



# Newton's Method for Solving Inclusions Using Set-Valued Approximations

Samir Adly, Radek Cibulka, Huynh van Ngai

## ► To cite this version:

Samir Adly, Radek Cibulka, Huynh van Ngai. Newton's Method for Solving Inclusions Using Set-Valued Approximations. SIAM Journal on Optimization, 2015, 25 (1), pp.159-184. 10.1137/130926730 . hal-01313138

**HAL Id: hal-01313138**

**<https://hal.science/hal-01313138>**

Submitted on 17 May 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# NEWTON'S METHOD FOR SOLVING INCLUSIONS USING SET-VALUED APPROXIMATIONS\*

SAMIR ADLY<sup>†</sup>, RADEK CIBULKA<sup>‡</sup>, AND HUYNH VAN NGAI<sup>§</sup>

**Abstract.** Results on stability of both local and global metric regularity under set-valued perturbations are presented. As an application, we study (super)linear convergence of a Newton-type iterative process for solving generalized equations. We investigate several iterative schemes such as the inexact Newton's method, the nonsmooth Newton's method for semismooth functions, the inexact proximal point algorithm, etc. Moreover, we also cover a forward-backward splitting algorithm for finding a zero of the sum of two multivalued (not necessarily monotone) operators. Finally, a globalization of the Newton's method is discussed.

**Key words.** generalized equation, metric regularity, semismooth function, Newton's method, forward-backward splitting, proximal point method

**1. Introduction.** Given two real Banach spaces  $X$  and  $Y$ , and multivalued mappings  $\Psi : X \rightrightarrows Y$  and  $F : X \rightrightarrows Y$ , we investigate the convergence properties of iterative processes for the following problem:

$$(1.1) \quad \text{Find } x \in X \text{ such that } 0 \in \Psi(x) + F(x).$$

Our aim is to derive a computational method to approximate a solution to (1.1). Namely, we study the following iterative process: *Choose a sequence of set-valued mappings  $A_k : X \times X \rightrightarrows Y$  and a starting point  $x_0 \in X$ , and generate a sequence  $(x_k)_{k \in \mathbb{N}}$  in  $X$  iteratively by taking  $x_{k+1}$  to be a solution to the auxiliary inclusion*

$$(1.2) \quad 0 \in A_k(x_{k+1}, x_k) + F(x_{k+1}) \quad \text{for each } k \in \mathbb{N}_0 := \{0, 1, 2, \dots\}.$$

Instead of considering a general inclusion

$$\text{find } x \in X \text{ such that } 0 \in \Gamma(x), \quad \text{with a given } \Gamma : X \rightrightarrows Y,$$

we focus on the case when the right-hand side  $\Gamma$  can be split into two parts as in (1.1). This might be useful in some applications since one can impose different assumptions on each part. Iterative schemes for solving inclusions with the general  $\Gamma$  can be found, for example, in [29]. In this paper, each result on convergence of the scheme (1.2) relies on metric regularity of an appropriate mapping which can be checked by using the graphical derivative criterion [8, Theorem 5.4.3] or the coderivative criterion [40,

---

<sup>†</sup>Laboratoire XLIM UMR-CNRS 6172, Université de Limoges, 87060 Limoges, France (samir.adly@unilim.fr).

<sup>‡</sup>Laboratoire XLIM UMR-CNRS 6172, Université de Limoges, 87060 Limoges, France, and Department of Mathematics, University of West Bohemia, Univerzitní 8, 306 14 Pilsen, Czech Republic (cibi@kma.zcu.cz).

<sup>§</sup>Department of Mathematics, University of Quynhon, 170 An Duong Vuong, Qui Nhon, Vietnam (ngaivn@yahoo.com). This author's research was supported by NAFOSTED under grant 101.99-2013.24.

Theorem 9.40]. When the mapping  $\Psi$  is single-valued, i.e.,  $\Psi = f$  for a given function  $f : X \rightarrow Y$ , one obtains a generalized equation, introduced by Robinson in [39], which reads as follows:

$$(1.3) \quad \text{Find } x \in X \text{ such that } 0 \in f(x) + F(x).$$

This model has been used to describe in a unified way various problems such as equations (when  $F \equiv 0$ ), inequalities (when  $Y = \mathbb{R}^n$  and  $F \equiv \mathbb{R}_+^n$ ), and variational inequalities (when  $F$  is the normal cone mapping corresponding to a closed convex subset of  $X$  or, more broadly, the subdifferential mapping of a convex function on  $X$ ). In particular, it covers optimality conditions, complementarity problems, and multiagent equilibrium problems (see [28] or [12]).

The case of single-valued approximations  $A_k : X \times X \rightarrow Y$  of the function  $f$  was studied in [10, 25, 26] and [12, Section 6C]. It is well known that specific choices of  $(A_k)_{k \in \mathbb{N}_0}$  lead to various methods for solving (1.3). Under the assumption that  $f$  is continuously differentiable (with the derivative  $f'$ ), taking  $A_k(x, u) = f(u) + f'(u)(x - u)$ ,  $(x, u) \in X \times X$  for each  $k \in \mathbb{N}_0$ , the iteration (1.2) becomes the *Newton's scheme*:

$$0 \in f(x_k) + f'(x_k)(x_{k+1} - x_k) + F(x_{k+1}) \quad \text{whenever } k \in \mathbb{N}_0.$$

When  $F$  coincides with the normal cone mapping with a closed convex subset of  $X$ , this scheme is known as the Josephy–Newton algorithm (see [27]). It is well known that if the starting point  $x_0$  is sufficiently close to some regular (in the sense of Robinson [39]) solution  $\bar{x}$ , then the Josephy–Newton method is well defined and converges superlinearly to  $\bar{x}$ .

On the other hand, when  $X = Y$ , considering a sequence  $(\lambda_k)_{k \in \mathbb{N}_0}$  in  $(0, +\infty)$  and taking  $A_k(x, u) = \lambda_k(x - u) + f(x)$ ,  $(x, u) \in X \times X$  for each  $k \in \mathbb{N}_0$ , one gets the *proximal point method*:

$$0 \in \lambda_k(x_{k+1} - x_k) + f(x_{k+1}) + F(x_{k+1}) \quad \text{for each } k \in \mathbb{N}_0.$$

In particular, if  $f \equiv 0$  one gets the proximal point algorithm for finding a solution to the inclusion

$$0 \in F(x).$$

The convergence properties of more general algorithms were studied in [3] and [4]. To be more precise, let  $(g_k)_{k \in \mathbb{N}_0}$  be a sequence of functions from  $X$  into  $Y$  such that each  $g_k$  is Lipschitz continuous near the origin and  $g_k(0) = 0$ . The authors of [3] and [4] consider the iteration scheme defined for each  $k \in \mathbb{N}_0$  by either

$$(1.4) \quad 0 \in g_k(x_{k+1} - x_k) + F(x_{k+1}) \quad \text{or} \quad 0 \in g_k(x_{k+1} - x_k) + F(x_{k+1}) - y,$$

where  $y \in Y$  is a given perturbation term nearby 0. Note that these algorithms cover the so-called *Mann's iteration* for solving the inclusion  $x \in T(x)$  with  $T : X \rightrightarrows X$  considered in [20]. The corresponding iteration schemes are, for each  $k \in \mathbb{N}_0$ , defined by

$$x_{k+1} \in (1 - \lambda_k)x_k + \lambda_k T(x_{k+1}) \quad \text{and} \quad x_{k+1} \in (1 - \lambda_k)x_k + \lambda_k(T(x_{k+1}) - y),$$

respectively, where the sequence  $(\lambda_k)_{k \in \mathbb{N}_0}$  in  $(0, 1)$  is nondecreasing and converges to 1, and  $y \in X$  is a given perturbation term in the vicinity of 0. Indeed, it suffices to

take  $Y = X$  and, for each  $x \in X$ , set  $F(x) = T(x) - x$  and  $g_k(x) = (\lambda_k - 1)x/\lambda_k$ ,  $k \in \mathbb{N}_0$  in (1.4).

From a numerical point of view, it is clear that the auxiliary inclusions in each of the above mentioned iteration schemes cannot be solved exactly because of the finite precision arithmetic and rounding errors. Hence various inexact methods were proposed in the literature to handle this issue. In [15], the authors suppose that there is a (known) sequence of error functions  $r_k : X \rightarrow Y$  and a sequence of single-valued approximations  $A_k : X \times X \rightarrow Y$  of  $f$ . Given  $x_k \in X$ , the next point  $x_{k+1}$  is required to be such that

$$(1.5) \quad 0 \in r_k(x_k) + A_k(x_{k+1}, x_k) + F(x_{k+1}).$$

At least for a solution  $\bar{x} \in X$  to (1.3), it is required to have  $r_k(\bar{x}) + A_k(\bar{x}, \bar{x}) = f(\bar{x})$  whenever  $k \in \mathbb{N}_0$  (which implies that  $r_k(\bar{x}) = 0$  provided that  $A_k(\bar{x}, \bar{x}) = f(\bar{x})$ ). The inexact Newton's method proposed in [2], and similarly the inexact Mann-type iteration in [20], require even to have  $r_k \equiv e_k$  for a given error sequence  $(e_k)_{k \in \mathbb{N}_0}$  in  $Y$  converging to 0. The (possibly) set-valued approximations can be used in investigating even more general error models. In [18], the author considered the local behavior of Newton-type algorithms for generalized equations with nonisolated solutions. The author used a set-valued approximation. The key assumption is the calmness of the solution mapping (which is equivalent to metric subregularity of the inverse mapping) of the perturbed generalized equation  $f(z) + F(z) + p \ni 0$ . In [26], the authors extended and improved the local convergence analysis of the Newtonian iterative methods developed earlier in [10, 18] by using both single-valued and set-valued approximations.

This paper is divided into eight parts. First, in section 2 we recall several notions from variational analysis. In the third section, we present a result concerning the stability of metric regularity under set-valued perturbations. In the fourth section, this statement is applied in the study of (super)linear convergence of a Newton-type iteration process (1.2). Also, the case when  $f$  can be approximated by a bunch of continuous linear operators around the reference point is investigated. This setting covers, in particular, the nonsmooth Newton's method for semismooth mappings between finite-dimensional spaces. In the fifth section, we discuss a forward-backward splitting algorithm for finding a zero of the sum of two multivalued operators (none of them necessarily monotone). In section 6, we investigate a globalization of the Newton's method in the sense that for any starting point  $x_0 \in X$ , the algorithm (1.2) produces a sequence converging to a solution  $\bar{x}$  of (1.3) which lies in a given neighborhood of  $x_0$ . A commentary and some concluding remarks can be found in sections 7 and 8.

**2. Mathematical background.** The recent book [12] by Dontchev and Rockafellar shows that the study of (equivalent) concepts of nonlinear analysis such as linear openness, metric regularity, and the inverse Aubin property is of a great importance from both the theoretical and the numerical point of view. We begin this section by fixing the notation. If we write  $a := b$ , we mean that  $a$  is defined by  $b$  without explicitly stating this in the text. Given a complete metric space  $(X, \varrho)$ , we denote by  $\mathbb{B}[x, r]$  and  $\mathbb{B}(x, r)$  the closed and open ball with the center  $x \in X$  and the radius  $r \geq 0$ , respectively. We set  $\mathbb{B}_X = \mathbb{B}[0, 1]$ . The closure, the interior, and the diameter of a subset  $K$  of  $X$  are denoted by  $\bar{K}$ ,  $\text{int } K$ , and  $\text{diam } K$ , respectively. The distance function generated by  $K$  is  $d(x, K) := \inf \{\varrho(x, k) : k \in K\}$ ,  $x \in X$ , with the usual convention  $d(x, \emptyset) := +\infty$ .

If  $X$  is a Banach space (always over  $\mathbb{R}$ ), then  $\|\cdot\|$  denotes its norm. By  $[x, y]$  we mean the closed line segment with endpoints  $x, y \in X$ , and the convex hull of  $K$  is denoted by  $\text{co } K$ . Given  $\alpha \in \mathbb{R}$ , we put  $\alpha K = \{\alpha k : k \in K\}$ . For the sets  $A$  and  $B$  in  $X$ , the set  $A + B := \{a + b : a \in A, b \in B\}$  is the *Minkowski sum* of  $A$  and  $B$  (we write  $a + B$  instead of  $\{a\} + B$ ). The *excess of  $A$  beyond  $B$*  is defined by

$$e(A, B) = \sup_{x \in A} d(x, B) = \inf\{\tau > 0 : A \subset B + \tau \mathbb{B}_X\},$$

with the conventions that  $e(\emptyset, B) := 0$  when  $B \neq \emptyset$  and that  $e(\emptyset, \emptyset) := +\infty$ . The *Pompeiu–Hausdorff distance between  $A$  and  $B$*  is then defined by

$$h(A, B) = \max\{e(A, B), e(B, A)\} = \inf\{\tau > 0 : A \subset B + \tau \mathbb{B}_X, B \subset A + \tau \mathbb{B}_X\}.$$

Note that

$$(2.1) \quad h(u + A, v + B) \leq \|u - v\| + h(A, B) \quad \text{for each } u, v \in X, A \subset X, \text{ and } B \subset X.$$

Let  $X$  and  $Y$  be Banach spaces. A set-valued mapping  $T : X \rightrightarrows Y$  is a mapping which assigns to each  $x \in X$  a (possibly empty) subset of  $Y$ . The *domain*, the *graph*, and the *range* of  $T$  are given respectively by  $\text{dom } T = \{x \in X : T(x) \neq \emptyset\}$ ,  $\text{gph } T = \{(x, y) \in X \times Y : y \in T(x)\}$ , and  $\text{rge } T = \bigcup_{x \in X} T(x)$ . By  $T^{-1}$  we denote the inverse of  $T$ , i.e.,  $x \in T^{-1}(y)$  if and only if  $y \in T(x)$ . If  $T(x)$  is a singleton for each  $x \in X$ , we identify the one-point set with its element,  $g(x)$  say. Then  $g : X \rightarrow Y$  denotes a single-valued mapping.

A mapping  $T : X \rightrightarrows Y$  is called *metrically regular at  $(\bar{x}, \bar{y}) \in \text{gph } T$  with a constant  $\kappa > 0$  on a neighborhood  $U \times V$  of  $(\bar{x}, \bar{y})$  in  $X \times Y$*  if

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

The mapping  $T$  is *metrically regular at  $(\bar{x}, \bar{y}) \in \text{gph } T$*  if there is a constant  $\kappa > 0$  along with neighborhoods  $U$  of  $\bar{x}$  in  $X$  and  $V$  of  $\bar{y}$  in  $Y$  such that  $T$  is metrically regular at  $(\bar{x}, \bar{y})$  with the constant  $\kappa$  on the neighborhood  $U \times V$ . If  $T$  is metrically regular at  $(\bar{x}, \bar{y})$  and  $T^{-1}$  has a localization at  $\bar{y}$  for  $\bar{x}$  that is nowhere multivalued, then  $T$  is called *strongly metrically regular at  $(\bar{x}, \bar{y})$* . In the case of a single-valued mapping  $g : X \rightarrow Y$  we simply speak about (strong) metric regularity at  $\bar{x}$ .

Further,  $T$  is (*Pompeiu–Hausdorff*) *upper semicontinuous at  $\bar{x} \in X$*  if

$$\lim_{x \rightarrow \bar{x}} e(T(\cdot), T(\bar{x})) = 0.$$

We say that  $T$  is *Lipschitz on  $U \subset X$*  provided that there is  $L > 0$  such that

$$h(T(x), T(y)) \leq L\|x - y\| \quad \text{whenever } x, y \in U.$$

Note that the previous property means that  $U \subset \text{dom } T$ . A mapping  $g : \text{dom } T \rightarrow Y$  is a *selection* for  $T$  provided that  $g(x) \in T(x)$  for each  $x \in \text{dom } T$ .

Finally,  $\mathcal{L}(X, Y)$  will denote the space of all continuous linear operators from  $X$  to  $Y$  equipped with the supremum norm.

**3. Stability of metric regularity under perturbation.** In this section, we prove a result on stability of the metric regularity under multivalued perturbation. Let  $X$  be a complete metric space, and let  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$  be an extended real-valued function. As usual,  $\text{dom } f := \{x \in X : f(x) < +\infty\}$  denotes the domain of  $f$ . We set

$$(3.1) \quad S := \{x \in X : f(x) \leq 0\}.$$

Given  $x \in X$ , the symbol  $[f(x)]_+$  denotes  $\max\{f(x), 0\}$ . First, we recall [33, Theorem 2.1] which allows us to estimate the distance  $d(\bar{x}, S)$  for a given point  $\bar{x} \notin S$ .

**THEOREM 3.1.** *Let  $(X, \varrho)$  be a complete metric space, let  $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$  be a lower semicontinuous function, and let  $\bar{x} \notin S$ . Then, setting*

$$(3.2) \quad m(\bar{x}) := \inf \left\{ \sup_{y \in X, y \neq \bar{x}} \frac{f(x) - [f(y)]_+}{\varrho(x, y)} : \begin{array}{l} \varrho(x, \bar{x}) < d(\bar{x}, S) \\ f(x) \leq f(\bar{x}) \end{array} \right\},$$

one has

$$(3.3) \quad m(\bar{x}) d(\bar{x}, S) \leq f(\bar{x}).$$

Now, we prove that the sum of a metrically regular mapping and a suitable multivalued perturbation remains metrically regular.

**THEOREM 3.2.** *Given Banach spaces  $X$  and  $Y$ , let  $\Phi : X \rightrightarrows Y$  be a set-valued mapping with closed graph, and let  $(\bar{x}, \bar{y}) \in \text{gph } \Phi$ . Suppose that  $\Phi$  is metrically regular at  $(\bar{x}, \bar{y})$  with a constant  $\kappa > 0$  on a neighborhood  $\mathbb{B}(\bar{x}, a) \times \mathbb{B}(\bar{y}, b)$  of  $(\bar{x}, \bar{y})$  for some  $a > 0$  and  $b > 0$ . Let  $\delta > 0$  and  $L \in (0, \kappa^{-1})$ , and set  $\tau = \kappa/(1 - \kappa L)$ . Let  $\alpha, \beta, \eta$  be positive constants satisfying*

$$(3.4) \quad 2\alpha + \eta/c + \beta\tau < \min\{a, \delta/2\}, \quad \beta(\tau + \kappa) < \delta, \quad 2(c\alpha + \eta) + \beta(1 + c\tau) < b,$$

with  $c := \max\{1, 1/\kappa\}$ . Then for any set-valued mapping  $G : X \rightrightarrows Y$  with closed graph such that

(i)  $G$  is Lipschitz on  $\mathbb{B}(\bar{x}, \delta)$  with the constant  $L$ ,

(ii)  $\text{diam } G(\bar{x}) \leq \eta$ ,

the set-valued mapping  $\Phi + G$  is metrically regular on  $\mathbb{B}(\bar{x}, \alpha) \times \mathbb{B}(\bar{y} + \bar{z}, \beta)$ , with the constant  $\tau$  for any  $\bar{z} \in G(\bar{x})$ .

*Proof.* Let  $f : X \times Y \times Y \rightarrow \mathbb{R}$  be defined by

$$f(x, z, y) := \begin{cases} \liminf_{u \rightarrow x} d(y, \Phi(u) + z) & \text{if } z \in G(x), \\ +\infty & \text{otherwise.} \end{cases}$$

Obviously,  $f$  is lower semicontinuous on  $X \times Y \times Y$ . For  $y \in Y$ , set

$$S(y) = \{(x, z) \in X \times Y : f(x, z, y) = 0\} = \{(x, z) \in X \times Y : z \in G(x), y - z \in \Phi(x)\}.$$

Let  $\varepsilon > 0$  be such that

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

Then pick  $\gamma \in (0, \varepsilon)$  such that

$$(3.5) \quad 1/(\kappa + \gamma) - L - \gamma > 1/(\tau + \varepsilon) \quad \text{and} \quad L + \gamma < c.$$

We define the (equivalent) metric  $\varrho : X \times Y \rightarrow \mathbb{R}$  on  $X \times Y$  by

$$\varrho((x_1, z_1), (x_2, z_2)) := \max\{\|x_1 - x_2\|, \|z_1 - z_2\|/c\}.$$

Let  $y \in \mathbb{B}(\bar{y} + \bar{z}, \beta)$  be given. One has

$$f(\bar{x}, \bar{z}, y) \leq \|y - \bar{y} - \bar{z}\| < \beta \leq \inf_{(x, z) \in X \times Y} f((x, z), y) + \beta.$$

By virtue of the Ekeland variational principle [16, Theorem 1.1], we can select a point  $(u, w) \in X \times Y$  satisfying

$$\varrho((u, w), (\bar{x}, \bar{z})) \leq \beta(\tau + \varepsilon) \quad \text{and} \quad f(u, w, y) \leq f(\bar{x}, \bar{z}, y)$$

such that

$$(3.6) \quad f(x, z, y) + \frac{1}{\tau + \varepsilon} \varrho((x, z), (u, w)) \geq f(u, w, y) \quad \text{for all} \quad (x, z) \in X \times Y.$$

Then  $w \in G(u)$ , and we have

$$\|u - \bar{x}\| \leq \beta(\tau + \varepsilon) < a, \quad \|y - w - \bar{y}\| \leq \|y - \bar{y} - \bar{z}\| + \|w - \bar{z}\| < \beta(1 + c(\tau + \varepsilon)) < b.$$

We prove that  $y - w \in \Phi(u)$ . Indeed, suppose that this is not the case. Let  $(u_n)_{n \in \mathbb{N}}$  be a sequence in  $X$  converging to  $u$  such that

$$(3.7) \quad \lim_{n \rightarrow +\infty} d(y, \Phi(u_n) + w) = f(u, w, y) > 0.$$

As  $\Phi$  has closed graph, going to a subsequence, if necessary, we may assume that  $u_n \notin \Phi^{-1}(y - w)$  for each  $n \in \mathbb{N}$ . By the metric regularity of  $\Phi$  on  $\mathbb{B}(\bar{x}, a) \times \mathbb{B}(\bar{y}, b)$ , for each  $n \in \mathbb{N}$ , we can find  $v_n \in \Phi^{-1}(y - w)$  such that

$$\|u_n - v_n\| < (1 + \gamma/\kappa) d(u_n, \Phi^{-1}(y - w)) \leq (\kappa + \gamma) d(y - w, \Phi(u_n)).$$

Then,  $\liminf_{n \rightarrow +\infty} \|u - v_n\| > 0$ . Since  $(u_n)_{n \in \mathbb{N}}$  converges to  $u$  and  $f(u, w, y) \leq f(\bar{x}, \bar{z}, y) < \beta$ , neglecting several leading terms, we may assume that  $u_n, v_n \in \mathbb{B}(\bar{x}, \delta)$  for each  $n \in \mathbb{N}$ . According to the Lipschitz property of  $G$  on  $\mathbb{B}(\bar{x}, \delta)$  and (3.5), there exists  $(w_n)_{n \in \mathbb{N}}$  such that

$$w_n \in G(v_n) \quad \text{and} \quad \|w - w_n\| \leq (L + \gamma)\|u - v_n\| < c\|u - v_n\| \quad \text{whenever} \quad n \in \mathbb{N}.$$

Therefore,  $\varrho((u, w), (v_n, w_n)) = \|u - v_n\|$  for any  $n \in \mathbb{N}$ . As  $y - w \in \Phi(v_n)$ , using also the relations (3.6), (3.7), and (3.5) we infer that

$$\begin{aligned} \frac{1}{\tau + \varepsilon} &\geq \limsup_{n \rightarrow +\infty} \frac{f(u, w, y) - f(v_n, w_n, y)}{\varrho((u, w), (v_n, w_n))} \geq \limsup_{n \rightarrow +\infty} \frac{d(y - w, \Phi(u_n)) - d(y - w_n, \Phi(v_n))}{\|u - v_n\|} \\ &\geq \limsup_{n \rightarrow +\infty} \frac{(\kappa + \gamma)^{-1} \|u_n - v_n\| - \|w - w_n\|}{\|u - v_n\|} \\ &\geq \limsup_{n \rightarrow +\infty} \frac{(\kappa + \gamma)^{-1} (\|u - v_n\| - \|u_n - u\|) - (L + \gamma)\|u - v_n\|}{\|u - v_n\|} \\ &= \lim_{n \rightarrow +\infty} \left( \frac{1}{\kappa + \gamma} - L - \gamma - \frac{\|u_n - u\|}{\|u - v_n\|(\kappa + \gamma)} \right) = \frac{1}{\kappa + \gamma} - L - \gamma > \frac{1}{\tau + \varepsilon}, \end{aligned}$$

a contradiction. Thus  $y - w \in \Phi(u)$ . Since  $w \in G(u)$ , we get that

$$(3.8) \quad S(y) \cap (\mathbb{B}[\bar{x}, \beta(\tau + \varepsilon)] \times \mathbb{B}[\bar{z}, \beta c(\tau + \varepsilon)]) \neq \emptyset.$$

Fix any  $(x, y) \in \mathbb{B}(\bar{x}, \alpha) \times \mathbb{B}(\bar{y} + \bar{z}, \beta)$ . We will show that

$$(3.9) \quad d(x, (\Phi + G)^{-1}(y)) \leq \tau d(y, \Phi(x) + G(x)).$$

Clearly, it suffices to prove the above inequality for the case that  $y \notin \Phi(x) + G(x)$ . Pick  $z \in G(x)$  such that

$$d(y, \Phi(x) + z) < (1 + \varepsilon) d(y, \Phi(x) + G(x)).$$

The Lipschitz property of  $G$  on  $\mathbb{B}(\bar{x}, \delta)$  implies that

$$(3.10) \quad \|z - \bar{z}\| \leq e(G(x), G(\bar{x})) + \text{diam } G(\bar{x}) \leq L\|x - \bar{x}\| + \eta < L\alpha + \eta < c\alpha + \eta.$$

Hence, by (3.8),

$$(3.11) \quad d((x, z), S(y)) \leq \varrho((x, z), (\bar{x}, \bar{z})) + d((\bar{x}, \bar{z}), S(y)) < \alpha + \eta/c + \beta(\tau + \varepsilon) < \delta/2.$$

Let us distinguish two cases.

*Case 1.*  $d(y, \Phi(x) + z) \geq \delta/(2\kappa)$ . As  $\kappa < \tau$ , by relation (3.8), one has

$$\begin{aligned} d(x, (\Phi + G)^{-1}(y)) &\leq \|x - \bar{x}\| + d(\bar{x}, (\Phi + G)^{-1}(y)) < \alpha + \beta(\tau + \varepsilon) < \delta/2 \\ &\leq \kappa d(y, \Phi(x) + z) < \tau(1 + \varepsilon) d(y, \Phi(x) + G(x)). \end{aligned}$$

*Case 2.*  $d(y, \Phi(x) + z) < \delta/(2\kappa)$ . Note that  $(x, z) \notin S(y)$ . We will show that

$$\begin{aligned} m(x, z) &:= \inf \left\{ \sup_{(u, w) \neq (v, w') \in X \times Y} \frac{f(u, w, y) - f(v, w', y)}{\varrho((u, w), (v, w'))} : \begin{array}{l} \varrho((u, w), (x, z)) < d((x, z), S(y)) \\ f(u, w, y) \leq f(x, z, y) \end{array} \right\} \\ &> \frac{1}{\tau + \varepsilon}. \end{aligned}$$

To see this, let  $(u, w) \in X \times Y$  be such that  $\varrho((u, w), (x, z)) < d((x, z), S(y))$  and  $f(u, w, y) \leq f(x, z, y)$ . Then  $w \in G(u)$ . As  $(u, w) \notin S(y)$ , we have  $u \notin \Phi^{-1}(y - w)$ . By (3.11), one has

$$\|u - \bar{x}\| \leq \|u - x\| + \|x - \bar{x}\| \leq 2\alpha + \eta/c + \beta(\tau + \varepsilon) < \min\{a, \delta/2\}.$$

Also, (3.10) and (3.11) imply that

$$\|y - w - \bar{y}\| \leq \|y - \bar{y} - \bar{z}\| + \|w - z\| + \|z - \bar{z}\| < 2(c\alpha + \eta) + \beta(1 + c(\tau + \varepsilon)) < b.$$

Thus,  $u \in \mathbb{B}(\bar{x}, \delta/2)$  and  $(u, y - w) \in \mathbb{B}(\bar{x}, a) \times \mathbb{B}(\bar{y}, b)$ . Let  $(u_n)_{n \in \mathbb{N}}$  be any sequence in  $X$  converging to  $u$  such that

$$(3.12) \quad \lim_{n \rightarrow +\infty} d(y, \Phi(u_n) + w) = f(u, w, y) > 0.$$

As  $\Phi$  has closed graph, going to a subsequence, if necessary, we may assume that  $u_n \notin \Phi^{-1}(y - w)$  for each  $n \in \mathbb{N}$ . By the metric regularity of  $\Phi$  on  $\mathbb{B}(\bar{x}, a) \times \mathbb{B}(\bar{y}, b)$  and (3.12), for each  $n \in \mathbb{N}$ , we can find  $v_n \in \Phi^{-1}(y - w)$  such that

$$\|u_n - v_n\| \leq (1 + \gamma/\kappa)d(u_n, \Phi^{-1}(y - w)) \leq (\kappa + \gamma)d(y - w, \Phi(u_n)) < (\kappa + \gamma)\delta/(2\kappa).$$

As  $u \in \mathbb{B}(\bar{x}, \delta/2)$ , we have  $v_n \in \mathbb{B}(\bar{x}, \delta)$  when  $n$  is sufficiently large and  $\gamma$  is sufficiently small (note that (3.5) remains true). Moreover,  $\liminf_{n \rightarrow +\infty} \|u - v_n\| > 0$  (because  $u \notin \Phi^{-1}(y - w)$ ). By the Lipschitz property of  $G$  on  $\mathbb{B}(\bar{x}, \delta)$ , there exists  $(w_n)_{n \in \mathbb{N}}$  such that

$$w_n \in G(v_n) \quad \text{and} \quad \|w - w_n\| \leq (L + \gamma)\|u - v_n\| < c\|u - v_n\| \quad \text{for each } n \in \mathbb{N}.$$

Therefore,  $\varrho((u, w), (v_n, w_n)) = \|u - v_n\|$  for any  $n \in \mathbb{N}$ . As  $y - w \in \Phi(v_n)$ , using



(3.12), we obtain

$$\begin{aligned}
& \limsup_{n \rightarrow +\infty} \frac{f(u, w, y) - f(v_n, w_n, y)}{\varrho((u, w), (v_n, w_n))} \geq \limsup_{n \rightarrow +\infty} \frac{d(y - w, \Phi(u_n)) - d(y - w_n, \Phi(v_n))}{\|u - v_n\|} \\
& \geq \limsup_{n \rightarrow +\infty} \frac{(\kappa + \gamma)^{-1} \|u_n - v_n\| - \|w - w_n\|}{\|u - v_n\|} \\
& \geq \limsup_{n \rightarrow +\infty} \frac{(\kappa + \gamma)^{-1} (\|u - v_n\| - \|u_n - u\|) - (L + \gamma) \|u - v_n\|}{\|u - v_n\|} \\
& = \lim_{n \rightarrow +\infty} \left( \frac{1}{\kappa + \gamma} - L - \gamma - \frac{\|u_n - u\|}{\|u - v_n\|(\kappa + \gamma)} \right) = \frac{1}{\kappa + \gamma} - L - \gamma > \frac{1}{\tau + \varepsilon}.
\end{aligned}$$

By virtue of Theorem 3.1, we derive that

$$\begin{aligned}
d(x, (\Phi + G)^{-1}(y)) & \leq d((x, z), S(y)) \leq (\tau + \varepsilon) f(x, z, y) \leq (\tau + \varepsilon) d(y, \Phi(x) + z) \\
& \leq (\tau + \varepsilon)(1 + \varepsilon) d(y, \Phi(x) + G(x)).
\end{aligned}$$

Hence, in both cases, taking  $\varepsilon \downarrow 0$ , we get (3.9). The proof is complete.  $\square$

*Remark 3.3.*

(i) Theorem 3.2 is valid for any complete metric space  $(X, d)$  and any linear metric space  $(Y, \delta)$  with a shift-invariant metric, i.e.,  $\delta(x + z, y + z) = \delta(x, y)$ , for any  $x, y$ , and  $z \in Y$ .

(ii) The conclusion of Theorem 3.2 fails (see [12, Example 5E.6] and also [9]) when condition (ii) of Theorem 3.2 on the diameter is omitted.

(iii) A result closely related to Theorem 3.2 was proved in [9], where the authors concentrate more on the size of the domain of the inverse of the sum. Under stronger assumptions, they obtained a stronger result: the Lipschitz continuity of  $(\Phi + G)^{-1}$ .

It turns out that the metric regularity of the resulting sum is not needed in some applications. Let us present a statement in which we do not insist on the uniform estimate for all points  $\bar{z} \in G(\bar{x})$ . Namely, without the assumption on the diameter of  $G(\bar{x})$ , one has the metric hemiregularity of the sum (see Remark 3.5(i)), given in the following theorem.

**THEOREM 3.4.** *Given Banach spaces  $X$  and  $Y$ , let  $\Phi : X \rightrightarrows Y$  be a set-valued mapping with closed graph, and let  $(\bar{x}, \bar{y}) \in \text{gph } \Phi$ . Suppose that  $\Phi$  is metrically regular at  $(\bar{x}, \bar{y})$  with a constant  $\kappa > 0$  on a neighborhood  $\mathbb{B}(\bar{x}, a) \times \mathbb{B}(\bar{y}, b)$  of  $(\bar{x}, \bar{y})$  for some  $a > 0$  and  $b > 0$ . Let  $\delta > 0$  and  $L \in (0, \kappa^{-1})$ , and set  $\tau = \kappa/(1 - \kappa L)$ . Let  $\beta > 0$  be such that*

$$(3.13) \quad \beta\tau < \min\{a, \delta/2\}, \quad \beta(\tau + \kappa) < \delta, \quad \beta(1 + c\tau) < \min\{b, \delta\},$$

*with  $c := \max\{1, 1/\kappa\}$ . Then for any set-valued mapping  $G : X \rightrightarrows Y$  with closed graph and a point  $\bar{z} \in G(\bar{x})$  such that*

$$(3.14) \quad e(G(x_1) \cap \mathbb{B}(\bar{z}, \delta), G(x_2)) \leq L\|x_1 - x_2\| \quad \text{whenever} \quad x_1, x_2 \in \mathbb{B}(\bar{x}, \delta),$$

*one has*

$$(3.15) \quad d(\bar{x}, (\Phi + G)^{-1}(y)) \leq \tau d(y, \Phi(\bar{x}) + \bar{z}) \quad \text{for all} \quad y \in \mathbb{B}(\bar{y} + \bar{z}, \beta).$$

*Proof.* The proof is quite similar to that of Theorem 3.2, and hence we mention only the needed changes. Let  $f$  and  $S(y)$ ,  $y \in Y$  be as before. Pick  $\varepsilon > 0$  such that

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

Fix any  $y \in \mathbb{B}(\bar{y} + \bar{z}, \beta)$ . Using exactly the same steps, one can prove that (3.8) is valid (it is only in the construction of  $(w_n)_{n \in \mathbb{N}}$  that one uses (3.14), together with the facts that  $w \in G(u) \cap \mathbb{B}(\bar{z}, \delta)$  and that  $(v_n)_{n \in \mathbb{N}}$  is in  $\mathbb{B}(\bar{x}, \delta)$  instead of the Lipschitz property of  $G$ ). Clearly, it suffices to prove (3.15) for the case that  $y - \bar{z} \notin \Phi(\bar{x})$ . By (3.8),

$$(3.16) \quad d((\bar{x}, \bar{z}), S(y)) \leq \beta(\tau + \varepsilon) < \delta/2.$$

Let us distinguish two cases.

*Case 1.*  $d(y, \Phi(\bar{x}) + \bar{z}) \geq \delta/(2\kappa)$ . As  $\kappa < \tau$ , by relation (3.16), one has

$$d(\bar{x}, (\Phi + G)^{-1}(y)) < \delta/2 \leq \kappa d(y, \Phi(\bar{x}) + \bar{z}) < \tau d(y, \Phi(\bar{x}) + \bar{z}).$$

*Case 2.*  $d(y, \Phi(\bar{x}) + \bar{z}) < \delta/(2\kappa)$ . One repeats exactly the same steps as in Case 2 in the proof of Theorem 3.2 with  $(x, z)$  replaced by  $(\bar{x}, \bar{z})$ . One uses (3.16) instead of (3.11) and the same construction of  $(w_n)_{n \in \mathbb{N}}$  mentioned above.  $\square$

*Remark 3.5.*

(i) Since  $\bar{y} \in \Phi(\bar{x})$ , the inequality (3.15) implies that

$$d(\bar{x}, (\Phi + G)^{-1}(y)) \leq \tau \|y - \bar{y} - \bar{z}\| \quad \text{for all } y \in \mathbb{B}(\bar{y} + \bar{z}, \beta).$$

The above property is sometimes called *metric hemiregularity* [1, Definition 2.4] and can be viewed as a counterpart to the traditional metric subregularity which means that

$$d(x, (\Phi + G)^{-1}(\bar{y})) \leq \tau d(\bar{y}, (\Phi + G)(x)) \quad \text{for all } x \text{ in the vicinity of } \bar{x}.$$

(ii) By setting  $\Phi = -F$ , problem (1.1) can be formulated as a coincidence problem:

$$\text{Find } \xi \in X \quad \text{such that} \quad \Phi(\xi) \cap \Psi(\xi) \neq \emptyset.$$

It is possible to obtain the conclusion of Theorem 3.4 from Theorem 2 in [7] by setting  $\Psi = -G(\cdot) + y$  with  $G$  Lipschitz. In Theorem 3.4, we assumed  $G$  to be pseudo-Lipschitz/Aubin continuous at the reference point (see conditions (3.14) and (5.6) for the precise definition), which is a weaker assumption. However, we note that the functional setting is not the same: we work in Banach spaces while results in [7] are for metric spaces. Concerning the coincidence points of mappings, we refer to [5, 6] and references therein.

**4. Convergence of Newton's method.** Let  $(x_k)_{k \in \mathbb{N}}$  be a sequence in a Banach space  $X$  which converges to a point  $x_0 \in X$ . Recall that  $(x_k)_{k \in \mathbb{N}}$  *converges linearly* if either  $\limsup_{k \rightarrow +\infty} \frac{\|x_{k+1} - x_0\|}{\|x_k - x_0\|} < 1$  when there is  $k_0 \in \mathbb{N}$  such that  $x_k \neq x_0$  whenever  $k > k_0$ , or there is  $k_0 \in \mathbb{N}$  such that  $x_k = x_0$  for each  $k > k_0$ . Similarly, we say that  $(x_k)_{k \in \mathbb{N}}$  *converges superlinearly* if either  $\lim_{k \rightarrow +\infty} \frac{\|x_{k+1} - x_0\|}{\|x_k - x_0\|} = 0$  when there is  $k_0 \in \mathbb{N}$  such that  $x_k \neq x_0$  whenever  $k > k_0$ , or there is  $k_0 \in \mathbb{N}$  such that  $x_k = x_0$  for each  $k > k_0$ . This type of convergence is often called *Q-linear* (*Q-superlinear*) in the literature. Throughout this section we will use the following premise.

**Standing assumptions.** Let  $X$  and  $Y$  be Banach spaces, let  $f : X \rightarrow Y$  be a continuous mapping, and let  $F : X \rightrightarrows Y$  be a set-valued mapping with closed graph. Assume that a point  $\bar{x} \in X$  is a solution to (1.3).

Theorem 3.4 can be used to derive a generalization of [12, Theorem 6C.1] when the iterative scheme (1.2) is considered.

**THEOREM 4.1.** *Suppose that  $f + F$  is metrically regular with a constant  $\kappa > 0$  on a neighborhood  $\mathbb{B}(\bar{x}, a) \times \mathbb{B}(0, b)$  of  $(\bar{x}, 0)$  for some  $a > 0$  and  $b > 0$ . Let sequences of positive scalars  $(\varepsilon_k)_{k \in \mathbb{N}_0}$  and  $(\mu_k)_{k \in \mathbb{N}_0}$  be such that*

$$L := \sup_k \mu_k < \kappa^{-1} \quad \text{and} \quad s := \sup_k \frac{\kappa \varepsilon_k}{1 - \kappa \mu_k} < 1.$$

*Then there exists a neighborhood  $U$  of  $\bar{x}$  such that for any sequence of multifunctions  $A_k : X \times X \rightrightarrows Y$  with closed graph satisfying the properties that*

(i)  $h(A_k(x, u) - f(x), A_k(x', u) - f(x')) \leq \mu_k \|x - x'\|$  whenever  $x, x', u \in \mathbb{B}(\bar{x}, a)$  and  $k \in \mathbb{N}_0$ ; and

(ii)  $d(f(\bar{x}), A_k(\bar{x}, u)) \leq \varepsilon_k \|u - \bar{x}\|$  for all  $u \in \mathbb{B}(\bar{x}, a)$  and all  $k \in \mathbb{N}_0$ ,

*and for any starting point  $x_0 \in U$ , there exists a sequence  $(x_k)_{k \in \mathbb{N}}$  generated by (1.2), and this sequence converges at least linearly to  $\bar{x}$ . Suppose, in addition, that  $x_k \neq \bar{x}$  for  $k$  large enough and that*

$$(4.1) \quad \lim_{k \rightarrow +\infty} \frac{d(f(\bar{x}), A_k(\bar{x}, x_k))}{\|x_k - \bar{x}\|} = 0;$$

*then  $(x_k)_{k \in \mathbb{N}}$  converges superlinearly to  $\bar{x}$ .*

*Proof.* Find  $\beta > 0$  such that (3.13) in Theorem 3.4 is valid for  $\Phi := f + F$  and  $\delta := a$ . Then take  $r > 0$  and  $\gamma \in (0, 1)$  such that

$$(4.2) \quad r(1 + \gamma) \sup_k \varepsilon_k < \beta \quad \text{and} \quad s + \gamma < 1.$$

For each  $k \in \mathbb{N}_0$ , set  $\kappa_k = \kappa / (1 - \mu_k \kappa)$  and then pick any  $\delta_k \in (0, \gamma / (1 + \varepsilon_k \kappa_k))$ .

Now, set  $U = \mathbb{B}(\bar{x}, r)$ , and let  $x_0$  be an arbitrary point in  $U$ . Assume that  $x_k \in U$  for some index  $k \in \mathbb{N}_0$  we will find  $x_{k+1} \in U$ , which verifies (1.2). If either  $x_k = \bar{x}$  or  $f(\bar{x}) \in A_k(\bar{x}, x_k)$ , then  $x_{k+1} := \bar{x}$  verifies (1.2) (in the first case according to (ii) because the set  $A_k(\bar{x}, \bar{x})$  is closed, and so it contains  $f(\bar{x})$ ). Now, suppose both that  $x_k \neq \bar{x}$  and  $f(\bar{x}) \notin A_k(\bar{x}, x_k)$ . Note that  $\kappa_k > 0$  and  $\delta_k < \gamma < 1$ . By (ii) and (4.2), there exists  $y_k \in A(\bar{x}, x_k)$  such that

$$\|y_k - f(\bar{x})\| < (1 + \delta_k) d(f(\bar{x}), A(\bar{x}, x_k)) \leq \varepsilon_k (1 + \delta_k) \|\bar{x} - x_k\| < \varepsilon_k (1 + \gamma) r < \beta.$$

Let us apply Theorem 3.4 with  $\bar{y} := 0$ ,  $\delta = a$ ,  $\Phi = f + F$ , and  $G := A_k(\cdot, x_k) - f$ . By assumption, the condition (3.14) is satisfied for  $\bar{z} := y_k - f(\bar{x}) \in G(\bar{x})$ , and  $L$  replaced by  $\mu_k$ . Note that (3.13) with  $\tau := \kappa_k$  is also valid, since  $\kappa / (1 - \kappa \mu_k) \leq \kappa / (1 - \kappa L)$ . Hence, (3.15) (with  $\tau$  replaced by  $\kappa_k$ ) holds for the multifunction

$$(\Phi + G)(x) = A_k(x, x_k) + F(x), \quad x \in X.$$

Note that  $\kappa_k \varepsilon_k \leq s$  and that  $y := 0 \in \mathbb{B}(\bar{z}, \beta)$ . As  $d(f(\bar{x}), A_k(\bar{x}, x_k)) > 0$ , and  $-f(\bar{x}) \in F(\bar{x})$ , relation (3.15) implies that

$$\begin{aligned} d(\bar{x}, (A_k(\cdot, x_k) + F(\cdot))^{-1}(0)) &\leq \kappa_k d(0, f(\bar{x}) + F(\bar{x}) + (y_k - f(\bar{x}))) \leq \kappa_k \|y_k - f(\bar{x})\| \\ &< \kappa_k (1 + \delta_k) d(f(\bar{x}), A(\bar{x}, x_k)) \leq \kappa_k \varepsilon_k (1 + \delta_k) \|x_k - \bar{x}\| \\ &< (\kappa_k \varepsilon_k + \delta_k (1 + \kappa_k \varepsilon_k)) \|x_k - \bar{x}\| < (s + \gamma) \|x_k - \bar{x}\| \\ &< r. \end{aligned}$$

Therefore, there exists  $x_{k+1} \in \mathbb{B}(\bar{x}, r) = U$  with  $0 \in A_k(x_{k+1}, x_k) + F(x_{k+1})$  such that

$$(4.3) \quad \|x_{k+1} - \bar{x}\| < \kappa_k(1 + \delta_k) d(f(\bar{x}), A(\bar{x}, x_k)) < (s + \gamma) \|x_k - \bar{x}\|.$$

We defined inductively the sequence  $(x_k)_{k \in \mathbb{N}}$  in  $U$  which verifies (4.3) for each  $k \in \mathbb{N}$ . Since  $s + \gamma < 1$ , this sequence converges linearly.

If  $x_k \neq \bar{x}$  for all  $k \in \mathbb{N}$ , then (4.3) implies that

$$0 \leq \frac{\|x_{k+1} - \bar{x}\|}{\|x_k - \bar{x}\|} < \kappa_k(1 + \delta_k) \frac{d(f(\bar{x}), A_k(\bar{x}, x_k))}{\|x_k - \bar{x}\|} \quad \text{whenever } k \in \mathbb{N}.$$

So (4.1) implies the superlinear convergence of  $(x_k)_{k \in \mathbb{N}}$ .  $\square$

*Remark 4.2.* Suppose that  $f$  is continuously differentiable at  $\bar{x}$ . In [14], the authors consider an inexact iterative process

$$(f(x_k) + f'(x_k)(x_{k+1} - x_k) + F(x_{k+1})) \cap R_k(x_k, x_{k+1}) \neq \emptyset, \quad k \in \mathbb{N}_0,$$

where  $R_k : X \times X \rightrightarrows Y$  is a sequence of set-valued mappings with closed graphs. Under the assumption that, for each  $k \in \mathbb{N}_0$ , the mapping  $(u, x) \mapsto R_k(u, x)$  is partially Aubin continuous with respect to  $x$  at  $\bar{x}$  for 0 uniformly in  $u$  around  $\bar{x}$  with a constant  $\mu > 0$ , i.e., there are  $\alpha > 0$  and  $\beta > 0$  such that

$$e(R_k(u, x_1) \cap \mathbb{B}(0, \beta), R_k(u, x_2)) \leq \mu \|x_1 - x_2\| \quad \text{for all } u, x_1, x_2 \in \mathbb{B}(\bar{x}, \alpha),$$

it is possible to derive a result similar to [14, Theorem 3.1] ensuring the existence and (super)linear convergence of some sequence satisfying the above scheme. Indeed, it suffices to consider

$$A_k(x, u) := f(u) + f'(u)(x - u) - R_k(u, x), \quad (x, u) \in X \times X, \quad k \in \mathbb{N}_0.$$

However, whenever  $0 \in R_k(u, x)$  for any  $(u, x)$  near  $(\bar{x}, \bar{x})$  and any  $k \in \mathbb{N}_0$ , this result would trivially follow from the one on the exact Newton method. But the only example given in [14] is  $R_k(x_k, x_{k+1}) := \mathbb{B}[0, \eta_k \|f(x_k)\|]$ ,  $k \in \mathbb{N}_0$ , with  $(\eta_k)_{k \in \mathbb{N}}$  being a given sequence of positive scalars.

Now, we are going to explore the iterative schemes when  $f$  is Lipschitz continuous in the vicinity of the reference point. In the previous statement, we assumed that the mapping  $f + F$  is metrically regular at the reference point. For a continuously differentiable mapping  $f$ , this is equivalent to the metric regularity of its “partial linearization”  $f(\bar{x}) + f'(\bar{x})(\cdot - \bar{x}) + F$ . However, for a nonsmooth  $f$  it can be too strong (see Remark 4.5(ii)). For  $F \equiv 0$ , the relationship between the metric regularity of all “partial linearizations” and the metric regularity of the original mapping  $f$  is discussed in Remark 4.4(ii). First, let us recall several notions.

Ioffe [21] initiated the use of the strict prederivatives to approximate a nonsmooth mapping  $g : X \rightarrow Y$  around a reference point. A positively homogeneous set-valued mapping  $G : X \rightrightarrows Y$  (i.e.,  $\text{gph } G$  is a cone in  $X \times Y$ ) is called the *strict prederivative* of  $g$  at  $x_0 \in X$  if for each  $c > 0$  there exists  $\delta > 0$  such that

$$(4.4) \quad g(x_1) \in g(x_2) + G(x_1 - x_2) + c \|x_1 - x_2\| \mathbb{B}_Y \quad \text{whenever } x_1, x_2 \in \mathbb{B}(x_0, \delta).$$

The strict prederivative is often generated by a family of continuous linear operators in such a way that there is a subset  $\mathcal{S}$  of  $\mathcal{L}(X, Y)$  such that  $G(x) = \{S(x) : S \in \mathcal{S}\}$  for each  $x \in X$ .

Let  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$  be locally Lipschitz continuous; i.e., for each  $x \in \mathbb{R}^n$  there are constants  $L_x > 0$  and  $\delta_x > 0$  such that  $\|g(u) - g(v)\| \leq L_x \|u - v\|$  whenever  $u, v \in \mathbb{B}(x, \delta_x)$ . For such a function  $g$  denote by  $E_g$  the set of points  $x \in \mathbb{R}^n$  at which the derivative  $g'(x)$  exists. By the Rademacher's theorem, the set  $\mathbb{R}^n \setminus E_g$  has Lebesgue measure zero. The  $B$ -subdifferential of  $g$  at  $u \in \mathbb{R}^n$ , denoted by  $\partial_B g(u)$ , is the set

$$\left\{ M \in \mathbb{R}^{m \times n} : M = \lim_{k \rightarrow +\infty} g'(u_k) \text{ for some } (u_k)_{k \in \mathbb{N}} \text{ in } E_g \text{ converging to } u \right\}.$$

Then  $\partial g(u) := \text{co } \partial_B g(u)$  denotes the *Clarke generalized Jacobian* of  $g$  at  $u$  (see also [11]). It is well known that  $\partial g(u)$  is a strict prederivative of  $g$  at  $u$  and  $x \rightrightarrows \partial g(x)$  has convex and compact values (see [21, Corollary 9.11]). On the other hand,  $x \rightrightarrows \partial_B g(x)$  is only compact-valued in general. However, both the mappings are upper semicontinuous.

*Assumption (A1).* In addition to the standing assumptions, suppose that there is a set-valued mapping  $\mathcal{H} : X \rightrightarrows \mathcal{L}(X, Y)$  such that

(A1.1)  $\mathcal{H}(\bar{x})$  is compact,

(A1.2) the mapping  $\mathcal{H}$  is upper semicontinuous at  $\bar{x} \in \text{int}(\text{dom } \mathcal{H})$ ,

(A1.3) there is a selection  $\psi : X \rightarrow \mathcal{L}(X, Y)$  for  $\mathcal{H}$  such that

$$\lim_{\bar{x} \neq x \rightarrow \bar{x}} \frac{\|f(x) - f(\bar{x}) - \psi(x)(x - \bar{x})\|}{\|x - \bar{x}\|} = 0.$$

Recall that a function  $g : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is called *semismooth* (see [17, section 7.4]) at  $u \in \mathbb{R}^n$  if it is locally Lipschitz continuous around  $u$ , is directionally differentiable near  $u$ , and satisfies the condition

$$(4.5) \quad \lim_{0 \neq v \rightarrow 0} \frac{\sup_{M \in \partial g(u+v)} \|g(u+v) - g(u) - Mv\|}{\|v\|} = 0.$$

Hence, if  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is semismooth at  $\bar{x}$ , then both  $\partial f$  and  $\partial_B f$  satisfy Assumption (A1).

**THEOREM 4.3.** *In addition to Assumption (A1), suppose that for each  $\bar{H} \in \mathcal{H}(\bar{x})$  the mapping*

$$\Phi_{\bar{H}}(x) := f(\bar{x}) + \bar{H}(x - \bar{x}) + F(x), \quad x \in X,$$

*is metrically regular at  $(\bar{x}, 0)$ . Then there exists a neighborhood  $U$  of  $\bar{x}$  such that for any starting point  $x_0 \in U$  there is a sequence  $(x_k)_{k \in \mathbb{N}}$  generated by*

$$(4.6) \quad 0 \in f(x_k) + \mathcal{H}(x_k)(x_{k+1} - x_k) + F(x_{k+1}) \quad \text{for each } k \in \mathbb{N}_0,$$

*and this sequence converges superlinearly.*

*Proof.* First, we claim that there are  $\kappa > 0$ ,  $a > 0$ , and  $b > 0$  such that for each  $\bar{H} \in \mathcal{H}(\bar{x})$  the mapping  $\Phi_{\bar{H}}$  is metrically regular at  $(\bar{x}, 0)$  with the constant  $\kappa$  on  $\mathbb{B}(\bar{x}, a) \times \mathbb{B}(0, b)$ . To see this, we will imitate the steps in the proof of Proposition 2.H3 in [13] which is adopted from [23]. Fix any  $\bar{H} \in \mathcal{H}(\bar{x})$ . Find  $\tau > 0$ ,  $\alpha > 0$ , and  $\beta > 0$  such that  $\Phi_{\bar{H}}$  is metrically regular at  $(\bar{x}, 0)$  with the constant  $\tau$  on  $\mathbb{B}(\bar{x}, \alpha) \times \mathbb{B}(0, \beta)$ . Fix any  $\varepsilon \in (0, 1/\tau)$  and any  $\bar{H}' \in \mathcal{H}(\bar{x})$  with  $\|\bar{H}' - \bar{H}\| < \varepsilon$ . Set  $G(x) := (\bar{H}' - \bar{H})(x - \bar{x})$ ,  $x \in X$ . Then  $\Phi_{\bar{H}'} = \Phi_{\bar{H}} + G$ . Moreover,  $G$  is single-valued and Lipschitz continuous with the constant  $\varepsilon$ , and  $G(\bar{x}) = 0$  (in particular  $\text{diam } G(\bar{x}) = 0$ ). Theorem 3.2 says that there are  $\alpha' > 0$  and  $\beta' > 0$  (independent of  $\bar{H}'$ ) such that  $\Phi_{\bar{H}'}$  is metrically

regular at  $(\bar{x}, 0)$  with the constant  $\tau/(1 - \tau\varepsilon)$  on  $\mathbb{B}(\bar{x}, \alpha') \times \mathbb{B}(0, \beta')$ . We showed that for any  $\bar{H} \in \mathcal{H}(\bar{x})$  there are  $\varepsilon_{\bar{H}} > 0$ ,  $\tau_{\bar{H}} > 0$ ,  $\alpha_{\bar{H}} > 0$ , and  $\beta_{\bar{H}} > 0$  such that  $\Phi_{\bar{H}'}$  is metrically regular at  $(\bar{x}, 0)$  with the constant  $\tau_{\bar{H}}$  on  $\mathbb{B}(\bar{x}, \alpha_{\bar{H}}) \times \mathbb{B}(0, \beta_{\bar{H}})$  for any  $\bar{H}' \in \mathcal{H}(\bar{x})$  with  $\|\bar{H}' - \bar{H}\| < \varepsilon_{\bar{H}}$ . By the compactness, from the open covering  $\cup_{\bar{H} \in \mathcal{H}(\bar{x})} \mathbb{B}(\bar{H}, \varepsilon_{\bar{H}})$  of  $\mathcal{H}(\bar{x})$ , we can choose a finite subcovering; i.e., there are  $k \in \mathbb{N}$ ,  $\bar{H}_i \in \mathcal{H}(\bar{x})$ , and  $\varepsilon_{\bar{H}_i} > 0$  such that  $\mathcal{H}(\bar{x})$  is covered with  $\mathbb{B}(\bar{H}_i, \varepsilon_{\bar{H}_i})$ ,  $i = 1, 2, \dots, k$ . Setting  $\kappa = \max_i \tau_{\bar{H}_i}$ ,  $a = \min_i \alpha_{\bar{H}_i}$ , and  $b = \min_i \beta_{\bar{H}_i}$ , one gets the claim.

Fix  $L \in (0, 1/\kappa)$ . By (A1.2), there is  $\delta > 0$  such that

$$(4.7) \quad \mathcal{H}(y) \subset \mathcal{H}(\bar{x}) + L\mathbb{B}_{\mathcal{L}(X,Y)} \quad \text{whenever} \quad y \in \mathbb{B}(\bar{x}, \delta) \subset \text{dom } \mathcal{H}.$$

Set  $\eta = 0$ . Find  $\alpha > 0$  and  $\beta > 0$  such that (3.4) is satisfied and also such that

$$\frac{2\kappa\beta}{1 - \kappa L} < 1.$$

Finally, use (A1.3) to find  $r \in (0, \min\{\delta, 1\})$  such that

$$\|f(y) - f(\bar{x}) - \psi(y)(y - \bar{x})\| \leq \beta\|y - \bar{x}\| \quad \text{whenever} \quad y \in \mathbb{B}(\bar{x}, r).$$

Set  $U = \mathbb{B}(\bar{x}, r)$ . Let  $x_0$  be an arbitrary point in  $U$ . Assume that  $x_k \in U$  for some index  $k \in \mathbb{N}_0$ . We will find  $x_{k+1} \in U$  which verifies (4.6). Set

$$y_k = f(x_k) - f(\bar{x}) - \psi(x_k)(x_k - \bar{x}).$$

If either  $x_k = \bar{x}$  or  $-y_k \in f(\bar{x}) + F(\bar{x})$ , then  $x_{k+1} := \bar{x}$  does the job. From now on, suppose that  $x_k \neq \bar{x}$  and that  $d(-y_k, f(\bar{x}) + F(\bar{x})) > 0$ . Since  $\psi$  is a selection for  $\mathcal{H}$ , it suffices to find  $x_{k+1} \in U$  solving

$$(4.8) \quad 0 \in f(x_k) + \psi(x_k)(x - x_k) + F(x).$$

By (4.7), there is  $\tilde{H} \in \mathcal{H}(\bar{x})$  and  $\tilde{H} \in \mathbb{B}_{\mathcal{L}(X,Y)}$  such that  $\psi(x_k) = \tilde{H} + L\tilde{H}$ . Set

$$G(u) := L\tilde{H}(u - \bar{x}) + y_k, \quad u \in X.$$

Clearly,  $G$  is single-valued and Lipschitz continuous with the constant  $L$ , and  $G(\bar{x}) = y_k$ . Moreover, given  $x \in X$ , we have that  $f(x_k) + \psi(x_k)(x - x_k) + F(x)$  is equal to

$$\begin{aligned} & f(x_k) - f(\bar{x}) - \psi(x_k)(x_k - \bar{x}) + f(\bar{x}) + \psi(x_k)(x - \bar{x}) + F(x) \\ &= y_k + f(\bar{x}) + (\tilde{H} + L\tilde{H})(x - \bar{x}) + F(x) = y_k + \Phi_{\tilde{H}}(x) + L\tilde{H}(x - \bar{x}) = (\Phi_{\tilde{H}} + G)(x). \end{aligned}$$

Theorem 3.2 implies that  $\Phi_{\tilde{H}} + G$  is metrically regular with the constant  $\kappa/(1 - \kappa L)$  on  $\mathbb{B}(\bar{x}, \alpha) \times \mathbb{B}(y_k, \beta)$ . Note that  $0 \in \mathbb{B}(y_k, \beta)$  because

$$\|y_k\| = \|f(x_k) - f(\bar{x}) - \psi(x_k)(x_k - \bar{x})\| \leq \beta\|x_k - \bar{x}\| < \beta r < \beta.$$

As  $0 \in f(\bar{x}) + F(\bar{x}) = \Phi_{\tilde{H}}(\bar{x})$ , we have  $y_k \in (\Phi_{\tilde{H}} + G)(\bar{x})$ , and therefore the above inequality implies that

$$\begin{aligned} d(\bar{x}, (\Phi_{\tilde{H}} + G)^{-1}(0)) &\leq \frac{\kappa}{1 - \kappa L} d(0, (\Phi_{\tilde{H}} + G)(\bar{x})) \leq \frac{\kappa}{1 - \kappa L} \|y_k\| \\ &< \frac{2\kappa}{1 - \kappa L} \|y_k\| \leq \frac{2\kappa\beta}{1 - \kappa L} \|x_k - \bar{x}\| \quad (< r). \end{aligned}$$

Therefore, there exists  $x_{k+1} \in \mathbb{B}(\bar{x}, r) = U$  with  $0 \in \Phi_{\bar{H}}(x_{k+1}) + G(x_{k+1})$  such that

$$(4.9) \quad \|x_{k+1} - \bar{x}\| < \frac{2\kappa}{1 - \kappa L} \|y_k\| \leq \frac{2\kappa\beta}{1 - \kappa L} \|x_k - \bar{x}\|.$$

We defined inductively the sequence  $(x_k)_{k \in \mathbb{N}}$  in  $U$  which verifies (4.8) for each  $k \in \mathbb{N}_0$ . By (4.9), it converges (linearly) to  $\bar{x}$ . If there is  $k_0 \in \mathbb{N}$  such that  $x_k \neq \bar{x}$  for each  $k > k_0$ , (A1.3) implies that

$$\lim_{k \rightarrow +\infty} \frac{\|y_k\|}{\|x_k - \bar{x}\|} = 0;$$

hence taking into account (4.9), we get the superlinear convergence of  $(x_k)_{k \in \mathbb{N}}$ .  $\square$

*Remark 4.4.*

(i) In the proof of the preceding result, twice we used Theorem 3.2 with a *single-valued* perturbation  $G$ . This theorem yields not only the stability of the metric regularity under a Lipschitz single-valued perturbation, which is well known, but also a uniform estimate on the sizes of the corresponding neighborhoods. Instead of this statement, one can use [13, Theorem 5G.3].

(ii) In addition to the standing assumptions, suppose that  $F \equiv 0$  and that  $f$  has the strict prederivative at  $\bar{x} \in X$  generated by a compact convex subset  $\mathcal{T}$  of  $\mathcal{L}(X, Y)$ . The *modulus of (linear) openness of  $\mathcal{T}$* , denoted by  $\sigma(\mathcal{T})$ , is defined by

$$\sigma(\mathcal{T}) = \inf \{ \sigma(T) : T \in \mathcal{T} \},$$

where  $\sigma(T) := \sup \{ r > 0 : \mathbb{B}[0, r] \subset T(\mathbb{B}_X) \}$  is the *modulus of (linear) openness of  $T$* . Páles [34] used this quantity to provide a condition guaranteeing the metric regularity at the reference point. If  $\sigma(\mathcal{T}) > 0$ , then [34, Theorem 2] implies that  $f$  is metrically regular at  $\bar{x}$  with a constant arbitrarily close to  $1/\sigma(\mathcal{T})$ . If, in addition, all the elements of  $\mathcal{T}$  are injective, then  $f$  is strongly metrically regular at  $\bar{x}$  by [34, Theorem 5].

If  $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$  is semismooth at  $\bar{x}$ , then a sufficient condition for the superlinear convergence of the nonsmooth Newton methods defined, for each  $k \in \mathbb{N}_0$ , by either

$$0 \in f(x_k) + \partial f(x_k)(x_{k+1} - x_k) \quad \text{or} \quad 0 \in f(x_k) + \partial_B f(x_k)(x_{k+1} - x_k),$$

is that all the matrices in  $\partial f(\bar{x})$ , respectively in  $\partial_B f(\bar{x})$ , have full rank. So we recover the results in [38, 41, 19] (note that the inexact algorithms therein can be treated similarly). In the latter two references, one can find also other suitable candidates for  $\mathcal{H}$ . From a numerical point of view, when this approximation is larger, one can compute an element of  $\mathcal{H}(x_k)$  more easily. However, this is contradictory to the assumption on surjectivity (nonsingularity) of all elements of  $\mathcal{H}(\bar{x})$ .

*Remark 4.5.*

(i) Formally, (A1.3) is less restrictive than the assumption that  $f$  is semismooth at the reference point. However, in practice, it is neither possible nor efficient to choose the “right” matrix in the corresponding subdifferential at each step (we just take some matrices in the subdifferential). Inspecting the proof of Theorem 4.3, we see that if (A1.3) is replaced by

$$\lim_{\bar{x} \neq x \rightarrow \bar{x}} \frac{\sup_{H \in \mathcal{H}(x)} \|f(x) - f(\bar{x}) - H(x - \bar{x})\|}{\|x - \bar{x}\|} = 0,$$

then one can choose any element  $H_k \in \mathcal{H}(x_k)$  at each step of (4.6). Indeed, it suffices to change  $\psi(x_k)$  to  $H_k$ .

(ii) The assumption on metric regularity of all  $\Phi_{\bar{H}}$  is less restrictive than to suppose that  $f + F$  is metrically regular (unless  $\mathcal{H}(\bar{x})$  is convex and generates the strict prederivative of  $f$  at  $\bar{x}$ ). Indeed, take  $f(x) := |x|$ ,  $x \in \mathbb{R}$ , and  $\mathcal{H} := \partial_B f$ . In [24], the authors supposed that all  $\Phi_{\bar{H}}$  are strongly metrically regular at the reference point which guarantees that the next iterate  $x_{k+1}$  in (4.6) is unique for each  $k \in \mathbb{N}_0$ . In [30], one can find an outstanding discussion on the attempts to generalize the Newton's method for continuous (not locally Lipschitz) functions. One can also use some generalized derivative to approximate  $f$  as [31] shows.

*Example 4.1.* In order to illustrate the result of this section, let us consider the following example in one dimension. Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  and  $F : \mathbb{R} \rightrightarrows \mathbb{R}$  be defined by

$$f(x) = \begin{cases} \ln(1+x) & \text{if } x \geq 0, \\ \max(x^2, -x^2 - 2x) & \text{if } x \leq 0, \end{cases} \quad F(x) = \begin{cases} \emptyset & \text{if } x < -2, \\ (-\infty, -x(2x+1)] \cup [x, +\infty) & \text{if } x \in [-2, -1], \\ (-\infty, x(2x+3)] \cup [x, +\infty) & \text{if } x \in [-1, 0], \\ (-\infty, -3x] \cup [\sqrt{x}, +\infty) & \text{if } x \geq 0. \end{cases}$$

We note that  $F^{-1}(0) = [-2, 0]$  and that the generalized equation  $f(x) + F(x) \ni 0$  has two solutions:  $\bar{x} = -1$  and  $\bar{x} = 0$  (the graphs of  $-f$  and  $F$  are plotted in Figure 1). We observe that the function  $f$  is semismooth on  $\mathbb{R}$  and that  $f + F$  is metrically regular (and not strongly metrically regular) at  $(-1, 0)$  and  $(0, 0)$ . By taking  $\mathcal{H} = \partial_B f$ , it is easy to show that for each  $\bar{H} \in \mathcal{H}(\bar{x})$  the mapping  $\Phi_{\bar{H}}(x) := f(\bar{x}) + \bar{H}(x - \bar{x}) + F(x)$  is metrically regular at  $(\bar{x}, 0)$  for  $\bar{x} = -1$  and  $\bar{x} = 0$ . The assumptions of Theorem 4.3 are satisfied. Hence the semismooth Newton's method (4.6) generates a superlinearly convergent sequence when started in a neighborhood of each solution  $\bar{x} = 0$  or  $\bar{x} = -1$ . Figures 2 (a) and (c) are a plot of  $\ln(e_k)$  versus  $k$  (with  $e_k = |x_k - \bar{x}|$ ) for different choices of the initial point  $x_0$ , while Figures 2 (b) and (d) are a plot of  $\ln(e_{k+1})$  versus  $\ln(e_k)$ . In both cases, we observe that the methods converge superlinearly: an estimation of the order of convergence  $p$  from the slopes of Figures 2 (b) and (d) gives us  $p = 1.73$  and  $p = 1.64$  for the solutions 0 and  $-1$ , respectively. This confirms the theoretical results of Theorem 4.3. The numerical results are shown in Table 1 for different starting points.

TABLE 1  
Numerical results for Example 4.1.

Iterations	$x_0 = 1$		$x_0 = -2$	
	$x_k$	$\frac{e_{k+1}}{e_k}$	$x_k$	$\frac{e_{k+1}}{e_k}$
1	$1.2876 \times 10^{-1}$	$4.9131 \times 10^{-2}$	$-1.3333$	$2.0000 \times 10^{-1}$
2	$6.3264 \times 10^{-3}$	$3.1171 \times 10^{-3}$	$-1.0667$	$5.8824 \times 10^{-2}$
3	$1.9720 \times 10^{-5}$	$9.8596 \times 10^{-6}$	$-1.0039$	$3.8911 \times 10^{-3}$
4	$1.9443 \times 10^{-10}$	$2.2676 \times 10^{-7}$	$-1.0000$	$1.5259 \times 10^{-5}$
5	$4.4090 \times 10^{-17}$	0.0000	$-1.0000$	0.0000
6	0.0000	—	$-1$	—

**5. Forward-backward splitting.** In this section, we will use the following premise.

*Assumption (A2).* Let  $X$  be a Banach space,  $Y := X$ , and suppose that



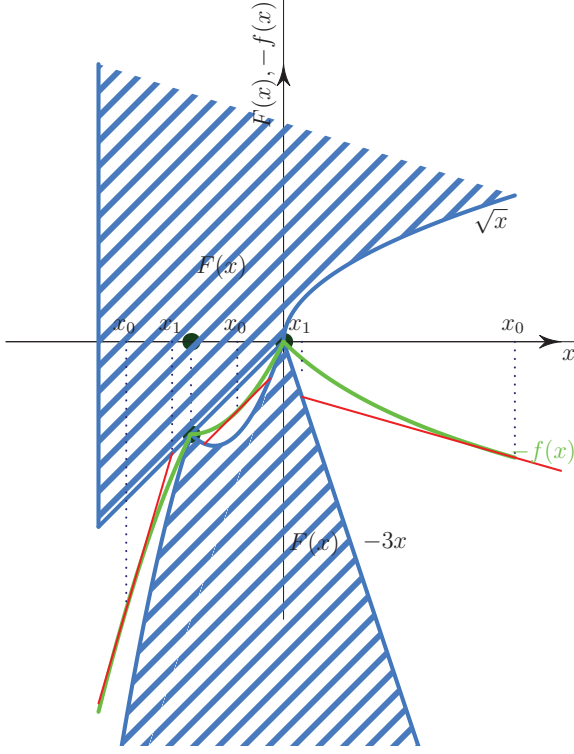
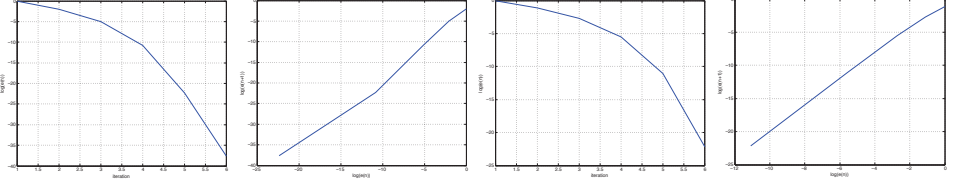


FIG. 1. The graph of  $-f$  and  $F$ .



(a)  $\ln e_k$  vs.  $k$  for 0. (b)  $\ln e_{k+1}$  vs.  $\log e_k$ . (c)  $\ln e_k$  vs.  $k$  for  $-1$ . (d)  $\ln e_{k+1}$  vs.  $\log e_k$ .

FIG. 2. Superlinear rate of convergence of the semismooth Newton's method (4.6) at 0 and  $-1$ .

- (A2.1) a point  $\bar{x} \in X$  is a solution to (1.1), and that  $\bar{y}_1 \in F(\bar{x})$  and  $\bar{y}_2 \in \Psi(\bar{x})$  are such that  $0 = \bar{y}_1 + \bar{y}_2$ ;
- (A2.2)  $F$  has closed graph and is metrically regular at  $(\bar{x}, \bar{y}_1)$  with a constant  $\kappa > 0$  on  $\mathbb{B}(\bar{x}, a) \times \mathbb{B}(\bar{y}_1, b)$  for some  $a > 0$  and  $b > 0$ ;
- (A2.3)  $\Psi$  has closed graph, and there is  $\varepsilon > 0$  such that

$$e(\Psi(\bar{x}), \Psi(u)) \leq \varepsilon \|u - \bar{x}\| \quad \text{whenever } u \in \mathbb{B}(\bar{x}, a).$$

**THEOREM 5.1.** *In addition to Assumption (A2), suppose that  $(\lambda_k)_{k \in \mathbb{N}_0}$  is a sequence in  $(0, 1)$  such that  $\varepsilon + 2 \sup_k \lambda_k < 1/\kappa$ . Then there exists a neighborhood  $U$  of  $\bar{x}$  such that for any starting point  $x_0 \in U$ , there is a sequence  $(x_k)_{k \in \mathbb{N}}$  generated by*

$$(5.1) \quad 0 \in \lambda_k(x_{k+1} - x_k) + F(x_{k+1}) + \Psi(x_k) \quad \text{for each } k \in \mathbb{N}_0,$$

and this sequence converges at least linearly to  $\bar{x}$ . If, in addition,

$$(5.2) \quad \lim_{k \rightarrow +\infty} \frac{d(\bar{y}_2, \lambda_k(\bar{x} - x_k) + \Psi(x_k))}{\|x_k - \bar{x}\|} = 0,$$

then  $(x_k)_{k \in \mathbb{N}}$  converges superlinearly.

*Proof.* For each  $k \in \mathbb{N}_0$ , set  $A_k(x, u) = \lambda_k(x - u) + \Psi(u)$ ,  $(x, u) \in X \times X$ . Let  $\delta := 2a$  and fix  $\lambda > \sup_k \lambda_k$  such that  $\varepsilon + 2\lambda < 1/\kappa$ . Find  $\beta > 0$  such that (3.13) in Theorem 3.4 for  $Y := X$ ,  $\Phi := F$ ,  $L := \lambda$  and  $\bar{y} := \bar{y}_1$  holds. Then take  $\gamma \in (0, 1)$  and  $r \in (0, a)$  such that

$$(5.3) \quad \varepsilon + 2\lambda + \gamma < 1/\kappa \quad \text{and} \quad r(\lambda + \varepsilon + \gamma) < \beta.$$

Fix any  $\xi \in (0, \gamma/(\lambda + \varepsilon))$ . Set  $U = \mathbb{B}(\bar{x}, r)$ . Let  $x_0$  be an arbitrary point in  $U$ . Assume that  $x_k \in U$  for some index  $k \in \mathbb{N}_0$ . We will find  $x_{k+1} \in U$  which verifies (5.1). If either  $x_k = \bar{x}$  or  $\bar{y}_2 \in A_k(\bar{x}, x_k)$ , then  $x_{k+1} := \bar{x}$  satisfies (5.1). From now on, suppose both that  $x_k \neq \bar{x}$  and  $\bar{y}_2 \notin A_k(\bar{x}, x_k)$ . Taking into account (5.3), there exists  $\bar{z} \in A_k(\bar{x}, x_k)$  such that

$$\begin{aligned} \|\bar{z} - \bar{y}_2\| &< (1 + \xi) d(\bar{y}_2, A_k(\bar{x}, x_k)) \leq (1 + \xi) e(\Psi(\bar{x}), \lambda_k(\bar{x} - x_k) + \Psi(x_k)) \\ &\leq (1 + \xi) (\lambda_k \|\bar{x} - x_k\| + e(\Psi(\bar{x}), \Psi(x_k))) < (1 + \xi)(\lambda + \varepsilon) \|\bar{x} - x_k\| \\ &< (\lambda + \varepsilon + \gamma) \|x_k - \bar{x}\| < \beta. \end{aligned}$$

Let  $\kappa_k := \kappa/(1 - \kappa\lambda_k)$ . Let us check the conditions of Theorem 3.4 with  $G := A_k(\cdot, x_k)$  and  $L := \lambda_k$ . Clearly, (3.13) is valid since  $\kappa_k < \kappa/(1 - \kappa\lambda)$ . Now, let  $u_1, u_2 \in \mathbb{B}(\bar{x}, \delta)$  be arbitrary. Then, by (2.1) we have that

$$h(G(u_1), G(u_2)) = h(\lambda_k(u_1 - x_k) + \Psi(x_k), \lambda_k(u_2 - x_k) + \Psi(x_k)) \leq \lambda_k \|u_1 - u_2\|.$$

Clearly, the above condition implies (3.14). Noting that  $\bar{y}_1 = -\bar{y}_2$ , we get that  $y := 0 \in \mathbb{B}(\bar{y}_1 + \bar{z}, \beta)$ . So the multifunction

$$(\Phi + G)(x) = A_k(x, x_k) + F(x), \quad x \in X,$$

satisfies (3.15) with  $\tau$  replaced by  $\kappa_k$ . Since  $d(\bar{y}_2, A_k(\bar{x}, x_k)) > 0$ , one gets that

$$\begin{aligned} d\left(\bar{x}, (A_k(\cdot, x_k) + F(\cdot))^{-1}(0)\right) &\leq \kappa_k d(0, F(\bar{x}) + \bar{z}) \leq \kappa_k \|\bar{y}_1 + \bar{z}\| \\ &< \frac{\kappa}{1 - \kappa\lambda} \|\bar{z} - \bar{y}_2\| \leq \frac{\kappa}{1 - \kappa\lambda} (1 + \xi) d(\bar{y}_2, A_k(\bar{x}, x_k)) \\ &< \frac{\kappa(\lambda + \varepsilon + \gamma)}{1 - \kappa\lambda} \|\bar{x} - x_k\|. \end{aligned}$$

Note that  $\nu := \kappa(\lambda + \varepsilon + \gamma)/(1 - \kappa\lambda) < 1$  by (5.3). Therefore, there exists  $x_{k+1} \in \mathbb{B}(\bar{x}, r) = U$  with  $0 \in A_k(x_{k+1}, x_k) + F(x_{k+1})$  such that

$$(5.4) \quad \|x_{k+1} - \bar{x}\| < \frac{\kappa}{1 - \kappa\lambda} (1 + \xi) d(\bar{y}_2, A_k(\bar{x}, x_k)) < \nu \|x_k - \bar{x}\|.$$

We defined inductively the sequence  $(x_k)_{k \in \mathbb{N}}$  in  $U$  which verifies (5.4) for each  $k \in \mathbb{N}$ , so it converges linearly. Clearly, (5.2) and (5.4) imply the superlinear convergence of  $(x_k)_{k \in \mathbb{N}}$ .  $\square$

*Remark 5.2.*

- (i) Inspecting the proof, one sees that if  $\lim_{k \rightarrow +\infty} \lambda_k = 0$  (which implies that  $\lambda$  can be chosen arbitrarily small), then the conclusion of the theorem remains true under the assumption that  $\varepsilon < 1/\kappa$ .
- (ii) Let  $(x_k)_{k \in \mathbb{N}_0}$  be a sequence generated by (5.1) which converges to  $\bar{x}$ . Then (5.2) holds true under the assumption that  $\lim_{k \rightarrow +\infty} \lambda_k = 0$  and that for each  $\varepsilon > 0$  there is  $r > 0$  such that

$$e(\Psi(\bar{x}), \Psi(u)) \leq \varepsilon \|u - \bar{x}\| \quad \text{whenever} \quad u \in \mathbb{B}(\bar{x}, r).$$

Indeed, let  $\gamma > 0$  be arbitrary. Find a neighborhood  $V_\gamma$  of  $\bar{x}$  in  $X$  such that

$$e(\Psi(\bar{x}), \Psi(u)) \leq \frac{\gamma}{2} \|u - \bar{x}\| \quad \text{whenever} \quad u \in V_\gamma.$$

Then there is  $k_0 \in \mathbb{N}$  such that  $\lambda_k < \gamma/2$  whenever  $k > k_0$ . Taking a larger  $k_0$ , if necessary, we may assume that  $x_k \in V_\gamma$  for each  $k > k_0$ . For any such point  $x_k$  we have that

$$\begin{aligned} d(\bar{y}_2, \lambda_k(\bar{x} - x_k) + \Psi(x_k)) &\leq e(\Psi(\bar{x}), \lambda_k(\bar{x} - x_k) + \Psi(x_k)) \\ &\leq (\lambda_k + \gamma/2) \|x_k - \bar{x}\| < \gamma \|x_k - \bar{x}\|. \end{aligned}$$

- (iii) Suppose that  $F$  and  $\Psi$  act between two different Banach spaces  $X$  and  $Y$ , and let  $g_k : X \rightarrow Y$  be a given sequence of functions with  $g_k(0) = 0$  for each  $k \in \mathbb{N}_0$ . Assume that there is a neighborhood  $W$  of 0 in  $X$  such that all the functions  $g_k$  are Lipschitz continuous on  $W$  with a constant  $\lambda_k > 0$ . Then the conclusion of the previous theorem remains true when one considers the iterative process

$$0 \in g_k(x_{k+1} - x_k) + F(x_{k+1}) + \Psi(x_k) \quad \text{for each} \quad k \in \mathbb{N}_0.$$

To conclude this section, let us discuss the relationship between the above theorem and the results in [36]. Recall that the concept of the strict prederivative can be generalized to the case of a set-valued mapping  $S : X \rightrightarrows Y$ . Namely, given  $x_0 \in X$ , one requests the existence of a positively homogeneous mapping  $H : X \rightrightarrows Y$  such that for each  $c > 0$  there exists  $\delta > 0$  such that

$$(5.5) \quad S(x_1) \subset S(x_2) + H(x_1 - x_2) + c\|x_1 - x_2\|\mathbb{B}_Y \quad \text{whenever} \quad x_1, x_2 \in \mathbb{B}(x_0, \delta).$$

Such a mapping  $H$  is then called the *strict prederivative* of  $S$  at  $x_0$ . Using terminology in [35], the mapping  $S$  is called *strictly  $H$ -differentiable* at  $x_0$ . The author also considered a weaker version of (5.5). Namely, given a positively homogeneous mapping  $H : X \rightrightarrows Y$ , we say that  $S$  is *pseudostrictly  $H$ -differentiable* at  $(x_0, y_0) \in \text{gph } S$  if for each  $c > 0$  there are  $\delta > 0$  and  $r > 0$  such that

$$S(x_1) \cap \mathbb{B}(y_0, r) \subset S(x_2) + H(x_1 - x_2) + c\|x_1 - x_2\|\mathbb{B}_Y \quad \text{whenever} \quad x_1, x_2 \in \mathbb{B}(x_0, \delta).$$

Clearly, if  $S$  is pseudostrictly  $H$ -differentiable at  $(x_0, y_0)$  and if there is  $\kappa > 0$  such that  $H(x) \subset \kappa\|x\|\mathbb{B}_Y$  for each  $x \in X$ , then for each  $l > \kappa$  there are neighborhoods  $V$  of  $x_0$  and  $W$  of  $y_0$  such that

$$(5.6) \quad S(x_1) \cap W \subset S(x_2) + l\|x_1 - x_2\|\mathbb{B}_Y \quad \text{whenever} \quad x_1, x_2 \in V.$$

A mapping  $S$  satisfying the above inclusion for some  $l > 0$ ,  $V$ , and  $W$  is called *pseudo-Lipschitz* (*Lipschitz-like* or *Aubin continuous*) at  $(x_0, y_0)$ . It is well known that  $S$  is pseudo-Lipschitz at  $(x_0, y_0)$  if and only if  $S^{-1}$  is metrically regular at  $(y_0, x_0)$ .

Our result implies [36, Corollarie 4.4.2]. It was assumed that  $\Psi(\bar{x}) = \{\bar{y}_2\}$  and that  $\Psi$  is outer Lipschitz continuous (calm) at  $\bar{x}$ ; i.e., there is a constant  $\varepsilon > 0$  along with a neighborhood  $U$  of  $\bar{x}$  in  $X$  such that  $U \subset \text{dom} \Psi$  and that

$$e(\Psi(x), \Psi(\bar{x})) \leq \varepsilon \|x - \bar{x}\| \quad \text{whenever } x \in U.$$

Indeed, fix any  $x \in U$ . Let  $\gamma > 0$  be arbitrary. If  $\Psi$  satisfies both the above properties, then

$$\Psi(x) \subset \Psi(\bar{x}) + (\varepsilon + \gamma)\|x - \bar{x}\|\mathbb{B}_X = \mathbb{B}[\bar{y}_2, (\varepsilon + \gamma)\|x - \bar{x}\|].$$

So,  $e(\Psi(\bar{x}), \Psi(x)) = d(\bar{y}_2, \Psi(x)) \leq (\varepsilon + \gamma)\|x - \bar{x}\|$ . Therefore the continuity assumption in (A2.3) is satisfied. Further, the assumptions that  $F^{-1}$  is pseudostrictly  $H$ -differentiable at  $(\bar{y}_1, \bar{x})$  and that there is  $\kappa > 0$  such that  $H(y) \subset \kappa\|y\|\mathbb{B}_X$  for each  $y \in X$  imply that  $F^{-1}$  is pseudo-Lipschitz at  $(\bar{y}_1, \bar{x})$ , and so  $F$  is metrically regular at  $(\bar{x}, \bar{y}_1)$ . Hence Theorem 5.1 implies also [36, Théorème 4.4.1].

**6. Global convergence.** Under the standing assumptions, we investigate the global convergence of the iterative scheme (1.2) in this section.

**DEFINITION 6.1.** *Given a set-valued mapping  $\Phi : X \rightrightarrows Y$ ,  $x_0 \in X$ ,  $r > 0$ , and  $s > 0$ , let*

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

*We say that  $\Phi$  is metrically regular on  $V(\Phi, x_0, r, s)$  with a constant  $\kappa > 0$  if*

$$d(x, \Phi^{-1}(y)) \leq \kappa d(y, \Phi(x)) \quad \text{for all } (x, y) \in V(\Phi, x_0, r, s).$$

First, let us present a global version of Theorem 3.2 (see [22] for the stability in the case that  $r = +\infty$ ).

**THEOREM 6.2.** *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  $\Phi$  is metrically regular on  $V(\Phi, x_0, r, s)$  with the constant  $\kappa$ . For  $L \in (0, \kappa^{-1})$ , set  $\tau = \kappa/(1 - \kappa L)$ . Then for any set-valued mapping  $G : X \rightrightarrows Y$  with closed graph such that  $G$  is Lipschitz continuous on  $\mathbb{B}[x_0, r]$  with the constant  $L$  and  $\Phi + G$  has closed graph, the set-valued mapping  $\Phi + G$  is metrically regular on  $V(\Phi + G, x_0, \frac{r}{4}, R)$ , where  $R = \min\{s, \frac{r}{5\tau}\}$ , with the constant  $\tau > 0$ .*

*Proof.* Consider the function  $f : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$  defined by

$$f(x, y) = \liminf_{u \rightarrow x} d(y, \Phi(u) + G(u)), \quad (x, y) \in X \times Y.$$

Obviously,  $f(\cdot, y)$  is lower semicontinuous for each  $y \in Y$ . Since  $\Phi + G$  has closed graph,

$$(\Phi + G)^{-1}(y) = \{x \in X : f(x, y) = 0\}, \quad y \in Y.$$

Let  $(\bar{x}, y) \in V(\Phi + G, x_0, \frac{r}{4}, R)$  with  $f(\bar{x}, y) > 0$  be given. Then

$$f(\bar{x}, y) < R \leq \inf_{(x, z) \in X \times Y} f(x, z) + R.$$

Fix any  $\varepsilon > 0$  with  $R(\tau + \varepsilon) < r/4$  and then find  $\gamma > 0$  such that

$$\frac{1}{\kappa + \gamma} - L - \gamma > \frac{1}{\tau + \varepsilon} \quad \text{and} \quad \kappa + \gamma < \tau.$$

By the Ekeland variational principle [16, Theorem 1.1], we can select  $u \in X$  satisfying

$$\|u - \bar{x}\| \leq R(\tau + \varepsilon) < r/4 \quad \text{and} \quad f(u, y) \leq f(\bar{x}, y) < R$$

such that

$$(6.1) \quad f(x, y) + \frac{1}{\tau + \varepsilon} \|x - u\| \geq f(u, y) \quad \text{for all } x \in X.$$

Then  $\|u - x_0\| \leq \|x_0 - \bar{x}\| + R(\tau + \varepsilon) < r/2$ . We claim that  $y \in \Phi(u) + G(u)$ . Indeed, if this fails, find a sequence  $(u_n)_{n \in \mathbb{N}}$  in  $X$  converging to  $u$  and a sequence  $(w_n)_{n \in \mathbb{N}}$  such that

$$w_n \in G(u_n) \quad \text{for each } n \in \mathbb{N} \quad \text{and} \quad \lim_{n \rightarrow +\infty} d(y, \Phi(u_n) + w_n) = f(u, y) > 0.$$

As  $R \leq s$ , neglecting several starting terms, if necessary, we may assume that  $0 < d(y - w_n, \Phi(u_n)) < s$  and  $u_n \in \mathbb{B}(\bar{x}, r/4)$  for each  $n \in \mathbb{N}$ . By the metric regularity of  $\Phi$  on  $V(\Phi, x_0, r, s)$ , for each  $n \in \mathbb{N}$ , we can find  $v_n \in \Phi^{-1}(y - w_n)$  such that

$$\|u_n - v_n\| < (\kappa + \gamma) d(y - w_n, \Phi(u_n)).$$

Hence, without loss of generality, we may assume that  $v_n \in \mathbb{B}(\bar{x}, r/2)$  for each  $n \in \mathbb{N}$ . For any  $n \in \mathbb{N}$ , according to the Lipschitz property of  $G$  on  $\mathbb{B}[x_0, r]$ , there exists  $z_n \in G(v_n)$  such that

$$\|w_n - z_n\| \leq (L + \gamma) \|u_n - v_n\|.$$

Note that  $\liminf_{n \rightarrow +\infty} \|u - v_n\| > 0$ . Indeed, if there is an infinite set  $N \subset \mathbb{N}$  such that  $\lim_{N \ni n \rightarrow +\infty} \|u - v_n\| = 0$ , then  $\lim_{N \ni n \rightarrow +\infty} \|z_n - w_n\| = 0$ . Since

$$y - w_n + z_n \in \Phi(v_n) + z_n \subset \Phi(v_n) + G(v_n) \quad \text{whenever } n \in \mathbb{N},$$

the closeness of the graph of  $\Phi + G$  would imply that  $y \in \Phi(u) + G(u)$ , which is not the case. Taking into account (6.1), we have

$$\begin{aligned} \frac{1}{\tau + \varepsilon} &\geq \limsup_{n \rightarrow +\infty} \frac{f(u, y) - f(v_n, y)}{\|u - v_n\|} \geq \limsup_{n \rightarrow +\infty} \frac{d(y - w_n, \Phi(u_n)) - d(y - z_n, \Phi(v_n))}{\|u - v_n\|} \\ &\geq \limsup_{n \rightarrow +\infty} \frac{(\kappa + \gamma)^{-1} \|u_n - v_n\| - \|z_n - w_n\|}{\|u - v_n\|} \\ &\geq \limsup_{n \rightarrow +\infty} \frac{(\kappa + \gamma)^{-1} \|u_n - v_n\| - (L + \gamma) \|u_n - v_n\|}{\|u - v_n\|} = \frac{1}{\kappa + \gamma} - L - \gamma > \frac{1}{\tau + \varepsilon}, \end{aligned}$$

a contradiction. The claim is proved.

Consequently,

$$(6.2) \quad d(\bar{x}, (\Phi + G)^{-1}(y)) < r/4.$$

According to Theorem 3.1, it suffices to show that  $m(\bar{x}) \geq 1/\tau$ , where

$$m(\bar{x}) := \inf \left\{ \sup_{z \in X, z \neq x} \frac{f(x, y) - f(z, y)}{\|x - z\|} : \begin{array}{l} \|x - \bar{x}\| < d(\bar{x}, (\Phi + G)^{-1}(y)), \\ f(x, y) \leq f(\bar{x}, y) \end{array} \right\}.$$

Fix any  $x \in X$  with

$$\|x - \bar{x}\| < d(\bar{x}, (\Phi + G)^{-1}(y)) \quad \text{and} \quad f(x, y) \leq f(\bar{x}, y).$$

Then  $y \notin \Phi(x) + G(x)$ , and by (6.2), we have  $(x, y) \in V(\Phi + G, x_0, r/2, R)$ . Take a sequence  $(x_n)_{n \in \mathbb{N}}$  in  $\mathbb{B}[x_0, r]$  converging to  $x$  and a sequence  $(w_n)_{n \in \mathbb{N}}$  with  $w_n \in G(x_n)$  for each  $n \in \mathbb{N}$  such that

$$\lim_{n \rightarrow +\infty} d(y, \Phi(x_n) + w_n) = \lim_{n \rightarrow +\infty} d(y, \Phi(x_n) + G(x_n)) = f(x, y) > 0.$$

Omitting several starting terms, we may assume that  $0 < d(y - w_n, \Phi(x_n)) < R \leq s$  and, consequently, that  $(x_n, y - w_n) \in V(\Phi, x_0, r, s)$  for all  $n \in \mathbb{N}$ . Pick a sequence  $(\varepsilon_n)_{n \in \mathbb{N}}$  in  $(0, 1)$  converging to 0 such that  $(\kappa + \varepsilon_n) d(y - w_n, \Phi(x_n)) < R\kappa$ . By the metric regularity of  $\Phi$  on  $V(\Phi, x_0, r, s)$ , for each  $n \in \mathbb{N}$ , we can find  $u_n \in \Phi^{-1}(y - w_n)$  such that

$$\|x_n - u_n\| < (\kappa + \varepsilon_n) d(y - w_n, \Phi(x_n)) < R\kappa < r/2.$$

Therefore, we may assume that

$$\|x_0 - u_n\| \leq \|x_0 - x_n\| + \|x_n - u_n\| < r \quad \text{for each } n \in \mathbb{N}.$$

By the Lipschitz property of  $G$  on  $\mathbb{B}[x_0, r]$ , for each  $n \in \mathbb{N}$ , we can find  $z_n \in G(u_n)$  such that

$$\|z_n - w_n\| \leq (L + \varepsilon_n) \|u_n - x_n\|.$$

Note that  $\liminf_{n \rightarrow +\infty} \|x - u_n\| > 0$ . Indeed, if there is an infinite set  $N \subset \mathbb{N}$  such that  $\lim_{N \ni n \rightarrow +\infty} \|x - u_n\| = 0$ , then  $\lim_{N \ni n \rightarrow +\infty} \|z_n - w_n\| = 0$ . Since

$$y - w_n + z_n \in \Phi(u_n) + z_n \subset \Phi(u_n) + G(u_n) \quad \text{whenever } n \in \mathbb{N},$$

the closeness of the graph of  $\Phi + G$  implies that  $y \in \Phi(x) + G(x)$ , which is not the case.

Now, we may estimate that

$$\begin{aligned} \limsup_{n \rightarrow +\infty} \frac{f(x, y) - f(u_n, y)}{\|x - u_n\|} &\geq \limsup_{n \rightarrow +\infty} \frac{d(y - w_n, \Phi(x_n)) - d(y - z_n, \Phi(u_n))}{\|x - u_n\|} \\ &\geq \limsup_{n \rightarrow +\infty} \frac{(\kappa + \varepsilon_n)^{-1} \|x_n - u_n\| - \|z_n - w_n\|}{\|x - u_n\|} \\ &\geq \limsup_{n \rightarrow +\infty} \frac{(\kappa + \varepsilon_n)^{-1} \|x_n - u_n\| - (L + \varepsilon_n) \|x_n - u_n\|}{\|x - u_n\|} \\ &\geq \lim_{n \rightarrow +\infty} \left( \frac{1}{\kappa + \varepsilon_n} - L - \varepsilon_n \right) = \frac{1}{\tau}. \end{aligned}$$

Hence, we conclude that  $m(\bar{x}) \geq 1/\tau$ .  $\square$

**THEOREM 6.3.** *Given  $r > 0$ ,  $s > 0$ ,  $\kappa > 0$ , and  $x_0 \in X$ , suppose that  $f + F$  is metrically regular on  $V(f + F, x_0, 5r, s)$  with the constant  $\kappa$ . Let a sequence of positive scalars  $(\mu_k)_{k \in \mathbb{N}_0}$  be such that*

$$(6.3) \quad \sup_k \mu_k < \kappa^{-1} \quad \text{and} \quad \sup_k \frac{3\kappa\mu_k}{1 - \kappa\mu_{k+1}} < 1.$$

*Assume that a sequence of multifunctions  $A_k : X \times X \rightrightarrows Y$ , such that both  $A_k$  and  $A_k + F$  have closed graphs, satisfies the following conditions:*

- (i)  $A_k(x, x) = f(x)$  for all  $x \in \mathbb{B}[x_0, r]$  and all  $k \in \mathbb{N}_0$ ;
- (ii)  $h(A_k(x, u) - f(x), A_k(x', u) - f(x')) \leq \mu_k \|x - x'\|$  whenever  $x, x', u \in \mathbb{B}[x_0, 5r]$  and  $k \in \mathbb{N}_0$ ;
- (iii)  $d(0, f(x_0) + F(x_0)) < \min\{s, r(1 - \mu_0\kappa)/\kappa\}$ .

*Then there exists a sequence  $(x_k)_{k \in \mathbb{N}}$  generated by (1.2) with the starting point  $x_0$  which converges at least linearly to a solution  $\bar{x} \in \mathbb{B}[x_0, r]$ .*

*Proof.* For  $k \in \mathbb{N}_0$ , set

$$\kappa_k = \frac{\kappa}{1 - \mu_k\kappa} \quad \text{and} \quad R_k = \min\left\{s, \frac{r}{\kappa_k}\right\}.$$

Take a sequence  $(\varepsilon_k)_{k \in \mathbb{N}_0}$  in  $(0, 1)$  converging to 0 such that

$$(6.4) \quad \sup_{k \in \mathbb{N}_0} \mu_k(\kappa_k + \varepsilon_k) < 1, \quad \alpha := \sup_{k \in \mathbb{N}} \mu_{k-1}(\kappa_k + \varepsilon_k) < \frac{1}{3}, \quad \text{and} \quad \varepsilon_k < \kappa_k, \quad k \in \mathbb{N}_0.$$

We claim that there is a sequence  $(x_i)_{i \in \mathbb{N}_0}$  in  $\mathbb{B}[x_0, r]$  with the starting point  $x_0$  such that, for all  $i \in \mathbb{N}_0$ , we have

- (a)  $x_{i+1} \in \mathbb{B}[x_i, \frac{r}{2^{i+1}}]$ ,
- (b)  $0 \in A_i(x_{i+1}, x_i) + F(x_{i+1})$ ,
- (c)  $d(0, f(x_i) + F(x_i)) < R_i$ ,
- (d)  $\|x_{i+1} - x_i\| \leq (\kappa_i + \varepsilon_i) d(0, f(x_i) + F(x_i)) \leq (\kappa_i + \varepsilon_i) \mu_{i-1} \|x_i - x_{i-1}\|$  when  $i \geq 1$ .

Clearly,  $x_0$  verifies (c) with  $i = 0$ . Now, assume that there is  $k \geq 1$  such that  $x_0, \dots, x_k$  have already been defined in such a way that they satisfy (a)–(d) for each  $0 \leq i \leq k - 1$ , and also (c) with  $i = k$ . We will find  $x_{k+1}$ . If  $0 \in f(x_k) + F(x_k)$ , i.e., (c) is satisfied with  $i = k + 1$  for  $x_{k+1} := x_k$ , then  $x_{k+1}$  also verifies (a), (b), and (d) with  $i = k$ , since  $A_k(x_k, x_k) = f(x_k)$ . From now on, suppose  $0 \notin f(x_k) + F(x_k)$ . Since (a) is satisfied for all  $i = 0, \dots, k - 1$ , we have  $x_k \in \mathbb{B}[x_0, r]$ . Theorem 6.2, with  $\Phi := f + F$  and  $G := A_k(\cdot, x_k) - f$ , implies that  $A_k(\cdot, x_k) + F$  is metrically regular with the constant  $\kappa_k$  on

$$V_k := \{(x, y) \in X \times Y : x \in \mathbb{B}[x_0, r], d(y, A_k(x, x_k) + F(x)) < R_k\}.$$

As (c) is true for  $i = k$ , the assumption (i) implies that  $(x_k, 0) \in V_k$ , and hence

$$d(x_k, (A_k(\cdot, x_k) + F)^{-1}(0)) \leq \kappa_k d(0, f(x_k) + F(x_k)) < r.$$

Hence, there exists  $x_{k+1} \in (A_k(\cdot, x_k) + F)^{-1}(0)$  such that

$$\|x_k - x_{k+1}\| < (\kappa_k + \varepsilon_k) d(0, f(x_k) + F(x_k)).$$

So (b) is true with  $i = k$ . As  $0 \in A_{k-1}(x_k, x_{k-1}) + F(x_k)$  and  $A_{k-1}(x_{k-1}, x_{k-1}) = f(x_{k-1})$ , the condition (ii) implies that

$$\begin{aligned} \|x_k - x_{k+1}\| &< (\kappa_k + \varepsilon_k) d(0, f(x_k) + F(x_k)) \\ &\leq (\kappa_k + \varepsilon_k) h(f(x_{k-1}) - A_{k-1}(x_{k-1}, x_{k-1}), f(x_k) - A_{k-1}(x_k, x_{k-1})) \\ &\leq (\kappa_k + \varepsilon_k) \mu_{k-1} \|x_k - x_{k-1}\| \leq \alpha \|x_k - x_{k-1}\|. \end{aligned}$$

Thus (d) and (a) are satisfied with  $i = k$ . Now, (b) with  $i = k$  implies that

$$\begin{aligned} d(0, f(x_{k+1}) + F(x_{k+1})) &\leq h(0, f(x_{k+1}) - A_k(x_{k+1}, x_k)) \leq \mu_k \|x_{k+1} - x_k\| \\ &\leq \mu_k (\kappa_k + \varepsilon_k) d(0, f(x_k) + F(x_k)). \end{aligned}$$

Since  $\mu_k(\kappa_k + \varepsilon_k) < 1$ , we have  $d(0, f(x_{k+1}) + F(x_{k+1})) \leq d(0, f(x_k) + F(x_k)) < s$ . Moreover, (6.3) implies that  $2\kappa_{k+1}\mu_k < 1$ , and therefore

$$d(0, f(x_{k+1}) + F(x_{k+1})) < \frac{(\kappa_k + \varepsilon_k)}{2\kappa_{k+1}} d(0, f(x_k) + F(x_k)) < \frac{r(\kappa_k + \varepsilon_k)}{2\kappa_k\kappa_{k+1}} < \frac{r}{\kappa_{k+1}}.$$

Therefore, (c) is satisfied with  $i = k + 1$ . The claim is proved.

Clearly,  $(x_k)_{k \in \mathbb{N}_0}$  is a Cauchy sequence, and therefore it converges to  $\bar{x} \in \mathbb{B}[x_0, r]$ , say. From (d), one obtains that  $0 \in f(\bar{x}) + F(\bar{x})$ , and also that, for all  $k \in \mathbb{N}_0$ , we have

$$\|x_{k+1} - \bar{x}\| = \lim_{m \rightarrow +\infty} \|x_k - x_{k+m}\| \leq \lim_{m \rightarrow +\infty} \sum_{i=0}^{m-1} \|x_{k+i} - x_{k+i+1}\| \leq \frac{\alpha}{1 - \alpha} \|x_{k+1} - x_k\|.$$

This implies directly the linear convergence of the sequence  $(x_k)_{k \in \mathbb{N}_0}$ . Indeed, as  $\alpha < 1/3$ , we have  $\nu := \alpha/(1 - 2\alpha) < 1$ . The above estimate yields that

$$\|x_{k+1} - \bar{x}\| \leq \frac{\alpha}{1 - \alpha} (\|x_{k+1} - \bar{x}\| + \|\bar{x} - x_k\|).$$

So  $\|x_{k+1} - \bar{x}\| \leq \nu \|x_k - \bar{x}\|$ .  $\square$

**7. Commentary.** In this section, we provide a comparison of our results with the existing literature on the same subject. In [10], Bonnans considered a condition of semistability to ensure the quadratic convergence of the Newton's method for variational inequalities and nonlinear programming. More precisely, he discussed the case where  $f$  is of class  $C^1$  and the set-valued part  $F = N_C$  coincides with the normal cone to a closed convex set. Under the conditions of semistability and hemistability (these two conditions are satisfied if Robinson's strong regularity holds at the solution), the author proved the superlinear convergence of the Newton's method (quadratic convergence if  $f$  is  $C^{1,1}$ ). We note that these results were generalized by Izmailov and Solodov [25] and Izmailov and Kurennoy [26] to the case  $f(x) + F(x) \ni 0$  with a smooth single-valued map  $f$  and a set-valued mapping  $F$  by using an inexact Josephy–Newton method. We note that in the papers [10, 25], the authors considered a single-valued approximation, and that in the current paper we authorized the approximation to be set-valued. Theorem 4.1 is more general and covers an inexact iterative process (see Remark 4.2). In [18], the author considered the local behavior of Newton-type algorithms for generalized equations with nonisolated solutions and used a set-valued approximation. The key assumption in [18] is the calmness of the solution mapping which assigns to a parameter  $p$  a solution of the perturbed



generalized equation  $f(z) + F(z) + p \ni 0$ . This assumption is equivalent to metric subregularity of  $f + F$  but does not guarantee the needed solvability of the partially linearized subproblems, which is assumed separately. The calmness assumption with a condition on the solvability of the auxiliary problem, together with a condition on the quality of the approximation, is used in [18] to show the superlinear convergence of the sequence to a solution. We impose the regularity assumption either on  $f + F$  (Theorem 4.1) or on its partial linearization (Theorem 4.3) only. Note that for a non-smooth  $f$  these two assumptions are not equivalent. The metric regularity concept may be stronger than the calmness but guarantees the solvability of each subproblem and provides some robustness, especially when dealing with numerical algorithms. In [26], the authors also used a set-valued approximation and improved some existing results in the literature [10, 18].

Finally, we note that in [10, 18, 24, 25, 26] all results are stated in finite-dimensional spaces, while the presented ones are valid in Banach spaces. There has been a number of developments in the last decade regarding Newton-type methods applied to nonsmooth equations in infinite-dimensional spaces, e.g., in PDE-constrained optimization; some of them are broadly covered in the recent book by Ulbrich [42].

**8. Concluding remarks.** The Newton’s method remains one of the most powerful tools used in numerical analysis and optimization for solving systems of nonlinear equations. It was generalized to many settings and has been widely used for solving nonlinear programming problems, nonlinear complementarity problems, and variational inequalities. In this paper, we have studied the superlinear, or the linear, convergence of the Newton-type method for finding the zero of the sum of two set-valued mappings. Our convergence results were proved by using the stability of the metric regularity under set-valued perturbation. We have also investigated a global convergence of this method in the sense that, starting at an arbitrary point, there exists a sequence which converges at least linearly to a solution of the initial problem. We have also studied a forward-backward splitting algorithm and proved its superlinear convergence under appropriate assumptions on the data of the problem.

It would be interesting to test numerically the proposed algorithms, and to compare them with others, on various problems coming from nonlinear programming and complementarity systems. As pointed out also by one of the referees, it would be interesting to characterize “good” iterative sequences via reasonable conditions on proximity to the solutions. As a perspective to this work, we plan to study some potential applications, in the field of nonregular electrical circuits, with meaningful algorithmic examples for which metric regularity does not imply strong regularity. We expect that applying our theorems (and the related analysis) to specific nonsmooth Newton’s methods in infinite dimensions in the spirit of [42] will provide valuable contributions to the area. This is beyond the scope of this paper and will be the subject of forthcoming research.

**Acknowledgments.** We thank Prof. Assen Dontchev for suggesting we prove a localized version of Theorem 3.2, and also for a fruitful discussion on an earlier version of the manuscript. We thank the anonymous referees for their valuable comments which led to several improvements.

## REFERENCES

- [1] M. APETRII, M. DUREA, AND R. STRUGARIU, *On subregularity properties of set-valued mappings*, Set-Valued Var. Anal., 21 (2013), pp. 93–126.

- [2] F. J. ARAGÓN ARTACHO, A. L. DONTCHEV, M. GAYDU, M. H. GEOFFROY, AND V. M. VELIOV, *Metric regularity of Newton's iteration*, SIAM J. Control Optim., 49 (2011), pp. 339–362.
- [3] F. J. ARAGÓN ARTACHO, A. L. DONTCHEV, AND M. H. GEOFFROY, *Convergence of the proximal point method for metrically regular mappings*, in CSVAA 2004—control set-valued analysis and applications, ESAIM Proc. 17, EDP Sci., Les Ulis, 2007, pp. 1–8.
- [4] F. J. ARAGÓN ARTACHO AND M. GAYDU, *A Lyusternik-Graves theorem for the proximal point method*, Comput. Optim. Appl., 52 (2011), pp. 785–803.
- [5] A. V. ARUTYUNOV, *An iterative method for finding coincidence points on two mappings*, Comput. Math. Math. Phys., 52 (2012), pp. 1483–1486.
- [6] A. V. ARUTYUNOV, *Coincidence points of two maps*, Funct. Anal. Appl., 48 (2014), pp. 72–75.
- [7] A. ARUTYUNOV, E. AVAKOV, B. GELMAN, A. DMITRUK, AND V. OBUKHOVSKII, *Locally covering maps in metric spaces and coincidence points*, J. Fixed Point Theory Appl., 5 (2009), pp. 106–127.
- [8] J. P. AUBIN AND H. FRANKOWSKA, *Set-Valued Analysis*, Modern Birkhäuser Classics, Birkhäuser Boston, Inc., Boston, MA, 2009.
- [9] M. BIANCHI, G. KASSAY, AND R. PINI, *An inverse map result and some applications to sensitivity of generalized equations*, J. Math. Anal. Appl., 399 (2013), pp. 279–290.
- [10] J. F. BONNANS, *Local analysis of Newton-type methods for variational inequalities and nonlinear programming*, Appl. Math. Optim., 29 (1994), pp. 161–186.
- [11] F. H. CLARKE, *Optimization and Nonsmooth Analysis*, Classics Appl. Math. 5, SIAM, Philadelphia, 1990.
- [12] A. L. DONTCHEV AND R. T. ROCKAFELLAR, *Implicit Functions and Solution Mappings. A View from Variational Analysis*, Springer Monogr. Math., Springer, Dordrecht, 2009.
- [13] A. L. DONTCHEV AND R. T. ROCKAFELLAR, *Implicit Functions and Solution Mappings. A View from Variational Analysis*, 2nd ed., Springer Ser. Oper. Res. Financial Engrg., Springer, Dordrecht, 2014.
- [14] A. L. DONTCHEV AND R. T. ROCKAFELLAR, *Convergence of inexact Newton methods for generalized equations*, Math. Program. Ser. B., 139 (2013), pp. 115–137.
- [15] A. L. DONTCHEV AND V. M. VELIOV, *Metric regularity under approximations*, Control Cybernet., 38 (2009), pp. 1283–1303.
- [16] I. EKELAND, *On the variational principle*, J. Math. Anal. Appl., 47 (1974), pp. 324–353.
- [17] F. FACCHINEI AND J.-S. PANG, *Finite-dimensional Variational Inequalities and Complementarity Problems*, Vol. II, Springer Ser. Oper. Res., Springer-Verlag, New York, 2003.
- [18] A. FISCHER, *Local behavior of an iterative framework for generalized equations with nonisolated solutions*, Math. Program., 94 (2002), pp. 91–124.
- [19] Y. GAO, *Newton methods for solving nonsmooth equations via a new subdifferential*, Math. Methods Oper. Res., 54 (2001), pp. 239–257.
- [20] M. H. GEOFFROY, *Stability of Mann's iterates under metric regularity*, Appl. Math. Comput., 215 (2009), pp. 686–694.
- [21] A. D. IOFFE, *Nonsmooth Analysis: Differential calculus of nondifferentiable mappings*, Trans. Amer. Math. Soc., 266 (1981), pp. 1–56.
- [22] A. D. IOFFE, *On perturbation stability of metric regularity*, Set-Valued Anal., 9 (2001), pp. 101–109.
- [23] A. F. IZMAILOV, *Strongly regular nonsmooth generalized equations*, Math. Program., 147 (2014), pp. 581–590.
- [24] A. F. IZMAILOV, A. S. KURENNOY, AND M. V. SOLODOV, *The Josephy-Newton method for semismooth generalized equations and semismooth SQP for optimization*, Set-Valued Var. Anal., 21 (2013), pp. 17–45.
- [25] A. F. IZMAILOV AND M. V. SOLODOV, *Inexact Josephy-Newton framework for generalized equations and its applications to local analysis of Newtonian methods for constrained optimization*, Comput. Optim. Appl., 46 (2010), pp. 347–368.
- [26] A. F. IZMAILOV AND A. S. KURENNOY, *Abstract Newtonian Frameworks and Their Applications*, [http://www.optimization-online.org/DB\\_HTML/2013/02/3760.html](http://www.optimization-online.org/DB_HTML/2013/02/3760.html) (2012).
- [27] N. H. JOSEPHY, *Newton's Method for Generalized Equations*, Technical Summary Report 1965, Mathematics Research Center, University of Wisconsin, Madison, 1979.
- [28] D. KLATTE AND B. KUMMER, *Nonsmooth Equations in Optimization. Regularity, Calculus, Methods and Applications*, Kluwer Academic Publishers, Dordrecht, 2002.
- [29] B. KUMMER, *Approximation of multifunctions and superlinear convergence*, in Recent Developments in Optimization, Lecture Notes in Econom. and Math. Systems 429, Springer, Berlin, 1995, pp. 243–251.
- [30] B. KUMMER, *Newton's Method for Continuous Functions?*, preprint, Institut für Mathematik, Humboldt-Universität zu Berlin, Berlin, Germany, 2012.

- [31] B. KUMMER, *Newton's method for non-differentiable functions*, in Advances in Mathematical Optimization, Ser. Math. Res. 45, J. Guddat et al., eds., Akademie Verlag, Berlin, 1988, pp. 114–125.
- [32] B. S. MORDUKHOVICH, *Variational Analysis and Generalized Differentiation. I. Basic Theory*, Grundlehren Math. Wiss. 330, Springer-Verlag, Berlin, 2006.
- [33] H. V. NGAI AND M. THÉRA, *Error bounds in metric spaces and application to the perturbation stability of metric regularity*, SIAM J. Optim., 19 (2008), pp. 1–20.
- [34] Z. PÁLES, *Inverse and implicit function theorems*, J. Math. Anal. Appl., 209 (1997), pp. 202–220.
- [35] C. H. J. PANG, *Generalized differentiation with positively homogeneous maps: Applications in set-valued analysis and metric regularity*, Math. Oper. Res., 36 (2011), pp. 377–397.
- [36] G. PASCALINE, *Convergence de Fisher et H-différentiabilité des applications multivoques*, Ph.D. thesis, Laboratoire de Mathématiques Informatique et Applications, Université des Antilles et de la Guyane, Pointe-à-Pitre, Guadeloupe, 2012.
- [37] J.-P. PENOT, *Calculus Without Derivatives*, Grad. Texts in Math. 266, Springer, New York, 2013.
- [38] L. QI AND J. SUN, *A nonsmooth version of Newton's method*, Math. Program., 58 (1993), pp. 353–367.
- [39] S. M. ROBINSON, *Strongly regular generalized equations*, Math. Oper. Res., 5 (1980), pp. 43–62.
- [40] R. T. ROCKAFELLAR AND R. J.-B. WETS, *Variational Analysis*, Springer-Verlag, Berlin, 1998.
- [41] H. XU, *Set-valued approximations and Newton's methods*, Math. Program. Ser. A, 84 (1999), pp. 401–420.
- [42] M. ULBRICH, *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*, MOS-SIAM Ser. Optim. 11, SIAM, Philadelphia, 2011.