NEWTON’S METHOD FOR SOLVING INCLUSIONS USING SET-VALUED APPROXIMATIONS

SAMIR ADLY*, RADEK CIBULKA †, AND HUYNH VAN NGAI ‡

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 non-smooth Newton’s method for semi-smooth functions, the inexact proximal point algorithm, etc. Moreover, the note also covers 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, semi-smooth function, Newton method, forward-backward splitting, proximal point method

AMS subject classifications. 49J52, 49J53, 90C30

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:

\begin{equation}
\text{Find } x \in X \text{ such that } 0 \in \Psi(x) + F(x).
\end{equation}

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

\begin{equation}
0 \in A_k(x_{k+1}, x_k) + F(x_{k+1}) \quad \text{for each} \quad k \in \mathbb{N}_0 := \{0, 1, 2, \ldots \}.
\end{equation}

Instead of considering a general inclusion

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

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 [27]. In this paper, each result on convergence of the scheme (1.2) relies on metric regularity of an appropriate mapping which can be checked (in Asplund spaces) via the well-known coderivative criterion [30, Theorem 4.18].

When the mapping $\Psi$ is single-valued, i.e. $\Psi = f$ for a given function $f : X \to Y$, one obtains a generalized equation, introduced by S. M. Robinson in [37], which reads as:

\begin{equation}
\text{Find } x \in X \text{ such that } 0 \in f(x) + F(x).
\end{equation}
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 $F = \mathbb{R}^n_+$), 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 multi-agent equilibrium problems (see [26] or [11]).

The case of single-valued approximations $A_k : X \times X \to Y$ of the function $f$ was studied in [8], [23], [24] and [11, 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*:

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

When $F$ coincides with the normal cone mapping to a closed convex subset of $X$, this scheme is known as the *Josephy-Newton algorithm* (see [25]). It is well-known that if the starting point $x_0$ is sufficiently close to some regular (in the sense of Robinson [37]) solution $\bar{x}$, then the Josephy-Newton method is well-defined and converges super-linearly 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*:

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

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 consider the iteration scheme defined for each $k \in \mathbb{N}_0$ either by

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

where $y \in Y$ is a given perturbation term nearby 0. Note that these algorithms cover so-called *Mann’s iteration* for solving the inclusion $x \in T(x)$ with $T : X \rightrightarrows X$ considered in [18]. 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 non-decreasing and converges to 1, and $y \in X$ is a given perturbation term in a 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 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 [13], the authors suppose that there
is a (known) sequence of error functions $r_k : X \to Y$ and a sequence of single-valued approximations $A_k : X \times X \to Y$ of $f$. Given $x_k \in X$, the next point $x_{k+1}$ is required to be such that

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

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 [18] require even to have $r_k \equiv \epsilon_k$ for a given error sequence $(\epsilon_k)_{k \in \mathbb{N}_0}$ in $Y$ converging to 0. The (possibly) set-valued approximations can be used in investigating even more general error models. In [16], the author considered the local behavior of Newton-type algorithms for generalized equations with non isolated solutions. The author used a set-valued approximation. The key assumption is the calmness of 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 non-smooth Newton method for semi-smooth 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 $x_0$.

2. Mathematical background. A recent book [11] by A. L. Dontchev and R. T. Rockafellar shows that the study of (equivalent) concepts of non-linear analysis such as linear openness, metric regularity, and inverse Aubin property, is of a great importance from both the theoretical and numerical point of view. We begin this section by fixing the notations. If we write $a := b$, we mean that $a$ is defined by $b$ without explicit saying in the text. Given a complete metric space $(X, \rho)$, we denote by $B[x, r]$ and $B(x, r)$ the closed and open ball with the center $x \in X$ and the radius $r \geq 0$, respectively. We set $B_X = B[0,1]$. The closure, the interior, and the diameter of a subset $K$ of $X$ is denoted by $\bar{K}$, int $K$, and diam $K$, respectively. The distance function generated by $K$ is $d(x, K) := \inf \{ d(x, k) : k \in K \}$, $x \in X$, 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 $co K$. 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

$$e(A, B) = \sup_{x \in A} d(x, B) = \inf \{ \tau > 0 : A \subset B + \tau B_X \},$$
with the convention that $e(0, B) := 0$ when $B \neq 0$, and that $e(0, 0) := +\infty$. The Pompeiu-Hausdorff distance between $A$ and $B$ is then defined by

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

Note that

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

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 singleton for each $x \in X$, we identify the one-point set with its element, $g(x)$ say. Then $g : X \to Y$ 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

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

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 $(x, y)$ 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 multi-valued, then $T$ is called strongly metrically regular at $(\bar{x}, \bar{y})$. In case of a single valued mapping $g : X \to Y$ we simply speak about (strong) metric regularity at $\bar{x}$.

Further, $T$ is (Pompeiu-Hausdorff) upper semi-continuous at $\bar{x} \in X$ if

$$\lim_{x \to \bar{x}} e(T(y), T(\bar{x})) = 0.$$ 

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

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

Note that the previous property means that $U \subset \text{dom} \ T$. A mapping $g : \text{dom} \ T \to Y$ 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 multi-valued perturbation. Let $X$ be a complete metric space and let $f : X \to \mathbb{R} \cup \{ +\infty \}$ 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 [31, Theorem 2.1] allowing us to estimate the distance $d(\bar{x}, S)$ for a given point $\bar{x} \notin S$.

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

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

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

Given $x \in X$, the symbol $[f(x)]_+$ denotes $\max \{ f(x), 0 \}$. First, we recall [31, Theorem 2.1] allowing us to estimate the distance $d(\bar{x}, S)$ for a given point $\bar{x} \notin S$.

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

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

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

Given $x \in X$, the symbol $[f(x)]_+$ denotes $\max \{ f(x), 0 \}$. First, we recall [31, Theorem 2.1] allowing us to estimate the distance $d(\bar{x}, S)$ for a given point $\bar{x} \notin S$.

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

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

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

Given $x \in X$, the symbol $[f(x)]_+$ denotes $\max \{ f(x), 0 \}$. First, we recall [31, Theorem 2.1] allowing us to estimate the distance $d(\bar{x}, S)$ for a given point $\bar{x} \notin S$.

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

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

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

Given $x \in X$, the symbol $[f(x)]_+$ denotes $\max \{ f(x), 0 \}$. First, we recall [31, Theorem 2.1] allowing us to estimate the distance $d(\bar{x}, S)$ for a given point $\bar{x} \notin S$.
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 \( B(\bar{x}, a) \times B(\bar{y}, b) \) 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 \( \alpha, \beta, \eta \) be positive constants satisfying

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

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 \( B(\bar{x}, \delta) \) with the constant \( L \); 
(ii) \( \text{diam} \, G(\bar{x}) \leq \eta, \)

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

**Proof.** Let \( f : X \times Y \times Y \to \mathbb{R} \) be defined by

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

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

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

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 \( g : X \times Y \to \mathbb{R} \) on \( X \times Y \) by

\[
    g((x_1, z_1), (x_2, z_2)) := \max \{\|x_1 - x_2\|, \|z_1 - z_2\|/c\}.
\]

Let \( y \in B(\bar{y} + \bar{z}, \beta) \) 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 [14, Theorem 1.1], we can select a point \((u, w) \in X \times Y\) satisfying

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

such that

\[(3.6) \quad f(x, z, y) + \frac{1}{\tau + \varepsilon} \, g((x, z), (u, w)) \geq f(u, w, y) \quad \text{for all} \quad (x, z) \in X \times Y.\]
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
\[
\lim_{n \to +\infty} d(y, \Phi(u_n) + w) = f(u, w, y) > 0.
\]
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 \( \mathbb{B}(\bar{x}, \alpha) \times \mathbb{B}(\bar{y}, \beta) \), 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, \( \liminf_{n \to +\infty} d(y - w, v_n) > 0 \). Since \((u_n)_{n \in \mathbb{N}}\) converges to \( u \) and \( f(u, w, y) \leq f(\bar{x}, \bar{z}, y, \Phi) < \beta \), 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
\[
\frac{1}{\tau + \varepsilon} \geq \limsup_{n \to +\infty} \frac{f(u, w, y) - f(v_n, w_n, y)}{\varrho((u, w, (v_n, w_n))} \geq \limsup_{n \to +\infty} \frac{d(y, \Phi(u_n)) - d(y - w_n, \Phi(v_n))}{\|u - v_n\|}
\]
\[
\geq \limsup_{n \to +\infty} \left( (\kappa + \gamma)^{-1}\|u_n - v_n\| - \|w - w_n\| \right)
\]
\[
\geq \limsup_{n \to +\infty} \left( (\kappa + \gamma)^{-1}(\|u - v_n\| - \|u_n - u\|) - (L + \gamma)\|u - v_n\| \right)
\]
\[
= \lim_{n \to +\infty} \left( \frac{1}{\kappa + \gamma} - L - \gamma - \frac{\|u_n - u\|}{\|u - v_n\|(\kappa + \gamma)} \right) = \frac{1}{\kappa + \gamma} - L - \gamma > \frac{1}{\tau + \varepsilon}.
\]
a contradiction. Thus \( y - w \in \Phi(u) \). Since \( w \in G(u) \), we get that
\[
S(y) \cap (\mathbb{B}[\bar{x}, \beta(\tau + \varepsilon)] \times \mathbb{B}[\bar{z}, \beta c(\tau + \varepsilon)]) \neq \emptyset.
\]
Fix any \((x, y) \in \mathbb{B}(\bar{x}, \alpha) \times \mathbb{B}(\bar{y}, \bar{z}, \beta) \). We will show that
\[
d(x, (\Phi + G)^{-1}(y)) \leq \tau d(y, \Phi(x) + G(x)).
\]
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
\[
\|z - \bar{z}\| \leq c(G(x), G(\bar{x})) + \text{diam } G(\bar{x}) \leq L\|x - \bar{x}\| + \eta < L\alpha + \eta < c\alpha + \eta.
\]
Hence, by (3.8),
\[
d((x, z), S(y)) \leq \varrho((x, z), (\bar{x}, \bar{z})) + d((\bar{x}, \bar{z}), S(y)) < \alpha + \eta/c + \beta(\tau + \varepsilon) < \delta/2.
\]
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

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

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

\[
m(x, z) := \inf \left\{ \sup_{(u, w) \in X \times Y} \frac{f(u, w, y) - f(v, w', y)}{\varrho(u, v, w, w')}: \varrho((u, w), (x, z)) < d((x, z), S(y)) \right\}
\]

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

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

Also, (3.10) and (3.11) imply that

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

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

\[
\lim_{n \to +\infty} d(y, \Phi(u_n) + w) = f(u, w, y) > 0.
\]

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

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

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 (3.12), we obtain

\[
\begin{align*}
\limsup_{n \to +\infty} \frac{f(u, w, y) - f(v_n, w_n, y)}{\varrho(u, w, (v_n, w_n))} &\geq \limsup_{n \to +\infty} \frac{d(y - w, \Phi(u_n)) - d(y - w, \Phi(v_n))}{\|u - v_n\|} \\
&\geq \limsup_{n \to +\infty} \frac{\kappa + \gamma)^{-1} \|u_n - v_n\| - \|w_n - w_n\|}{\|u - v_n\|} \\
&\geq \limsup_{n \to +\infty} \frac{\kappa + \gamma)^{-1} \|u_n - v_n\| - \|u_n - u\| - (L + \gamma) \|u_n - v_n\|}{\|u - v_n\|} \\
&= \lim_{n \to +\infty} \left( \frac{1}{\kappa + \gamma} - L - \gamma - \frac{\|u_n - u\|}{\|u - v_n\|((\kappa + \gamma)} \right) = \frac{1}{\kappa + \gamma} - L - \gamma > \frac{1}{\tau + \varepsilon}.
\end{align*}
\]
By virtue of Theorem 3.1, we derive that
\[
\begin{align*}
  d(x, (\Phi + G)^{-1}(y)) &\leq d(x, z), S(y)) \\
  &\leq (\tau + \varepsilon)f(x, z, y) \leq (\tau + \varepsilon) d(y, \Phi(x) + z) \\
  &\leq (\tau + \varepsilon)(1 + \varepsilon) d(y, \Phi(x) + G(x)).
\end{align*}
\]

Hence, in both the 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 [11, Example 5E.6] and also [7]) when the condition (ii) on diameter is omitted.

(iii) A closely related result to Theorem 3.2 was proved in [7]. The authors in [7] 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 when we do not insist on the uniform estimate for all points \(z \in G(\bar{x})\).

**Theorem 3.4.** Given Banach spaces \(X\) and \(Y\), let \(\Phi : X \Rightarrow Y\) 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 \(B(\bar{x}, a) \times B(\bar{y}, b)\) 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 \(\beta > 0\) be such that
\[
\begin{align*}
  \beta \tau &< \min\{a, \delta/2\}, \quad \beta(\tau + \kappa) < \delta, \quad \beta(1 + c\tau) < \min\{b, \delta\},
\end{align*}
\]
with \(c := \max\{1, 1/\kappa\}\). Then for any set-valued mapping \(G : X \Rightarrow Y\) with closed graph and a point \(\bar{z} \in G(\bar{x})\) such that
\[
\begin{align*}
  e(G(x_1) \cap B(\bar{z}, \delta), G(x_2)) &\leq L\|x_1 - x_2\| \quad \text{whenever} \quad x_1, x_2 \in B(\bar{x}, \delta),
\end{align*}
\]
one has
\[
\begin{align*}
  d(\bar{x}, (\Phi + G)^{-1}(y)) &\leq \tau d(y, \Phi(\bar{x}) + \bar{z}) \quad \text{for all} \quad y \in B(\bar{y} + \bar{z}, \beta).
\end{align*}
\]

**Proof.** The proof is quite similar to the one of Theorem 3.2, hence we mention the needed changes only. Let \(f\) and \(S(y), y \in Y\), be as before. Pick \(\varepsilon > 0\) such that
\[
\begin{align*}
  \beta(\tau + \varepsilon) &< \min\{a, \delta/2\}, \quad \beta(\tau + \kappa + 2\varepsilon) < \delta, \quad \beta(1 + c(\tau + \varepsilon)) < \min\{b, \delta\}.
\end{align*}
\]
Fix any \(y \in B(\bar{y} + \bar{z}, \beta)\). Using exactly the same steps, one can prove that (3.8) is valid (just in the construction of \((v_n)_{n \in \mathbb{N}}\) one uses (3.14) together with the fact that \(w \in G(u) \cap B(\bar{z}, \delta)\) and that \((v_n)_{n \in \mathbb{N}}\) is in \(B(\bar{x}, \delta)\) 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),
\[
\begin{align*}
  d((\bar{x}, \bar{z}), S(y)) &\leq \beta(\tau + \varepsilon) < \delta/2.
\end{align*}
\]
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
\[
\begin{align*}
  d(\bar{x}, (\Phi + G)^{-1}(y)) &< \delta/2 \leq \kappa d(y, \Phi(\bar{x}) + \bar{z}) < \tau d(y, \Phi(\bar{x}) + \bar{z}).
\end{align*}
\]
Case 2. $d(y, \Phi(\bar{x}) + \bar{z}) < \frac{\delta}{2\kappa}$. One repeats exactly the same steps as in the 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

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

The above property is sometimes called metric hemiregularity \cite{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:

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

It is possible to obtain the conclusion of Theorem 3.4 from Theorem 2 in \cite{6} by setting $\Psi = -G(.) + y$ with $G$ Lipschitz. In Theorem 3.4, we assumed $G$ to be pseudo Lipschitz/Aubin continuous at the reference point (see condition (3.14) and (5.6) for the precise definition), which is a weaker assumption.

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 $\limsup_{k \to +\infty} \frac{\|x_{k+1} - x_0\|}{\|x_k - x_0\|} < 1$ 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 super-linearly, if either $\lim_{k \to +\infty} \frac{\|x_{k+1} - x_0\|}{\|x_k - x_0\|} = 0$ 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 ($Q$-super-linear) in the literature. Throughout this section we will use the following premise.

Standing assumptions. Let $X$ and $Y$ be Banach spaces, let $f : X \to Y$ 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 [11, 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 $B(\bar{x}, a) \times B(0, b)$ 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

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

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 following properties

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

(ii) $d(f(\bar{x}), A_k(\bar{x}, u)) \leq \varepsilon_k \|u - \bar{x}\| \quad \text{for all} \ u \in B(\bar{x}, a)$ 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

$$\lim_{k \to +\infty} \frac{d(f(\bar{x}), A_k(\bar{x}, x_k))}{\|x_k - \bar{x}\|} = 0,$$
then \((x_k)_{k \in \mathbb{N}}\) converges super-linearly to \(\bar{x}\).

**Proof.** Find \(\beta > 0\) such that (3.13) in Theorem 3.4 is valid for \(\Phi := f + F\) and \(\mu := a\). 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 assumptions, 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 true 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

\[
d(\bar{x}, (A_k(\cdot, x_k) + F(\cdot))^{-1}(0)) \leq \kappa_k d(0, f(\bar{x}) + F(\bar{x}) + (y_k - f(\bar{x}))) \leq \kappa_k \|y_k - f(\bar{x})\|
\]

\[
< \kappa_k(1 + \delta_k) d(f(\bar{x}), A(\bar{x}, x_k)) \leq \kappa_k\varepsilon_k(1 + \delta_k)\|x_k - \bar{x}\|
\]

\[
< (\kappa_k\varepsilon_k + \delta_k(1 + \kappa_k\varepsilon_k))\|x_k - \bar{x}\| < (s + \gamma)\|x_k - \bar{x}\|
\]

\[
< r.
\]

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

\[
0 \leq \frac{\|x_{k+1} - \bar{x}\|}{\|x_k - \bar{x}\|} < \kappa_k(1 + \delta_k) \frac{d(f(\bar{x}), A_k(\bar{x}, x_k))}{\|x_k - \bar{x}\|}
\]

whenever \(k \in \mathbb{N}\).

So (4.1) implies the super-linear convergence of \((x_k)_{k \in \mathbb{N}}\).\(\square\)

**Remark 4.2.** Suppose that \(f\) is continuously differentiable at \(\bar{x}\). In [12], 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
The strict pre-derivative is often generated by a family of continuous linear operators (4.4) $g_{\text{pre-derivative}}$ of $g$ set-valued mapping $g$ non-smooth mapping $f$ differentiable mapping $f$. For each $x$ for each $x_0$ in $[12]$ 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.

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 [12] 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 with the metric regularity of its "partial linearization" $f(\bar{x}) + f'(\bar{x})(x - \bar{x}) + F$. However, for a non-smooth $f$ it can be too strong (see Remark 4.5 (ii)). For $F \equiv 0$, the relationship of 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.

A. D. Ioffe [19] initiated the use of the strict pre-derivatives to approximate a non-smooth mapping $g : X \rightrightarrows Y$ around a reference point. A positively homogeneous set-valued mapping $G : X \rightrightarrows Y$ (i.e. gph $G$ is a cone in $X \times Y$) is called the strict pre-derivative of $g$ at $x_0 \in X$ if for each $c > 0$ there exists $\delta > 0$ such that

(4.4) \hspace{1cm} g(x_1) \in g(x_2) + G(x_1 - x_2) + c\|x_1 - x_2\|B_Y \hspace{1cm} \text{whenever} \hspace{1cm} x_1, x_2 \in \mathbb{B}(x_0, \delta).

The strict pre-derivative 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 $g : \mathbb{R}^n \to \mathbb{R}^m$ 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

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

Then $\partial g(u) := \text{co} \partial_B g(u)$ denotes the Clarke generalized Jacobian of $g$ at $u$ (see also [9]). It is well known that $\partial g(u)$ is a strict pre-derivative of $g$ at $u$ and $x \rightrightarrows \partial g(x)$ has convex and compact values (see [19, 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 semi-continuous.

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 semi-continuous at $\bar{x} \in \text{int}(\text{dom} \mathcal{H})$;
(A1.3) there is a selection \( \psi : X \to \mathcal{L}(X,Y) \) for \( \mathcal{H} \) such that
\[
\lim_{\varepsilon \to 0} \frac{\|f(x) - f(x) - \psi(x)(x - \bar{x})\|}{\|x - \bar{x}\|} = 0.
\]

Recall that a function \( g : \mathbb{R}^n \to \mathbb{R}^m \) is called semi-smooth (see [15, Section 7.4]) at \( u \in \mathbb{R}^n \) if it is locally Lipschitz continuous around \( u \), directionally differentiable near \( u \), and it satisfies the condition
\[
\lim_{0 \neq v \to 0} \frac{\sup_{M \in \partial g(u + v)} \|g(u + v) - g(u) - M v\|}{\|v\|} = 0.
\]
Hence, if \( f : \mathbb{R}^n \to \mathbb{R}^m \) is semi-smooth at \( \bar{x} \) then both \( \partial f \) and \( \partial_B f \) satisfy the Assumption (A1).

**Theorem 4.3.** In addition to the Assumption (A1), suppose that for each \( \bar{H} \in \mathcal{H}((\bar{x},0) \subset X \) 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
\[
0 \in f(x_k) + \mathcal{H}(x_k)(x_{k+1} - x_k) + F(x_{k+1}) \quad \text{for each } k \in \mathbb{N},
\]
and this sequence converges super-linearly.

**Proof.** First, we claim that there are \( \kappa > 0 \), \( a > 0 \) and \( b > 0 \) such that for each \( \bar{H} \in \mathcal{H}((\bar{x},0) \subset X \) 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 the Proposition 2.3H3 in the forthcoming second edition of [12] which is adopted from [21]. Fix any \( \bar{H} \in \mathcal{H}((\bar{x},0) \subset X \) with the constant \( \tau > 0 \), \( \alpha > 0 \), and \( \beta > 0 \) such that \( \Phi_{\bar{H}} \) is metrically regular at \( (\bar{x},0) \) 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, 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 \( \mathbb{B}(\bar{x},a') \times \mathbb{B}(0,b') \). We showed that for any \( \bar{H} \in \mathcal{H}(\bar{x},0) \) 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 \( \bar{H}' \in \mathcal{H}(\bar{x},0) \) with \( \|\bar{H}' - \bar{H}\| < \varepsilon_{\bar{H}} \). By the compactness, from the open covering \( \bigcup_{\bar{H} \in \mathcal{H}(\bar{x},0)} \mathbb{B}(\bar{H},\varepsilon_{\bar{H}}) \) of \( \mathcal{H}(\bar{x},0) \), we can choose a finite sub-covering, i.e. there are \( k \in \mathbb{N}, \bar{H}_i \in \mathcal{H}(\bar{x},0) \), and \( \varepsilon_{\bar{H}_i} > 0 \) such that \( \mathcal{H}(\bar{x},0) \) is covered with \( \mathbb{B}(\bar{H}_i,\varepsilon_{\bar{H}_i}), \ i = 1, 2, \ldots, k \). Setting \( \kappa = \max_{i} \tau_{\bar{H}_i}, \ a = \min_{i} \alpha_{\bar{H}_i}, \) and \( \beta = \min_{i} \beta_{\bar{H}_i}, \) one gets the claim.

Fix \( L \in (0,1/\kappa) \). By (A1.2), there is \( \delta > 0 \) such that
\[
\mathcal{H}(y) \subset \mathcal{H}(\bar{x}) + L \mathbb{B}_{\mathcal{L}(X,Y)} \quad \text{whenever } y \in \mathbb{B}(\bar{x},\delta) \subset \text{dom } \mathcal{H}.
\]
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 } 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

\[
0 \in f(x_k) + \psi(x_k)(x - x_k) + F(x).
\]

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

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

Clearly, \( G \) is single-valued, 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

\[
f(x_k) - f(\bar{x}) - \psi(x_k)(x_k - \bar{x}) + f(\bar{x}) + \psi(x_k)(x - \bar{x}) + F(x) = y_k + f(\bar{x}) + (\bar{H} + L\bar{H})(x - \bar{x}) + F(x) = y_k + \Phi_{\bar{H}}(x) + L\bar{H}(x - \bar{x}) = (\Phi_{\bar{H}} + G)(x).
\]

Theorem 3.2 implies that \( \Phi_{\bar{H}} + G \) 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 \( 0 \in f(\bar{x}) + F(\bar{x}) = \Phi_{\bar{H}}(\bar{x}) \), we have \( y_k \in (\Phi_{\bar{H}} + G)(\bar{x}) \), therefore the above inequality implies that

\[
d\left(\bar{x}, (\Phi_{\bar{H}} + G)^{-1}(0)\right) \leq \frac{\kappa}{1 - \kappa L}d(0, (\Phi_{\bar{H}} + G)(\bar{x})) \leq \frac{\kappa}{1 - \kappa L}\|y_k\| < \frac{2\kappa}{1 - \kappa L}\|x_k - \bar{x}\| < r.
\]

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

\[
\|x_{k+1} - \bar{x}\| < \frac{2\kappa}{1 - \kappa L}\|y_k\| \leq \frac{2\kappa\beta}{1 - \kappa L}\|x_k - \bar{x}\|.
\]

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

\[
\lim_{k \to +\infty} \frac{\|y_k\|}{\|x_k - \bar{x}\|} = 0,
\]

hence, taking into account (4.9), we get the super-linear convergence of \( (x_k)_{k \in \mathbb{N}} \). \( \Box \)

**Remark 4.4.** (i) In the proof of the preceding result, we used twice 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 [11, Theorem 5G.3] which is present only in the forthcoming second edition of the book [11].
(ii) In addition to standing assumptions, suppose that $F \equiv 0$ and that $f$ has the strict pre-derivative at $\bar{x} \in X$ generated by a compact convex subset $T$ of $L(X,Y)$. The modulus of (linear) openness of $T$, denoted by $\sigma(T)$, is defined by

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

where $\sigma(T) := \sup \{ r > 0 : B[0,r] \subset T(B_X) \}$ is the modulus of (linear) openness of $T$. Z. Páles [32] used this quantity to provide a condition guaranteeing the metric regularity at the reference point. If $\sigma(T) > 0$ then [32, Theorem 2] implies that $f$ is metrically regular at $\bar{x}$ with a constant arbitrarily close to $1/\sigma(T)$. If, in addition, all the elements of $\mathcal{T}$ are injective then $f$ is strongly metrically regular at $\bar{x}$ by [32, Theorem 5].

If $f : \mathbb{R}^n \to \mathbb{R}^m$ is semi-smooth at $\bar{x}$ then a sufficient condition for the super-linear convergence of the non-smooth Newton methods defined, for each $k \in \mathbb{N}_0$, either by

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

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 [36], [38] and [17] (note that the inexact algorithms therein can be treated similarly). In the last two references, one can find also other suitable candidates for $\mathcal{H}$. From 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 (non-singularity) of all elements of $\mathcal{H}(\bar{x})$.

**Remark 4.5.** (i) Formally, (A1.3) is less restrictive than the assumption that $f$ is semi-smooth at the reference point. However, in practice, it is not possible and efficient to choose the "right" matrix in the corresponding subdifferential at each step (we just take some). Inspecting the proof of Theorem 4.3, we see that if (A1.3) is replaced by

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

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 $\Phi_H$ is less restrictive than to suppose that $f + F$ is metrically regular (unless $\mathcal{H}(\bar{x})$ is convex and generates the strict pre-derivative of $f$ at $\bar{x}$). Indeed, take $f(x) := |x|$, $x \in \mathbb{R}$, and $H := \partial_B f$. In [22], the authors supposed that all $\Phi_H$ 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 [28], one can find an outstanding discussion on the attempts to generalize Newton's method for continuous (not locally Lipschitz) functions. One can also use some generalized derivative to approximate $f$ as [29] shows.

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

(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 it is metrically regular at $(\bar{x}, \bar{y}_1)$ with a constant $\kappa > 0$ on $B(\bar{x}, a) \times B(\bar{y}_1, b)$ for some $a > 0$ and $b > 0$;
(A2.3) \( \Psi \) has closed graph and there is \( \varepsilon > 0 \) such that

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

**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

\[
0 \in \lambda_k (x_{k+1} - x_k) + F(x_{k+1}) + \Psi(x_k)
\]

for each \( k \in \mathbb{N}_0 \), and this sequence converges at least linearly to \( \bar{x} \). If, in addition,

\[
\lim_{k \to +\infty} \frac{d(\bar{y}_2, \lambda_k (\bar{x} - x_k) + \Psi(x_k))}{\|x_k - \bar{x}\|} = 0,
\]

then \( (x_k)_{k \in \mathbb{N}} \) converges super-linearly.

**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 true. Then take \( \gamma \in (0, 1) \) and \( r \in (0, a) \) such that

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

Fix any \( \xi \in (0, \gamma/(\lambda + \varepsilon)) \). Set \( U = B(\bar{x}, r) \). 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

\[
\|\bar{z} - \bar{y}_2\| < (1 + \xi) d(\bar{y}_2, A_k(\bar{x}, x_k)) \leq (1 + \xi) e(\Psi(\bar{x}), \lambda_k (\bar{x} - x_k) + \Psi(x_k))
\]

\[
\leq (1 + \xi)(\lambda_k \|\bar{x} - x_k\| + e(\Psi(\bar{x}), \Psi(x_k))) < (1 + \xi)(\lambda + \varepsilon)\|\bar{x} - x_k\|
\]

\[
< (\lambda + \varepsilon + \gamma) \|x_k - \bar{x}\| < \beta.
\]

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 \( u_1, u_2 \in B(\bar{x}, \delta) \) 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 \( y := 0 \in B(\bar{y}_1 + \bar{z}, \beta) \). 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

\[
d\left(\bar{x}, (A_k(\cdot, x_k) + F(\cdot))^{-1}(0)\right) \leq \kappa_k d(0, F(\bar{x}) + \bar{z}) \leq \kappa_k \|\bar{y}_1 + \bar{z}\|
\]

\[
< \frac{\kappa}{1 - \kappa \lambda} \|\bar{x} - \bar{y}_2\| \leq \frac{\kappa}{1 - \kappa \lambda} (1 + \xi) d(\bar{y}_2, A_k(\bar{x}, x_k))
\]

\[
< \frac{\kappa(\lambda + \varepsilon + \gamma)}{1 - \kappa \lambda} \|\bar{x} - x_k\|.
\]
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

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

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 super-linear convergence of \((x_k)_{k \in \mathbb{N}}\). \( \square \)

**Remark 5.2.**

(i) **Inspecting the proof, one sees that if** \( \lim_{k \to +\infty} \lambda_k = 0 \) (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** \( \lim_{k \to +\infty} \lambda_k = 0 \) **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 a point \( x_k \), we have that

\[
d(\bar{y}_2, \lambda_k(\bar{x} - x_k) + \Psi(x_k)) \leq e(\Psi(\bar{x}), \lambda_k(\bar{x} - x_k) + \Psi(x_k))
\]

\[
\leq (\lambda_k + \gamma/2) \|x_k - \bar{x}\| < \gamma \|x_k - \bar{x}\|.
\]

(iii) **Suppose that** \( F \) **and** \( \Psi \) **act between two different Banach spaces** \( X \) **and** \( Y \) **and let** \( g_k : X \to Y \) **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 a relationship of the above theorem with the results in [34]. Recall that the concept of the strict pre-derivative 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

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

Such a mapping \( H \) is then called the **strict pre-derivative** of \( S \) at \( x_0 \). Using terminology in [33], 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 pseudo **strictly \( H \)-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 pseudo strictly $H$-differentiable at $(x_0, y_0)$ and if there is $\kappa > 0$ such that $H(x) \subset \kappa \|x\|B_Y$ for each $x \in X$, then for each $l > \kappa$ there are neighborhoods $V$ of $x_0$ and $W$ of $y_0$ such that

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

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 [34, Corollary 4.4.2]. It was supposed 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

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

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

$$\Psi(y) \subset \Psi(\bar{x}) + (\varepsilon + \gamma)\|x - \bar{x}\|B_X = B[y_2, (\varepsilon + \gamma)\|x - \bar{x}\|].$$

So, $e(\Psi(x), \Psi(\bar{x})) = d(y_2, \Psi(x)) \leq (\varepsilon + \gamma)\|x - \bar{x}\|$. Therefore the continuity assumption in (4.2.3) is satisfied. Further, the assumption that $F^{-1}$ is pseudo strictly $H$-differentiable at $(\bar{y}_1, \bar{x})$ and that there is $\kappa > 0$ such that $H(y) \subset \kappa \|y\|B_X$ for each $y \in X$, implies 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 [34, 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

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

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

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

First, let us present a global version of Theorem 3.2 (see [20] 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, 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 $B[x_0, r]$ with the constant $L$, and that $\Phi + G$ has closed graph, the set-valued mapping $\Phi + G$ is metrically regular on $V(\Phi + G, x_0, \frac{\tau}{4}, R)$, where $R = \min\{s, \frac{r}{2}\}$, 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 \((\bar{x},y) \in V(\Phi + G, x_0, \frac{r}{2}, R)\) 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 [14, 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
\[
(6.1) \quad f(x, y) + \frac{1}{\tau + \varepsilon}\|x - u\| \geq f(u, y) \quad \text{for all} \quad x \in X.
\]
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
\[
w_n \in G(u_n) \quad \text{for each} \quad n \in \mathbb{N} \quad \text{and} \quad \lim_{n \to +\infty} d(y, \Phi(u_n) + w_n) = f(u, y) > 0.
\]
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 a 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 \(\liminf_{n \to +\infty} \|u - v_n\| > 0\). Indeed, if there is an infinite set \(N \subseteq \mathbb{N}\) such that \(\lim_{N \ni n \to +\infty} \|u - v_n\| = 0\), then \(\lim_{N \ni n \to +\infty} \|z_n - w_n\| = 0\). Since
\[
y - w_n + z_n \in \Phi(v_n) + z_n \subseteq \Phi(v_n) + G(v_n) \quad \text{whenever} \quad n \in \mathbb{N},
\]
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
\[
\frac{1}{\tau + \varepsilon} \geq \limsup_{n \to +\infty} \frac{f(u, y) - f(v_n, y)}{\|u - v_n\|} \geq \limsup_{n \to +\infty} \frac{d(y - w_n, \Phi(u_n)) - d(y - z_n, \Phi(v_n))}{\|u - v_n\|} \geq \limsup_{n \to +\infty} \frac{(\kappa + \gamma)^{-1}\|u_n - v_n\| - \|z_n - w_n\|}{\|u - v_n\|} \geq \limsup_{n \to +\infty} \frac{(\kappa + \gamma)^{-1}\|u_n - v_n\| - (L + \gamma)\|u_n - v_n\|}{\|u - v_n\|} = \frac{1}{\kappa + \gamma} - L - \gamma > \frac{1}{\tau + \varepsilon},
\]
a contradiction. The claim is proved.
According to Theorem 3.1, it suffices to show that \( m(\bar{x}) \geq 1/\tau \), where

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

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 \( B[x_0, r] \) 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

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

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 \( B[x_0, r] \), 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 \( \lim_{n \to +\infty} \|x - u_n\| > 0 \). Indeed, if there is an infinite set \( N \subset \mathbb{N} \) such that \( \lim_{N \ni n \to +\infty} \|x - u_n\| = 0 \), then \( \lim_{N \ni n \to +\infty} \|z_n - w_n\| = 0 \). 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 \( \Phi + G \) would imply that \( y \in \Phi(x) + G(x) \), which is not the case.

Now, we may estimate that

\[
\limsup_{n \to +\infty} \frac{f(x, y) - f(u_n, y)}{\|x - u_n\|} \geq \limsup_{n \to +\infty} \frac{d(y - w_n, \Phi(x_n)) - d(y - z_n, \Phi(u_n))}{\|x - u_n\|} \geq \limsup_{n \to +\infty} \frac{(\kappa + \varepsilon_n)^{-1}\|x_n - u_n\| - \|z_n - w_n\|}{\|x - u_n\|} \geq \lim_{n \to +\infty} \frac{(\kappa + \varepsilon_n)^{-1}\|x_n - u_n\| - (L + \varepsilon_n)\|x_n - u_n\|}{\|x - u_n\|} \geq \lim_{n \to +\infty} \left( \frac{1}{\kappa + \varepsilon_n} - L - \varepsilon_n \right) = \frac{1}{\tau}.
\]
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 \((\mu_k)_{k \in \mathbb{N}}\) 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

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

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) \( x_{i+1} \in \mathbb{B}[x_i, d(x_i, f(x_{i+1}))] \);
(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 \( x_0, \ldots, x_k \) 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 \( 0 \notin f(x_k) + F(x_k) \), 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 \( i = 0, \ldots, k - 1 \), 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

\[
\|x_k - x_{k+1}\| < (\kappa_k + \varepsilon_k) d(0, f(x_k) + F(x_k)) \\
\leq (\kappa_k + \varepsilon_k) h(f(x_k) - A_{k-1}(x_{k-1}, x_{k-1}), f(x_k) - A_{k-1}(x_k, x_{k-1})) \\
\leq (\kappa_k + \varepsilon_k) \mu_{k-1}\|x_k - x_{k-1}\| \leq \alpha\|x_k - x_{k-1}\|.
\]

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

\[
d(0, f(x_{k+1}) + F(x_{k+1})) < h(0, f(x_{k+1}) - A_k(x_{k+1}, x_k)) \leq \mu_k\|x_{k+1} - x_k\| \\
\leq \mu_k(\kappa_k + \varepsilon_k) d(0, f(x_k) + F(x_k)).
\]

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 $2\kappa_{k+1}\mu_k < 1$, 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, therefore it converges, to $\bar{x} \in B[x_0, r]$ 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

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

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. Concluding remarks. 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, nonlinear complementarity problems and variational inequalities. In this paper, we have studied the super-linear 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 super-linear 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 also by one of the referees, it will be interesting to characterize “good” iterative sequences via reasonable conditions on proximity to the solutions. This is out of the scope of this paper and will be the subject of a forthcoming research.
Acknowledgments. We thank to Assen Dontchev for suggesting to prove a localized version of Theorem 3.2 and also for a fruitful discussion on the early version of the manuscript. The authors are grateful to the anonymous referees for their valuable comments which led to several improvements of this paper.

REFERENCES


