(p)^n(p)\to 0 \quad \text{as } n\to\infty \tag{61}$$

where,

$$y_{n+1}=y_n-F(y_n). \tag{62}$$

That is $\lim\limits_{n\to\infty} y_n=x^{**}$, which shows the uniqueness of x^* in $U(p)$.

That completes the proof of Theorem 1. □

Remark 1. Under the hypotheses of Theorem 1 define

$$R_n(r)=(I-L_0'(|x_n|))^*S_n(r)+c_n, \tag{63}$$

where,

$$S_n(r)=L(|x_n|+r)-L(|x_n|)-L_0'(|x_n|)(r). \tag{64}$$

As in [1], [5] assume there exists $r_n\in C$ such that $|x_n|+r_n\leq p$, and

$$S_n(r_n)+L_0'(|x_n|)(r_n-c_n)\leq r_n-c_n \tag{65}$$

then $R_n\colon[0,r_n]\to[0,r_n]$ is well defined and monotone. Since

$$\begin{aligned} d_n+c_n\leq d_{n+1}\Rightarrow & L(p)-L(d_n)-L_0'(|x_n|)c_n \\ & \leq p-d_n-c_n, \end{aligned} \tag{66}$$

$$\Rightarrow s_n(p-d_n)+L_0'(|x_n|)(p-d_n-c_n)\leq p-d_n-c_n, \tag{67}$$

which implies that $r_n=p-d_n$ is a suitable r_n. That is using solutions of

$$R_n(r)\leq r \tag{68}$$

we obtain better a posteriori error estimates than

$$|x^*-x_n|\leq b-d_n\leq r-d_n \quad (n\geq 0). \tag{69}$$

The proof of the following propositions can be obtained from Lemmas 10, 11, 12 in [5] by simply replacing L' by L_0' in their proofs.

Proposition 1. *If $r \in [0, p - |x_n|]$ satisfies $R_n(r) \leq r$ then the following holds for all $n \geq 0$:*

$$c_n \leq R_n(r) = a \leq r \tag{70}$$

and

$$R_{n+1}(a - c_n) \leq a - c_n. \tag{71}$$

Proposition 2. *Assume hypotheses of Theorem 1 hold and let $r_n \in [0, p - |x_n|]$ be a solution (69). Then the following a posteriori estimates hold for all $n \geq 0$:*

$$|x^* - x_m| \leq q_m, \tag{72}$$

where,

$$q_n = r_n, \quad q_{m+1} = R_m(q_m) - c_m \quad (m \geq n). \tag{73}$$

Proposition 3. *Under the hypotheses of Theorem 1 any solution $r \in [0, p - |x_n|]$ of $R_n(r) \leq r$ yields the a posteriori estimate*

$$|x^* - x_n| \leq R_n^\infty(0) \leq r \quad (n \geq 0). \tag{74}$$

3. Monotone Convergence

Let X be partially ordered and set

$$X = V, \quad D = C^2 \quad \text{and} \quad |\cdot| = I. \tag{75}$$

We can show:

Theorem 2. *Let X be partially ordered, Y a Banach space, $G \in C^1(X_0 \subseteq X \to Y)$, $A \in L(Y \to X)$ and $x, y \in X$ such that:*

$$1. \quad [x, y] \in X_0, \tag{76}$$

$$2. \quad I - AG'(x) \in L_+(X), \tag{77}$$

$$3. \quad z_1 \leq z_2 \Rightarrow AG'(z_1) \leq AG'(z_2) \quad \textit{for all } z_1, z_2 \in [x, y], \tag{78}$$

$$4. \quad AG(x) \leq 0, \quad AG(y) \geq 0, \tag{79}$$

$$5. \quad (I - AG'(y))^n (y - x) \to 0 \quad \textit{as } n \to \infty. \tag{80}$$

Then sequence $\{y_n\}$ $(n \geq 0)$ generated by

$$y_0 = x, \quad y_{n+1} = y_n + AG'(y_n)^*[-AG(y_n)] \tag{81}$$

is well defined for all $n \geq 0$, monotone and converges to a unique zero y^* of AG in $[x, y]$.

Proof. Define

$$F(u) = AG(x + u), \quad M(r) = r - F(r), \quad p = y - x. \tag{82}$$

The result now follows from Theorem 1.

That completes the proof of Theorem 2. ∎

Remark 2. Conditions (28) and (32) can be replaced by:

$$\text{there exists } \theta \in (0,1)\colon M(p) \leq \theta p. \tag{83}$$

Then there exists $p' \in [0, \theta p]$ satisfying all hypotheses of Theorem 1. The solution $x^* \in U(p')$ and is unique in $U(p)$.

Remark 3. (a) If $L_0 = L = M$ the results obtained here reduce to the corresponding ones in [5] (see also [1], [2], [4]).

(b) If $L = M$, condition (33) holds automatically. Inequalities (29) and (30) hold in general. However if strict inequality holds in (29) and (30) then our error bounds are finer than the ones in Lemmas 9–12 [5] in this case, and under the same sufficient convergence conditions (see, e.g. (28) and (32)). Let us demonstrate this observation with one of the error bounds. Define as in [5] sequence $\{q_n\}$ $(n \geq 0)$ by

$$q_0 = 0, \quad q_{n+1} = L(q_n) + L'(|x_n|)(c_n). \tag{84}$$

Then we get

$$d_n < q_n \quad (n \geq 1). \tag{85}$$

(c) Estimate

$$M'(|x_n|)(\overline{d}_n) + |-F(x_n)| \leq \overline{d}_n \ \text{ for some } 0 \leq \overline{d}_n\colon \overline{d}_n + d_n \leq b \tag{33'}$$

or condition

$$x_n \text{ given by (2) is well defined for all } n \geq 1 \tag{33''}$$

can replace (33) in Theorem 1. Hypotheses $(33)''$ is the weakest of the three.

(d) (M, M') can be replaced by a sequence of in practice more precise pairs (M_n, M'_n) $(n \geq 0)$ in the hypotheses of Theorem 1 (see (e) that follows).

(e) If $L \neq M$ one would expect: (28) and (32) (or (83)) will be weaker than the corresponding ones in [5] given by

$$L(p) \leq p, \tag{86}$$

$$L'(p)^n(p) \to 0 \quad \text{as } n \to \infty \tag{87}$$

(or $L(p) \leq \theta p$) (because of (85)); and that (48) will be solvable in which case (33) can be dropped.

As an example of such a case we consider a Banach space X with a real norm $\|\cdot\|$. Assume $F'(0) = I$ and

$$\|F'(x) - F'(y)\| \leq \ell \|x - y\|, \tag{88}$$

$$\|F'(x) - F'(0)\| \leq \ell_0 \|x\| \tag{89}$$

for all $x, y \in U(r_0)$, and some $\ell_0, \ell \neq 0$ where

$$r_0 = \frac{1 - \sqrt{1 - 2\ell\eta}}{\ell} \tag{90}$$

and

$$\eta \geq \|F(0)\|. \tag{91}$$

Define operators L_0, L by

$$L_0(r) = \|F(0)\| + \frac{1}{2}\ell_0 r^2, \tag{92}$$

$$L(r) = \|F(0)\| + \frac{1}{2}\ell r^2, \tag{93}$$

sequences α_n, β_n, γ_n by

$$\alpha_0 = \beta_0 = \gamma_0 = 0, \quad \alpha_1 = \beta_1 = \alpha_1 = \eta, \tag{94}$$

$$\alpha_{n+1} = \alpha_n + \frac{\ell_0 \alpha_n^2}{2(1 - \ell_0 \alpha_n)}, \tag{95}$$

$$\beta_{n+1} = \beta_n + \frac{\ell \beta_n^2}{2(1 - \ell_0 \beta_n)}, \tag{96}$$

$$\gamma_{n+1} = \gamma_n + \frac{\ell \gamma_n^2}{2(1 - \ell \gamma_n)}, \tag{97}$$

and points α^*, β^*, γ^* by

$$\alpha^* = \frac{1 - \sqrt{1 - 2\ell_0 \eta}}{\ell_0}, \tag{98}$$

$$\beta^* = \lim_{n \to \infty} \beta_n, \tag{99}$$

$$\gamma^* = \frac{1 - \sqrt{1 - 2\ell \eta}}{\ell} \tag{100}$$

provided that

$$h = 2\ell\eta \leq 1. \tag{101}$$

Note that (101) is the famous Newton–Kantorovich hypothesis for the convergence of Newton's method [4], [5].

As it can easily be seen by (2), (34), (49), (52), (62), and (84) iterations (95), (96), (97) are the majorizing sequences (denoted by d_n in Theorem 1) corresponding to methods (62), (2) (via Theorem 1) and (2) (via Lemma 9 in [5] and (84)) respectively.

Condition $(33)''$ needed to solve (48) can be dropped since iterates say β_n (i.e. d_n) can be found for all $n > 0$.

The pair (M_n, M_n') is defined "discretely" for each $n \geq 0$ by sequence β_n which is monotonically convergence under (107). Therefore the verification of (28) and (32) is not needed since it is replaced by (107), which also makes the explicit construction of (M_n, M_n') unnecessary.

Then it can easily be seen

$$\alpha^* = L_0^\infty(0), \tag{102}$$

$$\gamma^* = L^\infty(0), \tag{103}$$

$$\alpha_n \le \beta_n \le \gamma_n \quad (n \ge 2), \tag{104}$$

and

$$\alpha^* \le \beta^* \le \gamma^*. \tag{105}$$

Note that in general

$$\ell_0 \le \ell. \tag{106}$$

Moreover if strict inequality holds in (106) then so does in (105). Furthermore it can easily be seen that (96) monotonically converges to β^* provided that

$$h_0 = (\ell + \ell_0)\eta \le 1, \tag{107}$$

which is a weaker hypothesis than (0). That is

$$h \le 1 \Rightarrow h_0 \le 1 \tag{108}$$

but not vice versa unless if $\ell = \ell_0$. That is in the Banach space case with real norm $\| \ \|$, our Theorem 1 guarantees convergence of Newton's method under (107), whereas Lemma 9 in [5] can only guarantee convergence under the stronger condition (101).

We complete this study with an example to show $\frac{\ell}{\ell_0}$ can be arbitrarily large, which implies that (106) holds a strict inequality and that (107) may hold but not (101).

Example 1. Let $X = Y = \mathbf{R}$, $x_0 = 0$, $A = G'(x_0)^{-1}$, and define function G on $\mathbf{R}$ by

$$G(x) = c_0 x + c_1 + c_2 \sin e^{c_3 x}, \tag{109}$$

where c_i, $i = 0, 1, 2, 3$ are given parameters. Using (109) it can easily be seen that for c_3 large and c_2 sufficiently small, $\frac{\ell}{\alpha_2}$ may be arbitrarily large. That is (107) can be satisfied but not (101).

REFERENCES

[1] Argyros, I.K., Newton methods on Banach spaces with a convergence structure and applications, *Computers and Math. with Applic. Intern. J.* **40**, 1 (2000), 37–48.

[2] Argyros, I.K., *Advances in the Efficiency of Computational Methods and Applications*, World Scientific Publ. Co., River Edge, NJ, 2000.

[3] Kantorovich, L.V. and Akilov, G.P., *Functional Analysis in Normed Spaces*, Pergamon Press, Oxford, 1982.

[4] Meyer, P.W., Newton's method in generalized Banach spaces, *Numer. Funct. Anal. and Optimization* **9** (1987), 249–259.

[5] Meyer, P.W., A unifying theorem on Newton's method, *Numer. Funct. Anal. and Optimization* **13** (5 and 6) (1992), 463–473.