Nonlinear Functional Analysis and Applications
Vol. 13, No. 1 (2008), pp. 43-59

http://nfaa.kyungnam.ac.kr/jour-nfaa.htm

CONCERNING THE CONVERGENCE OF NEWTON METHOD UNDER VERTGEIM-TYPE CONDITIONS

Ioannis K. Argyros

Cameron University, Department of Mathematical Sciences,
Lawton, OK 73505, USA
e-mail: ioannisa@cameron.edu

Abstract. The majorizing principle is used to show local and semilocal convergence of Newton methods to a locally unique solution of a nonlinear operator in a Banach space, when the Fréchet derivative of the operator involved satisfies a center-Hölder and a Hölder continuity condition. Then we investigate an unknown area ("terra incognita") between the convergence regions of Newton's method, and the corresponding modified Newton's method. Our approach compares favorably with other corresponding ones in this direction.

1. Introduction

In this study we are concerned with the problem of approximating a locally unique solution x^* of the nonlinear equation

$$F(x) = 0, \tag{1.1}$$

where F is a Fréchet-differentiable operator such that F' is a λ-Hölder continuous operator ($\lambda \in [0, 1]$) defined on an open subset D of a Banach space X with values in a Banach space Y.

A large number of problems in applied mathematics and also in engineering are solved by finding the solutions of certain equations. For example, dynamic systems are mathematically modelled by difference or differential equations, and their solutions usually represent the states of the systems. For the sake of simplicity, assume that a time-invariant system is driven by the equation $\dot{x} = Q(x)$ (for some suitable operator Q), where x is the state. Then the equilibrium states are determined by solving equation (1). Similar equations are

[0]Received December 28, 2005. Revised September 30, 2007.

[0]2000 Mathematics Subject Classification: 65B05, 65G99, 65J15, 47H17, 49M15.

[0]Keywords: Newton's method, modified Newton's method, Banach space, Hölder / Lipschitz continuity, Fréchet-derivative, majorizing principle, convergence region / radius.

used in the case of discrete systems. The unknowns of engineering equations can be functions (difference, differential, and integral equations), vectors (systems of linear or nonlinear algebraic equations), or real or complex numbers (single algebraic equations with single unknowns). Except in special cases, the most commonly used solution methods are iterative — when starting from one or several initial approximations a sequence is constructed that converges to a solution of the equation. Iteration methods are also applied for solving optimization problems. In such cases, the iteration sequences converge to an optimal solution of the problem at hand. Since all of these methods have the same recursive structure, they can be introduced and discussed in a general framework.

As in the elegant paper [5, Proposition 1.1] we study the convergence of Newton's method

$$x_{n+1} = x_n - F'(x_n)^{-1}F(x_n), \quad (n \geq 0) \quad (x_0 \in D) \tag{1.2}$$

with the convergence of the modified Newton's method

$$y_{n+1} = y_n - F'(y_0)^{-1}F(y_n) \quad (n \geq 0), \quad y_0 = x_0. \tag{1.3}$$

A survey of sufficient conditions for the local as well as the semilocal convergence of Newton-type methods as well as an error analysis for such methods can be found in [1]–[4], [8], [13] and the references there.

There is an unknown area, between the convergence regions ("terra incognita") of Newton's method, and the corresponding modified Newton's method. Here we show how to investigate this region and improve on earlier attempts in this direction [5, Proposition 1.1]–[7], [9], [11], [12].

2. Semilocal convergence analysis for methods (2) and (3)

To make the study as self-contained as possible we briefly reintroduce some results (until Remark 3) that can originally be found in [5, Proposition 1.1]–[7], [9], [11], [12].

Let $x_0 \in D$ be such that $F'(x_0)^{-1} \in L(Y, X)$ the space of bounded linear operators from Y into X. Assume F' satisfies a center-Hölder condition

$$\|F'(x_0)^{-1}(F'(x) - F'(x_0))\| \leq \ell_0 \|x - x_0\|^{\lambda_0} \tag{2.1}$$

and a Hölder condition

$$\|F'(x_0)^{-1}(F'(x) - F'(y))\| \leq \ell \|x - y\|^{\lambda} \tag{2.2}$$

for all $x, y \in U(x_0, R) = \{x \in X \mid \|x - x_0\| < R,\ R > 0\} \subseteq D$. We will assume throughout this study that $R \in (0, 1]$. The case $R > 1$ can be handled similarly.

Note that in general

$$\lambda_0 \geq \lambda, \text{ and } \ell_0 \leq \ell \tag{2.3}$$

hold. The results in [5, Proposition 1.1] were given in non-affine invariant form. Here we reproduce them in affine invariant form. The advantages of such an approach have been well explained in [2], [4].

Define:

$$\eta \geq \|F'(x_0)^{-1}F(x_0)\|, \tag{2.4}$$

$$h_0 = \ell_0 \eta^{\lambda_0}, \tag{2.5}$$

$$h = \ell \eta^{\lambda} \tag{2.6}$$

and function

$$\psi(r) = \frac{\ell}{1+\lambda} r^{1+\lambda} - r + \eta. \tag{2.7}$$

The first semilocal convergence result for methods (1.2) and (1.3) under Hölder conditions were given in [11], [12]:

Theorem 1. *Assume:*

$$h \leq \left(\frac{\lambda}{1+\lambda}\right)^{\lambda} \tag{2.8}$$

and

$$r^* \leq R, \tag{2.9}$$

where r^* *is the smallest positive zero of function* ψ. *Then sequence* $\{x_n\}$ $(n \geq 0)$ *generated by method (1.3) is well defined, remains in* $U(x_0, r^*)$ *for all* $n \geq 0$ *and converges to a unique solution* x^* *of equation (1.1) in* $U(x_0, r^*)$. *If* r^* *is the unique zero of* ψ *on* $[0, R]$ *and* $\psi(R) \leq 0$ *then* x^* *is unique in* $U(x_0, R)$.

Moreover, if

$$h \leq h_\nu, \tag{2.10}$$

where h_ν *is the unique solution in* $(0,1)$ *of equation*

$$\left(\frac{t}{1+\lambda}\right)^{\lambda} = (1-t)^{1+\lambda} \tag{2.11}$$

method (1.2) converges as well.

Therefore there is an unknown region, called "terra incognita" between the regions of convergence for methods (*1.2*) and (1.3). This obviously disappears in the Lipschitz case $\lambda = 1$, since then (2.8) reduces to the famous Newton–Kantorovich condition [8]:

$$h_{\mathrm{K}} = \ell \eta \leq \frac{1}{2}\,. \tag{2.12}$$

Theorem 1 holds [9] if condition (2.8) is replaced by the weaker

$$h \le 2^{\lambda-1}\left(\frac{\lambda}{1+\lambda}\right)^{\lambda}. \tag{2.13}$$

Later in [7] (2.13) was replaced by an even weaker condition

$$h \le \frac{1}{g(\lambda)}\left(\frac{\lambda}{1+\lambda}\right)^{\lambda}, \tag{2.14}$$

where,

$$g(\lambda) = \max_{t\ge 0} f(t), \tag{2.15}$$

$$f(t) = \frac{t^{1+\lambda} + (1+\lambda)t}{(1+t)^{1+\lambda} - 1} \tag{2.16}$$

with

$$g(\lambda) < 2^{1-\lambda} \quad \text{for all } \lambda \in (0,1). \tag{2.17}$$

Recently in [5] (2.14) was replaced by

$$h \le \frac{1}{a(\lambda)}\left(\frac{\lambda}{1+\lambda}\right)^{\lambda}, \tag{2.18}$$

where,

$$a(\lambda) = \min\left\{b \ge 1\colon \max_{0\le t\le t(b)} f(t) \le b\right\}, \tag{2.19}$$

$$t(b) = \frac{b\lambda^{\lambda}}{(1+\lambda)[b(1+\lambda)^{\lambda} - \lambda^{\lambda}]}. \tag{2.20}$$

The idea is to optimize b in the equation

$$\psi_b(r) = 0, \tag{2.21}$$

where,

$$\psi_b(r) = \frac{b\ell}{1+\lambda} r^{1+\lambda} - r + \eta \tag{2.22}$$

assuming

$$h \le \frac{1}{b}\left(\frac{\lambda}{1+\lambda}\right)^{\lambda}. \tag{2.23}$$

Note that condition (2.23) guarantees that equation (2.21) is solvable (see in [5, Proposition 1.1] or [7]).

With the above notation it was shown in [5, Proposition 1.1] (Theorem 2.2, p. 719):

Theorem 2. *Assume (2.18) holds and that $r^* \leq R$, where r^* is the smallest solution of the scalar equation*

$$\psi_a(r) = \frac{a(\lambda)\ell}{1+\lambda} r^{1+\lambda} - r + \eta = 0. \tag{2.24}$$

Then sequence $\{x_n\}$ $(n \geq 0)$ generated by Newton's method (1.2) is well defined, remains in $U(x_0, r^*)$ for all $n \geq 0$ and converges to a unique solution x^* of equation $F(x) = 0$ in $U(x_0, r^*)$.

Moreover if sequence r_n is defined by

$$r_0 = 0, \quad r_n = r_{n-1} - \frac{\psi_a(r_{n-1})}{\psi'(r_{n-1})} \quad (n \geq 1) \tag{2.25}$$

then the following error bounds hold for all $n \geq 1$:

$$\|x_n - x_{n-1}\| \leq r_n - r_{n-1} \tag{2.26}$$

and

$$\|x_n - x^*\| \leq r^* - r_n. \tag{2.27}$$

Remark 3. It was also shown in [5] (see Theorem 2.3) that

$$a(\lambda) < f(2) < g(\lambda) \text{ for all } \lambda \in (0,1), \tag{2.28}$$

which shows that (2.18) is a real improvement over (2.13) and (2.14).

We can summarize as follows:

$$\begin{aligned} h_\nu < 2^{\lambda-1}\left(\frac{\lambda}{1+\lambda}\right)^{\lambda} &< \frac{1}{g(\lambda)}\left(\frac{\lambda}{1+\lambda}\right)^{\lambda} \\ &< \frac{1}{a(\lambda)}\left(\frac{\lambda}{1+\lambda}\right)^{\lambda} \leq \left(\frac{\lambda}{1+\lambda}\right)^{\lambda} = h_{exi}. \end{aligned} \tag{2.29}$$

Below we present our contributions/improvements in the exploration of "terra incognita".

First of all we have observed that the Vertgeim result given in Theorem 1 holds under weaker conditions. Indeed:

Theorem 4. *Assume:*

$$h_0 \leq \left(\frac{\lambda_0}{1+\lambda_0}\right)^{\lambda_0} \tag{2.30}$$

replaces condition (2.8) in Theorem 1. Then under the rest of the hypotheses of Theorem 1, the conclusions for method (1.3) and equation (1.2) hold.

Proof. We note that (2.1) can be used instead of (2.2) in the proof of Theorem 1 given in [11]. That completes the proof of Theorem 4. □

Remark 5. Condition (2.30) is weaker than (2.8) if

$$h \le \left(\frac{\lambda}{1+\lambda}\right)^{\lambda} \Rightarrow h_0 \le \left(\frac{\lambda_0}{1+\lambda_0}\right)^{\lambda_0} \tag{2.31}$$

but not vice versa unless if $\lambda_0 = \lambda$, and $\ell = \ell_0$ (see also (2.3)). Therefore our Theorem 4 improves the convergence region for method (1.3) under weaker conditions and cheaper computational cost in this case. Note that in practice the computation of constant ℓ requires the computation of ℓ_0. Moreover the computation of ℓ_0 is less expensive than the computation of ℓ.

It turns out that we can improve on the error bounds given in Theorem 2 under the same hypotheses and computational cost. Indeed:

Theorem 6. *Assume hypotheses of Theorem 1 and condition (2.1) hold. Then sequence $\{x_n\}$ $(n \ge 0)$ generated by Newton's method (1.2) is well defined, remains in $U(x_0, r^*)$ for all $n \ge 0$, and converges to a unique solution x^* of equation $F(x) = 0$ in $U(x_0, r^*)$. Moreover, if scalar sequence s_n is defined by*

$$s_0 = 0, \quad s_n = s_{n-1} - \frac{\psi_a(s_{n-1})}{a(\lambda)\ell_0 s_{n-1}^{\lambda_0} - 1} \quad (n \ge 1) \tag{2.32}$$

then the following error bounds hold for all $n \ge 1$

$$\|x_n - x_{n-1}\| \le s_n - s_{n-1} \tag{2.33}$$

and

$$\|x_n - x^*\| \le s^* - s_n. \tag{2.34}$$

Furthermore if strict inequality holds in either of the inequalities (2.3), then we have:

$$s_n < r_n \quad (n \ge 2), \tag{2.35}$$

$$s_n - s_{n-1} < r_n - r_{n-1} \quad (n \ge 2), \tag{2.36}$$

and

$$s^* - s_n \le r^* - r_n \quad (n \ge 0). \tag{2.37}$$

where s^* is the limit of the sequence $\{s_n\}$.

Proof. We simply arrive at the more precise estimate

$$\|F'(x)^{-1}F'(x_0)\| \le [1 - \ell_0\|x - x_0\|^{\lambda_0}]^{-1} \tag{2.38}$$

instead of

$$\|F'(x)^{-1}F'(x_0)\| \le (1 - \ell\|x - x_0\|^{\lambda}) \tag{2.39}$$

used in the proof of Theorem 2 in [5, p. 720], for all $x \in U(x_0, R)$. Moreover note that because of (2.3) $\{s_n\}$ is a more precise majorizing sequence of $\{x_n\}$ that sequence $\{r_n\}$ (if strict inequality holds in (2.3) so otherwise $r_n = s_n$ $(n \ge 0)$). With the above changes the proof of Theorem 2 can be utilized so we can reach until (2.34).

Using (2.25), (2.32), and simple induction on n, we immediately obtain (2.35) and (2.36), whereas (2.37) is obtained from (2.36) by using standard majorization techniques [2], [4], [8], [13] (see also the proof of Proposition 10). That completes the proof of Theorem 6. □

Note also that the more precisely sequence $\{x_n\}$ remains in the smaller ball $U(x_0, s^*)$.

At this point we wonder if: (a) condition (2.18) can be weakened, by using more precise majorizing sequences along the lines of the proof of Theorem 4; (b) even more precise majorizing sequences than $\{s_n\}$ can be found.

It is convenient for us to define sequence $\{t_n\}\, m$ $(n \geq 0)$ by

$$t_0 = 0,\ t_1 = \eta,\ t_{n+2} = t_{n+1} + \frac{\ell\,(t_{n+1} - t_n)^{1+\lambda}}{(1+\lambda)\left[1 - \ell_0 t_{n+1}^{\lambda_0}\right]} \quad (n \geq 0)\,. \tag{2.40}$$

Iteration $\{t_n\}$ plays a crucial role as a majorizing sequence for $\{x_n\}$.

Clearly if

$$\ell_0 t_n^{\lambda_0} < 1 \quad (\ell_0 \neq 0, \lambda_0 \neq 0) \quad (n \geq 0) \tag{2.41}$$

holds for all $n \geq 0$, sequence $\{t_n\}$ is bounded above by $\ell_0^{-\frac{1}{\lambda_0}}$. Moreover by (2.40) it is also nondecreasing and as such it converges to some $t^* \in \left[0, \ell_0^{-\frac{1}{\lambda_0}}\right]$.

Next we provide conditions for the convergence of sequence $\{t_n\}$ to t^*. That is we show conditions (2.41).

Assume:

there exist parameters $\ell_0 > 0, \ell > 0, \eta > 0, \lambda_0 \in [0,1], \lambda \in [0,1], \gamma \geq 1$ such that

$$q_\gamma = \ell\eta^\lambda + (1+\lambda)\ell_0\gamma\eta^{\lambda_0} < 1 + \lambda. \tag{2.42}$$

Then interval

$$I = \left[1, \frac{1}{\ell_0\eta^{\lambda_0}} - \frac{\ell\eta^{\lambda-\lambda_0}}{(1+\lambda)\ell_0}\right] \neq \emptyset, \tag{2.43}$$

functions

$$c = c\,(\gamma) = \frac{\ell}{(1+\lambda)\,(1 - \ell_0\gamma\eta^{\lambda_0})}, \tag{2.44}$$

$$p_0 = p_0\,(\gamma) = c\,(\gamma)^{\frac{1}{\lambda}} \tag{2.45}$$

are well defined on I and

$$0 \leq c\eta^\lambda < 1. \tag{2.46}$$

Moreover assume

$$t_{n+1} \leq \gamma^{\frac{1}{\lambda_0}}\eta, \ \text{ for all } n \geq 0. \tag{2.47}$$

It then follows by (2.40)

$$\begin{aligned} t_{n+2}-t_{n+1} &= \frac{\ell}{(1+\lambda)\left(1-\ell_0 t_{n+1}{}^{\lambda_0}\right)}\left(t_{n+1}-t_n\right)^{1+\lambda} \le c\left(t_{n+1}-t_n\right)^{1+\lambda} \\ &\le c\left[c\left(t_n-t_{n-1}\right)^{1+\lambda}\right]^{1+\lambda} = c\cdot c^{1+\lambda}\left(t_n-t_{n-1}\right)^{(1+\lambda)^2} \\ &\le \ldots \le c^{\frac{(1+\lambda)^{n+1}-1}{1+\lambda-1}}\eta^{(1+\lambda)^{n+1}} = p_0^{-1}\left(p_0\eta\right)^{(1+\lambda)^{n+1}}. \end{aligned} \tag{2.48}$$

Let

$$d(\gamma) = \eta + \frac{1}{p_0}\left[\left(p_0\eta\right)^{(1+\lambda)^1} + \left(p_0\eta\right)^{(1+\lambda)^2} + \cdots + \left(p_0\eta\right)^{(1+\lambda)^n} + \cdots\right]. \tag{2.49}$$

Then d is a well defined function for all $\gamma \in I$.

Furthermore assume:

there exists $\gamma_0 \in I$ such that:

$$d(\gamma_0) \le \gamma_0^{\frac{1}{\lambda_0}}\eta. \tag{2.50}$$

Set

$$p = p_0(\gamma_0). \tag{2.51}$$

Under hypotheses (2.42), (2.47) and (2.50) sequence $\{t_n\}$ is nondecreasing and bounded above by $\gamma_0^{\frac{1}{\lambda_0}}\eta$ and as such it converges to some t^*. However it turns out hypothesis (2.47) can be dropped since it is implied by the other two. Indeed for all $n \ge 0$ we have

$$\ell_0 t_{n+1}^{\lambda_0} \le \ell_0 d^{\lambda_0}(\gamma_0) \le \ell_0\gamma_0\eta^{\lambda_0}, \tag{2.52}$$

which shows (2.41).

Hence, we showed:

Lemma 7. *Under the stated hypotheses:*

(a) condition (2.41) holds;

(b) sequence $\{t_n\}$ defined by (2.40) is nondecreasing and converges to some t^* such that

$$t_n \le t^* \le \left(\frac{1}{\ell_0}\right)^{\frac{1}{\lambda_0}} \quad (\lambda_0 \ne 0) \quad (n \ge 0) \tag{2.53}$$

Moreover the following error bounds hold for all $n \ge 0$:

$$0 \le t_{n+2}-t_{n+1} \le \tfrac{1}{p}\left(p\eta\right)^{(1+\lambda)^{n+1}} \le \tfrac{1}{p}q^{(1+\lambda)^{n+1}} \le \tfrac{1}{p}q^{(1+\lambda)(n+1)}, \tag{2.54}$$

$$0 \le t^*-t_n \le \tfrac{1}{p}\gamma_n \le \tfrac{1}{p}\overline{\gamma}_n \tag{2.55}$$

where, $q = p\eta$,

$$\gamma_n = \lim_{k\to\infty}\left\{(p\eta)^{(1+\lambda)^{n+k-1}} + \cdots + (p\eta)^{(1+\lambda)^n}\right\}$$
$$\leq \lim_{k\to\infty}\frac{(p\eta)^{(1+\lambda)^n}\left[1-(p\eta)^{(1+\lambda)k}\right]}{1-(p\eta)^{1+\lambda}}$$
$$\leq \frac{(p\eta)^{(1+\lambda)^n}}{1-(p\eta)^{1+\lambda}}, \tag{2.56}$$

and

$$\overline{\gamma}_n = \lim_{k\to\infty}\left[q^{(1+\lambda)^{n+k-1}} + \cdots + q^{(1+\lambda)^n}\right]$$
$$\leq \lim_{k\to\infty}\frac{q^{(1+\lambda)^n}\left(1-q^{(1+\lambda)k}\right)}{1-q^{1+\lambda}} \leq \frac{q^{(1+\lambda)^n}}{1-q^{1+\lambda}}. \tag{2.57}$$

We can show the main semilocal convergence theorem for Newton's method (1.2):

Theorem 8. *Let* $F\colon D \subseteq X \to Y$ *be a Fréchet-differentiable operator. Assume:*

there exist a point $x_0 \in D$ *and parameters* $\eta \geq 0$, $\ell_0 \geq 0$, $\ell \geq 0$, $\lambda \in [0,1]$, $q \in [0,1)$, $\overline{\delta} \in [0,1]$, $R > 0$ *such that: conditions (2.1), (2.2), (2.4), and (2.41) or hypotheses of Lemma 7 hold, and*

$$\overline{U}(x_0, t^*) \subseteq U(x_0, R). \tag{2.58}$$

Then, $\{x_n\}$ $(n \geq 0)$ *generated by Newton's method (1.2) is well defined, remains in* $\overline{U}(x_0, t^*)$ *for all* $n \geq 0$ *and converges to a unique solution* $x^* \in \overline{U}(x_0, t^*)$ *of equation* $F(x) = 0$.

Moreover the following error bounds hold for all $n \geq 0$*:*

$$\|x_{n+2} - x_{n+1}\| \leq \frac{\ell\|x_{n+1} - x_n\|^{1+\lambda}}{(1+\lambda)[1-\ell_0\|x_{n+1} - x_0\|^{\lambda_0}]} \leq t_{n+2} - t_{n+1} \tag{2.59}$$

and

$$\|x_n - x^*\| \leq t^* - t_n, \tag{2.60}$$

where iteration $\{t_n\}$ $(n \geq 0)$ *and point* t^* *are given in Lemma 7.*

Furthermore, if there exists $R > t^*$ *such that*

$$R_0 \leq R \tag{2.61}$$

and

$$\ell_0 \int_0^1 [\theta t^* + (1-\theta)R]^{\lambda_0} d\theta \leq 1, \tag{2.62}$$

the solution x^* *is unique in* $U(x_0, R_0)$.

Proof. We shall prove:

$$\|x_{k+1} - x_k\| \le t_{k+1} - t_k \tag{2.63}$$

and

$$\overline{U}(x_{k+1}, t^* - t_{k+1}) \subseteq \overline{U}(x_k, t^* - t_k) \tag{2.64}$$

hold for all $n \ge 0$.

For every $z \in \overline{U}(x_1, t^* - t_1)$

$$\|z - x_0\| \le \|z - x_1\| + \|x_1 - x_0\| \le t^* - t_1 + t_1 = t^* - t_0$$

implies $z \in \overline{U}(x_0, t^* - t_0)$. Since also

$$\|x_1 - x_0\| = \|F'(x_0)^{-1}F(x_0)\| \le \eta = t_1$$

(2.63) and (2.64) hold for $n = 0$. Given they hold for $n = 0, 1, \ldots, k$ then

$$\|x_{k+1} - x_0\| \le \sum_{i=1}^{k+1} \|x_i - x_{i-1}\| \le \sum_{i=1}^{k+1}(t_i - t_{i-1}) = t_{k+1} - t_0 = t_{k+1} \tag{2.65}$$

and

$$\|x_k + \theta(x_{k+1} - x_k) - x_0\| \le t_k + \theta(t_{k+1} - t_k) < t^*, \quad \theta \in [0, 1]. \tag{2.66}$$

Using (1.2) we obtain the approximation

$$\begin{aligned} F(x_{k+1}) &= F(x_{k+1}) - F(x_k) - F'(x_k)(x_{k+1} - x_k) \\ &= \int_0^1 [F'(x_k + \theta(x_{k+1} - x_k)) - F'(x_k)](x_{k+1} - x_k) d\theta \end{aligned} \tag{2.67}$$

and by (2.2)

$$\begin{aligned} &\|F'(x_0)^{-1}F(x_{k+1})\| \le \\ &\le \int_0^1 \|F'(x_0)^{-1}[F'(x_k + \theta(x_{k+1} - x_k)) - F'(x_k)]\| d\theta \|x_{k+1} - x_k\| \\ &\le \frac{\ell}{1+\lambda}\|x_{k+1} - x_k\|^{1+\lambda}. \end{aligned} \tag{2.68}$$

By (2.1), the estimate

$$\|F'(x_0)^{-1}[F'(x_{k+1}) - F'(x_0)]\| \le \ell_0 \|x_{k+1} - x_0\|^{\lambda_0} \le \ell_0 t_{k+1}^{\lambda_0} < 1 \quad \text{(by (3))}$$

and the Banach Lemma on invertible operators [8] $F'(x_{k+1})^{-1}$ exists and

$$\|F'(x_0)F'(x_{k+1})^{-1}\| \le \frac{1}{1 - \ell_0\|x_{k+1} - x_0\|^{\lambda_0}} \le \frac{1}{1 - \ell_0 t_{k+1}^{\lambda_0}}. \tag{2.69}$$

Therefore, by (1.2), (2.42), (2.68) and (2.69) we obtain in turn

$$\begin{aligned}\|x_{k+2}-x_{k+1}\| &= \|F'(x_{k+1})^{-1}F(x_{k+1})\| \\ &\leq \|F'(x_{k+1})^{-1}F'(x_0)\|\cdot\|F'(x_0)^{-1}F(x_{k+1})\| \\ &\leq \frac{\ell\|x_{k+1}-x_k\|^{1+\lambda}}{(1+\lambda)[1-\ell_0\|x_{k+1}-x_0\|^{\lambda_0}]} \\ &\leq \frac{\ell(t_{k+1}-t_k)^{1+\lambda}}{(1+\lambda)[1-\ell_0 t_{k+1}^{\lambda_0}]} = t_{k+2}-t_{k+1}. \qquad (2.70)\end{aligned}$$

Thus for every $z \in \overline{U}(x_{k+2}, t^*-t_{k+2})$ we have

$$\|z-x_{k+1}\| \leq \|z-x_{k+2}\|+\|x_{k+2}-x_{k+2}\| \leq t^*-t_{k+2}+t_{k+2}-t_{k+2} = t^*-t_{k+1}.$$

That is

$$z \in \overline{U}(x_{k+1}, t^*-t_{k+1}). \qquad (2.71)$$

Estimates (2.70) and (2.71) imply that (2.63) and (2.64) hold for $n = k+1$. By induction the proof of (2.63) and (2.64) is completed.

Lemma 7 implies that $\{t_n\}$ $(n \geq 0)$ is a Cauchy sequence. From (2.63) and (2.64) $\{x_n\}$ $(n \geq 0)$ becomes a Cauchy sequence too, and as such it converges to some $x^* \in \overline{U}(x_0, t^*)$ (since $\overline{U}(x_0, t^*)$ is a closed set) so that (2.60) holds.

The combination of (2.70) and (2.70) yields $F(x^*) = 0$. Finally to show uniqueness let y^* be a solution of equation $F(x) = 0$ in $U(x_0, R)$. It follows from (2.1), the estimate

$$\begin{aligned}&\left\|F'(x_0)^{-1}\int_0^1 [F'(y^*+\theta(x^*-y^*))-F'(x_0)]\right\| d\theta \\ &\leq \ell_0 \int_0^1 \|y^*+\theta(x^*-y^*)-x_0\|^{\lambda_0} d\theta \\ &\leq \ell_0 \int_0^1 [\theta\|x^*-x_0\|+(1-\theta)\|y^*-x_0\|]^{\lambda_0} d\theta \\ &< \ell_0 \int_0^1 [\theta t^*+(1-\theta)R_0]^{\lambda_0} d\theta \leq 1 \quad \text{(by (65))} \qquad (2.72)\end{aligned}$$

and the Banach Lemma on invertible operators that linear operator

$$L = \int_0^1 F'(y^*+\theta(x^*-y^*))d\theta \qquad (2.73)$$

is invertible. □

Using the identity

$$0 = F(y^*)-F(x^*) = L(x^*-y^*) \qquad (2.74)$$

we deduce $x^* = y^*$. To show uniqueness in $\overline{U}(x_0, t^*)$, using (4) we get:

$$\|F'(x_0)^{-1}(L - F'(x_0))\| \leq \frac{\ell_0}{1+\lambda_0}(t^*)^{1+\lambda_0} < 1 \quad \text{(by Lemma 1)},$$

which implies again $x^* = y^*$.

That completes the proof of Theorem 8.

Note that upper bounds on the distances $t_{n+1} - t_n$, $t^* - t_n$ are given in Lemma 7.

Remark 9. In the result that follows we show that our error bounds on the distances involved are finer and the location of the solution x^* at least as precise.

Proposition 10. *Under hypotheses of Theorems 6 and 8 with* $\ell_0 < \ell$ *the following error bounds hold:*

$$r_0 = t_0 = s_0 = 0, \quad r_1 = t_1 = s_1 = \eta$$

$$t_{n+1} < s_{n+1} < r_{n+1} \quad (n \geq 1), \tag{2.75}$$

$$t_{n+1} - t_n < s_{n+1} - s_n < r_{n+1} - r_n \quad (n \geq 1), \tag{2.76}$$

$$t^* - t_n \leq s^* - s_n \leq r^* - r_n \quad (n \geq 0) \tag{2.77}$$

and

$$t^* \leq s^* \leq r^*. \tag{2.78}$$

Proof. We use induction on the integer k to show the left hand sides of (2.75) and (2.76) first. By (2.32) and (2.42) we obtain

$$t_2 - t_1 = \frac{\ell\eta^{1+\lambda}}{(1+\lambda)[1-\ell_0\eta^{\lambda_0}]} < \frac{\psi_a(s_1)}{(1+\lambda)[1-\ell_0\eta^{\lambda_0}]} = s_2 - s_1$$

and

$$t_2 < s_2.$$

Assume:

$$t_{k+1} < s_{k+1}, \quad t_{k+1} - t_k < s_{k+1} - s_k \quad (k \leq n). \tag{2.79}$$

Using (2.32), and (2.42) we get

$$t_{k+2} - t_{k+1} = \frac{\ell(t_{k+1}-t_k)^{1+\lambda}}{(1+\lambda)[1-\ell_0 t_{k+1}^{\lambda_0}]} < \frac{\ell(s_{k+1}-s_k)^{1+\lambda}}{(1+\lambda)[1-\ell t_{k+1}^{\lambda_0}]} \leq s_{k+2} - s_{k+1},$$

(by the proof of Theorem 2.2 in [[5], end of page 720 and first half of page 721) and

$$t_{k+2} < s_{k+2}.$$

Let $m \geq 0$, we can obtain

$$\begin{aligned} t_{k+m} - t_k &< (t_{k+m} - t_{k+m-1}) + (t_{k+m-1} - t_{k+m-2}) + \cdots + (t_{k+1} - t_k) \\ &< (s_{k+m} - s_{k+m-1}) + (s_{k+m-1} - s_{k+m-2}) + \cdots + (s_{k+1} - s_k) \\ &= s_{k+m} - s_k. \end{aligned} \tag{2.80}$$

By letting $m \to \infty$ in (2.80) we obtain (2.77). For $n = 1$ in (2.77) we get (2.78).

That completes the proof of Proposition 10, since the right-hand side estimates in (2.76)–(2.78) were shown in Theorem 6. □

In the next remark we also show that our sufficient convergence conditions are weaker in general than the earlier ones.

Remark 11. (a) The Lipschitz case:

Set $\lambda_0 = \lambda = 1$. Then we showed in [3] that crucial condition (2.41) holds if

$$h_A = (\ell_0 + \ell)\,\eta \leq 1. \tag{2.81}$$

Moreover condition (2.18) reduces to (2.12) in this case. It follows from (2.12) and (2.81) that

$$h_K \leq 1 \Longrightarrow h_A \leq 1 \tag{2.82}$$

but not vice versa unless equality holds in (2.3).

Finer error bounds on the distances $\|x_{n+1} - x_n\|$, $\|x_n - x^*\|$ and a more precise information on the location of the solution x^* are also obtained under (2.81) [3].

(b) The Hölder case: $\lambda_0, \lambda \in (0,1)$.

Since $\frac{\lambda}{\lambda_0}, \frac{\ell}{\ell_0}$ can be arbitrarily large [3], clearly our conditions (2.41) or (2.42), (2.43), and (2.48) are weaker in general than all the ones by others already mentioned above. Moreover by Proposition 10, the remark made at the end of case (a) above also holds true for the Hölder case.

3. Local convergence analysis of Newton's method (2)

We state the following local convergence result for Newton's method (1.2).

Theorem 12. *Let $F\colon D \subseteq X \to Y$ be a Fréchet-differentiable operator. Assume:*

(a) there exist a simple zero $x^* \in D$ of equation $F(x) = 0$, parameters $\overline{\ell}_0 \geq 0$, $\ell \geq 0$, $\mu_0, \mu \in [0,1]$ not all zero at the same time such that:

$$\|F'(x^*)^{-1}[F'(x) - F'(y)]\| \leq \overline{\ell}\|x - y\|^{\mu}, \tag{3.1}$$

$$\|F'(x^*)^{-1}[F'(x) - F'(x^*)]\| \leq \overline{\ell}_0\|x - x^*\|^{\mu_0} \tag{3.2}$$

for all $x, y \in \overline{U}(x_0, R) \subseteq D$ $(R \geq 0)$*;*

(b) Equation:

$$\overline{\ell} r^{\mu} + (1+\mu)\left[\overline{\ell}_0 r^{\mu_0} - 1\right] = 0, \tag{3.3}$$

has a minimal solution δ *satisfying*

$$0 \leq \delta \leq R. \tag{3.4}$$

Then, Newton's method $\{x_n\}$ $(n \geq 0)$ *generated by (1.2) is well defined, remains in* $U(x^*, q)$ *for all* $n \geq 0$ *and converges to* x^**, provided that* $x_0 \in U(x^*, \delta)$*. Moreover the following error bounds hold for all* $n \geq 0$

$$\|x_{n+1} - x^*\| \leq \frac{\overline{\ell}\|x_n - x^*\|^{1+\mu}}{(1+\mu)[1 - \overline{\ell}_0\|x_n - x^*\|^{\mu_0}]}\,. \tag{3.5}$$

Proof. Inequality (3.5) follows from the approximation

$$\begin{aligned} &x_{n+1} - x^* \\ &= x_n - x^* - F'(x_n)^{-1}F(x_n) \\ &= -[F'(x_n)^{-1}F'(x^*)]\left\{F'(x^*)^{-1}\int_0^1 [F'(x^*+t(x_n - x^*)) - F'(x_n)](x_n - x^*)dt\right\}, \end{aligned} \tag{3.6}$$

and estimates

$$\|F'(x_n)^{-1}F'(x^*)\| \leq [1 - \overline{\ell}_0\|x_n - x^*\|^{\mu_0}]^{-1} \quad \text{(see (2.72))} \tag{3.7}$$

$$\begin{aligned} &\left\|F'(x^*)^{-1}\int_0^1 [F'(x^* + t(x_n - x^*)) - F'(x_n)](x_n - x^*)dt\right\| \\ &\leq \frac{\overline{\ell}}{1+\mu}\|x_n - x^*\|^{1+\mu}, \quad \text{(see (2.71))}. \end{aligned} \tag{3.8}$$

The rest follows using induction on the integer n, (3.6)–(3.8) and along the lines of the proof of Theorem 8.

That completes the proof of Theorem 8. ☐

The corresponding local result for the modified Newton method (1.3) can be:

Remark 13. Using only condition (3.1) and the approximation

$$y_{n+1} - x^* = F'(y_0)^{-1}\int_0^1 [F'(x^* + t(y_n - x^*)) - F'(y_0)](y_n - x^*)dt \tag{3.9}$$

as in the proof of Theorem 12 we obtain the convergence radius

$$q_0 = \begin{cases} \left[\dfrac{1+\mu}{(2^{1+\mu}-1)\overline{\ell}}\right]^{1/\mu}, & \overline{\ell} \neq 0,\ \mu \neq 0 \\ R, & \mu = 0, \end{cases} \tag{3.10}$$

and the corresponding error bounds

$$\begin{aligned} \|y_{n+1} - x^*\| &\leq \overline{\ell} \int_0^1 [\|x^* - y_0\| + t\|y_n - x^*\|]^{\mu} dt \|y_n - x^*\| \\ &\leq \frac{\overline{\ell}(2^{1+\mu}-1)}{1+\mu} q_0^{\mu} \|y_n - x^*\| \quad (n \geq 0). \end{aligned} \tag{3.11}$$

Remark 14. As noted in [2], [4] and [14] the local results obtained here can be used for projection methods such as Arnoldi's, the generalized minimum residual method (GMRES), the generalized conjugate residual method (GCR), for combined Newton/finite-difference projection methods and in connection with the mesh independence principle in order to develop the cheapest mesh refinement strategies.

Remark 15. The local results obtained here can also be used to solve equations of the form $F(x) = 0$, where F' satisfies the autonomous differential equation [2], [4], [8]:

$$F'(x) = T(F(x)), \tag{3.12}$$

where, $T\colon Y \to X$ is a known continuous operator. Since $F'(x^*) = T(F(x^*)) = T(0)$, we can apply the results obtained here without actually knowing the solution x^* of equation (1.1).

We complete this section with a numerical example to show that through Theorem 6 we can obtain a wider choice of initial guesses x_0 than before.

Example 16. Let $X = Y = \mathbf{R}$, $D = U(0,1)$ and define function F on D by

$$F(x) = e^x - 1. \tag{3.13}$$

Then it can easily be seen that we can set $T(x) = x + 1$ in (3.12). Since $F'(x^*) = 1$, we get $\|F'(x) - F'(y)\| \leq e\|x-y\|$. Hence we set $\overline{\ell} = e$, $\mu_0 = \mu = 1$. Moreover since $x^* = 0$ we obtain in turn

$$\begin{aligned} F'(x) - F'(x^*) &= e^x - 1 = x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \cdots \\ &= \left(1 + \frac{x}{2!} + \cdots + \frac{x^{n-1}}{n!} + \cdots\right)(x - x^*) \end{aligned}$$

and for $x \in U(0,1)$,

$$\|F'(x) - F'(x^*)\| \leq (e-1)\|x - x^*\|.$$

That is, $\bar{\ell}_0 = e - 1$. Using (3.3) we obtain

$$r^* = .254028662$$

Rheinboldt's radius is given by

$$p = \frac{2}{3\bar{\ell}}\,.$$

Note that

$$p < r^* \quad (\bar{\ell} < \ell).$$

In particular in this case we obtain

$$p = .245252961.$$

That is our convergence radius r^* is larger than the corresponding p due to Rheinboldt [10]. This observation is very important in computational mathematics (see Remark 15). Note also that local results were not given in [5]–[7].

The case $\mu \in [0,1)$ was not covered in [5]–[7], [9], [11], [12]. The "terra incognita" can be examined along the lines of the semilocal case studied above. However we leave the details to the motivated reader.

References

[1] J. Appel, E. DePascale, Ju. V. Lysenko and P. P. Zabrejko, *New results on Newton–Kantorovich approximations with applications to nonlinear integral equations*, Numer. Funct. Anal. and Optimiz. **18**, (1 and 2) (1997), 1–18.

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

[3] I. K. Argyros, *An improved convergence analysis and applications for Newton-like methods in Banach space*, Numer. Funct. Anal. and Optimiz. **24**, (7 and 8) (2003), 653–672.

[4] I. K. Argyros and F. Szidarovszky, *The Theory and Applications of Iteration Methods*, C. R. C. Press, Boca Raton, Florida, 1993.

[5] F. Cianciaruso and E. DePascale, *Newton–Kantorovich approximations when the derivative is Hölderian: Old and new results*, Numer. Funct. Anal. and Optimiz., **24** (7 and 8) (2003), 713–723.

[6] N. T. Demidovich, P. P. Zabreiko and Ju. V. Lysenko, *Some remarks on the Newton–Kantorovich method for nonlinear equations with Hölder continuous linearizations*, Izv. Akad. Nauk. Beloruss, **3** (1993), 22–26 (in Russian).

[7] E. DePascale and P. P. Zabreiko, *The convergence of the Newton–Kantorovich mehtod under Vertgeim conditions: A new improvement*, Z. Anal. Anwend., **17**, 2 (1998), 271–280.

[8] L. V. Kantorovich and G. P. Akilov, *Functional Analysis in Normed Spaces*, Pergamon Press, Oxford, 1964.

[9] J. V. Lysenko, *Conditions for the convergence of the Newton–Kantorovich method for nonlinear equations with Hölder linearizations*, Dokl. Akad. Nauk USSR, **38** (1994), 20–24 (in Russian).
[10] W. C. Rheinboldt, *An adaptive continuation process for solving systems of nonlinear equations*, Polish Academy of Science, Banach Ctr. Publ. **3** (1977), 129–142.
[11] B. A. Vertgeim, *On conditions for the applicability of Newton's method*, Dokl. Akad. Nauk SSSR **110** (1956), 719–722 (in Russian).
[12] B. A. Vertgeim, *On some methods of the approximate solution of nonlinear functional equations in Banach spaces*, Amer. Math. Soc. Transl. **16**, 2 (1957), 378–382.
[13] T. Yamamoto, *Historical developments in the convergence analysis for Newton's and Newton-like methods*, J. Comput. Appl. Math. **124** (2000), 1–23.
[14] T. J. Ypma, *Local convergence of inexact Newton methods*, SIAM J. Numer. Anal. **21**, 3 (1984), 583–590.