Newton's method for solving generalized equations: Kantorovich's and Smale's approaches

Samir Adly, Huynh van Ngai, van Vu Nguyen

Newton's method for solving generalized equations: Kantorovich's and Smale's approaches

Samir Adly [a,*], Huynh Van Ngai [b], Van Vu Nguyen [a]

[a] *Laboratoire Xlim, Université de Limoges, 123 Avenue Albert Thomas, 87060 Limoges, France*
[b] *Department of Mathematics, University of Quy Nhon, 170 An Duong Vuong, Quy Nhon, Viet Nam*

In this paper, we study Newton-type methods for solving generalized equations involving set-valued maps in Banach spaces. Kantorovich-type theorems (both local and global versions) are proved as well as the quadratic convergence of the Newton sequence. We also extend Smale's classical (α, γ)-theory to generalized equations. These results are new and can be considered as an extension of many known ones in the literature for classical nonlinear equations. Our approach is based on tools from variational analysis. The metric regularity concept plays an important role in our analysis.

1. Introduction

It is well-known in the literature of applied mathematics, engineering and sciences that the classical Newton method and its generalizations are among the most famous and effective methods for numerically solving the nonlinear equation $f(x) = 0$, for a given function $f : \mathbb{R}^m \longrightarrow \mathbb{R}^m$. This success is due to its quadratic rate of convergence under some suitable assumptions on the problem data and the choice of the initial point. The classical convergence results state that Newton's method is only locally convergent. More precisely, if the function f is sufficiently smooth and its Jacobian $\nabla f(x^*)$ is nonsingular at the solution x^*, then by choosing an initial point x_0 in a neighborhood of this solution x^*, the convergence of the sequence generated by Newton's method is guaranteed and the rate of convergence is at least quadratic. Much more has been written about Newton's (or Simpson–Raphson–Newton's) method and a classical reference is the book by Ortega and Rheinboldt [15].

In 1948, L.V. Kantorovich published a famous paper (see [14,11]) about the extension of Newton's method to functional spaces. This result obtained by L.V. Kantorovich can be regarded both as an existence result (of a zero of f) and a convergence result (of the associated iterative process). The assumptions used essentially focus on the values of the function f and its first derivative f' at the starting point x_0 as well as the behavior of the derivative f' in a neighborhood of x_0. Kantorovich's theorem requires the knowledge of a local Lipschitz constant for the derivative. In the same spirit, another fundamental result on Newton's method is the well-known point estimation theory of Smale [19], based on the α-theory and γ-theory for analytical functions. The α-theory uses information about all derivatives of the function f at the initial point x_0 in order to give the size of the attraction's basin around the zero of the function f.

In 1980, S.M. Robinson [17] studied generalized equations with parameters: namely parameterized variational inequalities. He proved an implicit function theorem that can be used to study the local convergence of Newton's method for generalized equations (or variational inclusions). In his Ph.D. thesis, Josephy [10] used this theorem to investigate the local convergence of the Newton and quasi-Newton methods for the variational inequality of the form $f(x) + N_C(x) \ni 0$, where C is a closed convex subset of $\mathbb{R}^m$ and N_C is the forward-normal cone of C. He proved that if the solution x^* is regular (in the sense of Robinson [17]) and if x_0 is in a neighborhood of x^*, then the Josephy–Newton method is well-defined and converges superlinearly to x^*. Further generalization of Newton's method was considered by many authors. For example J.F. Bonnans [3] obtained a local convergence result of Newton's method for variational inequalities under weaker assumptions than the one's required by Robinson's theorem. More precisely, Bonnans proved that under the condition of semi-stability and hemistability (these two conditions are satisfied if Robinson's strong regularity holds at the solution), superlinear convergence of Newton's method (quadratic convergence if f is $C^{1,1}$) holds.

The Josephy–Newton method for set-valued inclusions of the form $f(x)+F(x) \ni 0$, where $f : \mathbb{R}^m \longrightarrow \mathbb{R}^m$ is a single-valued mapping of class C^1 and $F : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ is a set-valued map, was also considered by A. Dontchev [5–7]. In this case, the algorithm starts from some point x_0 near a solution and generates a sequence (x_n) defined by solving the following auxiliary problem

$$0 \in f(x_n) + Df(x_n)(x_{n+1} - x_n) + F(x_{n+1}). \tag{1.1}$$

Here (and in what follows), Df indicates the first order derivative of f. In [6,7], Dontchev proved the local Q-quadratic convergence of Josephy–Newton method under the assumptions that f is of class C^1 and F has closed graph with $(f + F)^{-1}$ being Aubin continuous at $(x^*, 0)$ (see [8] for more details).

The present paper is concerned with Newton–Kantorovich and Newton–Smale approaches for generalized equations of the form

$$0 \in f(x) + F(x), \tag{1.2}$$

where $f : X \longrightarrow Y$ is a single-valued mapping defined between the Banach spaces X and Y, supposed to be of class C^2 on an open set U in X, and $F : X \rightrightarrows Y$ is a set-valued mapping with a closed graph.

The paper is organized as follows. In section 2, we recall some preliminary results and backgrounds that will be used in the rest of the paper. Section 3 is devoted to the proof of a Kantorovich-type theorem for the generalized equation (1.2). In section 4, we prove two theorems corresponding to Smale's α and γ-theory for problem (1.2) in the case where f is analytic.

2. Preliminaries

The capital letters $X, Y, \ldots$ will denote Banach spaces. By X^* we mean the dual space associated with X. Throughout the paper, we will use the common notation $\|\cdot\|$ for the norm on some arbitrary Banach space X and $\langle \cdot, \cdot \rangle$ for the canonical duality pairing on $X^* \times X$. We denote by $\mathbb{B}_X(x, r)$ (resp. $\mathbb{B}_X[x, r]$) the open (resp. closed) ball with center $x \in X$ and radius $r > 0$. The unit ball in X will be defined by $\mathbb{B}_X$ (or simply

$\mathbb{B}$ when X is clear). If $x \in X$ and $K \subset X$, the notation $d(x, K)$ stands for the *distance function* from x to K, defined by

$$d(x, K) := \inf_{z \in K} \|x - z\|.$$

For two subsets C, D of X, we define the *excess* of C beyond D by

$$e(C, D) := \inf \{t > 0 : C \subset D + t\mathbb{B}_X\},$$

and the Hausdorff distance between them as

$$d^H(C, D) := \max \big\{e(C, D), e(D, C)\big\}.$$

A *set-valued mapping* (or multifunction) $T : X \rightrightarrows Y$ is a correspondence, assigning to each $x \in X$ a subset $T(x) \subset Y$. The *domain* and *graph* of mapping T are defined respectively by $\mathrm{dom}(T) = \big\{x \in X : T(x) \neq \emptyset\big\}$ and $\mathrm{gph}(T) = \big\{(x, y) \in X \times Y : y \in T(x)\big\}$. For the given Banach spaces X, Y, Z and two mappings $T_1 : X \rightrightarrows Y$, $T_2 : Y \rightrightarrows Z$ the *composition* of T_1 and T_2 is a mapping $T : X \rightrightarrows Z$ defined by $T(x) = \bigcup_{y \in T_1(x)} T_2(y)$. Finally, the *inverse* of the mapping T is the set-valued mapping $T^{-1} : Y \rightrightarrows X$ such that $x \in T^{-1}(y)$ if and only if $y \in T(x)$.

The key concept in our analysis is *the metric regularity property* [1,8,12,13,18]. Recall that a mapping $T : X \rightrightarrows Y$ is *metrically regular* at $(\bar{x}, \bar{y}) \in \mathrm{gph}(T)$ with a modulus $\kappa > 0$ if there are some neighborhoods U of $\bar{x}$ in X and V of $\bar{y}$ in Y so that

$$d\big(x, T^{-1}(y)\big) \leqslant \kappa d\big(y, T(x)\big), \quad \text{whenever} \quad (x, y) \in U \times V.$$

The infimum of all such constants κ is called the *regularity modulus* of T at $\bar{x}$ for $\bar{y}$, usually denoted by $\mathrm{reg}(T; (\bar{x}, \bar{y}))$.

The metric regularity of a mapping T at some point $(\bar{x}, \bar{y})$ in $\mathrm{gph}(T)$ is known to be equivalent to the so-called Lipschitz-like property (or Aubin property) of the inverse T^{-1} at $(\bar{y}, \bar{x})$ (see [8,12]). Recall that a mapping $S : Y \rightrightarrows X$ is Lipschitz-like at $(\bar{y}, \bar{x}) \in \mathrm{gph}(S)$ with constant κ provided that there exists a neighborhood $V \times U$ of $(\bar{y}, \bar{x})$ in $Y \times X$ such that

$$e\big(S(y) \cap U, S(y')\big) \leqslant \kappa \|y - y'\|, \quad \text{for all} \quad y, y' \in V.$$

In the sequel, we make use of the following theorem proved in [1], which gives a sufficient condition for the stability of metric regularity property under a suitable perturbation. Recall that a mapping $S : X \rightrightarrows Y$ is called Lipschitz continuous on a set D of X iff there exists a constant $L > 0$ (Lipschitz constant) such that

$$d^H\big(S(x), S(x')\big) \leqslant L \|x - x'\|, \quad \text{for all} \quad x, x' \in D.$$

Theorem 2.1. *Given Banach spaces X and Y, let $\Phi : X \rightrightarrows Y$ be a set-valued mapping with closed graph and let $(\bar{x}, \bar{y}) \in \mathrm{gph}(\Phi)$. Suppose that Φ is metrically regular at $(\bar{x}, \bar{y})$ with modulus $\kappa > 0$ on a neighborhood $\mathbb{B}_X(\bar{x}, a) \times \mathbb{B}_Y(\bar{y}, b)$ of $(\bar{x}, \bar{y})$ for some $a > 0$ and $b > 0$. Let $\delta > 0$, $L \in (0, \kappa^{-1})$, and set $\tau = \kappa/(1 - \kappa L)$. Let α, β be positive constants satisfying*

$$2\alpha + \beta\tau < \min\{a, \delta/2\}, \beta(\tau + \kappa) < \delta, 2c\alpha + \beta(1 + c\tau) < b,$$

with $c := \max\big\{1, \kappa^{-1}\big\}$. If $G : X \longrightarrow Y$ is Lipschitz continuous on $\mathbb{B}_X(\bar{x}, \delta)$ with the constant L, and the sum $\Phi + G$ has closed graph, then $\Phi + G$ is metrically regular on $\mathbb{B}_X(\bar{x}, \alpha) \times \mathbb{B}_Y(\bar{y} + G(\bar{x}), \beta)$ with modulus τ.

Remark 2.2. The proof of Theorem 2.1 is a direct consequence of Theorem 3.2 in [1].

Definition 2.3. Given a set-valued mapping $\Phi : X \rightrightarrows Y$, a point $x_0 \in X$ and some positive constants $r > 0$, $s > 0$, we define

$$V(\Phi, x_0, r, s) = \{(x, y) \in X \times Y : x \in \mathbb{B}_X[x_0, r], d(y, \Phi(x)) < s\}. \tag{2.1}$$

One says that the mapping Φ is metrically regular on the set $V(\Phi, x_0, r, s)$ with a modulus $\tau > 0$ if

$$d\big(x, \Phi^{-1}(y)\big) \leqslant \tau d\big(y, \Phi(x)\big) \quad \text{for all} \quad (x, y) \in V(\Phi, x_0, r, s). \tag{2.2}$$

When the mapping Φ satisfies Definition 2.3, we denote by $\operatorname{reg}(\Phi, x_0, r, s)$ the infimum of all $\tau > 0$ for which (2.2) holds. Otherwise, we set $\operatorname{reg}(\Phi, x_0, r, s) = \infty$.

A global version of Theorem 2.1 is stated in the following theorem.

Theorem 2.4. *Let $\Phi : X \rightrightarrows Y$ be a set-valued mapping with closed graph, and let $x_0 \in X$, $r > 0$, $s > 0$, and $\kappa > 0$ be such that Φ is metrically regular on $V(\Phi, x_0, r, s)$ with modulus κ. For $L \in (0, \kappa^{-1})$ set $\tau = \kappa/(1 - \kappa L)$. If $G : X \longrightarrow Y$ is Lipschitz continuous on $\mathbb{B}_X[x_0, r]$ with constant L, and $\Phi + G$ has closed graph, then the set-valued mapping $\Phi + G$ is metrically regular on $V(\Phi + G, x_0, \frac{r}{4}, R)$ with the modulus τ, where $R = \min\left\{s, \frac{r}{5\tau}\right\}$.*

Remark 2.5. The proof of Theorem 2.4 is a direct consequence of Theorem 6.2 in [1].

3. Theorems of Kantorovich's type

Let us first establish a local convergence theorem for Newton's iteration (1.1).

Theorem 3.1. *Let $f : X \longrightarrow Y$ be a function of class C^2 on an open set U in X and let $F : X \rightrightarrows Y$ be a set-valued mapping with closed graph. Let $\xi \in U$ be a solution of problem (1.2) and $\eta = Df(\xi)(\xi) - f(\xi) \in Y$. Suppose that the set-valued mapping $\Phi(\cdot) = Df(\xi)(\cdot) + F(\cdot)$ is metrically regular on neighborhood $V = \mathbb{B}_X[\xi, r] \times \mathbb{B}_Y[\eta, \rho]$ of (ξ, η) with modulus τ such that $\mathbb{B}_X[\xi, r] \subset U$. Define*

$$K(\tau, \xi, r) := \tau \sup_{\|z - \xi\| \leqslant r} \left\|D^2 f(z)\right\|, \ \textit{and} \quad \varepsilon = \min\left\{r, \rho, \tau\rho\right\},$$

where $D^2 f$ stands for the second order derivative of f. If $2K(\tau, \xi, r)r < 1$, then for all $x_0 \in \mathbb{B}_X(\xi, \varepsilon)$, there exists a Newton iterative sequence (x_n) generated by (1.1), converging quadratically to ξ:

$$\|x_{n+1} - \xi\| < \frac{1}{2r} \|x_n - \xi\|^2, \quad n = 0, 1, 2, \dots \tag{3.1}$$

In order to prove Theorem 3.1, the following lemma will be useful.

Lemma 3.2. *With f, r, τ and $K(\tau, \xi, r)$ as in the theorem, for any $x, x' \in \mathbb{B}_X[\xi, r]$, one has:*

(i) $\|Df(x) - Df(x')\| \leqslant \frac{K(\tau,\xi,r)}{\tau} \|x - x'\|$,
(ii) $\|f(x') - f(x) - Df(x)(x' - x)\| \leqslant \frac{K(\tau,\xi,r)}{2\tau} \|x' - x\|^2$.

Proof. The proof is evident from the mean value theorem and definition of $K(\tau, \xi, r)$. $\square$

Proof of Theorem 3.1. Let $L = \frac{K(\tau,\xi,r)r}{\tau} \leqslant \frac{1}{2\tau}$, and set $\bar{\tau} = \frac{\tau}{1-\tau L}$, $\nu = \frac{\varepsilon}{4\tau}$. Then,

$$\nu\bar{\tau} = \nu\frac{\tau}{1-\tau L} \leqslant 2\nu\tau = \frac{\varepsilon}{2} < r/2.$$

Moreover, since $\tau L \leqslant \frac{1}{2}$, one has

$$\nu(\bar{\tau}+\tau) = \nu\tau\left(\frac{1}{1-\tau L}+1\right) \leqslant 3\tau\nu = \frac{3\varepsilon}{4} < r.$$

Set $c = \max\left\{1, \tau^{-1}\right\}$. If $\tau \geqslant 1$, then $c = 1$, and in this case, we have

$$\nu(1+c\bar{\tau}) \leqslant \nu(\tau+\bar{\tau}) \leqslant 3\tau\nu < \rho.$$

Otherwise, $\tau < 1$, then $c = \tau^{-1}$. In this case, it holds that

$$\nu(1+c\bar{\tau}) = \nu(1+\tau^{-1}\bar{\tau}) = \nu\left(1+\frac{1}{1-\tau L}\right) \leqslant 3\nu < \rho.$$

Let us take $\mu > 0$ such that

$$2\mu + \nu\bar{\tau} < r/2, \nu(\bar{\tau}+\tau) < r, 2c\mu + \nu(1+c\bar{\tau}) < \rho.$$

Applying Theorem 2.1 with $a = r$, $b = \rho$ and $\delta = r$, for any linear and continuous map $A : X \longrightarrow Y$ having $\|A\| \leqslant L$ the sum $\Phi_A = A + \Phi$ is metrically regular on $\mathbb{B}_X[\xi,\mu] \times \mathbb{B}_Y[\eta + A(\xi),\nu]$ with a modulus $\bar{\tau}$.

Let $x_0 \in \mathbb{B}_X(\xi,\varepsilon)$, then by Lemma 3.2, we get

$$\|Df(x_0) - Df(\xi)\| \leqslant \frac{K(\tau,\xi,r)}{\tau}\|x_0 - \xi\| \leqslant \frac{K(\tau,\xi,r)\varepsilon}{\tau} \leqslant \frac{K(\tau,\xi,r)r}{\tau} = L.$$

Hence, $\Phi_0 = Df(x_0) + F = [Df(x_0) - Df(\xi)] + \Phi$ is metrically regular on the neighborhood $\mathbb{B}_X[\xi,\mu] \times \mathbb{B}_Y[y_0,\nu]$ of (ξ, y_0), with modulus $\tau_0 = \frac{\tau}{1-\tau L}$, where $y_0 = Df(x_0)(\xi) - f(\xi)$. For $z_0 = Df(x_0)(x_0) - f(x_0)$, we have

$$\begin{aligned}
\|z_0 - y_0\| &= \|Df(x_0)(x_0) - f(x_0) - [Df(x_0)(\xi) - f(\xi)]\| \\
&= \|f(\xi) - f(x_0) - Df(x_0)(\xi - x_0)\| \leqslant \frac{K(\tau,\xi,r)}{2\tau}\|\xi - x_0\|^2 \\
&\leqslant \frac{K(\tau,\xi,r)}{2\tau}r\varepsilon \leqslant \frac{\varepsilon}{4\tau} = \nu,
\end{aligned}$$

which implies $z_0 \in \mathbb{B}_Y[y_0,\nu]$. Using the fact that Φ_0 is metrically regular on $\mathbb{B}_X[\xi,\mu] \times \mathbb{B}_Y[y_0,\nu]$, we obtain

$$\begin{aligned}
d\big(\xi, \Phi_0^{-1}(z_0)\big) &\leqslant \tau_0 d\big(z_0, \Phi_0(\xi)\big) \leqslant \tau_0 d\big(z_0, y_0\big) \leqslant \tau_0\frac{K(\tau,\xi,r)}{2\tau}\|\xi - x_0\|^2 \\
&= \frac{K(\tau,\xi,r)}{2(1-\tau L)}\|\xi - x_0\|^2.
\end{aligned}$$

Since $L \leqslant \frac{1}{2\tau}$ and $rK(\tau,\xi,r) < 1/2$, we deduce

$$d\big(\xi, \Phi_0^{-1}(z_0)\big) < \frac{1}{2r}\|\xi - x_0\|^2.$$

Thus, we can select $x_1 \in \Phi_0^{-1}(z_0)$ verifying

$$\|\xi - x_1\| < \frac{1}{2r} \|\xi - x_0\|^2 .$$

Observe that, as

$$Df(x_0)(x_0) - f(x_0) = z_0 \in \Phi_0(x_1) = Df(x_0)(x_1) + F(x_1),$$

x_1 is a Newton iteration generated by (1.1) with $n = 0$. On the other hand, the choice of x_0 implies

$$\|\xi - x_1\| < \frac{1}{2r} \|\xi - x_0\|^2 \leqslant \frac{1}{2} \|\xi - x_0\| < \varepsilon.$$

That is, $x_1 \in \mathbb{B}_X(\xi, \varepsilon)$. Consequently, instead of x_0, now we can consider x_1 as a new starting point and continue the process. Repeating this procedure, we obtain a sequence (x_n) generated by (1.1) for which relation (3.1) holds. By using repeatedly this inequality, we obtain

$$\|x_n - \xi\| < \left(\frac{1}{2r} \|x_0 - \xi\| \right)^{2^n - 1} \|x_0 - \xi\| \leqslant \left(\frac{1}{2} \right)^{2^n - 1} \|x_0 - \xi\| .$$

Combining the preceding estimation with (3.1), the quadratic convergence is shown. The proof of Theorem 3.1 is thereby completed. □

Remark 3.3. A similar result to Theorem 3.1 was obtained by A. Dontchev in [6, Theorem 1]. Our assumptions are slightly different than the ones used in [6, Theorem 1] and the conclusion of Theorem 3.1 is more precise in the sense that it gives an explicit region for starting points to ensure the convergence of the algorithm.

The next theorem states a global result ensuring the existence of a solution as well as the rate of convergence for the Newton iterative sequence. The assumptions used here are based on a classical Kantorovich theorem presented in [4]. An optimal error bound estimate is obtained by using the technique in [9].

Theorem 3.4. *Given two Banach spaces X and Y, let $f : X \longrightarrow Y$ be a function of class C^2 on an open set $U \subset X$ and $F : X \rightrightarrows Y$ be a set-valued mapping with closed graph. For $\tau > 0$, $\varepsilon > 0$ and $y \in U$ with $\mathbb{B}[y, \varepsilon] \subset U$ we define*

$$\beta(\tau, y) := \tau d\big(0, f(y) + F(y)\big), \quad K(\tau, y, \varepsilon) := \tau \sup_{\|z - y\| \leqslant \varepsilon} \left\| D^2 f(z) \right\| ,$$

where $D^2 f$ stands for the second-order derivative of f. Let $x \in U$, $\alpha \in (0, 1]$ and $r > 0$, $s > 0$ such that the following conditions are satisfied.

1. *$\Phi = Df(x) + F$ is metrically regular on $V = V(\Phi, x, 4r, s)$ with a modulus $\tau > \mathrm{reg}(\Phi, x, 4r, s)$,*
2. *$d\big(0, G(x)\big) < s$, where $G = f + F$,*
3. *$2\beta(\tau, x)K(\tau, x, r) \leqslant \alpha$,*
4. *$2\eta\beta(\tau, x) \leqslant r$, with $\eta = \frac{1 - \sqrt{1-\alpha}}{\alpha} = \frac{1}{1 + \sqrt{1-\alpha}}$.*

Then there exists a solution $\xi \in U$ of the generalized equation (1.2) such that

$$\|x - \xi\| \leqslant 2\eta\beta(\tau, x) \leqslant r. \tag{3.2}$$

Moreover, there is a sequence (x_n) generated by the Newton method (1.1) which starts at x and converges to ξ. For this sequence, the following error bounds hold:

- *if $\alpha < 1$, then*

$$\|x_n - \xi\| \leqslant \frac{4\sqrt{1-\alpha}}{\alpha} \frac{\theta^{2^n}}{1-\theta^{2^n}} \beta(\tau, x), \text{ with } \theta = \frac{1-\sqrt{1-\alpha}}{1+\sqrt{1-\alpha}}; \tag{3.3}$$

- *if $\alpha = 1$, then*

$$\|x_n - \xi\| \leqslant 2^{-n+1} \beta(\tau, x). \tag{3.4}$$

Let us first note that if $\beta(\tau, x) = 0$, then the conclusion of Theorem 3.4 is trivially valid. Indeed, in that case, $d\big(0, f(x) + F(x)\big) = \tau^{-1}\beta(\tau, x) = 0$, and hence $0 \in f(x) + F(x)$ since $\operatorname{gph}(F)$ is closed. This means that x is a solution of problem (1.2), and we simply set $x_n = x$ for all n. Assuming now $\beta(\tau, x) > 0$, set $\beta = \beta(\tau, x) > 0$, $K = (2\beta)^{-1}\alpha > 0$. Consider the majorizing function given by

$$\omega(t) = Kt^2 - 2t + 2\beta. \tag{3.5}$$

The proof will be performed by induction, and it is based on the following lemma.

Lemma 3.5. *Consider the Newton sequence (t_n) associated with the scalar equation $\omega(t) = 0$:*

$$\begin{cases} t_0 = 0, \\ t_{n+1} = t_n - \omega'(t_n)^{-1}\omega(t_n), n = 0, 1, \dots \end{cases} \tag{3.6}$$

Then this sequence is well-defined, strictly increasing and converges to the smallest root $t^* = \frac{1-\sqrt{1-\alpha}}{K}$ of ω. When $\alpha < 1$ we have

$$\begin{cases} t^* - t_n \leqslant \frac{4\sqrt{1-\alpha}}{\alpha} \frac{\theta^{2^n}}{1-\theta^{2^n}} (t_1 - t_0) = \frac{4\sqrt{1-\alpha}}{\alpha} \frac{\theta^{2^n}}{1-\theta^{2^n}} \beta, \\ \frac{2(t_{n+1}-t_n)}{1+\sqrt{1+4\theta^{2^n}(1+\theta^{2^n})^{-2}}} \leqslant t^* - t_n \leqslant \theta^{2^{n-1}} (t_n - t_{n-1}). \end{cases} \tag{3.7}$$

In the case $\alpha = 1$, one has

$$\begin{cases} t^* - t_n \leqslant 2^{-n+1}(t_1 - t_0) = 2^{-n+1}\beta, \\ 2\left(\sqrt{2} - 1\right)(t_{n+1} - t_n) \leqslant t^* - t_n \leqslant t_n - t_{n-1}. \end{cases} \tag{3.8}$$

Proof. It is sufficient to verify the monotonicity of the sequence defined in (3.6); the existence of (t_n) as well as the error bounds in (3.7), (3.8) follow directly from [9]. The scalar function defined by $\varphi : t \in (-\infty, t^*) \longmapsto t - \omega'(t)^{-1}\omega(t)$ satisfies

$$\varphi'(t) = \frac{\omega''(t)\omega(t)}{[\omega'(t)]^2} = 2K \frac{\omega(t)}{[\omega'(t)]^2}.$$

Let us observe that ω is positive on $I = (-\infty, t^*)$, so $\varphi'(t) > 0$. This means that φ is strictly increasing. Since $t_{k+1} = \varphi(t_k)$, the proof is completed by induction on n. $\square$

Remark 3.6. It follows immediately from the proof of Lemma 3.5 that

$$t_n < t^* = \frac{1-\sqrt{1-\alpha}}{K} = \frac{1-\sqrt{1-\alpha}}{\alpha}2\beta = 2\eta\beta \leqslant r \tag{3.9}$$

for all n.

Now, we return to the main proof of Theorem 3.4.

Proof of Theorem 3.4. We set $G := f+F$ and rewrite the Newton's iteration (1.1) in the following equivalent form:

$$-f(x_n) + Df(x_n)x_n \in \big[Df(x_n) + F\big](x_{n+1}). \tag{3.10}$$

We will construct by induction a sequence (x_n) such that

$$\|x_n - x_{n+1}\| < t_{n+1} - t_n \tag{3.11}$$

where (t_n) defined in Lemma 3.5.

Let us put $x_0 = x$, $\Phi_0 = \Phi$, and $\tau_0 = \tau$. Since $\tau_0 > \operatorname{reg}(\Phi_0, x_0, 4r, s)$, then Φ_0 is metrically regular on $V(\Phi_0, x_0, 4r, s)$ with some modulus $\bar{\tau}_0 < \tau_0$. By setting $y_0 = -f(x_0) + Df(x_0)(x_0)$, we have

$$\begin{aligned} d\big(y_0, \Phi_0(x_0)\big) &= d\big(-f(x_0) + Df(x_0)(x_0), [Df(x_0) + F](x_0)\big) \\ &= d\big(0, G(x_0)\big) < s. \end{aligned}$$

Hence $(x_0, y_0) \in V(\Phi_0, x_0, 4r, s)$. By using the metric regularity property of Φ_0, the following evaluation holds

$$\begin{aligned} d\big(x_0, \Phi_0^{-1}(y_0)\big) &\leqslant \bar{\tau}_0 d\big(y_0, \Phi_0(x_0)\big) = \bar{\tau}_0 d\big(0, G(x_0)\big) < \tau_0 d\big(0, G(x_0)\big) \\ &= \beta(\tau_0, x_0) = \beta. \end{aligned}$$

Thus, we can select x_1 in $\Phi_0^{-1}(y_0)$ with $\|x_0 - x_1\| < \beta$. Due to Lemma 3.5, it is easy to deduce that $\beta = t_1 - t_0$. Therefore, (3.11) is satisfied for $n = 0$.

Suppose that for some $n \geqslant 1$ there exist $x_1, x_2, \ldots, x_n$ such that

- $x_{k+1} \in \Phi_j^{-1}(y_k)$, where $\Phi_k(\cdot) = Df(x_k)(\cdot) + F(\cdot)$ and $y_k = [-f + Df(x_k)](x_k)$;
- $\|x_k - x_{k+1}\| < t_{k+1} - t_k = \delta_k$ for all $k = 0, 1, \ldots, n-1$.

If x_n is a solution of problem (1.2), then the proof is done. Otherwise, using (3.9) and the triangle inequality we have

$$\big\|x_0 - x_n\big\| \leqslant \sum_{j=0}^{n-1} \big\|x_j - x_{j+1}\big\| < \sum_{j=0}^{n-1}(t_{j+1} - t_j) = t_n - t_0 = t_n < r.$$

Let $K_0 = K(\tau_0, x_0, r) \leqslant \frac{\alpha}{2\beta} = K$; a similar argument as in Lemma 3.2 gives us

$$\|Df(x_n) - Df(x_0)\| \leqslant \frac{K_0}{\tau_0}\|x_0 - x_n\| \leqslant \frac{1}{\tau_0}Kt_n.$$

We introduce the notation $L_n := \frac{1}{\tau_0} K t_n$. The fact that $t_n < t^* = \frac{1-\sqrt{1-\alpha}}{K}$ in Remark 3.6 implies

$$L_n \bar{\tau}_0 \leqslant L_n \tau_0 = K t_n < K t^* = 1 - \sqrt{1-\alpha} \leqslant 1.$$

We define $\bar{\tau}_n = \frac{\bar{\tau}_0}{1-L_n\bar{\tau}_0}$, $\tau_n = \frac{\tau_0}{1-L_n\tau_0}$, $r_n = r/4$ and $s_n = \min\left\{s, \frac{4r}{5\tau_n}\right\}$. Due to Theorem 2.4, the mapping $\Phi_n(\cdot) = [Df(x_n) - Df(x_0)](\cdot) + \Phi_0(\cdot)$ is metrically regular on $V_n = V(\Phi_n, x_0, 4r_n, s_n)$ with modulus $\bar{\tau}_n < \tau_n$. Setting $y_n = [-f + Df(x_n)](x_n)$, we get $d\big(y_n, \Phi_n(x_n)\big) = d\big(0, G(x_n)\big)$. The choice of x_n gives us $y_{n-1} \in \Phi_{n-1}(x_n)$. Consequently,

$$z_{n-1} := f(x_n) - f(x_{n-1}) - Df(x_{n-1})(x_n - x_{n-1}) \in G(x_n),$$

and then, by using Taylor's expansion, we obtain the estimate

$$\begin{aligned} d\big(0, G(x_n)\big) &\leqslant \|z_{n-1}\| \\ &= \left\| \int_0^1 D^2 f(x_{n-1} + s(x_n - x_{n-1}))(x_n - x_{n-1})^2 (1-s)\, ds \right\| \\ &\leqslant \int_0^1 \left\| D^2 f(x_{n-1} + s(x_n - x_{n-1})) \right\| \|x_n - x_{n-1}\|^2 (1-s)\, ds \\ &\leqslant \frac{K_0}{\tau_0} \frac{\delta_{n-1}^2}{2} \leqslant \frac{K}{\tau_0} \frac{\delta_{n-1}^2}{2}. \end{aligned} \tag{3.12}$$

We are going to establish the following inequalities:

$$\frac{K}{\tau_0} \frac{\delta_{n-1}^2}{2} < s, \quad n = 1, 2, \ldots \tag{3.13}$$

Indeed, when $n = 1$ the left-hand side of (3.13) is equal to

$$\frac{K}{\tau_0} \frac{\delta_0^2}{2} = \frac{K}{\tau_0} \frac{\beta^2}{2} = \frac{K\beta}{\tau_0} \frac{\tau_0 d\big(0, G(x_0)\big)}{2} = \frac{1}{4} \alpha d\big(0, G(x_0)\big) \leqslant \frac{1}{4} d\big(0, G(x_0)\big),$$

which affirms (3.13). Consider the situation for which $n \geqslant 2$. If $\alpha = 1$, error bounds in (3.8) supply us with

$$\delta_{n-1} = t_n - t_{n-1} \leqslant \frac{1}{2\left(\sqrt{2}-1\right)} (t^* - t_{n-1}) \leqslant \frac{1}{2\left(\sqrt{2}-1\right)} 2^{-n+2} \beta \leqslant \frac{\sqrt{2}+1}{2} \beta.$$

When $\alpha < 1$, paying attention to (3.7), we find

$$\begin{aligned} \delta_{n-1} = t_n - t_{n-1} &\leqslant \frac{1}{2} \left(1 + \sqrt{1 + \frac{4\theta^{2^{n-1}}}{\left(1 + \theta^{2^{n-1}}\right)^2}} \right) (t^* - t_{n-1}) \\ &\leqslant \frac{1+\sqrt{2}}{2} (t^* - t_{n-1}) \leqslant \frac{1+\sqrt{2}}{2} \frac{4\sqrt{1-\alpha}}{\alpha} \frac{\theta^{2^{n-1}}}{1 - \theta^{2^{n-1}}} \beta \\ &\leqslant \frac{1+\sqrt{2}}{2} \frac{4\sqrt{1-\alpha}}{\alpha} \frac{\theta^2}{1-\theta^2} \beta = \frac{1+\sqrt{2}}{2} \frac{\alpha}{\left(1+\sqrt{1-\alpha}\right)^2} \beta. \end{aligned}$$

In summary, the inequality $\delta_{n-1} \leqslant \frac{1+\sqrt{2}}{2} \frac{\alpha}{(1+\sqrt{1-\alpha})^2} \beta$ is true with $n \geqslant 2$. And this yields

$$
\begin{aligned}
\frac{K}{\tau_0}\frac{\delta_{n-1}^2}{2} &\leqslant \frac{1}{2}\frac{K}{\tau_0}\left(\frac{1+\sqrt{2}}{2}\right)^2\beta^2 = \frac{1}{2\tau_0}\left(\frac{1+\sqrt{2}}{2}\right)^2\frac{\alpha}{2}\tau_0 d\big(0,G(x_0)\big)\\
&\leqslant \left(\frac{1+\sqrt{2}}{4}\right)^2\alpha d\big(0,G(x_0)\big) < s.
\end{aligned}
$$

Hence, (3.13) is proved.

Returning to the main proof, combining (3.12) and (3.13) one has

$$
d\big(0,G(x_n)\big) \leqslant \frac{K}{\tau_0}\frac{\delta_{n-1}^2}{2} < s.
$$

On the other hand, thanks to Taylor's expansion, we can write

$$
\omega(t_n) = \omega(t_{n-1}) + \omega'(t_{n-1})\delta_{n-1} + \omega''(t_{n-1})\frac{\delta_{n-1}^2}{2} = K\delta_{n-1}^2.
$$

As $\tau_n = \frac{\tau_0}{1-L_n\tau_0} = \frac{\tau_0}{1-Kt_n}$ and $2\eta\beta = 2\eta\beta(\tau,x) \leqslant r$, it follows from (3.12) that

$$
\begin{aligned}
d\big(0,G(x_n)\big) &\leqslant \frac{K}{\tau_0}\frac{\delta_{n-1}^2}{2} = \frac{1}{2}\frac{1}{\tau_n(1-Kt_n)}\omega(t_n) = \frac{1}{\tau_n}\big[-\omega'(t_n)^{-1}\omega(t_n)\big] = \frac{1}{\tau_n}\delta_n\\
&= \frac{\beta}{\tau_n}\left(\frac{\delta_n}{\beta}\right) \leqslant \frac{r}{2\eta\tau_n}\left(\frac{\delta_n}{\beta}\right) = \frac{4r}{5\tau_n}\frac{5\left(1+\sqrt{1-\alpha}\right)}{8}\left(\frac{\delta_n}{\beta}\right).
\end{aligned}
$$

We knew $\delta_n \leqslant \frac{1+\sqrt{2}}{2}\frac{\alpha}{(1+\sqrt{1-\alpha})^2}\beta$ in the preceding part of this proof. Therefore,

$$
\begin{aligned}
d\big(0,G(x_n)\big) &\leqslant \frac{4r}{5\tau_n}\frac{5\left(1+\sqrt{1-\alpha}\right)}{8}\left(\frac{\delta_n}{\beta}\right)\\
&\leqslant \frac{4r}{5\tau_n}\frac{5\left(1+\sqrt{1-\alpha}\right)}{8}\frac{1+\sqrt{2}}{2}\frac{\alpha}{\left(1+\sqrt{1-\alpha}\right)^2} < \frac{4r}{5\tau_n},
\end{aligned}
$$

which tells us $d\big(y_n,\Phi_n(x_n)\big) = d\big(0,G(x_n)\big) < s_n$. Because of $\|x_0 - x_n\| \leqslant r = 4r_n$, the conclusion $(x_n,y_n) \in V_n$ is clear. Thanks to the metric regularity property of Φ_n, we obtain

$$
d\big(x_n,\Phi_n^{-1}(y_n)\big) \leqslant \bar{\tau}_n d\big(y_n,\Phi_n(x_n)\big) = \bar{\tau}_n d\big(0,G(x_n)\big) < \tau_n d\big(0,G(x_n)\big) = \beta_n.
$$

Let us select $x_{n+1} \in \Phi_n^{-1}(y_n)$ such that $\|x_n - x_{n+1}\| < \beta_n$. Taking into account (3.12) one has

$$
\beta_n = \tau_n d\big(0,G(x_n)\big) \leqslant \frac{\tau_0}{1-Kt_n}\frac{K}{\tau_0}\frac{\delta_{n-1}^2}{2} = -\omega'(t_n)^{-1}\omega(t_n) = t_{n+1} - t_n.
$$

Consequently, we derive

$$
\big\|x_n - x_{n+1}\big\| < t_{n+1} - t_n.
$$

Hence, the construction of the sequence (x_n) by induction is now completed.

Since (t_n) is convergent, it is easy to see that (x_n) is a Cauchy sequence in the Banach space X. Therefore, the sequence (x_n) is also convergent. Let $\xi = \lim\limits_{n\to\infty} x_n$. We show that ξ is a solution of problem (1.2). In fact, for each n there is some $w_n \in F(x_{n+1})$ such that

$$
f(x_n) + Df(x_n)(x_{n+1} - x_n) + w_n = 0. \tag{3.14}
$$

Observing that f is of class C^2 and $x_n \to \xi$, and passing to the limit in (3.13), we get

$$w_n = -f(x_n) - Df(x_n)(x_{n+1} - x_n) \to -f(\xi) \ \text{ as } \ n \to \infty.$$

Since $\mathrm{gph}(F)$ is closed and $(x_{n+1}, w_n) \in \mathrm{gph}(F)$, we have

$$\lim_{n\to\infty} \big(x_{n+1}, w_n\big) = (\xi, -f(\xi)) \in \mathrm{gph}(F).$$

Thus $0 \in f(\xi) + F(\xi)$, which means that ξ solves (1.2).

As $\lim\limits_{n\to\infty} x_n = \xi$, we deduce

$$\sum_{j=n}^{\infty} (x_j - x_{j+1}) = x_n - \xi,$$

for each $n \geqslant 0$. So,

$$\|x_n - \xi\| \leqslant \sum_{j=n}^{\infty} \|x_j - x_{j+1}\| < \sum_{j=n}^{\infty} (t_{j+1} - t_j) = t^* - t_n. \tag{3.15}$$

Note that $t^* = 2\eta\beta$ and $t_0 = 0$, by letting $n = 0$ in (3.15), we get (3.2). Finally, taking into account (3.7), (3.8), and (3.15), we obtain (3.3) and (3.4). The proof is thereby completed. □

Remark 3.7. The assumptions and the conclusion of Theorem 3.4 are different from the ones obtained in [6, Theorem 2]. Our assumptions concern only the starting point, while the assumptions in [6, Theorem 2] depend on the next iteration.

Remark 3.8. Kantorovich-type results were also presented in [16, Theorem 3.2]. The difference between Theorem 3.4 and [16, Theorem 3.2] lies essentially on the used assumptions. In fact, the involved parameters are completely different as well as the region of the metric regularity. For Theorem 3.4, one needs the metric regularity property of $\Phi := Df(x) + F$ on the set $V(\Phi, x, r, s)$, where x is the starting point. On the other hand, in [16, Theorem 3.2], the authors used the Lipschitz-like property assumption for $Q_{\bar{x}}^{-1}$ (which is equivalent to the metric regularity of $Q_{\bar{x}}$ around $(\bar{x}, \bar{y}) \in \mathrm{gph}(f+F)$), where $Q_{\bar{x}}(\cdot) = f(\bar{x}) + Df(\cdot - \bar{x}) + F(\cdot)$, as well as the condition $\lim\limits_{x\to\bar{x}} d\big(\bar{y}, f(x) + F(x)\big) = 0$ (a kind of lower semicontinuity of $f + F$, which is not necessary in our analysis). The following example compares the applicability of the two results.

Example 3.9. Consider the mappings $f(x) = \frac{1}{3}x^3 - x + 1$ and $F(x) = \mathbb{R}_+$, $x \in \mathbb{R}$. By a direct computation, one has

$$d\big(0, f(x) + F(x)\big) = \max\left\{0, \frac{1}{3}x^3 - x + 1\right\}, \beta(\tau, x) = \tau \max\left\{0, \frac{1}{3}x^3 - x + 1\right\},$$
$$K(\tau, x, r) = \tau \sup_{|u-x|\leqslant r} |f''(u)| \leqslant 2\tau(|x| + r).$$

By setting $x = x_0 = -2$, we get

$$d\big(0, f(x_0) + F(x_0)\big) = \frac{1}{3}, \beta(\tau, x_0) = \frac{\tau}{3}, K(\tau, x_0, r) \leqslant 2\tau(2 + r).$$

Let $\Phi_{x_0}(u) := f'(x_0)u + F(u)$, $u \in \mathbb{R}$, and let $r = 0.5$, $s = 1$. For each $(u, v) \in V\big(\Phi_{x_0}, x_0, 4r, s\big)$ one obtains

$$\begin{aligned} d\big(u, \Phi_{x_0}^{-1}(v)\big) &= \max\left\{0, u - \frac{v}{f'(x_0)}\right\} = \max\left\{0, u - \frac{1}{3}v\right\} \\ &= \max\left\{0, \frac{1}{3}(3u - v)\right\} = \frac{1}{3}\max\left\{0, f'(x_0)u - v\right\} = \frac{1}{3}d\big(v, \Phi_{x_0}(u)\big). \end{aligned}$$

This relation shows that $\operatorname{reg}\big(\Phi_{x_0}, x_0, 4r, s\big) = \frac{1}{3}$. Let us choose $\tau = 0.5 > \frac{1}{3}$, which yields $\beta(\tau, x_0) = \frac{1}{6}$, $K(\tau, x_0, r) \leqslant \frac{5}{6}$. By taking $\alpha = \frac{5}{18} < 1$, then $2\eta\beta = \frac{1}{3}\frac{1}{1+\sqrt{1-\alpha}} < \frac{1}{3} < r$. Thus, all conditions of Theorem 3.4 are satisfied for starting point $x_0 = -2$.

Let us check if the assumptions of Theorem 3.2 in [16] hold for $x_0 = -2$. One has $Q_{x_0}(u) = f(x_0) + f'(x_0)(u - x_0) + \mathbb{R}_+$. We use the same notation for the constants M, L, r_{x_0}, r_{y_0}, r_0, δ, η and $y_0 \in Q_{x_0}(x_0)$ as in Theorem 3.2 of [16]. These constants must satisfy the following constraints

$$\begin{cases} r_0 = \min\left\{r_{y_0}, -2Lr_{x_0}^2, \frac{r_{x_0}(1-MLr_{x_0})}{4M}\right\}, \\ \delta \leqslant \min\left\{\frac{r_{x_0}}{4}, \frac{r_{y_0}}{11L}, 6r_0, 1\right\}, \\ (M+1)L(\eta\delta + 2r_{x_0}) \leqslant 2, \\ |y_0| < \frac{L\delta^2}{4}. \end{cases} \tag{3.16}$$

We note that $y_0 \in f(x_0) + \mathbb{R}_+ = \left[\frac{1}{3}, +\infty\right)$, $M \geqslant \operatorname{reg}(Q_{x_0}, (x_0, y_0)) = \frac{1}{3}$ and

$$L = \sup_{|u-x_0|\leqslant r_{x_0}/2} |f''(u)| = \sup_{|u-x_0|\leqslant r_{x_0}/2} |2u| = 2\,|x_0| + r_{x_0} = 4 + r_{x_0}.$$

From the second and the last inequalities in (3.16), one gets

$$\frac{1}{3} \leqslant |y_0| < \frac{L\delta^2}{4} \leqslant \frac{L}{64}r_{x_0}^2 = \frac{r_{x_0}^2(4 + r_{x_0})}{64}. \tag{3.17}$$

By using the third inequality of (3.16), we deduce

$$(4 + r_{x_0})r_{x_0} = Lr_{x_0} < \frac{1}{M+1} \leqslant \frac{3}{4}. \tag{3.18}$$

Nevertheless, (3.17) and (3.18) are not simultaneously valid. Hence, the conditions of Theorem 3.2 in [16] can not be satisfied for the starting point $x_0 = -2$.

4. Theorems of Smale's type

In this section, we consider problem (1.2) by assuming that the function f is analytic in an open subset U of the Banach space X.

Definition 4.1. A function $f : X \longrightarrow Y$ is called analytic at a point $x \in X$ if all derivatives $D^k f(x)$ exist, and there is a neighborhood $\mathbb{B}_X(x, \varepsilon)$ of x such that

$$f(y) = \sum_{k=0}^{\infty} \frac{D^k f(x)}{k!}(y - x)^k, \tag{4.1}$$

for all $y \in \mathbb{B}_X(x, \varepsilon)$, where $\frac{D^k f(x)}{k!}v^k$ stands for the value of the k-multilinear operator $\frac{D^k f(x)}{k!}$ at the k-multiple $(v, \ldots, v)$ in X^k. We will say that f is analytic on an open set $U \subset X$ if f is analytic at every point in U.

For an analytic function f and for each $x \in U$, the radius of convergence $R(f,x)$ of the Taylor's series in (4.1) is defined by (cf. [4])

$$R(f,x)^{-1} = \limsup_{k\to\infty} \left\| \frac{D^k f(x)}{k!} \right\|^{\frac{1}{k}}. \tag{4.2}$$

Theorem 4.2. *Let $\xi \in U$ be a solution of problem (1.2) and $\eta = Df(\xi)(\xi) - f(\xi)$. Suppose that the mapping $\Phi(\cdot) = Df(\xi)(\cdot) + F(\cdot)$ is metrically regular on the neighborhood $V = \mathbb{B}_X[\xi, r] \times \mathbb{B}_Y[\eta, r']$ of (ξ, η) with modulus $\tau > 0$. Define*

$$\gamma = \gamma(\tau, f, \xi) = \sup_{k\geqslant 2} \left\{ \left[\tau \left\| \frac{D^k f(\xi)}{k!} \right\| \right]^{\frac{1}{k-1}} \right\}.$$

Let $\rho = \frac{3-\sqrt{7}}{2} \approx 0.17712\ldots$ be the smallest solution of the equation

$$2t - \psi(t) = 0, \quad \psi(t) = 2t^2 - 4t + 1,$$

and $\theta = \frac{(1-\rho)^2}{\psi(\rho)} > 1$. Pick

$$0 < \varepsilon < \min\left\{ \frac{1}{2\theta} r, \frac{1}{1+\theta} r', \frac{\tau}{1+\theta} r' \right\},$$

such that $\mathbb{B}_X[\xi, \varepsilon] \subset U$. Then for any $x_0 \in \mathbb{B}_X[\xi, \varepsilon]$ satisfying

$$\|x_0 - \xi\|\, \gamma < \rho,$$

there exists a Newton sequence generated by (1.1), which converges quadratically to ξ:

$$\|x_n - \xi\| \leqslant \frac{\gamma}{\psi(\rho)} \|x_{n-1} - \xi\|^2, \quad n = 1, 2, \ldots. \tag{4.3}$$

The proof of Theorem 4.2 needs some auxiliary lemmas whose proofs are similar to the classical ones and will be omitted in this paper (see e.g. [4] for more details).

Lemma 4.3. *For any $x \in \mathbb{B}_X[\xi, r] \cap U$ with $r\gamma < 1$, the following Taylor's series is convergent to $f(x)$:*

$$\sum_{k=0}^{\infty} \frac{D^k f(\xi)}{k!} (x - \xi)^k = f(x).$$

Lemma 4.4. *Let $x \in U$ be such that $s = \gamma \|x - \xi\| < 1$. Then for all $k \geqslant 1$, one has*

$$\left\|D^k f(x) - D^k f(\xi)\right\| \leqslant k! \tau^{-1} \gamma^{k-1} \left[\frac{1}{(1-s)^{k+1}} - 1 \right]. \tag{4.4}$$

In particular,

$$\|Df(x) - Df(\xi)\| \leqslant \tau^{-1} \gamma \frac{2-s}{(1-s)^2} \|x - \xi\|. \tag{4.5}$$

Lemma 4.5. *Let r be a positive real number such that $\gamma r < 1$ and let $x \in \mathbb{B}_X[\xi, r] \cap U$. One has*

$$\|f(\xi) - f(x) - Df(x)(\xi - x)\| \leqslant \tau^{-1}\gamma \frac{1}{(1 - \gamma\|x - \xi\|)^2} \|x - \xi\|^2 .$$

Let us now prove Theorem 4.2.

Proof of Theorem 4.2. As in the previous section, we can rewrite Newton's iteration in the following form

$$x_{n+1} \in \Phi_n^{-1}(y_n) \quad \text{with} \quad \Phi_n = Df(x_n) + F \;\text{ and }\; y_n = [Df(x_n) - f](x_n).$$

Set $L = \tau^{-1}\frac{\rho(2-\rho)}{(1-\rho)^2} < \tau^{-1}$, $\bar{\tau} = \frac{\tau}{1-L\tau}$, $c = \max\left\{1, \tau^{-1}\right\}$ and pick $\nu = \tau^{-1}\varepsilon > 0$. In order to use Theorem 2.1, let us check first that

$$\nu\bar{\tau} < \frac{r}{2}, \nu(\bar{\tau} + \tau) < r, \nu(1 + c\bar{\tau}) < r'.$$

From the choice of ν, we deduce

$$\nu\bar{\tau} = \nu\frac{\kappa}{1 - L\kappa} = \frac{\varepsilon}{1 - \frac{\rho(2-\rho)}{(1-\rho)^2}} = \frac{(1-\rho)^2}{\psi(\rho)}\varepsilon = \theta\varepsilon < \frac{r}{2}.$$

Moreover, we have

$$\nu(\bar{\tau} + \tau) = \nu\tau\left(\frac{1}{1 - L\kappa} + 1\right) = \varepsilon\left[\frac{(1-\rho)^2}{\psi(\rho)} + 1\right] = \varepsilon(\theta + 1) < \frac{r}{2\theta}(1 + \theta) \leqslant r.$$

On the other hand, if $\tau < 1$, then $c = \tau^{-1}$. In this case, one obtains

$$\begin{aligned} \nu(1 + c\bar{\tau}) &= \nu\left(1 + \frac{1}{1 - L\tau}\right) = \nu(1 + \theta) = \tau^{-1}\varepsilon(1 + \theta) \\ &< \tau^{-1}\frac{\tau}{1 + \theta}r'(1 + \theta) = r'. \end{aligned}$$

Otherwise, if $\tau \geqslant 1$, then $c = 1$. Thus,

$$\nu(1 + c\bar{\tau}) \leqslant \nu(\tau + \bar{\tau}) = \varepsilon(\theta + 1) < r'.$$

Now, we take $\mu > 0$ such that

$$2\mu + \nu\bar{\tau} < r/2, \nu(\bar{\tau} + \tau) < r, 2c\mu + \nu(1 + c\bar{\tau}) < r'.$$

Applying Theorem 2.1 with $a = r$, $\delta = r$ and $b = r'$, for any linear and continuous map $A : X \longrightarrow Y$ having $\|A\| \leqslant L$, the sum $\Phi_A = A + \Phi$ is metrically regular on $\mathbb{B}_X[\xi, \mu] \times \mathbb{B}_Y[\eta + A(\xi), \nu]$ together with modulus $\bar{\tau}$.

Let $x_0 \in \mathbb{B}_X[\xi, \varepsilon]$ such that $\gamma\|x_0 - \xi\| < \rho$. If $x_0 = \xi$ then we set $x_n = x_0$ and stop. Otherwise, put $z_0 = [Df(x_0) - f](\xi)$, $y_0 = [Df(x_0) - f](x_0)$ and $\Phi_0(\cdot) = Df(x_0)(\cdot) + F(\cdot)$. Denoting $s_0 = \gamma\|x_0 - \xi\| < \rho$ and applying Lemma 4.4, we have

$$\begin{aligned} \|Df(x_0) - Df(\xi)\| &\leqslant \tau^{-1}\gamma\frac{2 - s_0}{(1 - s_0)^2}\|x_0 - \xi\| \leqslant \tau^{-1}\left(\gamma\|x_0 - \xi\|\right)\frac{2 - \rho}{(1 - \rho)^2} \\ &\leqslant \tau^{-1}\frac{\rho(2 - \rho)}{(1 - \rho)^2} = L. \end{aligned}$$

Hence, $\Phi_0 = [Df(x_0) - Df(\xi)] + \Phi$ is metrically regular on $\mathbb{B}_X[\xi,\mu] \times \mathbb{B}_Y[z_0,\nu]$ with modulus $\bar\tau$. Since $s_0 = \gamma\|x_0 - \xi\| < \rho$, using Lemma 4.5, we get

$$\begin{aligned}\|y_0 - z_0\| = \|f(\xi) - f(x_0) - Df(x_0)(\xi - x_0)\| &\leqslant \tau^{-1}\gamma\frac{1}{(1-s_0)^2}\|x_0-\xi\|^2 \\ &\leqslant \tau^{-1}\frac{\rho}{(1-\rho)^2}\|x_0-\xi\| \leqslant \tau^{-1}\varepsilon = \nu.\end{aligned}$$

Thanks to the metric regularity property of Φ_0 on $\mathbb{B}_X[\xi,\mu] \times \mathbb{B}_Y[z_0,\nu]$, we obtain

$$\begin{aligned}d\big(\xi,\Phi_0^{-1}(y_0)\big) \leqslant \bar\tau d\big(y_0,\Phi_0(\xi)\big) \leqslant \bar\tau\|y_0-z_0\| &\leqslant \bar\tau\tau^{-1}\gamma\frac{1}{(1-s_0)^2}\|x_0-\xi\|^2 \\ &= \frac{\tau}{1-\tau L}\tau^{-1}\gamma\frac{1}{(1-s_0)^2}\|x_0-\xi\|^2 = \frac{\gamma}{(1-\tau L)(1-s_0)^2}\|x_0-\xi\|^2 \\ &< \frac{\gamma}{(1-\tau L)(1-\rho)^2}\|x_0-\xi\|^2 = \sigma\|x_0-\xi\|^2,\end{aligned}$$

where $\sigma = \frac{\gamma}{(1-\tau L)(1-\rho)^2} = \frac{\gamma}{\psi(\rho)}$. Therefore, there exists a point, say x_1, belonging to $\Phi_0^{-1}(y_0)$ and satisfying the following relation

$$\|\xi - x_1\| < \sigma\|x_0-\xi\|^2.$$

Observe that

$$\sigma\|x_0-\xi\| = \frac{\gamma\|x_0-\xi\|}{\psi(\rho)} \leqslant \frac{\rho}{\psi(\rho)} = \frac{1}{2}, \tag{4.6}$$

which implies $x_1 \in \mathbb{B}_X[\xi,\varepsilon]$ since $\|x_0 - \xi\| \leqslant \varepsilon$. In addition, the inclusion $x_1 \in \Phi_0^{-1}(y_0)$ shows that x_1 satisfies Newton's iteration (1.1).

Repeating the previous procedure, we can construct a sequence (x_n) satisfying (1.1) for which the estimate in (4.3) is valid as well. Taking into account (4.3), we deduce

$$\|x_n-\xi\| \leqslant \left(\frac{\gamma}{\psi(\rho)}\|x_0-\xi\|\right)^{2^n-1}\|x_0-\xi\| \leqslant \left(\frac{1}{2}\right)^{2^n-1}\|x_0-\xi\|,$$

which yields $x_n \to \xi$ quadratically. This completes the proof of Theorem 4.2. □

Remark 4.6. For the nonlinear equation case $f(x) = 0$, i.e. $F = 0$, the metric regularity property of $\Phi(\cdot) = Df(\xi)(\cdot) + F(\cdot)$ in Theorem 4.2 is equivalent to the surjectivity of $Df(\xi)$. In such a situation, the property of metric regularity holds on the whole space. Hence, the radius of the starting domain ε is simply chosen such that $\varepsilon\gamma < \rho$ and $\mathbb{B}_X[\xi,\varepsilon] \subset U$.

The following theorem is a result of α-theory type for the set-valued generalized equations of the form (1.2).

Theorem 4.7. *Consider problem* (1.2) *in the case where f is analytic on open subset $U \subset X$. For $\tau > 0$ and $z \in U$ we define*

$$\begin{aligned}\beta(\tau,z) &= \tau d\big(0, f(z)+F(z)\big), \\ \gamma(\tau,f,z) &= \sup_{k\geqslant 2}\left\{\left[\tau\left\|\frac{D^kf(z)}{k!}\right\|\right]^{\frac{1}{k-1}}\right\},\end{aligned}$$

$$\alpha(\tau, f, z) = \beta(\tau, z)\gamma(\tau, f, z).$$

Let $\psi(t) = 2t^2 - 4t + 1$ *and* $\alpha \approx 0.1307169\ldots$ *be the smallest real root of the equation*

$$2t - [\psi(t)]^2 = 0.$$

Let $x \in U$ *and* $r, s > 0$ *such that* $\mathbb{B}_X[x, r] \subseteq U$ *and assume the following conditions are fulfilled:*

(1) $\Phi(\cdot) = Df(x)(\cdot) + F(\cdot)$ *is metrically regular on* $V = V(\Phi, x, 4r, s)$ *with a modulus* $\tau > \operatorname{reg}(\Phi, x, 4r, s)$,
(2) $d\big(0, f(x) + F(x)\big) < s$,
(3) $\eta\beta(\tau, x) \leqslant r$, *where* $\eta = \frac{\alpha+1-\sqrt{\alpha^2-6\alpha+1}}{4\alpha}$,
(4) $\alpha(\tau, f, x) \leqslant \alpha$.

Then, there exists a solution ξ *of the generalized equation* (1.2) *such that*

$$\|x - \xi\| \leqslant \eta\beta(\tau, x) \leqslant r. \tag{4.7}$$

In addition, there is a sequence (x_n) *generated by Newton's method which starts at* x *and satisfies the estimate*

$$\|x_n - \xi\| \leqslant \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^{2^k-1} [\psi(\alpha)]^n \left(\frac{1}{2}\right)^{2^n-1} \beta(\tau, x). \tag{4.8}$$

As similar as in Theorem 3.4, we have only to consider the case where $\beta(\tau, x) > 0$. The proof will need some complement statements.

Lemma 4.8. *Suppose that* f *is analytic on* $U \subset X$. *With* x, r, s, τ *and* $\gamma(\tau, f, x)$ *as in the previous theorem, if* $r\gamma(\tau, f, x) < 1$ *and* $\|y - x\| \leqslant r$, *the Taylor series*

$$\sum_{k=0}^{\infty} \frac{D^k f(x)}{k!}(y - x)^k$$

is convergent.

Lemma 4.9. *Suppose that* f *is analytic on an open* $U \subset X$. *Let* $\bar{x} \in U$ *and* $r > 0$ *such that* $\mathbb{B}_X[\bar{x}, r] \subset U$. *Then for* $x \in \mathbb{B}_X[\bar{x}, r]$ *and* $x' \in \mathbb{B}_X[x, \varrho] \cap U$ *with* $\gamma(\tau, f, x)\varrho < 1$, *one has*

$$\left\|\frac{D^k f(x')}{k!}\right\| \leqslant \tau^{-1} \frac{\gamma(\tau, f, x)^{k-1}}{[1 - \varrho\gamma(\tau, f, x)]^{k+1}}. \tag{4.9}$$

The proofs of Lemmas 4.8 and 4.9 are analogous to the classical ones used by Smale [4,19], where the quantity $\tau \left\|\frac{D^k f(x)}{k!}\right\|$ is replaced by $\left\|Df(x)^{-1}\frac{D^k f(x)}{k!}\right\|$.

Lemma 4.10. *Let* $x \in U$, $r > 0$, $s > 0$ *and* $\tau > 0$ *be such that* $\operatorname{reg}(\Phi, x, r, s) < \tau$ *with* $\Phi(\cdot) = Df(x)(\cdot) + F(\cdot)$. *Let* x' *be such that* $u = \|x' - x\|\gamma(\tau, f, x) < 1 - \frac{\sqrt{2}}{2}$. *Define* $\Phi'(\cdot) = Df(x')(\cdot) + F(\cdot)$ *and* $\tau' = \tau\frac{(1-u)^2}{\psi(u)}$, *where the function* ψ *is defined as in Theorem 4.7. If setting* $r' = r/4$ *and* $s' = \min\left\{s, \frac{r}{5\tau'}\right\}$, *then for all* $\mu \geqslant \tau'$ *the mapping* Φ' *is metrically regular on* $V(\Phi', x, r', s')$ *with modulus* μ.

Proof. According to the assumptions of Lemma 4.10, the mapping Φ is metrically regular on $V = V(\Phi, x, r, s)$ with some modulus $\kappa < \tau$. Let us denote $\alpha = \alpha(\tau, f, x)$, $\beta = \beta(\tau, x)$ and $\gamma = \gamma(\tau, f, x)$. Since $\|x' - x\|\gamma(\tau, f, x) < 1$, using Taylor's expansion and Lemma 4.9 we have

$$\begin{aligned}
\|Df(x') - Df(x)\| &= \left\| \int_0^1 D^2 f(x + t(x' - x))(x' - x)\, dt \right\| \\
&\leqslant \int_0^1 \left\|D^2 f(x + t(x' - x))\right\| \|x' - x\|\, dt \\
&\leqslant \int_0^1 \tau^{-1} \frac{2\gamma}{(1 - t\|x' - x\|\gamma)^3} \|x' - x\|\, dt \\
&\leqslant \tau^{-1}\left[\frac{1}{(1 - \|x' - x\|\gamma)^2} - 1\right] = \tau^{-1}\left[\frac{1}{(1 - u)^2} - 1\right].
\end{aligned}$$

Therefore, the linear perturbation $g = Df(x') - Df(x)$ is Lipschitz continuous on $\mathbb{B}_X[x, r]$ with constant $L = \tau^{-1}\left[\frac{1}{(1-u)^2} - 1\right]$ satisfying

$$\kappa L \leqslant \tau L = \frac{1}{(1 - u)^2} - 1 < 1.$$

Set $\kappa' = \frac{\kappa}{1-\kappa L}$, $\tau' = \frac{\tau}{1-\tau L}$, $r' = \frac{r}{4}$ and $s' = \min\left\{s, \frac{r}{5\tau'}\right\}$. Then $\Phi' = g + \Phi$ is metrically regular on $V' = V(\Phi', x, r', s')$ with a modulus $\kappa' < \tau' \leqslant \mu$ (see Theorem 2.4). A simple computation shows that $\tau' = \tau\frac{(1-u)^2}{\psi(u)}$. This completes the proof of Lemma 4.10. □

Lemma 4.11. *Let $\beta = \beta(\tau, x) > 0$ and $\gamma = \gamma(\tau, f, x) > 0$. Consider the majorizing function defined by*

$$\omega(t) = \frac{1}{1 - \gamma t} - 2\gamma t + \alpha - 1, \quad \textit{for} \quad t < \gamma^{-1}, \tag{4.10}$$

where α is given in the statement of Theorem 4.7. The Newton sequence obtained from the equation $\omega(t) = 0$ has the following form

$$\begin{cases} t_0 = 0, \\ t_{n+1} = t_n - \omega'(t_n)^{-1}\omega(t_n), n = 0, 1, \ldots \end{cases} \tag{4.11}$$

Then, (t_n) is well-defined, increasing and converges to the smallest zero t^* of ω. Moreover, one has

$$\delta_n = t_{n+1} - t_n \leqslant [\psi(\alpha)]^n \left(\frac{1}{2}\right)^{2^n - 1} \beta. \tag{4.12}$$

Proof. Let ψ be the function defined as in Theorem 4.7. We have

$$\begin{cases} \omega'(t) = \frac{\gamma}{(1-\gamma t)^2} - 2\gamma = -\gamma\frac{\psi(\gamma t)}{(1-\gamma t)^2}, \\ \omega''(t) = \frac{2\gamma^2}{(1-\gamma t)^3}, \\ \omega^{(k)}(t) = \frac{k!\gamma^k}{(1-\gamma t)^{k+1}}, k \geqslant 2. \end{cases} \tag{4.13}$$

Let

$$\beta(\omega,t)=\begin{cases}-\omega'(t)^{-1}\omega(t), & \text{if} \quad \omega'(t)\neq 0,\\ \infty, & \text{otherwise,}\end{cases}$$

$$\gamma(\omega,t)=\begin{cases}\sup_{k\geqslant 2}\left(\left|\omega'(t)^{-1}\frac{\omega^{(k)}(t)}{k!}\right|^{\frac{1}{k-1}}\right), & \text{if} \quad \omega'(t)\neq 0,\\ \infty, & \text{otherwise,}\end{cases}$$

$$\alpha(\omega,t)=\beta(\omega,t)\gamma(\omega,t).$$

It is easy to see that $\beta(\omega,0)=\beta$, $\gamma(\omega,0)=\gamma$, and $\alpha(\omega,0)=\alpha$. By (4.11), it holds that

$$t_1=t_0+\beta(\omega,t_0)=\beta(\omega,0)=\beta>t_0.$$

A simple computation shows that

$$t^*=\frac{\alpha+1-\sqrt{\alpha^2-6\alpha+1}}{4\gamma}<\left(1-\frac{\sqrt{2}}{2}\right)\gamma^{-1}$$

is the smallest solution of the equation $\omega(t)=0$. From the choice of α, we have

$$\alpha<\frac{\alpha+1-\sqrt{\alpha^2-6\alpha+1}}{4}=t^*\gamma,$$

which implies that $t_1=\beta=\gamma^{-1}\alpha<t^*$.

Suppose that for some $n\geqslant 1$ there exist $t_1<\ldots<t_n<t^*$ satisfying $t_{j+1}=t_j-\omega'(t_j)^{-1}\omega(t_j)$. By definition of ψ, we have $0<\psi(\gamma t)\leqslant 1$ for any t with $\gamma t<1-\frac{1}{\sqrt{2}}$. Since $\gamma t_n<\gamma t^*<1-\frac{1}{\sqrt{2}}$, it is obvious that $\psi(\gamma t_n)>0$. Hence, $\omega'(t_n)=-\gamma\frac{\psi(\gamma t)}{(1-\gamma t)^2}<0$, and so $t_{n+1}=t_n-\omega'(t_n)^{-1}\omega(t_n)$ makes sense. Because ω has no zero in the interval $(-\infty,t^*)$, and $\omega(0)=\alpha>0$, we deduce that $\omega(t)>0$ whenever $t<t^*$. Thus, from $t_n<t^*$ we get

$$t_{n+1}=t_n-\omega'(t_n)^{-1}\omega(t_n)>t_n.$$

To show that $t_{n+1}<t^*$, we consider the following function

$$\phi(t)=t-\omega'(t)^{-1}\omega(t),\quad t<\gamma^{-1}.$$

We have

$$\phi'(t)=\frac{\omega(t)\omega''(t)}{[\omega'(t)]^2}=\frac{2(1-\gamma t)}{[\psi(\gamma t)]^2}\omega(t)>0,\quad \text{for each}\quad t<t^*.$$

Therefore, the function ϕ is increasing on $(-\infty,t^*]$. So, the monotonicity of ϕ tells us

$$t_{n+1}=\phi(t_n)<\phi(t^*)=t^*.$$

By induction, the sequence (t_n) is well-defined, increasing and bounded from above by t^*. Let $\bar{t}=\lim\limits_{n\to\infty}t_n$; then $\bar{t}\leqslant t^*$. Passing to the limit as $n\to\infty$ in (4.11) we find $\omega(\bar{t})=0$, which implies $\bar{t}=t^*$.

Now, for each n, we set $\gamma_n=\gamma(\omega,t_n)$, $\delta_n=t_{n+1}-t_n=\beta(\omega,t_n)$ and $\alpha_n=\alpha(\omega,t_n)$. Recall that $\alpha_0=\alpha<1-\frac{\sqrt{2}}{2}$. As a result of [4, Lemme 133], we obtain

$$\delta_{n+1} \leqslant \frac{1-\alpha_n}{\psi(\alpha_n)}\alpha_n\delta_n, \alpha_{n+1} \leqslant \min\left\{\alpha_0, \frac{1}{[\psi(\alpha_n)]^2}\alpha_n^2\right\} \tag{4.14}$$

via inductive arguments. The second inequality in (4.14) yields

$$\begin{aligned}\alpha_n &\leqslant \frac{1}{[\psi(\alpha_n)]^2}\alpha_n^2 \leqslant \frac{1}{[\psi(\alpha_0)]^2}\alpha_n^2 \leqslant \cdots \leqslant \left[\frac{\alpha_0}{\psi(\alpha_0)^2}\right]^{2^n} [\psi(\alpha_0)]^2 \\ &= \left(\frac{1}{2}\right)^{2^n} [\psi(\alpha)]^2.\end{aligned}$$

Finally,

$$\begin{aligned}\delta_n &\leqslant \frac{1-\alpha_{n-1}}{\psi(\alpha_{n-1})}\alpha_{n-1}\delta_{n-1} \leqslant \frac{1}{\psi(\alpha)}\alpha_{n-1}\delta_{n-1} \leqslant \cdots \leqslant \\ &\leqslant \left[\prod_{i=0}^{n-1}\left(\frac{1}{\psi(\alpha)}\alpha_i\right)\right]\delta_0 \leqslant \left[\prod_{i=0}^{n-1}\left(\frac{1}{\psi(\alpha)}\left(\frac{1}{2}\right)^{2^i}[\psi(\alpha)]^2\right)\right]\delta_0 \\ &= [\psi(\alpha)]^n\left(\frac{1}{2}\right)^{2^n-1}\delta_0 = [\psi(\alpha)]^n\left(\frac{1}{2}\right)^{2^n-1}\beta,\end{aligned}$$

which completes the proof of Lemma 4.11. □

We are now in a position to prove Theorem 4.7.

Proof of Theorem 4.7. We rewrite (1.1) in the following form

$$x_{n+1} \in \Phi_n^{-1}(y_n), \Phi_n = Df(x_n) + F, y_n = [-f + Df(x_n)](x_n). \tag{4.15}$$

Let us begin with x satisfying the assumptions of Theorem 4.7 and $\beta = \beta(\tau, x) > 0$. Set $x_0 = x$, $\Phi_0 = \Phi$ and $\tau_0 = \tau$. Due to the fact that $\tau_0 > \operatorname{reg}(\Phi_0, x_0, 4r, s)$, we can pick $\kappa_0 < \tau_0$ for which Φ_0 is metrically regular on $V(\Phi_0, x_0, 4r, s)$ with modulus κ_0. By taking $y_0 = [-f + Df(x_0)](x_0)$, we get

$$d\big(y_0, \Phi_0(x_0)\big) = d\big(0, f(x_0) + F(x_0)\big) < s,$$

which means that $(x_0, y_0) \in V(\Phi_0, x_0, 4r, s)$. Therefore,

$$\begin{aligned}d\big(x_0, \Phi_0^{-1}(y_0)\big) &\leqslant \kappa_0 d\big(y_0, \Phi_0(x_0)\big) = \kappa_0 d\big(0, f(x_0) + F(x_0)\big) \\ &< \tau_0 d\big(0, f(x_0) + F(x_0)\big) = \beta(\tau_0, x_0) = \beta.\end{aligned}$$

Choose $x_1 \in \Phi_0^{-1}(y_0)$ such that $\|x_0 - x_1\| < \beta$. With the notation of Lemma 4.11 we have $t_1 - t_0 = \beta$, which yields

$$\|x_0 - x_1\| < t_1 - t_0.$$

Suppose that, for some $n \geqslant 1$, there are $x_1, \ldots, x_n$ having the following properties

- $x_{k+1} \in \Phi_k^{-1}(y_k)$, for $\Phi_k(\cdot) = Df(x_k)(\cdot) + F(\cdot)$, $y_k = [-f + Df(x_k)](x_k)$;
- $\|x_k - x_{k+1}\| < t_{k+1} - t_k = \delta_k$.

The case $0 \in f(x_n) + F(x_n)$ being trivial, we will focus on the remain one. By the triangle inequality, we have

$$\|x - x_k\| = \|x_0 - x_k\| \leqslant \sum_{j=0}^{k-1} \|x_j - x_{j+1}\| < \sum_{j=0}^{k-1} (t_{j+1} - t_j) = t_k < t^* \tag{4.16}$$

for all $k \leqslant n$. Let $\gamma_0 = \gamma(\tau_0, f, x_0) = \gamma$, and denote $u_n = \|x_0 - x_n\| \gamma_0$. Inequality (4.16) tells us $u_n < t_n\gamma < t^*\gamma < 1 - \frac{\sqrt{2}}{2}$. Set $\kappa_n = \kappa_0 \frac{(1-\gamma t_n)^2}{\psi(\gamma t_n)}$, $\tau_n = \tau_0 \frac{(1-\gamma t_n)^2}{\psi(\gamma t_n)}$, $r_n = r/4$ and $s_n = \min\left\{s, \frac{4r}{5\tau_n}\right\}$, so $\kappa_n > \kappa_0 \frac{(1-u_n)^2}{\psi(u_n)}$. Hence, according to Lemma 4.10, the mapping $\Phi_n(\cdot) = Df(x_n)(\cdot) + F(\cdot)$ is metrically regular on $V_n = V(\Phi_n, x_0, 4r_n, s_n)$ with modulus $\kappa_n < \tau_n$. Pick $y_n = [-f + Df(x_n)](x_n)$, we will claim that $(x_n, y_n) \in V_n$. Indeed, including (4.16), we get

$$\|x_0 - x_n\| < t_n \leqslant t^* = \eta\beta \leqslant r = 4r_n.$$

It is easy to check that $d\big(y_n, \Phi_n(x_n)\big) = d\big(0, f(x_n) + F(x_n)\big)$. Furthermore, as a consequence of $x_n \in \Phi_{n-1}^{-1}(y_{n-1})$, we can write

$$z_{n-1} = f(x_n) - [f(x_{n-1}) + Df(x_{n-1})(x_n - x_{n-1})] \in f(x_n) + F(x_n).$$

Hence, by using Taylor's expansion and (4.9) in Lemma 4.9, we find

$$\begin{aligned}
\|z_{n-1}\| &= \left\| \int_0^1 \Big\{ (1-s) \left[D^2 f(sx_n + (1-s)x_{n-1})(x_n - x_{n-1})^2 \right] \Big\} \, ds \right\| \\
&\leqslant \int_0^1 (1-s) \left\| D^2 f(sx_n + (1-s)x_{n-1}) \right\| \|x_n - x_{n-1}\|^2 \, ds \\
&\leqslant \int_0^1 (1-s) \left[\tau_0^{-1} \frac{2\gamma_0 \|x_{n-1} - x_n\|^2}{[1 - \gamma_0 \|sx_n + (1-s)x_{n-1}\|]^3} \right] ds \\
&\leqslant \tau_0^{-1} \int_0^1 (1-s) \frac{2\gamma_0 \|x_{n-1} - x_n\|^2}{[1 - \gamma_0 (s\|x_n - x_0\| + (1-s)\|x_{n-1} - x_0\|)]^3} \, ds \\
&\leqslant \tau_0^{-1} \int_0^1 (1-s) \frac{2\gamma \delta_{n-1}^2}{[1 - \gamma(st_n + (1-s)t_{n-1})]^3} \, ds \\
&= \tau_0^{-1}\gamma^{-1} \int_0^1 (1-s) \Big[\omega''(t_{n-1} + s\delta_{n-1})\delta_{n-1}^2 \Big] \, ds \\
&= \tau_0^{-1}\gamma^{-1} \big\{ \omega(t_n) - [\omega(t_{n-1}) + \omega'(t_{n-1})\delta_{n-1}] \big\} \\
&= \tau_0^{-1}\gamma^{-1}\omega(t_n) = \tau_0^{-1} \frac{\psi(\gamma t_n)}{(1-\gamma t_n)^2} \big[-\omega'(t_n)^{-1}\omega(t_n) \big] \\
&= \tau_0^{-1} \frac{\psi(\gamma t_n)}{(1-\gamma t_n)^2} \delta_n.
\end{aligned}$$

Therefore,

$$d\big(y_n, \Phi_n(x_n)\big) = d\big(0, f(x_n) + F(x_n)\big) \leqslant \|z_{n-1}\| \leqslant \tau_0^{-1} \frac{\psi(\gamma t_n)}{(1-\gamma t_n)^2} \delta_n. \tag{4.17}$$

Since $\psi(\gamma t_n) - (1-\gamma t_n)^2 = (\gamma t_n)^2 - 2(\gamma t_n) \leqslant 0$, the estimate in (4.17) shows that $d\big(y_n, \Phi_n(x_n)\big) \leqslant \tau_0^{-1}\delta_n$. Thanks to Lemma 4.11 and recalling that $\beta = \beta(\tau, x) = \beta(\tau_0, x_0)$, we obtain

$$\delta_n \leqslant [\psi(\alpha)]^n \left(\frac{1}{2}\right)^{2^n-1} \delta_0 \leqslant \frac{1}{2}\delta_0 = \frac{1}{2}\beta = \frac{1}{2}\tau_0 d\big(0, f(x_0) + F(x_0)\big) < \tau_0 s,$$

which implies $d\big(y_n, \Phi_n(x_n)\big) < s$. On the other hand, from the definitions of τ_n and α, combined with (4.17) and $\eta\beta = \eta\beta(\tau, x) \leqslant r$, we deduce

$$\begin{aligned} d\big(y_n, \Phi_n(x_n)\big) &\leqslant \tau_0^{-1}\frac{\psi(\gamma t_n)}{(1-\gamma t_n)^2}\delta_n = \frac{1}{\tau_n}\delta_n \leqslant \frac{1}{\tau_n}\frac{1}{2}\beta \leqslant \frac{1}{\tau_n}\frac{1}{2}\eta^{-1} r \\ &= \frac{1}{2\tau_n}\frac{4\alpha}{\alpha + 1 - \sqrt{\alpha^2 - 6\alpha + 1}} r \\ &= \frac{4r}{5\tau_n}\frac{5\left(\alpha + 1 + \sqrt{\alpha^2 - 6\alpha + 1}\right)}{16} < \frac{4r}{5\tau_n}. \end{aligned}$$

Thus, $d\big(y_n, \Phi_n(x_n)\big) < s_n$ and then, $(x_n, y_n) \in V_n$. By virtue of the metric regularity property of Φ_n, the following evaluation is valid

$$d\big(x_n, \Phi_n^{-1}(y_n)\big) \leqslant \kappa_n d\big(y_n, \Phi_n(x_n)\big) = \kappa_n d\big(0, G(x_n)\big) < \tau_n d\big(0, G(x_n)\big),$$

in which $G(x) := f(x) + F(x)$. Hence, there exists $x_{n+1} \in \Phi_n^{-1}(y_n)$ such that $\|x_n - x_{n+1}\| < \tau_n d\big(0, G(x_n)\big)$. Taking into account (4.17), we get

$$\begin{aligned} \|x_n - x_{n+1}\| &< \tau_n d\big(0, G(x_n)\big) \leqslant \tau_0 \frac{(1-\gamma t_n)^2}{\psi(\gamma t_n)} \|z_{n-1}\| \\ &\leqslant \tau_0 \frac{(1-\gamma t_n)^2}{\psi(\gamma t_n)} \tau_0^{-1} \frac{\psi(\gamma t_n)}{(1-\gamma t_n)^2}\delta_n = \delta_n = t_{n+1} - t_n. \end{aligned}$$

Therefore, Newton's sequence (x_n) for solving (1.2) is completely determined and satisfies

$$\|x_n - x_{n+1}\| < t_{n+1} - t_n, \quad \text{for all} \quad n = 0, 1, \ldots \tag{4.18}$$

The rest of proof is similar to the one in Theorem 3.4. Indeed, a same argument as in the proof of Theorem 3.4 affirms the convergence of the sequence (x_n). Let $\xi = \lim_{n\to\infty} x_n$, for each n one has $\sum_{k=n}^{\infty}(x_k - x_{k+1}) = x_n - \xi$. This permits for writing

$$\big\|x_n - \xi\big\| = \left\|\sum_{k=n}^{\infty}(x_k - x_{k+1})\right\| \leqslant \sum_{k=n}^{\infty}\big\|x_k - x_{k+1}\big\| < \sum_{k=n}^{\infty}(t_{k+1} - t_k) = t^* - t_n.$$

Taking $n = 0$ in the preceding estimate, the following evaluation holds true

$$\big\|x_0 - \xi\big\| < t^* - t_0 = t^* = \frac{\alpha + 1 - \sqrt{\alpha^2 - 6\alpha + 1}}{4\gamma} = \eta\beta = \eta\beta(\tau, x) \leqslant r.$$

Finally,

$$\begin{aligned}\|x_n-\xi\| &< \sum_{k=n}^{\infty}(t_{k+1}-t_k)=\sum_{k=n}^{\infty}\delta_k \leqslant \sum_{k=n}^{\infty}[\psi(\alpha)]^k\left(\frac{1}{2}\right)^{2^k-1}\beta \\ &\leqslant [\psi(\alpha)]^n\left(\frac{1}{2}\right)^{2^n-1}\sum_{k=0}^{\infty}\left(\frac{1}{2}\right)^{2^k-1}\beta(\tau,x)\end{aligned}$$

and the proof of Theorem 4.7 is thereby completed. □

Remark 4.12. It is theoretically possible to improve the value of constant α in the proof of Theorem 4.7. Indeed, the best theoretical value might be given by

$$\alpha=\sup\left\{a\in\left(0,1-\frac{1}{\sqrt{2}}\right):\sup_{0\leqslant t\leqslant a}\frac{t}{[\psi(t)]^2}=q(a)<1\right\}.$$

In [20], Wang Xinghua used the notion of Lipschitz condition with L-average to study the behavior of such a method for nonlinear equation $f(x)=0$. And he obtained a better constant $\alpha\leqslant 3-2\sqrt{2}\approx 0.1715729\ldots$ by considering another majorizing function $h(t)=\beta-t+\frac{\gamma t^2}{1-\gamma t}$.

Remark 4.13. Theorem 3.1 involves informations of the second derivative D^2f on a neighborhood of the solution ξ, while Theorem 4.2 needs informations of high-order derivatives only at such a solution ξ. The same comparison is valid for Theorems 3.4 and 4.7 by replacing the starting point x_0 with ξ.

Remark 4.14. Problem (1.2) subsumes as a particular case of the nonlinear equation $f(x)=0$ by taking $F\equiv 0$. In this context, our results can be applied by requiring only the metric regularity of the function f. While, in the classical ones, the invertibility of the first derivative of f at the reference point x is crucial (cf. [2,4,19]). However, with the assumption of the invertibility for $Df(x)$, it will be not possible to recover respectively Kantorovich's and Smale's classical results by setting $F\equiv 0$. This is due to the fact that our involved constants are larger than the classical ones.

5. Concluding remarks

In this paper, we investigated the Newton-type method for generalized equations in Banach spaces. We extended both Kantorovich-type theorems and Smale's classical (α,γ)-theory to this kind of problem and showed the quadratic convergence of the Newton sequence under the metric regularity assumption. Many issues remain to be investigated such as applying the algorithm studied in this paper to concrete examples. There are many problems in nonlinear programming, complementarity systems and differential variational inequalities that can be formulated as a generalized equation. It will be interesting to test these algorithms numerically and to compare them with others in the literature. This will be the subject of future works.

Acknowledgments

The authors thank Professor Michael Overton for reading the revised version of this manuscript. We are grateful to the anonymous referees for their insightful comments, which contribute to the improvement of the initial versions of this manuscript. The research of Huynh Van Ngai is supported by NAFOSTED.

References

[1] S. Adly, R. Cibulka, H.V. Ngai, Newton's method for solving inclusions using set-valued approximations, SIAM J. Optim. 25 (1) (2015) 159–184.
[2] L. Blum, F. Cucker, M. Shub, S. Smale, Complexity and Real Computation, Springer, New York, 1998.

[3] J.F. Bonnans, Local analysis of Newton-type methods for variational inequalities and nonlinear programming, Appl. Math. Optim. 29 (1994) 161–186.
[4] J.-P. Dedieu, Points Fixes, Zéros et la Méthode de Newton, Mathématiques & Applications, vol. 54, Springer, Berlin, Heidelberg, 2006.
[5] A.L. Dontchev, Local convergence of the Newton method for generalized equations, C. R. Acad. Sci. Paris Sér. I Math. 322 (4) (1996) 327–331.
[6] A.L. Dontchev, Local analysis of Newton-type method based on partial linearization, in: Letters in Applied Mathematics, vol. 32, Amer. Math. Soc., 1996, pp. 295–306.
[7] A.L. Dontchev, Uniform convergence of the Newton method for Aubin continuous maps, Serdica Math. J. 22 (3) (1996) 283–296.
[8] A.L. Dontchev, R.T. Rockafellar, Implicit Functions and Solution Mappings: A View from Variational Analysis, Springer Monographs in Mathematics, Springer, New York, 2009.
[9] W.B. Gragg, R.A. Tapia, Optimal error bounds for the Newton–Kantorovich theorem, SIAM J. Numer. Anal. 11 (1) (1974) 10–13.
[10] N.H. Josephy, Newton's method for generalized equations, Technical summary report, Mathematics Research Center, University of Wisconsin, Madison, 1979.
[11] L.V. Kantorovich, G.P. Akilov, Analyse Fonctionnelle, Mir, Moscou, 1981, translated from Russian.
[12] B.S. Mordukhovich, Variational Analysis and Generalized Differentiation I: Basic Theory, Grundlehren der Mathematischen Wissenschaften (A Series of Comprehensive Studies in Mathematics), vol. 330, Springer, Berlin, Heidelberg, 2006.
[13] H.V. Ngai, M. Théra, Error bounds on metric spaces and application to the perturbation stability of metric regularity, SIAM J. Optim. 19 (1) (2008) 1–20.
[14] J.M. Ortega, The Newton–Kantorovich theorem, Amer. Math. Monthly 75 (1968) 658–660.
[15] J.M. Ortega, W.C. Rheinboldt, Iterative Solution of Nonlinear Equations in Several Variables, Classics in Applied Mathematics, vol. 30, Society for Industrial and Applied Mathematics, 2000.
[16] M.H. Rashid, S.H. Yu, C. Li, S.Y. Wu, Convergence analysis of the Gauss–Newton-type method for Lipschitz-like mappings, J. Optim. Theory Appl. 158 (2013) 216–233.
[17] S.M. Robinson, Strongly regular generalized equations, Math. Oper. Res. 5 (1) (1980) 43–62.
[18] R.T. Rockafellar, R.J.-B. Wets, Variational Analysis, Grundlehren der Mathematischen Wissenschaften (A Series of Comprehensive Studies in Mathematics), vol. 317, Springer, Berlin, Heidelberg, 1998.
[19] S. Smale, Newton's method estimates from data at one point, in: The Merging of Disciplines: New Directions in Pure, Applied, and Computational Mathematics, Springer, New York, 1986, pp. 185–196.
[20] X. Wang, Convergence of Newton's method and inverse function theorem in Banach space, Math. Comp. 68 (225) (1999) 169–186.