Computational and Applied Mathematics (2023) 42:76
https://doi.org/10.1007/s40314-023-02203-2

Extended Kantorovich theory for solving nonlinear equations with applications

Samundra Regmi[1] · Ioannis K. Argyros[2] · Santhosh George[3] · Michael Argyros[4]

Received: 24 August 2022 / Revised: 13 December 2022 / Accepted: 14 January 2023 /
Published online: 8 February 2023

Abstract

The Kantorovich theory plays an important role in the study of nonlinear equations. It is used to establish the existence of a solution for an equation defined in an abstract space. The solution is usually determined by using an iterative process such as Newton's or its variants. A plethora of convergence results are available based mainly on Lipschitz-like conditions on the derivatives, and the celebrated Kantorovich convergence criterion. But there are even simple real equations for which this criterion is not satisfied. Consequently, the applicability of the theory is limited. The question there arises: is it possible to extend this theory without adding convergence conditions? The answer is, Yes! This is the novelty and motivation for this paper. Other extensions include the determination of better information about the solution, i.e. its uniqueness ball; the ratio of quadratic convergence as well as more precise error analysis. The numerical section contains a Hammerstein-type nonlinear equation and other examples as applications.

Keywords Nonlinear equation · Criterion · Integral equation · Convergence

Communicated by Andreas Fischer.

✉ Ioannis K. Argyros
iargyros@cameron.edu

Samundra Regmi
sregmi5@uh.edu

Santhosh George
sgeorge@nitk.edu.in

Michael Argyros
michael.i.argyros-1@ou.edu

1 Department of Mathematics, University of Houston, Houston, TX 772005, USA

2 Department of Computing and Mathematical Sciences, Cameron University, Lawton, OK 73505, USA

3 Department of Mathematical and Computational Sciences, National Institute of Technology Karnataka, Mangaluru 575 025, India

4 Department of Computer Science, University of Oklahoma, Norman, OK 73019, USA

Mathematics Subject Classification 65G99 · 65H10 · 65N12 · 58C15

1 Introduction

Let $\mathcal{U}$, $\mathcal{V}$, be Banach spaces and $L(\mathcal{U}, \mathcal{V})$ stand for the space of all continuous linear operators mapping $\mathcal{U}$ into $\mathcal{V}$. Consider, a differentiable mapping as per Fréchet $\mathcal{L} : \Omega \subseteq \mathcal{U} \longrightarrow \mathcal{V}$, and its corresponding nonlinear equation

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

with Ω denoting a non-empty open set. The task of determining a solution $x_* \in \Omega$ is very challenging but important since applications from numerous computational disciplines are brought in the form (1.1) (Argyros and Magréñan 2018; Argyros 2004a, b; Ezquerro et al. 2010; Ortega and Rheinboldt 1970; Verma 2019). The analytic form of x_* is rarely attainable. That is why mainly iterative processes are used to generate approximations to the solution x_*.

Among these processes, the most widely used is Newton's and its variants. In particular, Newton's process (NP) is defined by

$$x_0 \in \Omega,\ x_{n+1} = x_n - \mathcal{L}'(x_n)^{-1}\mathcal{L}(x_n) \text{ for } n = 0, 1, 2, \ldots \tag{1.2}$$

There exists a plethora of results related to the study of NP (Argyros and Magréñan 2018; Argyros 2004a; Argyros and Hilout 2010). These studies are based on the theory inaugurated by Kantorovich and its variants (Argyros 2021, 2022, 2004a, b; Argyros and Magréñan 2018; Argyros and Hilout 2010; Dennis 1968; Ezquerro et al. 2010; Gragg and Tapia 1974; Hernandez 2001; Ezquerro and Hernandez 2018; Kantorovich and Akilov 1982; Ortega and Rheinboldt 1970; Potra and Pták 1980, 1984; Proinov 2010; Rheinboldt 1968; Tapia 1971).

The following conditions (A) are used in non-affine or affine invariant form.

Suppose: (A1) $\exists$ point $x_0 \in \Omega$ and parameter $\lambda \geq 0 : \mathcal{L}'(x_0)^{-1} \in L(\mathcal{V}, \mathcal{U})$ and

$$\|\mathcal{L}'(x_0)^{-1}\mathcal{L}(x_0)\| \leq \lambda.$$

(A2) $\exists$ parameter $M_1 > 0$: Lipschitz condition

$$\|\mathcal{L}'(x_0)^{-1}(\mathcal{L}'(w_1) - \mathcal{L}'(w_2))\| \leq M_1\|w_1 - w_2\|)$$

holds $\forall w_1 \in \Omega$ and $w_2 \in \Omega$.

(A3)

$$\lambda \leq \frac{1}{2M_1}. \tag{1.3}$$

(A4) $B[x_0, \rho] \subset \Omega$, where parameter $\rho > 0$ is given later.

Let us denote $B[x_0, r] := \{x \in \Omega : \|x - x_0\| \leq r\}$ for $r > 0$. Set $\rho = r_1 = \frac{1-\sqrt{1-2M_1\lambda}}{M_1}$.

There are many variants of Kantorovich's convergence result for NP. One of those follows (Chen and Yamamoto 1989; Deuflhard 2004; Kantorovich and Akilov 1982).

Theorem 1.1 *Under the conditions A for $\rho = r_1$; the NP is contained in $B(x_0, r_1)$, convergent to a solution $x_* \in B[x_0, r_1]$ of Eq.* (1.1) *and*

$$\|x_{n+1} - x_*\| \leq t_{n+1} - t_n = \frac{M_1(t_n - t_{n-1})^2}{2(1 - M_1 t_n)},$$

where the scalar sequence $\{t_n\}$ *is given by*

$$t_0 = 0,\ t_1 = \lambda,\ t_{n+1} = t_n + \frac{M_1(t_n - t_{n-1})^2}{2(1 - M_1 t_n)}.$$

Moreover, the convergence is linear if $\lambda = \frac{1}{2M_1}$ *and quadratic if* $\lambda < \frac{1}{2M_1}$. *Furthermore, the solution is unique* $B[x_0, r_1]$ *in the first case and in* $B(x_0, r_2)$ *in the second, where* $r_2 = \frac{1+\sqrt{1-2M_1\lambda}}{M_1}$.

A plethora of studies has used conditions A (Argyros and Magréñan 2018; Argyros 2004a; Argyros and Hilout 2010).

Example 1.2 Consider cubic polynomial

$$c(x) = x^3 - \mu.$$

Let $\Omega = B(x_0, 1-\mu)$ for some parameter $\mu \in (0, \frac{1}{2})$. Choose $x_0 = 1$. Then, the conditions A are verified for $\lambda = \frac{1-\mu}{3}$ and $M_1 = 2(2-\mu)$. It follows that the estimate

$$\frac{1-\mu}{3} > \frac{1}{4(2-\mu)}$$

holds $\forall \mu \in (0, \frac{1}{2})$. That is condition (A3) is not satisfied. Therefore, the convergence is not assured by this theorem also used in Chen and Yamamoto (1989), Dennis (1968), Deuflhard (2004), Ezquerro et al. (2010), Gragg and Tapia (1974), Hernandez (2001), Ezquerro and Hernandez (2018), Kantorovich and Akilov (1982), Ortega and Rheinboldt (1970), Potra and Pták (1980, 1984), Proinov (2010), Rheinboldt (1968), Tapia (1971), Yamamoto (1987a, b, 2000) and Zabrejko and Bnuen (1987). But the NP converges. Hence, clearly, there is a need to improve the results based on condition A, which is only sufficient but not necessary.

In this paper, several avenues are presented for achieving this goal. The idea is to replace the Lipschitz parameter M_1 with smaller ones.

Consider the center Lipschitz condition

$$\|\mathcal{L}'(x_0)^{-1}(\mathcal{L}'(w_1) - \mathcal{L}'(x_0))\| \le M_0\|w_1 - x_0\| \ \ \forall w_1 \in \Omega, \tag{1.4}$$

the set $\Omega_1 = B[x_0, \frac{1}{M_0}] \cap \Omega$ and the Lipschitz -2 condition

$$\|\mathcal{L}'(x_0)^{-1}(\mathcal{L}'(w_1) - \mathcal{L}'(w_2))\| \le M\|w_1 - w_2\| \ \ \forall w_1, w_2 \in \Omega_1. \tag{1.5}$$

Notice that by the definition of the set Ω_1

$$\Omega_1 \subset \Omega_0. \tag{1.6}$$

Then, the Lipschitz parameters are related by

$$M_0 \le M_1, \tag{1.7}$$

and

$$M \le M_1. \tag{1.8}$$

Notice also since parameters M_0 and M are specializations of parameter M_1, $M_1 = M_1(\Omega)$, $M_0 = M_0(\Omega)$, but $M = M(\Omega_1)$, where by $M_1(\Omega)$ we mean that the parameter M_1 depends on the set Ω. Therefore, no additional work is required to find M_0 and M (see also the numerical examples).

Moreover, the ratio $\frac{M_0}{M}$ can be very small (arbitrarily).

Example 1.3 Define scalar function

$$\mathcal{L}(x) = b_0 x + b_1 + b_2 \sin e^{b_3 x},$$

for $x_0 = 0$, where $b_j,\ j = 0, 1, 2, 3$ are real parameters. It follows by this definition that for b_3 sufficiently large and b_2 sufficiently small, $\frac{M_0}{M_1}$ can be small (arbitrarily), i.e., $\frac{M_0}{M_1} \longrightarrow 0$.

Other extensions involve tighter majorizing sequences for NP (see Sect. 2) and improved uniqueness results for solution x_* (Sect. 3). The applications appear in Sect. 4 followed by and the conclusions in Sect. 5.

2 Real sequences

Let K_0, M_0, K, M and λ be positive parameters. An important role in the study of NM is played by the majorizing sequence $\{s_n\}$ defined for $s_0 = 0,\ s_1 = \lambda$, as

$$s_2 = s_1 + \frac{K(s_1 - s_0)^2}{2(1 - K_0 s_1)},\ s_{n+2} = s_{n+1} + \frac{M(s_{n+1} - s_n)^2}{2(1 - M_0 s_{n+1})}. \tag{2.1}$$

That is why some convergence results are listed for it in what follows in this section.

Lemma 2.1 *Suppose conditions $K_0\lambda < 1$ and $s_{n+1} < \frac{1}{M_0}$ hold for all $n = 1, 2, \ldots$ Then, the following assertions hold*

$$s_n < s_{n+1} < \frac{1}{M_0}, \text{ for all } n = 0, 1, 2, \ldots$$

and there exists $s_* \in [\lambda, \frac{1}{M_0}]$ such that $\lim_{n\longrightarrow\infty} s_n = s_*$.

Proof The definition of sequence $\{s_n\}$ and the conditions of the Lemma imply the assertion and $\lim_{n\longrightarrow\infty} s_n = s_* \in [\lambda, \frac{1}{M_0}]$. Notice that s_* is the least upper bound (unique) of the sequence $\{s_n\}$. □

Next, criteria stronger than those in Lemma 2.1 are developed for the convergence of the sequence (2.1). However, these criteria are easier to verify than those of the Lemma 2.1.

Define parameter γ by

$$\gamma = \frac{2M}{M + \sqrt{M^2 + 8M_0 M}}.$$

This parameter plays a role in the study of NP.

Suppose from now on that $K_0 \leq M_0$. Define the real quadratic polynomials q, q_1, q_2 by

$$q(t) = M_0(K - 2K_0)t^2 + 2M_0 t - 1,$$
$$q_1(t) = (MK + 2\gamma M_0(K - 2K_0))t^2 + 4\gamma(M_0 + K_0)t - 4\gamma,$$

and

$$q_2(t) = M_0(K - 2(1-\gamma)K_0)t^2 + 2(1-\gamma)(M_0 + K_0)t - 2(1-\gamma).$$

The discriminants D, D_1, D_2 of these polynomials can be given as

$$D = 4M_0(M_0 + K - 2K_0) > 0,$$
$$D_1 = 16\gamma(\gamma(M_0 - K_0)^2 + (M + 2\gamma M_0)K) > 0$$

and

$$D_2 = 4(1-\gamma)((1-\gamma)(M_0 - K_0)^2 + 2M_0 K) > 0,$$

respectively. It follows by the definition of γ, q_1 and q_2 that

$$M = \frac{2M_0\gamma^2}{1-\gamma},\ MK + 2\gamma M_0(K - 2K_0) = \frac{2M_0\gamma}{1-\gamma}(K - 2(1-\gamma)K_0),$$

since after multiplying the polynomial q_2 by $\frac{2M_0\gamma}{1-\gamma}$, we obtain the polynomial q_1 i.e.,

$$q_1(t) = \frac{2M_0\gamma}{1-\gamma} q_2(t).$$

That is polynomials q_1 and q_2 have the same roots. Denote by $\frac{1}{2r_1}$ the unique positive root of polynomial q. This root is given explicitly by the quadratic formula and can be written as

$$\frac{1}{2r_1} = \frac{1}{M_0 + \sqrt{M^2 + M_0(K - 2K_0)}}.$$

Moreover, denote by $\frac{1}{2r_2}$ the common positive root of polynomials q_1 and q_2. This root can also be written as

$$\frac{1}{2r_2} = \frac{2}{\gamma(M_0 + K_0) + \sqrt{(\gamma(M_0 + K_0))^2 + \gamma(MK + 2\gamma M_0(K - 2K_0)}}.$$

Define parameter N by

$$N^{-1} = \min\left\{\frac{1}{2r_1}, \frac{1}{2r_2}\right\}.$$

Suppose

$$\lambda \le \frac{1}{2N}. \tag{2.2}$$

By the choice of parameters r_1, r_2 polynomials q, q_1, q_2 and condition (2.2), it follows that $M_0 s_2 < 1$, since $q(\lambda) < 0$, $K_0\lambda < 1$, $q_1(\lambda) \le 0$ and $q_2(\lambda) \le 0$. Furthermore, the following estimate holds

$$\gamma_0 \le \gamma \le 1 - \frac{M_0(s_2 - s_1)}{1 - M_0 s_1}, \tag{2.3}$$

where the parameter $\gamma_0 = \frac{M(s_2 - s_1)}{2(1 - M_0 s_2)}$. Indeed, the left hand side inequality reduces to $q_1(\lambda) \le 0$ and the right hand side to $q_2(\lambda) \le 0$. These assertions are true by the choice of λ.

Lemma 2.2 *Under condition (2.2), sequence $\{s_n\}$ satisfies*

$$s_n < s_{n+1} \le \bar{s}_{**} = \lambda + \left(1 + \frac{\gamma_0}{1-\gamma}\right)\frac{K\lambda^2}{2(1 - K_0\lambda)}, \tag{2.4}$$

$$0 < s_{n+2} - s_{n+1} \le \gamma_0\gamma^{n-1}\frac{K\lambda^2}{2(1 - K_0\lambda)} \quad \text{for all } n = 1, 2, \ldots \tag{2.5}$$

and is convergent to its least upper bound $s_* \in (\lambda, \bar{s}_{**}]$ so that

$$s_* - s_n \le \frac{\gamma_0(s_2 - s_1)\gamma^{n-2}}{1-\gamma} \quad \text{for all } n = 2, 3, \ldots \tag{2.6}$$

Proof Induction is used to show the estimate

$$0 < \frac{M(s_{k+1} - s_k)}{2(1 - M_0 s_{k+1})} \leq \gamma \ \forall\ n = 1, 2, 3, \ldots \tag{2.7}$$

It follows by the definition of roots $\frac{1}{2r_1}$, $\frac{1}{2r_2}$ and polynomial g_1 that estimate (2.7) holds for $k = 1$. Using the Definition (2.1) of sequence $\{s_n\}$ and parameter γ_0

$$\begin{aligned}
0 < s_3 - s_2 \leq \gamma_0(s_2 - s_1) &\Rightarrow s_3 \leq s_2 + \gamma_0(s_2 - s_0)\\
&\Rightarrow s_3 \leq s_2 + (1 + \gamma_0)(s_2 - s_1) - (s_2 - s_1)\\
&\Rightarrow s_3 \leq s_1 + \frac{1 - \gamma_0^2}{1 - \gamma_0}(s_2 - s_1) < \bar{s}_{**}.
\end{aligned}$$

Suppose that estimate (2.7) holds for $k = 1, 2, \ldots, n - 1$. Then, similarly by (2.7) and the induction hypotheses, we obtain in turn

$$\begin{aligned}
s_{k+2} &\leq s_{k+1} + \gamma_0\gamma^{k-1}(s_2 - s_1)\\
&\leq s_k + \gamma_0\gamma^{k-2}(s_2 - s_1) + \gamma_0\gamma^{k-1}(s_2 - s_1)\\
&\leq s_1 + (1 + \gamma_0(1 + \gamma + \cdots + \gamma^{k-1}))(s_2 - s_1)\\
&= \lambda + \left(1 + \gamma_0\frac{1 - \gamma^k}{1 - \gamma}\right)(s_2 - s_1)\\
&< \bar{s}_{**}.
\end{aligned}$$

It follows by the definition (2.1) of sequence $\{s_n\}$, estimate (2.3) and induction hypothesis (2.7) that

$$0 < s_{k+2} - s_{k+1} \leq \gamma_0\gamma^{k-1}(s_2 - s_1) \leq \gamma^k(s_2 - s_1).$$

Then, the estimate (2.7) for $k + 1$ replacing k holds, if

$$\frac{M}{2}(s_{k+2} - s_{k+1}) \leq \gamma(1 - M_0 s_{k+1}),$$

or

$$\frac{M}{2}(s_{k+2} - s_{k+1}) + \gamma M_0 s_{k+1} - \gamma \leq 0,$$

or

$$\frac{M}{2}\gamma^k(s_2 - s_1) + \gamma M_0\left(\lambda + \frac{1 - \gamma^{k+1}}{1 - \gamma}(s_2 - s_1)\right) - \gamma \leq 0,$$

or

$$p_k(t) \leq 0 \text{ at } t = \gamma, \tag{2.8}$$

where, the polynomial $p_k : [0, 1) \longrightarrow \mathbb{R}$ is defined by

$$p_k(t) = \frac{M}{2}(s_2 - s_1)t^{k+1} + tM_0(1 + t + \cdots + t^k)(s_2 - s_1) - (1 - M_0 s_1)t. \tag{2.9}$$

There is a connection between consecutive polynomials:

$$\begin{aligned}
p_{k+1}(t) - p_k(t) = {}&\frac{M}{2}(s_2 - s_1)^{k+2} + tM_0(1 + t + \cdots + t^{k+1})(s_2 - s_1)\\
&-(1 - M_0\lambda)t - \frac{M}{2}(s_2 - s_1)t^k
\end{aligned}$$

$$-tM_0(1+t+\cdots+t^k)(s_2-s_1)+(1-M_0\lambda)t$$
$$=\frac{1}{2}(2M_0t^2+Mt-1)t^k(s_2-s_1).$$

It follows that

$$p_{k+1}(t)=p_k(t)+\frac{1}{2}q_3(t)t^k(s_2-s-1),$$

where

$$q_3(t)=2M_0t^2+Mt-M.$$

Notice that $q_3(\gamma)=0$ by the definition of γ. Then, in particular

$$p_{k+1}(\gamma)=p_k(\gamma).$$

Define function $p_\infty:[0,1)\longrightarrow\mathbb{R}$ by

$$p_\infty(t)=\lim_{k\longrightarrow\infty}p_k(t).$$

By this definition and polynomials p_k

$$p_\infty(t)=\gamma(\frac{M_0}{1-\gamma}(s_2-s_1)+M_0s_1-1). \tag{2.10}$$

Consequently, assertion (2.8) holds if

$$\frac{1}{\gamma}p_\infty(t)\leq 0 \text{ at } t=\gamma,$$

which is true by the choice of parameter $\frac{1}{2r_2}$ and polynomial p_2. The induction for assertion (2.7) is terminated leading to the conclusions. □

Remark 2.3 The linear convergence of sequence $\{s_n\}$ is shown under the condition (2.2). This condition provides the smallness of λ to force convergence. The quadratic convergence of sequence $\{s_n\}$ can be shown if λ is chosen to be bounded above from by a smaller parameter than $\frac{1}{2N}$. Moreover, under condition (2.2) an upper bound on iterate s_k is obtained which can then be used in the proof to show quadratic convergence.

Lemma 2.4 *Under condition (2.10) further suppose that for some $\epsilon>0,\ \beta=\frac{\epsilon}{\epsilon+1}$*

$$M_0\left(\frac{\gamma_0(s_2-s_1)}{1-\gamma}+\lambda+s_2-s_1\right)\leq\beta \tag{2.11}$$

and

$$\lambda<\frac{2}{(1+\epsilon)M}. \tag{2.12}$$

Then, the conclusions of Lemma 2.2 hold for sequence $\{s_n\}$,

$$s_{n+1}-s_n\leq\frac{M}{2}(1+\epsilon)(s_n-s_{n-1})^2 \tag{2.13}$$

and

$$0<s_{n+1}-s_n\leq\frac{1}{b}(b\lambda)^{2^n}, \tag{2.14}$$

where $b=\frac{M}{2}(1+\epsilon)$ and $b\lambda<1$.

Proof Assertion (2.14) certainly holds if the following estimate is shown

$$0 < \frac{M}{2(1 - M_0 s_{k+1})} \leq \frac{M}{2}(1 + \epsilon). \tag{2.15}$$

The estimate (2.15) holds for $k = 1$, since it is equivalent to $M_0 \lambda \leq \beta$. But this is true by $M_0 \leq 2N$, condition (2.2) and inequality $\frac{\epsilon M_0}{2(1+\epsilon)N} \leq \beta$.

Define polynomials $g_n : [0, 1) \longrightarrow \mathbb{R}$ by

$$g_n(t) = (1 + \epsilon)M_0\gamma_0(1 + t + \cdots + t^{n-1})(s_2 - s_1) + (1 + \epsilon)M_0(\lambda + s_2 - s_1) - \epsilon. \tag{2.16}$$

It follows from this definition that

$$g_{n+1}(t) - g_n(t) = (1 + \epsilon)M_0\gamma_0(s_2 - s_1)t^n > 0,$$

Evidently, estimate (2.15) holds. Define function $g_\infty : (0, 1) \longrightarrow \mathbb{R}$ by

$$g_\infty(t) = \lim_{k \longrightarrow \infty} g_k(t).$$

Hence, we get $g_\infty(t) = \frac{(1+\epsilon)M_0\gamma_0(s_2-s_1)}{1-t} + (1 + \epsilon)M_0(\lambda + s_2 - s_1) - \epsilon$. Evidently, the estimate

$$g_n(t) \leq 0 \text{ at } t = \gamma$$

holds if instead

$$g_\infty(t) \leq 0 \text{ at } t = \gamma.$$

But this is identical to condition (2.11). The induction for the assertion (2.15). Then, it follows by estimate (2.15) and the definition of sequence $\{s_n\}$ that assertion (2.13) holds. Using the definition of parameter b and estimate (2.13)

$$\begin{aligned} b(s_{k+1} - s_k) &\leq (b(s_k - s_{k-1}))^2 = (b(s_k - s_{k-1})^2 \\ &\leq b^2(b(s_{k-1} - s_{k-2})^2)^2 = b^2b^2(s_{k-1} - s_{k-2})^{2^2} \\ &\leq b^2b^2b^2(s_{k-2} - s_{k-3})^{2^3} \leq \ldots, \end{aligned}$$

thus,

$$\begin{aligned} s_{k+1} - s_k &\leq b^{1+2+2^2+\cdots+2^{k-1}}\lambda^{2^k} \\ &= b^{\frac{2^k-1}{2-1}}\lambda^{2^k} \\ &= b^{-1}b^{2^k}\lambda^{2^k} = \frac{(b\lambda)^{2^k}}{b}. \end{aligned}$$

Notice that $0 < b\lambda < 1$ by the condition (2.12). Hence, sequence $\{s_k\}$ converges quadratically to t_*. □

Remark 2.5 Condition (2.11) is left uncluttered. It can be expressed as a function of λ by

$$\varphi(\lambda) = \frac{M_0\gamma_0(s_2 - s_1)}{1 - \gamma} + M_0(\lambda + s_2 - s_1) - \beta.$$

Suppose

$$\lambda < \frac{\epsilon}{(\epsilon + 1)M_0}. \tag{2.17}$$

It follows by the definition of polynomial φ and condition (2.17) that $\varphi(0) = -\beta < 0$. Moreover, $\varphi(t) \longrightarrow \infty$ as $t \longrightarrow r_3 = \min\{\frac{1}{K_0}, \frac{1}{2r_2}\}$. Hence, function φ_1 has zeros in interval $(0, r_3)$ as a consequence of the intermediate value theorem. Let λ_0 stand for the smallest such zero. Then, conditions (2.2), (2.11), and (2.12) are condensed as

$$\lambda \leq \frac{1}{2N_0} := \min\left\{\frac{1}{N}, \frac{2}{(1+\epsilon)M}, \frac{\epsilon}{(1+\epsilon)M_0}, \lambda_0\right\}. \tag{2.18}$$

If $\frac{1}{2N_0} = \frac{\epsilon}{(1+\epsilon)M_0}$ or $\frac{1}{2N_0} = \frac{2}{(1+\epsilon)M}$, then condition (2.18) should hold as a strict inequality.

3 Convergence of NP

The Lipschitz parameters are associated with operator $\mathcal{L}$ and its derivatives.

Suppose there exist parameters $K_0 > 0$, $K > 0$ such that

$$\|\mathcal{L}'(x_0)^{-1}(\mathcal{L}'(x_1) - \mathcal{L}'(x_0))\| \leq K_0\|x_1 - x_0\|, \tag{3.1}$$

$$\|\mathcal{L}'(x_0)^{-1}(\mathcal{L}'(x_0 + \xi(x_1 - x_0)) - \mathcal{L}'(x_0))\| \leq K\xi\|x_1 - x_0\|, \tag{3.2}$$

for $x_1 = x_0 - \mathcal{L}'(x_0)^{-1}\mathcal{L}(x_0)$ and each $\xi \in [0, 1]$ and

$$B[x_0, t_*] \subset \Omega. \tag{3.3}$$

Conditions (A1), (1.4), (1.5), (3.1)–(3.3) and those of Lemma 2.1 or Lemma 2.2 are summarized by (H).

Next, under conditions H, we show the main convergence result for NP.

Theorem 3.1 *Under conditions H sequence NP is convergent to a solution $x_* \in B[x_0, t_*]$ of equation $\mathcal{L}(x) = 0$. Moreover, upper bounds*

$$\|x_* - x_i\| \leq s_* - s_i \tag{3.4}$$

hold $\forall\ i = 0, 1, 2, \ldots$

Proof The assertions

$$\|x_{j+1} - x_j\| \leq s_{j+1} - s_j \tag{3.5}$$

and

$$B[x_{j+1}, s_* - s_{j+1}] \subset B[x_j, s_* - s_j] \tag{3.6}$$

are proven by induction $\forall\ j = 0, 1, 2, \ldots$ Using (A1)

$$\|x_1 - x_0\| + \|\mathcal{L}'(x_0)^{-1}\mathcal{L}(x_0)\| \leq \lambda = s_1 - s_0.$$

Let $u \in B[x_1, s_* - s_1]$. It follows by condition (A1)

$$\|u - x_0\| \leq \|u - x_1\| + \|x_1 - x_0\| \leq s_* - s_1 + s_1 - s_0 = s_*,$$

so $u \in B[x_1, s_* - s_1]$. That is assertions (3.5) and (3.6) hold if $j = 0$. Assume these assertions hold if $j = 0, 1, 2, \ldots, n$. It follows for each $\xi \in [0, 1)$

$$\|x_j + \xi(x_{j+1} - x_j) - x_0\| \leq s_j + \xi(s_{j+1} - s_j) \leq t_*,$$

and

$$\|x_{j+1} - x_0\| \leq \sum_{i=1}^{j+1}\|x_i - x_{i-1}\| \leq \sum_{i=1}^{j+1}(s_i - s_{i-1}) = s_{j+1}.$$

It follows by induction hypotheses, the Lemmas and conditions (3.1) and (1.1)

$$\|\mathcal{L}'(x_0)^{-1}(\mathcal{L}'(x_{j+1}-\mathcal{L}'(x_0))\| \le \bar{K}\|x_{j+1}-x_0\| \le \bar{K}(s_{j+1}-s_0) \le \bar{K}s_{j+1} < 1,$$

where $\tilde{K} = \begin{cases} K_0, & j=0 \\ M_0, & j=1,2,\ldots \end{cases}$ Hence, the inverse of linear operator $\mathcal{L}'(x_{j+1})$ exists. Notice that if $j=0$, K_0 can be used, whereas if $j=1,2,\ldots$, then M_0 is utilized.

$$\|\mathcal{L}'(x_{j+1})^{-1}\mathcal{L}'(x_0)\| \le \frac{1}{1-\tilde{K}s_{j+1}}, \tag{3.7}$$

as a consequence of a lemma on linear operators that are invertible due to Banach's perturbation lemma (Argyros and Magréñan 2018; Argyros 2004a, b; Argyros and Hilout 2010).

The identity can be given by NP

$$\mathcal{L}(x_{n+1}) = \int_0^1 (\mathcal{L}'(x_j+\xi(x_{j+1}-x_n)) - \mathcal{L}'(x_j))(x_{j+1}-x_j)\mathrm{d}\xi, \tag{3.8}$$

since

$$\mathcal{L}(x_{j+1}) = \mathcal{L}(x_{j+1}) - \mathcal{L}(x_j) - \mathcal{L}'(x_j)(x_{j+1}-x_j).$$

Then, using induction hypotheses, identity (3.8) and condition (1.5)

$$\|\mathcal{L}'(x_0)^{-1}\mathcal{L}(x_{j+1})\| \le \tilde{M}\int_0^1 \xi\|x_{j+1}-x_j\|^2\mathrm{d}\xi \le \frac{\tilde{M}}{2}(s_{j+1}-s_j)^2, \tag{3.9}$$

where $\tilde{M} = \begin{cases} K, & j=0 \\ M, & j=1,2,\ldots \end{cases}$

It follows by NP, estimates (3.7), (3.9) and the definition (2.1) of sequence $\{s_n\}$

$$\begin{aligned}\|x_{j+2}-x_{j+1}\| &= \|\mathcal{L}'(x_{j+1})^{-1}\mathcal{L}(x_{j+1})\| \\ &= \|\mathcal{L}'(x_{j+1})^{-1}\mathcal{L}'(x_0)\mathcal{L}'(x_0)^{-1}\mathcal{L}(x_{j+1})\| \\ &\le \|\mathcal{L}'(x_{j+1})^{-1}\mathcal{L}'(x_0)\|\|\mathcal{L}'(x_0)^{-1}\mathcal{L}(x_{j+1})\| \\ &\le \frac{\bar{K}}{2}\frac{(s_{j+1}-s_j)^2}{1-\bar{M}s_{j+1}} = s_{j+2}-s_{j+1},\end{aligned}$$

where $\bar{K} = \begin{cases} K, & j=0 \\ M, & j=1,2,\ldots \end{cases}$ and $\bar{M} = \begin{cases} K_0, & j=0 \\ M_0, & j=1,2,\ldots \end{cases}$

Moreover, if $v \in B[x_{j+2}, s_* - s_{j+2}]$, it follows

$$\begin{aligned}\|v-x_{j+1}\| &\le \|v-x_{j+2}\| + \|x_{j+2}-x_{j+1}\| \\ &\le s_* - s_{j+2} + s_{j+2} - s_{j+1} = s_* - s_{j+1}.\end{aligned}$$

Thus, the element $v \in B[x_{j+1}, s_* - s_{j+1}]$ completing the induction for assertions (3.5) and (3.6). Notice that scalar majorizing sequence $\{s_j\}$ is fundamental as convergent. Hence, the sequence $\{x_j\}$ is also convergent to some $x_* \in B[x_0, s_*]$. Furthermore, let $j \longrightarrow \infty$ in estimate (3.9) to conclude $\mathcal{L}(x_*) = 0$. Finally, the proof of assertion (3.4) using estimate (3.5) as standard is omitted (Yamamoto 1987b). □

Next, the uniqueness ball for the solution x_* is presented. Notice that not all conditions S are used.

Proposition 3.2 *Under center-Lipschitz condition (1.4) further assume the existence of a solution $p \in B(x_0, R) \subset \Omega$ of equation $\mathcal{L}(x) = 0$ such that linear operator $\mathcal{L}'(p)$ is invertible for some $R > 0$; a parameter $R_1 > R$ given by*

$$R_1 = \frac{2}{M_0} - R. \tag{3.10}$$

Then, the element p solves uniquely equation $\mathcal{L}(x) = 0$ in the set $T = B(x_0, R_1) \cap \Omega$.

Proof Define linear operator $Q = \int_0^1 \mathcal{L}'(\bar{p}+\xi(p-\bar{p}))d\xi$ for some element $\bar{p} \in T$ satisfying $F(\bar{p}) = 0$. By using the definition of parameter R_1, set T and condition (1.4)

$$\begin{aligned} \|\mathcal{L}'(x_0)^{-1}(\mathcal{L}'(x_0) - Q)\| &\leq \| \int_0^1 \mathcal{L}'(x_0)^{-1}(\mathcal{L}'(p + \xi(p - \bar{p})) - \mathcal{L}'(x_0))d\xi \| \\ &\leq M_0 \int_0^1 ((1-\xi)\|p - x_0\| + \xi\|\bar{p} - x_0\|)d\xi \\ &< \frac{M_0}{2}(R + R_1) = 1. \end{aligned} \tag{3.11}$$

Then, the estimate (4.5) and the Banach lemma on linear operators (Argyros and Magréñan 2018; Argyros 2004a, b; Argyros and Hilout 2010) with inverses, imply the invertablility of linear operator Q. Moreover, by the approximation $0 = \mathcal{L}(p) - \mathcal{L}(\bar{p}) = Q(p - \bar{p})$, we deduce $\bar{p} = p$. □

Remark 3.3 Notice that not all conditions of Theorem 3.1 are used in Proposition 3.2. But if they were, then, we can set $p = x^*$ and $R = s^*$.

4 Special cases and Examples

It turns out that conditions of Theorem 3.1 reduce to the ones given by the earlier studies. But first we have the observations

Remark 4.1 Let us compare conditions H to conditions A:

It follows by these conditions that $K_0 \leq K \leq M_0$. Consequently, replacing M_0 or M_1 by these tighter parameters gives previously mentioned benefits. Moreover, notice that parameters K_0, K, M_0, M are specializations of the originally used M_1. Hence, no additional cost is required in their computation.

(1) The condition (A1) is common.
(2) The condition (A2) always implies the conditions (3.1) and (3.2). However, the converse implication does not hold necessarily, unless $K_0 = K = M_1$.
(3) The new majorizing sequence $\{s_n\}$ is tighter than the sequence $\{t_n\}$ used by Kantorovich. In particular, under the conditions of the Theorem 1.1 and the Theorem 3.1, a simple inductive argument gives

$$\begin{aligned} 0 \leq s_n \leq t_n, \\ 0 \leq s_{n+1} - s_n \leq t_{n+1} - t_n \end{aligned}$$

and

$$s_* \leq \rho.$$

Notice also:

Table 1 Sequence (2.1)

n	1	2	3	4	5	6	7
s_{n+1}	0.2000	0.2594	0.2690	0.2693	0.2693	0.2693	0.2693

(4) The conditions of Lemma 2.2 are stronger than those of the Lemma 2.1.
(5) The conditions of the Kantorovich Theorem 1.1 imply

$$t_n < \frac{1}{M_1}.$$

This inequality implies the one by our Lemma 2.1 but not vice versa unless if $M_0 = M_1$.
(6) Next, a comparison between Lemma 2.2 and the corresponding one in Theorem 1.1 follows.

Case $K_0 = M_0 = K = M$.
(i) It follows by the definition of N that $N = M$ and condition (2.2) reduces to

$$M\lambda \leq \frac{1}{2}. \tag{4.1}$$

Furthermore, if $M = M_1$ it reduces to the Kantorovich condition (A3) in the conditions A. But by estimate (1.8) it follows that if $M < M_1$ then condition (A3) implies (4.1) but not vice versa. Hence, the new convergence criterion (4.1) weakens the Kantorovich criterion (A3) (see also the examples where $M < M_1$). (ii) The majorizing sequence reads

$$u_0 = o, u_1 = \lambda,\ u_{n+2} = u_{n+1} + \frac{M(u_{n+1} - u_n)^2}{2(1 - M_0 u_{n+1})}.$$

This sequence is more precise than $\{t_n\}$ but not necessarily $\{s_n\}$, unless if $K = M$ and $K_0 = M_0$.
(iii) The uniqueness ball is extended, since M_0 is used for M_1 in the formula. (see also Proposition 3.2).
Other specializations of the Lipschitz conditions give similar benefits (Table 1).

Example 4.2 The parameters using the example of the introduction are $K_0 = \frac{\mu+5}{3}$, $K = M_0 = \frac{\mu+11}{6}$. Moreover,$\Omega_0 = B(1, 1-\mu) \cap B(1, \frac{1}{M_0}) = B(1, \frac{1}{M_0})$. Set $M = 2(1 + \frac{1}{3-\mu})$ $M_0 < M_1$ and $M < M_1$ for all $\mu \in (0, 0.5)$. Criterion (2.2) is then satisfied if $\mu \in S_0 :=$ $[0.42, 0.5)$. Hence, the range of values for μ for which NP converges is extended. Interval S_0 can be enlarged if the condition of Lemma 2.1 is verified. Then, for $\mu = 0.4$, we have the following $\frac{1}{M_0} = 0.5263$,

Hence, the conditions of Lemma 2.1 hold. For $\epsilon = 0.63$, we have

$$M_0\left(\frac{\gamma_0(s_2 - s_1)}{1-\gamma} + \lambda + s_2 - s_1\right) = 0.1143 \leq \beta = 0.3865$$

and

$$\lambda = 0.2000 < \frac{2}{(1+\epsilon)M} = 0.4431.$$

Thus, the conditions (2.11) and (2.12) hold, and the interval S_0 is further enlarged.

Example 4.3 Let $\mathcal{U} = \mathcal{V} = C[0, 1]$ be the space of continuous real functions on the interval $[0, 1]$. The max-norm is used. Set $\Omega = B[x_0, 3]$. Define Hammerstein nonlinear integral operator $\mathcal{L}$ on Ω as

$$\mathcal{L}(v)(w) = v(w) - y(w) - \int_0^1 G(w, t)v^3(t)\mathrm{d}t, \;\; v \in C[0, 1], w \in [0, 1]. \tag{4.2}$$

where function $y \in C[0, 1]$, and G is a kernel related by Green's function

$$G(w, t) = \begin{cases} (1 - w)t, \;\; t \leq w \\ w(1 - t), \;\; w \leq t. \end{cases} \tag{4.3}$$

It follows by this definition that $\mathcal{L}'$ is defined by

$$[\mathcal{L}'(v)(z)](w) = z(w) - 3\int_0^1 G(w, t)v^2(t)z(t)\mathrm{d}t \tag{4.4}$$

$z \in C[0, 1], w \in [0, 1]$. Pick $x_0(w) = y(w) = 1$. It then follows from (4.2)–(4.4) that $\mathcal{L}'(x_0)^{-1} \in L(\mathcal{V}, \mathcal{U})$,

$$\|I - \mathcal{L}'(x_0)\| < 0.375, \;\; \|\mathcal{L}'(x_0)^{-1}\| \leq 1.6,$$
$$\lambda = 0.2, \;\; M_0 = 2.4, \;\; M_1 = 3.6,$$

and $\Omega_0 = B(x_0, 3) \cap B(x_0, 0.4167) = B(x_0, 0.4167)$, so $M = 1.5$. Notice that $M_0 < M_1$ and $M < M_1$. Set $K_0 = K = M_0$. The Kantorovich convergence criterion (A3) is not satisfied, since $2M_1\lambda = 1.44 > 1$. Therefore convergence of NP is not assured. But our condition is satisfied, since $2M\lambda = 0.6 < 1$.

Remark 4.4 (i) Under conditions H, set $p = x_*$ and $R = s_*$.

(ii) Lipschitz condition (1.5) can be replaced by

$$\|\mathcal{L}'(x_0)^{-1}(\mathcal{L}'(w_1) - \mathcal{L}'(w_2))\| \leq d\|w_1 - w_2\|, \tag{4.5}$$

$\forall \; w_1 \in \Omega_0$ and $w_2 = w_1 - \mathcal{L}'(w_1)^{-1}\mathcal{L}(w_1) \in \Omega_0$. This, even smaller parameter d can replace M in the aforementioned results. The existence of iterate w_2 is assured by (1.4).

(iii) Another way to reduce Lipschitz constant M is given as follows: Suppose Lipschitz condition (1.5) is replaced by

$$\|\mathcal{L}'(x_0)^{-1}(\mathcal{L}'(w_1) - \mathcal{L}'(w_2))\| \leq d_0\|w_1 - w_2\|, \tag{4.6}$$

$\forall w_1 \in T_1$ and $w_2 = w_1 - \mathcal{L}'(w_1)^{-1}\mathcal{L}(w_1) \in T_1$, where $T_1 = B(x_1, \frac{1}{M_0} - \lambda)$ provided that $\lambda\mu < 1$. Notice that $d_0 \leq d \leq M$, since $T_1 \subset \Omega_0 \subset \Omega_1$. In the case of example 4.1, the parameters are $d_0 = \frac{5(4-\mu)^3+\mu(3-\mu)^3}{3(3-\mu)(4-\mu)^2} < d = \frac{6+2(3-\mu)(1+2\mu)}{3(3-\mu)} < M \;\forall \mu \in (0, 0.5)$.

5 Conclusion

A new methodology extends the applicability of NP. The new results are finer than the earlier ones. Therefore, they can replace them. No additional conditions are used. The methodology is very general. Consequently, it can be applied to extend other procedures (Argyros 2004b; Chen and Yamamoto 1989; Deuflhard 2004; Yamamoto 1987b).

References

Argyros IK (2004a) A unifying local-semilocal convergence analysis and applications for two-point Newton-like methods in Banach space. J Math Anal Appl 228:374–397

Argyros IK (2004b) On the Newton–Kantorovich hypothesis for solving equations. J Comput Appl Math 169:315–332

Argyros IK (2021) Unified convergence criteria for iterative Banach space valued methods with applications. Mathematics 9(16):1942. https://doi.org/10.3390/math9161942

Argyros IK (2022) The theory and applications of iteration methods. Engineering series, 2nd edn. CRC Press, Taylor and Francis Group, Boca Raton

Argyros IK, Hilout S (2010) Extending the Newton–Kantorovich hypothesis for solving equations. J Comput Appl Math 234:2993–3006

Argyros IK, Magréñan AA (2018) A contemporary study of iterative schemes. Elsevier (Academic Press), New York

Chen X, Yamamoto T (1989) Convergence domain of certain iterative methods for solving nonlinear equations. Numer Funct Anal Optim 10:37–48

Dennis JE Jr (1968) On Newton-like methods. Numer Math 11:324–330

Deuflhard P (2004) Newton methods for nonlinear problems. Affine invariance and adaptive algorithms. Springer Series in Computational Mathematics, vol 35. Springer, Berlin

Ezquerro JA, Hernandez MA (2018) Newton's scheme: an updated approach of Kantorovich's theory. Cham, Geneva

Ezquerro JA, Gutiérrez JM, Hernandez MA, Romero N, Rubio MJ (2010) The Newton method: from Newton to Kantorovich (Spanish). Gac R Soc Mat Esp 13:53–76

Gragg WB, Tapia RA (1974) Optimal error bounds for the Newton–Kantorovich theorem. SIAM J Numer Anal 11:10–13

Hernandez MA (2001) A modification of the classical Kantorovich conditions for Newton's method. J Comput Appl Math 137:201–205

Kantorovich LV, Akilov GP (1982) Functional analysis. Pergamon Press, Oxford

Ortega LM, Rheinboldt WC (1970) Iterative solution of nonlinear equations in several variables. Academic Press, New York

Potra FA, Pták V (1980) Sharp error bounds for Newton's process. Numer Math 34:63–72

Potra FA, Pták V (1984) Nondiscrete induction and iterative processes. Research notes in mathematics, vol 103. Pitman (Advanced Publishing Program), Boston

Proinov PD (2010) New general convergence theory for iterative processes and its applications to Newton–Kantorovich type theorems. J Complex 26:3–42

Rheinboldt WC (1968) A unified convergence theory for a class of iterative processes. SIAM J Numer Anal 5:42–63

Tapia RA (1971) Classroom notes: the Kantorovich theorem for Newton's method. Am Math Mon 78:389–392

Verma R (2019) New trends in fractional programming. Nova Science Publisher, New York

Yamamoto T (1987a) A method for finding sharp error bounds for Newton's method under the Kantorovich assumptions. Numer Math 49:203–320

Yamamoto T (1987b) A convergence theorem for Newton-like methods in Banach spaces. Numer Math 51:545–557

Yamamoto T (2000) Historical developments in convergence analysis for Newton's and Newton-like methods. J Comput Appl Math 124:1–23

Zabrejko PP, Bnuen DF (1987) The majorant method in the theory of Newton–Kantorovich approximations and the Pták error estimates. Numer Funct Anal Optim 9:671–684