Pure and Applied Functional Analysis
Volume 5, Number 3, 2020, 591–619

Yokohama Publishers
ISSN 2189-3764 ONLINE JOURNAL

CONVERGENCE ANALYSIS OF THE GAUSS-NEWTON METHOD FOR CONVEX INCLUSION PROBLEMS AND CONVEX COMPOSITE OPTIMIZATION

CHONG LI AND KUNG FU NG

Abstract. Using the convex process theory we study the convergence issues of the iterative sequences generated by the Gauss-Newton method for the convex inclusion problem defined by a cone C and a Fréchet differentiable function F (the derivative is denoted by F'). The restriction in our consideration is minimal and, even in the classical case (the initial point x_0 is assumed to satisfy the following two conditions: F' is Lipschitz around x_0 and the convex process T_{x_0}, defined by $T_{x_0}\cdot = F'(x_0)\cdot - C$, is surjective), our results are new in giving sufficient conditions (which are weaker than the known ones) ensuring the convergence of the iterative sequence with initial point x_0. When F is analytic, we study point estimate conditions similar to Smale's conditions for nonlinear analytic equations. The same study is also made for the so-called convex-composite optimization problem (with objective function given as the composite of a convex function with a Fréchet differentiable map).

1. Introduction

In this paper, we consider a pair of two closely related problems. One is known as the convex inclusion problem

$$F(x) \in C, \tag{1.1}$$

where F is a map from a Euclidian space $\mathbb{R}^v$ (or a finite dimensional space) to another $\mathbb{R}^m$ and C is a closed convex set in $\mathbb{R}^m$. The other problem to be considered is

$$\min_{x\in\mathbb{R}^v}(h \circ F)(x), \tag{1.2}$$

where h is a real-valued convex function on $\mathbb{R}^m$ and F is as in problem (1.1). If $h := d(\cdot, C)$, the distance function associated to C, then (1.2) reduces to (1.1) (provided that the latter is solvable). Many problems in optimization theory, such as minimax problems, penalization methods and goal programming, can be cast as problem (1.1) and/or (1.2); see [3, 4, 6, 7, 8, 12, 14, 17] and [23] for many such examples. In particular, in the case when C is the negative cone $\mathbb{R}^m_-$ in $\mathbb{R}^m$ and $F := (f_i)$ with each $f_i : \mathbb{R}^v \to \mathbb{R}$, problem (1.1) is reduced to the well-known

2010 *Mathematics Subject Classification.* 47J15, 65H10; Secondary 41A29.
Key words and phrases. The Gauss-Newton method, convex process, convex composite optimization, the weak-Robinson condition, the weak-Smale condition, majorizing function, convergence criterion.

feasibility problem, which has been studied extensively, especially in the case when each f_i is convex; see for example the survey [1] by Bauschke and Borwein. In [19], Robinson proposed the following algorithm (which is called the extended Newton method) for solving (1.1) (assuming that C is a closed convex cone) with starting point x_0:

Algorithm $\mathbf{A}(x_0)$. For $k = 0, 1, \ldots$, having x_k, determine x_{k+1} as follows.

If $\mathfrak{D}_\infty(x_k) \neq \emptyset$, choose $d_k \in \mathfrak{D}_\infty(x_k)$ to satisfy $\|d_k\| = \min_{d \in \mathfrak{D}_\infty(x_k)} \|d\|$, and set $x_{k+1} = x_k + d_k$, where $\mathfrak{D}_\infty(x)$ is defined by

$$\mathfrak{D}_\infty(x) := \{d \in \mathbb{R}^v : \ F(x) + F'(x)d \in C\} \quad \text{for each } x \in \mathbb{R}^v. \tag{1.3}$$

Since $\mathfrak{D}_\infty(x)$ may be empty for some $x \in \mathbb{R}^v$, the above algorithm is not necessarily well defined in some unfavorable cases (we say that an algorithm is well defined if it generates at least one sequence). Robinson made two important assumptions in [19]. One is

$$\operatorname{Range}(T_{x_0}) = \mathbb{R}^m, \tag{1.4}$$

where T_{x_0} is the convex process defined by

$$T_{x_0} d = F'(x_0)d - C \quad \text{for each } d \in \mathbb{R}^v. \tag{1.5}$$

The second assumption is that F' is Lipschitz continuous (say with modulus K). Under these assumptions (so in particular, $T_{x_0}^{-1}$ is normed: $\|T_{x_0}^{-1}\| < \infty$), it was proved in [19] that a sequence $\{x_k\}$ generated by Algorithm $\mathbf{A}(x_0)$ converges to a solution x^* satisfying $F(x^*) \in C$ provided that the following "convergence criterion" is satisfied:

$$\|x_1 - x_0\| \leq \frac{1}{2K\|T_{x_0}^{-1}\|}. \tag{1.6}$$

In the present paper, we prove the same result with a sharper convergence criterion and under weaker assumptions (allowing particularly that $\|T_{x_0}^{-1}\| = \infty$). Similarly, we establish a convergence result regarding an algorithm in the Gauss-Newton method for solving problem (1.2). This algorithm has been studied in [5, 10, 16, 29] and in a recent work [11] of ours. The main feature of our present approach is that the norm of $T_{x_0}^{-1}$ is allowed to be infinite. Moreover, our convergence criterion for the convergence of a sequence generated by the algorithm is not only sharper than the earlier results but also has the so-called affine invariant property, namely it is unchanged if $f = h \circ F$ is also represented as $f = \tilde{h} \circ \tilde{F}$, where $\tilde{h} = h \circ A^{-1}$, $\tilde{F} = A \circ F$ and A is an inversible transformation.

The paper is organized as follows. In section 2, we list some basic concepts and known facts needed in the sequel. We introduce in section 3 the new notion of the weak-Robinson condition for convex processes and prove some related results for use of the proof of our main result, which is given in section 4. Applications to two special and important cases (Kantorovich's type condition and Smale's condition) are provided in section 5, where we also present a kind of point estimate results for (1.1) and (1.2) which are inspired by the corresponding results of Smale [2, 24, 25] for analytic equations. We end section 5 with some examples about the comparison of the present paper with the known ones.

2. Preliminaries

Let $\mathbf{B}(x,r)$ stand for the open unit ball in $\mathbb{R}^v$ or $\mathbb{R}^m$ with center x and radius r while the corresponding closed ball is denoted by $\overline{\mathbf{B}(x,r)}$. The closed unit ball in $\mathbb{R}^v$ is denoted by $\overline{\mathbf{B}}_{\mathbb{R}^v}$. Let S be a closed convex subset of $\mathbb{R}^v$ or $\mathbb{R}^m$. We use $d(x,S)$ to denote the distance from x to S. Let $h:\mathbb{R}^m\to\mathbb{R}$ be a convex function, F a nonlinear Fréchet differentiable map from $\mathbb{R}^v$ to $\mathbb{R}^m$, and C the set of minimum points of h. We begin with the Gauss-Newton method. Let $\Delta\in(0,+\infty]$, $x\in\mathbb{R}^v$ and let $D_\Delta(x)$ represent the set of all $d\in\mathbb{R}^v$ satisfying $\|d\|\le\Delta$ and

$$h(F(x)+F'(x)d)=\min\{h(F(x)+F'(x)d'): d'\in\mathbb{R}^v, \|d'\|\le\Delta\}. \tag{2.1}$$

Clearly, $d\in D_\Delta(x)$ if and only if d is a solution of the convex minimization problem:

$$\min\{h(F(x)+F'(x)d'): d'\in\mathbb{R}^v, \|d'\|\le\Delta\}. \tag{2.2}$$

Let

$$\mathfrak{D}_\Delta(x)=\{d\in\mathbb{R}^v:\ \|d\|\le\Delta, F(x)+F'(x)d\in C\}. \tag{2.3}$$

Since C is the set of minimum points of h, we note that $\mathfrak{D}_\Delta(x)\subseteq D_\Delta(x)$ for each $x\in\mathbb{R}^v$.

Remark 2.1. (a) If $\Delta<+\infty$, then $D_\Delta(x)\neq\emptyset$ for each $x\in\mathbb{R}^v$.

(b) If $F(x^*)\in C$ then x^* solves (1.2).

(c) Suppose that $\mathfrak{D}_\Delta(x)\neq\emptyset$. Then for each $d\in\mathbb{R}^v$ with $\|d\|\le\Delta$, the following equivalences hold.

$$d\in D_\Delta(x)\Longleftrightarrow d\in\mathfrak{D}_\Delta(x)\Longleftrightarrow d\in\mathfrak{D}_\infty(x)\Longleftrightarrow d\in D_\infty(x). \tag{2.4}$$

Following [5, 10, 16, 29], we consider the following algorithm (which is called the Gauss-Newton method) for solving (1.2); let $\eta\ge1,\Delta\in(0,+\infty]$ and $x_0\in\mathbb{R}^v$.

Algorithm $\mathbf{A}(\eta,\Delta,x_0)$. For $k=0,1,\dots$, having x_k, determine x_{k+1} as follows.

If $h(F(x_k))=\min\{h(F(x_k)+F'(x_k)d): d\in\mathbb{R}^v, \|d\|\le\Delta\}$, then stop; otherwise, choose $d_k\in D_\Delta(x_k)$ to satisfy $\|d_k\|\le\eta d(0,D_\Delta(x_k))$, and set $x_{k+1}=x_k+d_k$.

Throughout, unless explicitly mentioned otherwise, we use L to denote a positive-valued increasing absolutely continuous function on $[0,\Lambda)$ such that $\Lambda\le+\infty$ and $\int_0^\Lambda L(\tau)\ \mathrm{d}\tau=+\infty$. For $\alpha>0$, let $r_\alpha\in(0,\Lambda)$ and $b_\alpha>0$ be defined by

$$\alpha\int_0^{r_\alpha}L(\tau)\ \mathrm{d}\tau=1 \quad\text{and}\quad b_\alpha=\alpha\int_0^{r_\alpha}L(\tau)\tau\ \mathrm{d}\tau; \tag{2.5}$$

thus (see[11, p.615])

$$b_\alpha<r_\alpha. \tag{2.6}$$

Fix a constant $\xi\ge0$, and define

$$\phi_\alpha(t)=\xi-t+\alpha\int_0^t L(\tau)(t-\tau)\ \mathrm{d}\tau \quad\text{for each } t\in[0,\Lambda). \tag{2.7}$$

Thus

$$\phi'_\alpha(t)=-1+\alpha\int_0^t L(\tau)\ \mathrm{d}\tau,\quad \phi''_\alpha(t)=\alpha L(t)\quad\text{for each } t\in[0,\Lambda)$$

and $\phi'''_\alpha(t)$ exists almost everywhere thanks to the assumption that L is absolutely continuous. Let $t_{\alpha,n}$ denote the sequence generated by Newton's method for ϕ_α with initial point $t_{\alpha,0} = 0$:

$$t_{\alpha,n+1} = t_{\alpha,n} - \omega_\alpha(t_{\alpha,n}) \quad \text{for each } n = 0, 1, \ldots, \tag{2.8}$$

where $\omega_\alpha : [0, \Lambda) \to \mathbb{R}$ is defined by

$$\omega_\alpha(t) := \phi'_\alpha(t)^{-1}\phi_\alpha(t) \quad \text{for each } t \in [0, \Lambda). \tag{2.9}$$

In particular, by (2.7) and (2.8),

$$t_{\alpha,1} = \xi. \tag{2.10}$$

The following lemmas are known; see for example [11, 26].

Lemma 2.2. *Suppose that $0 < \xi \le b_\alpha$. Then the following assertions hold.*

(i) *ϕ_α is strictly decreasing on $[0, r_\alpha]$ and strictly increasing on $[r_\alpha, \Lambda)$ with*

$$\phi_\alpha(\xi) > 0, \quad \phi_\alpha(r_\alpha) = \xi - b_\alpha \le 0, \quad \lim_{t \to \Lambda+} \phi_\alpha(t) \ge \xi > 0.$$

Moreover, if $\xi < b_\alpha$, ϕ_α has two zeros, denoted respectively by r^*_α and r^{**}_α, such that

$$\xi < r^*_\alpha < \frac{r_\alpha}{b_\alpha}\xi < r_\alpha < r^{**}_\alpha, \tag{2.11}$$

and, if $\xi = b_\alpha$, ϕ_α has a unique zero r^*_α in (ξ, Λ) (in fact $r^*_\alpha = r_\alpha$).

(ii) *$\{t_{\alpha,n}\}$ is strictly monotonically increasing and converges to r^*_α.*

(iii) *The convergence of $\{t_{\alpha,n}\}$ is of quadratic rate if $\xi < b_\alpha$, and linear if $\xi = b_\alpha$.*

(iv) *ω_α is increasing on $[0, r^*_\alpha)$.*

Lemma 2.3. *Let r_α, b_α and ϕ_α be defined by (2.5) and (2.7). Let $\alpha' > \alpha$ with the corresponding $\phi_{\alpha'}$. Then the following assertions hold.*

(i) *The functions $\alpha \mapsto r_\alpha$ and $\alpha \mapsto b_\alpha$ are strictly decreasing on $(0, +\infty)$.*

(ii) *$\phi_\alpha < \phi_{\alpha'}$ on $(0, \Lambda)$.*

(iii) *The function $\alpha \mapsto r^*_\alpha$ is strictly increasing on the interval $I(\xi)$, where $I(\xi)$ denotes the set of all $\alpha > 0$ such that $\xi \le b_\alpha$.*

Lemma 2.4. *Let $0 \le c < \Lambda$. Define*

$$\chi(t) = \frac{1}{t^2}\int_0^t L(c+\tau)(t-\tau)\ \mathrm{d}\tau \quad \textit{for each } t \in [0, \Lambda - c).$$

Then χ is increasing on $[0, \Lambda - c)$.

3. Convex process and the weak-Robinson condition

The concept of convex process (which was introduced by Rockafeller [21, 22] for convexity problems) plays a key role in the study of this section.

Definition 3.1. A set-valued mapping $T : \mathbb{R}^v \to 2^{\mathbb{R}^m}$ is called a convex process from $\mathbb{R}^v$ to $\mathbb{R}^m$ if it satisfies

(a) $T(x + y) \supseteq Tx + Ty$ for all x, $y \in \mathbb{R}^v$;

(b) $T(\lambda x) = \lambda Tx$ for all $\lambda > 0$, $x \in \mathbb{R}^v$;

(c) $0 \in T0$.

Thus $T : \mathbb{R}^v \to 2^{\mathbb{R}^m}$ is a convex process if and only if its graph $Gr(T)$ is a convex cone in $\mathbb{R}^v \times \mathbb{R}^m$. As usual, the domain, range and inverse of a convex process T are respectively denoted by $\mathrm{D}(T)$, $\mathrm{R}(T)$ and T^{-1}; i.e.,

$$\mathrm{D}(T) = \{x \in \mathbb{R}^v : Tx \neq \emptyset\}, \quad \mathrm{R}(T) = \bigcup\{Tx : x \in \mathrm{D}(T)\}$$

and

$$T^{-1}y = \{x \in \mathbb{R}^v : y \in Tx\}.$$

Obviously T^{-1} is a convex process from $\mathbb{R}^m$ to $\mathbb{R}^v$. Furthermore, for a nonempty set A in $\mathbb{R}^v$ or $\mathbb{R}^v$, it would be convenient to use the notation $\|A\|$ to denote its distance to the origin, that is,

$$\|A\| := \inf\{\|a\| : \ a \in A\}. \tag{3.1}$$

We also make the convention that $A + \emptyset = \emptyset$ for each set A. For the whole paper, we always assume that the domain of any convex process is nonempty.

Definition 3.2. Let T be a convex process. Define

$$\|T\| = \sup\{\|Tx\| : \ x \in \mathrm{D}(T), \ \ \|x\| \leq 1\} \leq +\infty.$$

If $\|T\| < +\infty$, we say that the convex process T is normed.

Let $T,\ S : \mathbb{R}^v \to 2^{\mathbb{R}^m}$ and $Q : \mathbb{R}^m \to 2^{\mathbb{R}^l}$ be convex processes. Recall that $T \subseteq S$ means that $Gr(T) \subseteq Gr(S)$, that is, $Tx \subseteq Sx$ for each $x \in \mathrm{D}(T)$. By definition, one can verify easily that $\|T\| \geq \|S\|$ if $T \subseteq S$ and $\mathrm{D}(T) = \mathrm{D}(S)$. Moreover, $T \subseteq S$ if and only if $T^{-1} \subseteq S^{-1}$. The sum $T + S$, composite $Q\,S$ and multiple λT (with $\lambda \in \mathbb{R}$) are processes defined respectively by

$$(T + S)(x) = Tx + Sx \quad \text{for each } x \in \mathbb{R}^v,$$

$$Q\,S(x) = Q(S(x)) = \bigcup_{y \in S(x)} Q(y) \quad \text{for each } x \in \mathbb{R}^v$$

and

$$(\lambda T)(x) = \lambda(Tx) \quad \text{for each } x \in \mathbb{R}^v.$$

It is well known (and easy to verify) that $T + S$, $Q\,S$, λT are still convex processes and the following assertions hold:

$$\|T + S\| \leq \|T\| + \|S\|, \qquad \|Q\,S\| \leq \|Q\|\,\|S\| \qquad \text{and} \qquad \|\lambda T\| = |\lambda|\|T\|.$$

We also require two propositions below: the first one is known in [20] while the second is a direct consequence of the first one and [19, Theorem 5].

Proposition 3.3. *Suppose that T is a convex process from $\mathbb{R}^v$ to $\mathbb{R}^m$. If $\mathrm{D}(T) = \mathbb{R}^v$, then T is normed. Consequently, T^{-1} is normed if $\mathrm{R}(T) = \mathbb{R}^m$.*

Proposition 3.4. *Let S_1 and S_2 be convex processes from $\mathbb{R}^v$ to $\mathbb{R}^m$ with $\mathrm{D}(S_1) = \mathrm{D}(S_2) = \mathbb{R}^v$ and $\mathrm{R}(S_1) = \mathbb{R}^m$. Suppose that $\|S_1^{-1}\|\|S_2\| < 1$ and that $(S_1 + S_2)(x)$ is closed for each $x \in \mathbb{R}^v$. Then $\mathrm{R}(S_1 + S_2) = \mathbb{R}^m$ and $\|(S_1 + S_2)^{-1}\| \leq \frac{\|S_1^{-1}\|}{1 - \|S_1^{-1}\|\|S_2\|}$.*

The following definition is a modified version of the corresponding notions in [11]. Let L and Λ be as in section 2 and let $L(\mathbb{R}^v, \mathbb{R}^m)$ denote the Banach space of all linear operators from $\mathbb{R}^v$ to $\mathbb{R}^m$. Let $x_0 \in \mathbb{R}^v$ and $r \in (0, +\infty)$.

Definition 3.5. Let $T:\mathbb{R}^m\to 2^{\mathbb{R}^l}$ be a convex process and $H:\mathbb{R}^v\to L(\mathbb{R}^v,\mathbb{R}^m)$ be a mapping. Let $0<r\le\Lambda$. The pair (T,H) is said to satisfy

(a) the weak L-average Lipschitz condition on $\mathbf{B}(x_0,r)$ if

$$\|T(H(x)-H(x_0))\|\le\int_0^{\|x-x_0\|}L(\tau)\,\mathrm{d}\tau\quad\text{for each } x\in\mathbf{B}(x_0,r);$$

(b) the L-average Lipschitz condition on $\mathbf{B}(x_0,r)$, if

$$\|T(H(x)-H(x'))\|\le\int_{\|x'-x_0\|}^{\|x-x'\|+\|x'-x_0\|}L(\tau)\,\mathrm{d}\tau$$

for all $x,\,x'\in\mathbf{B}(x_0,r)$ with $\|x-x'\|+\|x'-x_0\|<r$.

Moreover, in the case when L is a positive constant, we say that (T,H) is

(c) Lipschitz continuous on $\mathbf{B}(x_0,r)$ with modulus L if

$$\|T(H(x)-H(y))\|\le L\|x-y\|\quad\text{for all } x,y\in\mathbf{B}(x_0,r),$$

that is $T\circ H$ is Lipschitz continuous on $\mathbf{B}(x_0,r)$ in the usual sense.

Note that, when L is a positive constant, (b) and (c) are mutually equivalent. An important class of (T,H) satisfying (b) arises from our attempt to extend Smale's α-theory to the inclusion problem (1.1) and the optimization problem (1.2) instead of his nonlinear analytic equations; see section 5.

Lemma 3.6. *Let $g:[0,1]\to\mathbb{R}$ and $G:[0,1]\to\mathbb{R}^m$ be continuous functions, and let $T:\mathbb{R}^m\to 2^{\mathbb{R}^l}$ be a convex process such that* $\mathrm{D}(T)\supseteq\mathrm{R}(G)$. *Then $\int_0^1 G(\tau)\,\mathrm{d}\tau\in\mathrm{D}(T)$. Suppose in addition that*

$$\|TG(t)\|\le g(t)\quad\textit{for each } t\in[0,1]. \tag{3.2}$$

Then

$$\left\|T\int_0^1 G(\tau)\,\mathrm{d}\tau\right\|\le\int_0^1 g(\tau)\,\mathrm{d}\tau. \tag{3.3}$$

Proof. Note first that the convex hull $\mathrm{co}\,(\mathrm{R}(G))$ of $\mathrm{R}(G)$ is contained in $\mathrm{D}(T)$. Let $0\le a<b\le 1$. Since G is a continuous and $[a.b]$ is compact, $\mathrm{R}(G)$ is compact in $\mathbb{R}^m$ and so is $\mathrm{co}\,(\mathrm{R}(G))$. Moreover, we have that

$$\frac{1}{b-a}\int_a^b G(\tau)\,\mathrm{d}\tau=\lim_{k\to\infty}\frac{1}{k}\sum_{i=1}^{k}G(a+\frac{i}{k}(b-a)).$$

This implies that $\frac{1}{b-a}\int_a^b G(\tau)\,\mathrm{d}\tau\in\mathrm{co}\,(\mathrm{R}(G))\subseteq\mathrm{D}(T)$. In particular, $\int_0^1 G(\tau)\,\mathrm{d}\tau\in\mathrm{D}(T)$. Furthermore, there exist $\{\tau_i\}_{i=1}^{l+1}\subseteq[a,b]$ and $\{\alpha_i\}_{i=1}^{l+1}\subseteq[0,1]$ with $\sum_{i=1}^{l+1}\alpha_i=1$ such that

$$\frac{1}{b-a}\int_a^b G(\tau)\,\mathrm{d}\tau=\sum_{i=1}^{l+1}\alpha_i G(\tau_i).$$

Then

$$T\int_a^b G(\tau)\,\mathrm{d}\tau=(b-a)T\left(\frac{1}{b-a}\int_a^b G(\tau)\,\mathrm{d}\tau\right)\supseteq(b-a)\sum_{i=1}^{l+1}\alpha_i TG(\tau_i),$$

and it follows from (3.2) that

$$\left\|T\int_a^b G(\tau)\,\mathrm{d}\tau\right\| \le (b-a)\sum_{i=1}^{l+1}\alpha_i\|TG(\tau_i)\| \le (b-a)\sum_{i=1}^{l+1}\alpha_i g(\tau_i).$$

Hence, by the mean valued theorem,

$$\left\|T\int_a^b G(\tau)\,\mathrm{d}\tau\right\| \le (b-a)g(t) \quad \text{for some } t\in[a,b]. \tag{3.4}$$

In particular, for each $k=1,2,\dots$ and $i=1,2,\dots,k$, we apply the above discussion to $[\frac{i-1}{k},\frac{i}{k}]$ in place of $[a,b]$ and so there exists $t_i^k\in[\frac{i-1}{k},\frac{i}{k}]$ such that

$$\left\|T\int_{\frac{i-1}{k}}^{\frac{i}{k}} G(\tau)\,\mathrm{d}\tau\right\| \le \frac{1}{k}g(t_i^k) \quad \text{for each } k=1,2,\dots \text{ and } i=1,2,\dots,k.$$

Consequently, for each $k=1,2,\dots,$

$$\begin{aligned}\left\|T\textstyle\int_0^1 G(\tau)\,\mathrm{d}\tau\right\| &= \left\|T\textstyle\sum_{i=1}^k\int_{\frac{i-1}{k}}^{\frac{i}{k}} G(\tau)\,\mathrm{d}\tau\right\|\\ &\le \textstyle\sum_{i=1}^k\left\|T\int_{\frac{i-1}{k}}^{\frac{i}{k}} G(\tau)\,\mathrm{d}\tau\right\|\\ &\le \tfrac{1}{k}\textstyle\sum_{i=1}^k g(t_i^k).\end{aligned}$$

Letting $k\to+\infty$, (3.3) holds and the proof is complete. □

For the remainder of the present paper, we shall always assume that C is a nonempty closed convex cone in $\mathbb{R}^m$, and that $F:\mathbb{R}^v\to\mathbb{R}^m$ is a smooth function, that is, its Fréchet derivative is continuous. For $x\in\mathbb{R}^v$ and, we define a convex process T_x by

$$T_x d = F'(x)d - C \quad \text{for each } d\in\mathbb{R}^v. \tag{3.5}$$

Note that $\mathrm{D}(T_x)=\mathbb{R}^v$, and T_x^{-1} is given by

$$T_x^{-1}y = \{d\in\mathbb{R}^v:\ F'(x)d\in y+C\} \quad \text{for each } y\in\mathbb{R}^m. \tag{3.6}$$

Moreover,

$$\mathfrak{D}_\infty(x) = T_x^{-1}(-F(x)) = T_x^{-1}(-F(x)+C) \tag{3.7}$$

(since $C+C=C$). Recall that r_1 is defined by (2.5) with $\alpha=1$, that is,

$$\int_0^{r_1} L(\tau)\,\mathrm{d}\tau = 1. \tag{3.8}$$

Lemma 3.7. *Let $x_0\in\mathbb{R}^v$ and $0<r\le r_1$. Suppose that*

$$\mathrm{R}(F'(x)) \subseteq \mathrm{R}(T_{x_0}) \quad \textit{for each } x\in\mathbf{B}(x_0,,r) \tag{3.9}$$

and that $(T_{x_0}^{-1},F')$ satisfies the weak L-average Lipschitz condition on $\mathbf{B}(x_0,r)$:

$$\|T_{x_0}^{-1}(F'(x)-F'(x_0))\| \le \int_0^{\|x-x_0\|} L(\tau)\mathrm{d}\tau \quad \textit{for each } x\in\mathbf{B}(x_0,r). \tag{3.10}$$

Then, for each $x \in \mathbf{B}(x_0,,r)$*, it holds that* $\mathrm{R}(T_{x_0}) \subseteq \mathrm{R}(T_x)$*,*

$$\text{(3.11)}\quad \mathrm{D}(T_x^{-1}F'(x_0)) = \mathbb{R}^v \quad \textit{and} \quad \|T_x^{-1}F'(x_0)\| \leq \left(1 - \int_0^{\|x-x_0\|} L(\tau)\,\mathrm{d}\tau\right)^{-1}.$$

Proof. Let $x \in \mathbf{B}(x_0,,r)$. Let $S_1 = \mathbf{I}$ (the identity map on $\mathbb{R}^v$) and let $S_2 = T_{x_0}^{-1}(F'(x) - F'(x_0))$. By (3.9), $\mathrm{R}(F'(x) - F'(x_0)) \subseteq \mathrm{R}(T_{x_0})$ and so $\mathrm{D}(S_2) = \mathbb{R}^v$. Note further that S_2 is a normed convex process with closed graph and that

$$\|S_2\| \leq \int_0^{\|x-x_0\|} L(\tau)\,\mathrm{d}\tau < \int_0^{r_1} L(\tau)\,\mathrm{d}\tau = 1$$

(by (3.10) and (3.8)). Thus, by Proposition 3.4, $\mathrm{R}(\mathbf{I} + S_2) = \mathbb{R}^v$, and

$$\text{(3.12)}\qquad \|\,(\mathbf{I} + S_2)^{-1}\,\| \leq \frac{\|\mathbf{I}^{-1}\|}{1 - \|\mathbf{I}^{-1}\|\|S_2\|} \leq \frac{1}{1 - \int_0^{\|x-x_0\|} L(\tau)\,\mathrm{d}\tau}.$$

Further, since $T_{x_0}^{-1}F'(x_0) \supseteq F'(x_0)^{-1}F'(x_0) \supseteq \mathbf{I}$ and $T_{x_0}^{-1}F'(x) \supseteq T_{x_0}^{-1}(F'(x) - F'(x_0)) + T_{x_0}^{-1}F'(x_0)$, it follows that

$$\text{(3.13)}\qquad T_{x_0}^{-1}F'(x) \supseteq S_2 + \mathbf{I}.$$

So $\mathrm{R}(T_{x_0}^{-1}F'(x)) \supseteq \mathrm{R}(S_2 + \mathbf{I}) = \mathbb{R}^v$ and

$$\text{(3.14)}\qquad \begin{array}{rcl} \|- \left(T_{x_0}^{-1}F'(x)\right)^{-1}\| & = & \|\left(T_{x_0}^{-1}F'(x)\right)^{-1}\| \\ & \leq & \|\,(\mathbf{I} + S_2)^{-1}\,\| \\ & \leq & \frac{1}{1 - \int_0^{\|x-x_0\|} L(\tau)\,\mathrm{d}\tau}. \end{array}$$

Moreover, for any $y, z \in \mathbb{R}^v$, the following equivalences are valid:

$$\begin{array}{rcl} z \in -\left(T_{x_0}^{-1}F'(x)\right)^{-1} y & \Longleftrightarrow & y \in T_{x_0}^{-1}F'(x)(-z) \\ & \Longleftrightarrow & F'(x_0)y \in F'(x)(-z) + C \\ & \Longleftrightarrow & F'(x)z \in (-F'(x_0)y) + C \\ & \Longleftrightarrow & z \in T_x^{-1}(-F'(x_0))y. \end{array}$$

Then $T_x^{-1}(-F'(x_0)) = -\left(T_{x_0}^{-1}F'(x)\right)^{-1}$. Hence

$$\mathrm{D}(T_x^{-1}F'(x_0)) = \mathrm{R}(T_{x_0}^{-1}F'(x)) = \mathbb{R}^v,$$

and (3.14) implies that

$$\text{(3.15)}\qquad \|T_x^{-1}(-F'(x_0))\| \leq \frac{1}{1 - \int_0^{\|x-x_0\|} L(\tau)\,\mathrm{d}\tau};$$

thus (3.11) is shown (since $F'(x_0)$ is linear, it is evident that $\|T_x^{-1}(-F'(x_0))\| = \|T_x^{-1}F'(x_0)\|$).

To prove the inclusion $\mathrm{R}(T_{x_0}) \subseteq \mathrm{R}(T_x)$, let $y \in F'(x_0)u - C$ for some $u \in \mathbb{R}^v$. Then, by what we have already proved, there exists $w \in \mathbb{R}^v$ such that $-u \in T_{x_0}^{-1}F'(x)w$, that is, $F'(x_0)(-u) \in F'(x)w + C$. Then $F'(x_0)u \in F'(x)(-w) - C$. Since C is a (convex) cone, it follows that $y \in F'(x_0)u - C \subseteq F'(x)(-w) - C \subseteq \mathrm{R}(T_x)$. This proves that $\mathrm{R}(T_{x_0}) \subseteq \mathrm{R}(T_x)$ and completes the proof. $\square$

In his study of the convex inclusion problem (1.1), Robinson imposed an important condition that T_{x_0} is surjective (henceforth to be referred to as the Robinson condition; see [11]). In light of the preceding lemma, we put forward the following definition giving conditions weaker than the Robinson condition.

Definition 3.8. Let $x_0 \in \mathbb{R}^v$ and $r > 0$. The inclusion (1.1) is said to satisfy the (first) weak-Robinson condition at x_0 on $\mathbf{B}(x_0, r)$ if

$$-F(x_0) \in \mathrm{R}(T_{x_0}) \quad \text{and} \quad \mathrm{R}(F'(x)) \subseteq \mathrm{R}(T_{x_0}) \quad \text{for each } x \in \mathbf{B}(x_0, r). \tag{3.16}$$

Similarly, the inclusion (1.1) is said to satisfy the second weak-Robinson condition at x_0 on $\mathbf{B}(x_0, r)$ if F is C^2 and

$$-F(x_0) \in \mathrm{R}(T_{x_0}) \quad \text{and} \quad \mathrm{R}(F''(x)) \subseteq \mathrm{R}(T_{x_0}) \quad \text{for each } x \in \mathbf{B}(x_0, r). \tag{3.17}$$

To study the relationship between these two notions, we first verify below a lemma, which will be also used in Section 5.

Lemma 3.9. *Let $k \geq 2$, $\delta > 0$ and assume that F is C^k. Suppose that $\mathrm{R}(T_{x_0})$ is closed and that the inclusion (1.1) satisfies the weak-Robinson condition at x_0 on $\mathbf{B}(x_0, \delta)$. Then, for each $i \in \overline{1, k}$,*

$$\mathrm{R}(F^{(i)}(x)) \subseteq \mathrm{R}(T_{x_0}) \quad \textit{for all } x \in \mathbf{B}(x_0, \delta). \tag{3.18}$$

Proof. We proceed by mathematical induction. By the assumed weak-Robinson condition, the result (3.18) holds for $i = 1$. Assume that (3.18) holds for $i = j < k$. Let $x \in \mathbf{B}(x_0, \delta)$ and $z_1, z_2, \ldots, z_{j+1} \in \mathbb{R}^v$. Then, by (3.18), there exists $\delta_0 > 0$ such that $\mathrm{R}(F^{(j)}(x + tz_{j+1})) \subseteq \mathrm{R}(T_{x_0})$ for all t with $|t| \leq \delta_0$. In particular, $F^{(j)}(x + tz_{j+1})(\pm z_1, z_2, \ldots, z_j) \in \mathrm{R}(T_{x_0})$ and so

$$-F^{(j)}(x)(z_1, z_2, \ldots, z_j) = F^{(j)}(x)((-z_1), z_2, \ldots, z_j) \in \mathrm{R}(T_{x_0}).$$

Since $\mathrm{R}(T_{x_0})$ is a cone, it follows that

$$\frac{F^{(j)}(x + tz_{j+1})(z_1, z_2, \ldots, z_j) - F^{(j)}(x)(z_1, z_2, \ldots, z_j)}{t} \in \mathrm{R}(T_{x_0})$$

for all t with $|t| \leq \delta_0$. Passing to the limits and since $\mathrm{R}(T_{x_0})$ is closed, one has $F^{(j+1)}(x)(z_1, z_2, \ldots, z_{j+1}) \in \mathrm{R}(T_{x_0})$ and (3.18) is shown. □

Proposition 3.10. *Suppose that the inclusion (1.1) satisfies the second weak-Robinson condition at x_0 on $\mathbf{B}(x_0, r)$. Then the inclusion (1.1) satisfies the weak-Robinson condition at x_0 on $\mathbf{B}(x_0, r)$. The converse is also true if $\mathrm{R}(T_{x_0})$ is closed.*

Proof. Let $x \in \mathbf{B}(x_0, r)$. By (3.17), we have for each $t \in [0, 1]$ that $\mathrm{R}(F''(x_0 + t(x - x_0))) \subseteq \mathrm{R}(T_{x_0})$ and it follows from Lemma 3.6 that

$$\mathrm{R}\left(\int_0^1 F''(x_0 + t(x - x_0))\mathrm{d}t\right) \subseteq \mathrm{R}(T_{x_0})$$

and hence that

$$\mathrm{R}\left(F'(x) - F'(x_0)\right) = \mathrm{R}\left(\int_0^1 F''(x_0 + t(x - x_0))(x - x_0)\mathrm{d}t\right) \subseteq \mathrm{R}(T_{x_0}).$$

Since $\mathrm{R}(T_{x_0})$ is a convex cone containing $\mathrm{R}(F'(x_0))$, this implies that $\mathrm{R}(F'(x))$ is contained in $\mathrm{R}(T_{x_0})$ and the first assertion of the proposition is clear. The second assertion follows directly from Lemma 3.9. $\square$

We remark that $\mathrm{R}(T_{x_0})$ is closed if $\|T_{x_0}^{-1}\| < \infty$; see [12, Fact 4.1].

Remark 3.11. Let $A^{\ominus}$ denote the negative polar of the subset A of $\mathbb{R}^m$:

$$A^{\ominus} := \{z \in \mathbb{R}^m : \langle z, a\rangle \le 0 \text{ for all } a \in A\}.$$

Following [5] and [11] respectively, $x_0 \in \mathbb{R}^v$ is called

(a) a regular point of the inclusion (1.1) if

$$\ker(F'(x_0)^T) \cap (C - F(x_0))^{\ominus} = \{0\}; \tag{3.19}$$

(b) a quasi-regular point of the inclusion (1.1) if there exist $r \in (0, +\infty]$ and an increasing positive-valued function β on $[0, r)$ such that $\mathfrak{D}_\infty(x) \neq \emptyset$ and

$$d(0, \mathfrak{D}_\infty(x)) \le \beta(\|x - x_0\|)\, d(F(x), C) \text{ for all } x \in \mathrm{B}(x_0, r). \tag{3.20}$$

Furthermore, let $\mathbf{r}_{x_0}$ denote the supremum $\mathbf{r}_{x_0}$ of r such that (3.20) holds for some increasing positive-valued function β on $[0, r)$, and β_{x_0} the infimum of β such that (3.20) holds on $[0, \mathbf{r}_{x_0})$. We call $\mathbf{r}_{x_0}$ and β_{x_0} respectively the quasi-regular radius and the quasi-regular bound function of the quasi-regular point x_0. Then from [11], the following implications hold for the inclusion (1.1) when C is a closed convex cone:

$$\begin{array}{ccc} \text{Robinson condition at } x_0 & \Longrightarrow & x_0 \text{ is a regular point} \\ \Downarrow & & \Downarrow \\ \text{weak-Robinson condition at } x_0 & & x_0 \text{ is a quasi-regular point.} \end{array}$$

Moreover, the converse of each implication above is not true (see [11]).

The following proposition establishes the relationship between the weak-Robinson condition and the quasi-regularity. To verify this proposition, we need first a lemma, which will also be used in the next section. The pair L, Λ are as explained in section 2.

Lemma 3.12. *Let $x_0, x, x' \in \mathbb{R}^v$ be such that $\|x - x'\| + \|x' - x_0\| < \Lambda$ and $\mathrm{R}(F'(z)) \subseteq \mathrm{R}(T_{x_0})$ for each z in the line-segment $[x', x]$. Suppose that*

$$\begin{aligned} &\|T_{x_0}^{-1}(F'(z) - F'(x'))\| \\ \le\ &\int_{\|x'-x_0\|}^{\|z-x'\|+\|x'-x_0\|} L(\tau)\,\mathrm{d}\tau \quad \text{for each } z \in [x', x]. \end{aligned} \tag{3.21}$$

Then $T_{x_0}^{-1}\int_0^1 (F'(x' + \tau(x - x')) - F'(x'))(x' - x)\,\mathrm{d}\tau \neq \emptyset$ and

$$\begin{aligned} &\left\| T_{x_0}^{-1}\int_0^1 (F'(x' + \tau(x - x')) - F'(x'))(x' - x)\,\mathrm{d}\tau \right\| \\ \le\ &\int_0^{\|x-x'\|} L(\|x' - x_0\| + \tau)(\|x - x'\| - \tau)\,\mathrm{d}\tau. \end{aligned} \tag{3.22}$$

Proof. Define G and g respectively by

$$G(t) := (F'(x' + t(x - x')) - F'(x'))(x' - x) \quad \text{for each } t \in [0, 1]$$

and

$$g(t) := \int_{\|x'-x_0\|}^{t\|x-x'\|+\|x'-x_0\|} L(\tau)\|x - x'\|\,\mathrm{d}\,\tau \quad \text{for each } t \in [0, 1].$$

Then G and g are continuous on $[0,1]$, and, by (3.21), (3.2) holds with T replaced by $T_{x_0}^{-1}$. Note further that $\mathrm{D}(T_{x_0}^{-1}) \supseteq \mathrm{R}(G)$. Thus, Lemma 3.6 is applicable and it follows by elementary calculus that

$$\begin{array}{rl} & \left\| T_{x_0}^{-1} \int_0^1 (F'(x'+\tau(x-x')) - F'(x'))(x'-x) \right\| \\ \leq & \int_0^1 \mathrm{d}\, t \int_{\|x'-x_0\|}^{t\|x-x'\|+\|x'-x_0\|} L(\tau)\|x-x'\| \,\mathrm{d}\,\tau \\ = & \int_0^{\|x-x'\|} L(\|x'-x_0\|+\tau)(\|x-x'\|-\tau)\,\mathrm{d}\tau. \end{array}$$

The proof is complete. □

Proposition 3.13. *Let $x_0 \in \mathbb{R}^v$ and $0 < r \leq r_1$. Suppose that (1.1) satisfies the weak-Robinson condition at x_0 on $\mathbf{B}(x_0, r)$ and that $(T_{x_0}^{-1}, F')$ satisfies the weak L-average Lipschitz condition on $\mathbf{B}(x_0, r)$. Then the following assertions hold.*

(i) *For each $x \in \mathbf{B}(x_0, r)$,*

$$\mathfrak{D}_\infty(x) \neq \emptyset. \tag{3.23}$$

(ii) *If $F(x_0) \notin C$, then x_0 is a quasi-regular point.*

(iii) *If $T_{x_0}^{-1}$ is normed, then x_0 is a quasi-regular point with the quasi-regular radius $\mathbf{r}_{x_0}$ and the quasi-regular bound function β_{x_0} satisfying $\mathbf{r}_{x_0} \geq r$ and*

$$\beta_{x_0}(t) \leq \|T_{x_0}^{-1}\| \left(1 - \int_0^t L(\tau)\,\mathrm{d}\tau\right)^{-1} \quad \textit{for each } t \in [0, r].$$

Proof. (i). By a straightforward verification and making use of the fact that $C+C = C$, one has that

$$T_x^{-1} F'(x_0) T_{x_0}^{-1} \subseteq T_x^{-1} \quad \text{for each } x \in X. \tag{3.24}$$

Let $x \in \mathbf{B}(x_0, r)$. Thanks to the given assumptions, Lemmas 3.7 and 3.12 are applicable to $[x_0, x]$ in place of $[x', x]$. Hence,

$$T_x^{-1} F'(x_0)(x_0 - x) \neq \emptyset \tag{3.25}$$

and

$$T_{x_0}^{-1} \int_0^1 (F'(x_0 + t(x-x_0)) - F'(x_0))\,(x_0 - x)\mathrm{d}\,t \neq \emptyset.$$

This together with (3.11) implies that

$$T_x^{-1} F'(x_0) T_{x_0}^{-1} \int_0^1 (F'(x_0 + t(x-x_0)) - F'(x_0))\,(x_0 - x)\mathrm{d}\,t \neq \emptyset. \tag{3.26}$$

Since T_x^{-1} is a convex process and

$$\begin{array}{rcl} F(x_0) - F(x) & = & \int_0^1 F'(x_0 + t(x-x_0))\,(x_0 - x)\mathrm{d}\,t \\ & = & \int_0^1 (F'(x_0 + t(x-x_0)) - F'(x_0))\,(x_0 - x)\mathrm{d}\,t + F'(x_0)(x_0 - x), \end{array}$$

it follows (3.24) that

$$\begin{array}{rl} & T_x^{-1}(F(x_0) - F(x)) \\ \supseteq & T_x^{-1} F'(x_0) T_{x_0}^{-1} \left(\int_0^1 (F'(x_0 + t(x-x_0)) - F'(x_0))\,(x_0 - x)\mathrm{d}\,t \right) \\ & +(T_x^{-1} F'(x_0))(x_0 - x) \neq \emptyset, \end{array} \tag{3.27}$$

where the nonemptiness assertion holds by (3.25) and (3.26). Similarly, by (3.11), (3.16) and (3.24) again, we have that

$$\emptyset \neq T_x^{-1}F'(x_0)T_{x_0}^{-1}(-F(x_0)) \subseteq T_x^{-1}(-F(x_0)). \tag{3.28}$$

From the convex process property,

$$T_x^{-1}(-F(x)) \supseteq T_x^{-1}(-F(x_0)) + T_x^{-1}(F(x_0) - F(x)), \tag{3.29}$$

we make use of (3.27) and (3.28) to conclude that $T_x^{-1}(-F(x)) \neq \emptyset$, that is, (3.23) holds (because of (3.7)).

(ii). Assume that $F(x_0) \notin C$. Then there exists $0 < \bar{r} < r$ such that $F(x) \notin C$ for each $x \in \overline{\mathbf{B}(x_0, \bar{r})}$. Set $\rho := \min\{d(F(x), C) : \ x \in \overline{\mathbf{B}(x_0, \bar{r})}\}$. Then $\rho > 0$. By (i), $\mathfrak{D}_\infty(x) \neq \emptyset$ for each $x \in \mathbf{B}(x_0, \bar{r})$. Below we will show that there exists a constant $\theta > 0$ such that

$$d(0, \mathfrak{D}_\infty(x)) \leq \theta \quad \text{for each } x \in \mathbf{B}(x_0, \bar{r}). \tag{3.30}$$

Granting this, one sees that

$$d(0, \mathfrak{D}_\infty(x)) \leq \frac{\theta}{\rho} d(F(x), C) \quad \text{for each } x \in \mathbf{B}(x_0, \bar{r}),$$

and so x_0 is a quasi-regular point. To verify (3.30), let $x \in \mathbf{B}(x_0, \bar{r})$. By (3.27),

$$\begin{aligned} & \|T_x^{-1}(F(x_0) - F(x))\| \\ \leq \ & \|T_x^{-1}F'(x_0)\| \left(\left\| T_{x_0}^{-1} \int_0^1 (F'(x_0 + t(x - x_0)) - F'(x_0))\,(x_0 - x) \mathrm{d}\, t \right\| + \bar{r} \right) \\ \leq \ & \|T_x^{-1}F'(x_0)\| \left(\int_0^{\bar{r}} L(\tau)(\bar{r} - \tau)\,\mathrm{d}\,\tau + \bar{r} \right), \end{aligned} \tag{3.31}$$

where the last inequality holds because, by (3.22) (applied to $[x_0, x]$ in place of $[x', x]$),

$$\left\| T_{x_0}^{-1} \int_0^1 (F'(x_0 + t(x - x_0)) - F'(x_0))\,(x_0 - x) \mathrm{d}\, t \right\| \leq \int_0^{\|x - x_0\|} L(\tau)(\|x - x_0\| - \tau)\,\mathrm{d}\,\tau.$$

Further, by (3.28),

$$\|T_x^{-1}(-F(x_0))\| \leq \|T_x^{-1}F'(x_0)\| \|T_{x_0}^{-1}(-F(x_0))\|. \tag{3.32}$$

By (3.29), (3.31) and (3.32), we have that

$$\|T_x^{-1}(-F(x))\| \leq \theta, \tag{3.33}$$

where

$$\theta := \|T_x^{-1}F'(x_0)\| \left(\|T_{x_0}^{-1}(-F(x_0))\| + \int_0^{\bar{r}} L(\tau)(\bar{r} - \tau)\,\mathrm{d}\,\tau + \bar{r} \right).$$

Note that $\theta < +\infty$ by (3.11)), (3.16) and the fact that $\|x - x_0\| \leq \bar{r} < r \leq r_1$. By (3.7), (3.33) means that $d(0, \mathfrak{D}_\infty(x)) \leq \theta$ and so (3.30) is shown.

(iii). Assume that T_{x_0} is normed. Then, by (3.7), (3.24) and Lemma 3.7, one has that, for each $x \in \mathrm{B}(x_0, r)$,

$$\begin{aligned} d(0, \mathfrak{D}_\infty(x)) &= \|T_x^{-1}(C - F(x))\| \\ &\leq \|T_x^{-1}\| d(F(x), C) \\ &\leq \|T_x^{-1} F'(x_0) T_{x_0}^{-1}\| d(F(x), C) \\ &\leq \|T_{x_0}^{-1}\| \left(1 - \int_0^{\|x - x_0\|} L(\tau)\,\mathrm{d}\tau\right)^{-1} d(F(x), C). \end{aligned}$$

Recalling the definition of β defined in Remark 3.11, we complete the proof. □

Remark 3.14. (i) In general, the quasi-regularity at point x_0 doesn't imply the weak-Robinson condition at x_0 even in the case when $(T_{x_0}^{-1}, F')$ is Lipschitz continuous, see [11, Example 6.1].

(ii) It may happen that $\|T_{x_0}^{-1}\| = \infty$; see [12, Example 4.2].

4. Convergence criteria

This section is devoted to establish our main two convergence results in the Gauss-Newton method: the first concerns with the Algorithm $\mathbf{A}(\eta, \Delta, x_0)$ for problem (1.2) while the second concerns with $\mathbf{A}(x_0)$ for problem (1.1).

For the remainder of this paper, we make the following blanket arrangement on notations. Fix a point $x_0 \in \mathbb{R}^v$ and constants $\eta \geq 1$, $\Delta \in (0, +\infty]$. Define ξ and α by

$$\xi = \eta \|T_{x_0}^{-1}(-F(x_0))\| \quad \text{and} \quad \alpha = \frac{\eta}{1 + (\eta - 1)\int_0^\xi L(\tau)\,\mathrm{d}\tau}. \tag{4.1}$$

Let b_α be defined as in (2.5) while r_α^* denotes the smaller zero of ϕ_α defined by (2.7).

For simplicity of statements, C will always denote a closed convex cone in $\mathbb{R}^m$, and when the convex-composite minimization (1.2) or the Algorithms $\mathbf{A}(x_0, \eta, \Delta)$ are considered such as in Theorems 4.1, 5.2, 5.7 and 5.9, we assume further that $C := \operatorname{argmin} h$.

Theorem 4.1. *Let $\{x_n\}$ be a sequence generated by Algorithm $\mathbf{A}(\eta, \Delta, x_0)$. Suppose that the inclusion (1.1) satisfies the weak-Robinson condition at x_0 on $\mathbf{B}(x_0, r_\alpha^*)$ and that $(T_{x_0}^{-1}, F')$ satisfies the L-average Lipschitz condition on $\mathbf{B}(x_0, r_\alpha^*)$. Assume that*

$$\xi \leq \min\{b_\alpha, \Delta\}. \tag{4.2}$$

Then $\{x_n\}$ converges to some x^* such that $F(x^*) \in C$, and the following assertions hold for each $n = 1, 2, \ldots$:

$$\|x_{n+1} - x_n\| \leq (t_{\alpha,n+1} - t_{\alpha,n}) \left(\frac{\|x_n - x_{n-1}\|}{t_{\alpha,n} - t_{\alpha,n-1}}\right)^2, \tag{4.3}$$

$$\|x_n - x_{n-1}\| \leq t_{\alpha,n} - t_{\alpha,n-1}, \tag{4.4}$$

$$F(x_{n-1}) + F'(x_{n-1})(x_n - x_{n-1}) \in C \tag{4.5}$$

and

$$\|x_{n-1} - x^*\| \leq r_\alpha^* - t_{\alpha,n-1}. \tag{4.6}$$

Proof. Since (4.6) follows directly from (4.4), it suffices to show (4.3)-(4.5). Let us first note that, for each $n \geq 1$,

$$\xi \leq t_{\alpha,n} < r_\alpha^* \leq r_1. \tag{4.7}$$

In fact, since $\xi \leq r_\alpha^* \leq r_\alpha$, it follows from (2.5) that

$$\int_0^\xi L(\tau)\,\mathrm{d}\tau \leq \int_0^{r_\alpha} L(\tau)\mathrm{d}\tau = \frac{1}{\alpha}.$$

This together with the definition of α implies that $\alpha \geq \frac{\eta}{1+(\eta-1)\frac{1}{\alpha}}$. This means that $\alpha \geq 1$ and so $r_\alpha^* \leq r_\alpha \leq r_1$ by Lemma 2.3 (i). Hence (4.7) holds thanks to Lemma 2.2. We shall use mathematical induction to verify (4.3)-(4.5). For this end, let $k \geq 1$ and use $\overline{1,k}$ to denote the set of all integers n satisfying $1 \leq n \leq k$. We first verify the following implication:

$$\left\{\begin{array}{l}\text{(4.4) holds for all } n \in \overline{1,k}, \\ \text{(4.5) holds for } n = k\end{array}\right. \Longrightarrow \left\{\begin{array}{l}\text{(4.3) holds for } n = k, \\ \text{(4.5) holds for } n = k+1.\end{array}\right. \tag{4.8}$$

To do this, suppose that (4.4) holds for each $n \in \overline{1,k}$ and (4.5) holds for $n = k$. Write

$$x_k^\tau = \tau x_k + (1-\tau)x_{k-1} \quad \text{for each } \tau \in [0,1]. \tag{4.9}$$

Note that

$$\|x_k - x_0\| \leq \sum_{i=1}^k \|x_i - x_{i-1}\| \leq \sum_{i=1}^k (t_{\alpha,i} - t_{\alpha,i-1}) = t_{\alpha,k} \tag{4.10}$$

and

$$\|x_{k-1} - x_0\| \leq t_{\alpha,k-1} \leq t_{\alpha,k}. \tag{4.11}$$

It follows from (4.9) and (4.7) that $x_k^\tau \in \mathbf{B}(x_0, r_\alpha^*) \subseteq \mathbf{B}(x_0, r_1)$ for each $\tau \in [0,1]$. Note in particular that, by Remark 2.1 and (3.23) in Proposition 3.13 (applied to r_α^* in place of r),

$$D_\infty(x_k) = \mathfrak{D}_\infty(x_k) \neq \emptyset \tag{4.12}$$

(where $D_\infty(x_k)$ and $\mathfrak{D}_\infty(x_k)$ are defined by (2.1) and (2.3) respectively). By (4.10) and (3.11) in Lemma 3.7 (applied to x_k, r_α^* in place of x, r), we have

$$\|T_{x_k}^{-1} F'(x_0)\| \leq \left(1 - \int_0^{\|x_k - x_0\|} L(\tau)\,\mathrm{d}\tau\right)^{-1} \leq \left(1 - \int_0^{t_{\alpha,k}} L(\tau)\,\mathrm{d}\tau\right)^{-1}. \tag{4.13}$$

We claim that

$$\emptyset \neq (T_{x_k}^{-1} F'(x_0)) T_{x_0}^{-1}[-F(x_k) + F(x_{k-1}) + F'(x_{k-1})(x_k - x_{k-1})] \subseteq D_\infty(x_k). \tag{4.14}$$

By (3.11), to prove the above nonemptiness assertion, it is sufficient to show that

$$T_{x_0}^{-1}[-F(x_k) + F(x_{k-1}) + F'(x_{k-1})(x_k - x_{k-1})] \neq \emptyset. \tag{4.15}$$

Since

$$-F(x_k) + F(x_{k-1}) = \int_0^1 F'(x_{k-1} + \tau(x_k - x_{k-1}))(x_{k-1} - x_k)\,\mathrm{d}\tau,$$

(4.15) follows from Lemma 3.12 (applied to $[x_{k-1}, x_k]$ in place of $[x', x]$ and noting that $\|x_k - x_{k-1}\| + \|x_{k-1} - x_0\| \le t_{\alpha,k} < r_\alpha^* < \Lambda$). This and (3.22) imply that

$$\begin{array}{rl} & \|T_{x_0}^{-1}[-F(x_k) + F(x_{k-1}) + F'(x_{k-1})(x_k - x_{k-1})]\| \\ = & \left\|T_{x_0}^{-1} \int_0^1 \left(F'(x_k^\tau) - F'(x_{k-1})(x_{k-1} - x_k)\right) d\tau\right\| \\ \le & \int_0^{\|x_k - x_{k-1}\|} L(\|x_{k-1} - x_0\| + \tau)(\|x_k - x_{k-1}\| - \tau)\, \mathrm{d}\tau \\ \le & \int_0^{\|x_k - x_{k-1}\|} L(t_{\alpha,k-1} + \tau)(\|x_k - x_{k-1}\| - \tau)\, \mathrm{d}\tau \end{array} \tag{4.16}$$

(recalling that L is increasing and $\|x_{k-1} - x_0\| \le t_{\alpha,k-1}$ by (4.11)). Similar but using (2.7), (2.8) and (2.8), we have that

$$\begin{array}{rl} \phi_\alpha(t_{\alpha,k}) & = \alpha \int_0^{t_{\alpha,k} - t_{\alpha,k-1}} L(t_{\alpha,k-1} + \tau)(t_{\alpha,k} - t_{\alpha,k-1} - \tau)\, \mathrm{d}\tau \\ & \ge \alpha \frac{(t_{\alpha,k} - t_{\alpha,k-1})^2}{\|x_k - x_{k-1}\|^2} \int_0^{\|x_k - x_{k-1}\|} L(t_{\alpha,k-1} + \tau)(\|x_k - x_{k-1}\| - \tau)\, \mathrm{d}\tau \end{array}$$

(see Lemma 2.4), and it follows from (4.16) that

$$\left\|T_{x_0}^{-1}[-F(x_k) + F(x_{k-1}) + F'(x_{k-1})(x_k - x_{k-1})]\right\| \le \frac{\phi_\alpha(t_{\alpha,k})}{\alpha} \left(\frac{\|x_k - x_{k-1}\|}{t_{\alpha,k} - t_{\alpha,k-1}}\right)^2. \tag{4.17}$$

We next show the inclusion in (4.14). Let $z = -F(x_k) + F(x_{k-1}) + F'(x_{k-1})(x_k - x_{k-1})$ and $d \in (T_{x_k}^{-1} F'(x_0)) T_{x_0}^{-1}(z)$, that is, $d \in (T_{x_k}^{-1} F'(x_0))u$ for some $u \in T_{x_0}^{-1}(z)$. We have to show that $d \in D_\infty(x_k)$. Note that $F'(x_k)d \in F'(x_0)u + C$ and $F'(x_0)u \in z + C$, so $F'(x_k)d \in z + C + C = z + C$, since C is a convex cone. Since (4.5) holds for $n = k$, it follows from the definition of z that

$$F(x_k) + F'(x_k)d \in F(x_{k-1}) + F'(x_{k-1})(x_k - x_{k-1}) + C \subseteq C + C = C,$$

that is $d \in D_\infty(x_k)$ as required to show. Therefore, (4.14) is valid and it follows from (4.13) and (4.17) that

$$\begin{array}{rl} & d(0, D_\infty(x_k)) \\ \le & \|\left(T_{x_k}^{-1} F'(x_0)\right) T_{x_0}^{-1}[-F(x_k) + F(x_{k-1}) + F'(x_{k-1})(x_k - x_{k-1})]\| \\ \le & \left(1 - \int_0^{t_{\alpha,k}} L(\tau)\, \mathrm{d}\tau\right)^{-1} \frac{\phi_\alpha(t_{\alpha,k})}{\alpha} \left(\frac{\|x_k - x_{k-1}\|}{t_{\alpha,k} - t_{\alpha,k-1}}\right)^2 \\ \le & -\frac{1}{\eta} \frac{\phi_\alpha(t_{\alpha,k})}{\phi'_\alpha(t_{\alpha,k})} \left(\frac{\|x_k - x_{k-1}\|}{t_{\alpha,k} - t_{\alpha,k-1}}\right)^2, \end{array} \tag{4.18}$$

where the last inequality holds because, by (4.7) and (4.1),

$$1 \le \eta = \alpha \left(1 + (\eta - 1) \int_0^\xi L(\tau)\, \mathrm{d}\tau\right) \le \alpha \left(1 + (\eta - 1) \int_0^{t_{\alpha,k}} L(\tau)\, \mathrm{d}\tau\right)$$

(since $t_{\alpha,k} \ge t_{\alpha,1} = \xi$), and so

$$\frac{\eta}{\alpha} \left(1 - \int_0^{t_{\alpha,k}} L(\tau)\, \mathrm{d}\tau\right)^{-1} \le \left(1 - \alpha \int_0^{t_{\alpha,k}} L(\tau)\, \mathrm{d}\tau\right)^{-1} = -\phi'_\alpha(t_{\alpha,k})^{-1}.$$

By (4.18) and (2.8), we have

$$\begin{array}{rcl} \eta d(0, D_\infty(x_k)) & \le & -\frac{\phi_\alpha(t_{\alpha,k})}{\phi'_\alpha(t_{\alpha,k})} \left(\frac{\|x_k - x_{k-1}\|}{t_{\alpha,k} - t_{\alpha,k-1}}\right)^2 \\ & = & (t_{\alpha,k+1} - t_{\alpha,k}) \left(\frac{\|x_k - x_{k-1}\|}{t_{\alpha,k} - t_{\alpha,k-1}}\right)^2. \end{array} \tag{4.19}$$

Since $\left(\frac{\|x_k-x_{k-1}\|}{t_{\alpha,k}-t_{\alpha,k-1}}\right)^2 \le 1$ by assumptions of (4.8), it follows from (4.19) that

$$d(0, D_\infty(x_k)) \le \eta d(0, D_\infty(x_k)) \le t_{\alpha,k+1} - t_{\alpha,k}. \tag{4.20}$$

Noting that, by Lemma 2.2(iv),

$$t_{\alpha,k+1} - t_{\alpha,k} = -\phi'_\alpha(t_{\alpha,k})^{-1}\phi_\alpha(t_{\alpha,k}) \le -\phi'_\alpha(t_{\alpha,0})^{-1}\phi_\alpha(t_{\alpha,0}) = \xi \le \Delta,$$

it follows that $d(0, D_\infty(x_k)) \le \Delta$, which together with (4.12) implies that there exists $d_0 \in \mathbb{R}^v$ with $\|d_0\| \le \Delta$ such that $F(x_k) + F'(x_k)d_0 \in C$. Consequently, by Remark 2.1,

$$D_\Delta(x_k) = \mathfrak{D}_\Delta(x_k) \ne \emptyset$$

and

$$d(0, D_\Delta(x_k)) = d(0, D_\infty(x_k)). \tag{4.21}$$

In particular, in view of Algorithm $\mathbf{A}(\eta, \Delta, x_0)$, (4.5) holds for $n = k+1$. Moreover, (4.21) together with (4.19) implies that

$$\|x_{k+1} - x_k\| = \|d_k\| \le \eta d(0, D_\Delta(x_k)) \le (t_{\alpha,k+1} - t_{\alpha,k}) \left(\frac{\|x_k - x_{k-1}\|}{t_{\alpha,k} - t_{\alpha,k-1}}\right)^2.$$

That shows that (4.3) holds for $n = k$ and implication (4.8) is proved.

Now we are ready to prove that (4.3)-(4.5) hold for each $n = 1, 2 \dots$. By the weak-Robinson condition assumption, we have from (3.7) that

$$D_\infty(x_0) = \mathfrak{D}_\infty(x_0) = T_{x_0}^{-1}(-F(x_0)) \ne \emptyset. \tag{4.22}$$

Hence, by (4.1) and (2.10),

$$\eta d(0, D_\infty(x_0)) \le \eta \|T_{x_0}^{-1}(-F(x_0))\| = \xi = t_{\alpha,1} - t_{\alpha,0} \le \Delta. \tag{4.23}$$

Since $\eta \ge 1$, it follows from (4.22) that there exists $d \in D_\infty(x_0)$ such that $\|d\| \le \Delta$. Thus, $d(0, D_\Delta(x_0)) = d(0, D_\infty(x_0))$ and, by Remark 2.1,

$$D_\Delta(x_0) = \mathfrak{D}_\Delta(x_0) \ne \emptyset.$$

In particular, it follows from Algorithm $\mathbf{A}(\eta, \Delta, x_0)$ that $F(x_0) + F'(x_0)d_1 \in C$ and so (4.5) holds for $n = 1$. Furthermore, by (4.23) and Algorithm $\mathbf{A}(\eta, \Delta, x_0)$, one has that $\|d_1\| \le \eta d(0, D_\infty(x_0)) \le t_{\alpha,1} - t_{\alpha,0}$, i.e., $\|x_1 - x_0\| \le t_{\alpha,1} - t_{\alpha,0}$. This shows that (4.4) holds for $n = 1$. Thus implication (4.8) is applicable to concluding that (4.3) holds for $n = 1$. Assume that (4.4), (4.3) and (4.5) hold for all $n \in \overline{1,k}$. Then, we have that

$$\|x_{k+1} - x_k\| \le (t_{\alpha,k+1} - t_{\alpha,k}) \left(\frac{\|x_k - x_{k-1}\|}{t_{\alpha,k} - t_{\alpha,k-1}}\right)^2 \le t_{\alpha,k+1} - t_{\alpha,k}.$$

This shows that (4.4) holds for $n = k+1$. Furthermore, we apply (4.8) to get that (4.5) holds for $n = k+1$. Finally, applying (4.8) to $k+1$ in place of k, one has that (4.3) holds for $n = k+1$. This completes the proof. $\square$

Recall that b_1 and the function ϕ_1 are defined respectively by (2.5) and (2.7) with $\alpha = 1$. Let r_1^* be the smaller zero point of the function ϕ_1 (see the opening paragraph of this section for notation arrangements).

Theorem 4.2. *Suppose that the inclusion* (1.1) *satisfies the weak-Robinson condition at* x_0 *on* $\mathbf{B}(x_0, r_1^*)$ *and that* $(T_{x_0}^{-1}, F')$ *satisfies the* L*-average Lipschitz condition on* $\mathbf{B}(x_0, r_1^*)$. *Let* ξ *be given by* (4.1) *with* $\eta = 1$ *and assume that*

$$\xi \le b_1. \tag{4.24}$$

Then, Algorithm $\mathbf{A}(x_0)$ *is well-defined and any sequence* $\{x_n\}$ *so generated converges to a solution* x^* *of* (1.1) *satisfying* (4.3)-(4.6) *with* $\alpha = 1$ *for each* $n = 1, 2, \ldots$. *Moreover, any sequence generated by Algorithm* $\mathbf{A}(x_0)$ *is also a sequence generated by Algorithm* $\mathbf{A}(1, +\infty, x_0)$ *and vice versa.*

Proof. Let h be the distance function of C defined by

$$h(y) := d(y, C) = \inf_{z \in C} \|y - z\| \quad \text{for each } y \in \mathbb{R}^m. \tag{4.25}$$

Let $\Delta = +\infty$ and $\eta = 1$ (so $\alpha = 1$ by (4.1)). Since $\mathfrak{D}_\infty(x_0) \neq \emptyset$ by the weak-Robinson condition, there exists $x_1 \in \mathbb{R}^v$ such that $d_1 := x_1 - x_0 \in \mathfrak{D}_\infty(x_0)$ and $\|d_1\| = d(0, \mathfrak{D}_\infty(x_0))$. Noting that $\mathfrak{D}_\infty(x_0) = D_\infty(x_0)$ by Remark 2.1 (c), x_1 can be regarded as a point obtained by Algorithm $\mathbf{A}(\eta, \Delta, x_0)$ at its first iteration. Then Theorem 4.1 is applicable; it follows from (4.5) and Remark 2.1 that there exists $x_2 \in \mathbb{R}^v$ such that $d_2 := x_2 - x_1 \in \mathfrak{D}_\infty(x_1) = D_\infty(x_1)$ with the minimal norm. Hence, x_2 is also a point obtained by Algorithm $\mathbf{A}(\eta, \Delta, x_0)$ at its second iteration. Inductively, we see that, for each k, $\emptyset \neq \mathfrak{D}_\infty(x_k) = D_\infty(x_k)$, and this means that Algorithm $\mathbf{A}(x_0)$ is well-defined and any sequence $\{x_k\}$ so generated is also a sequence generated by Algorithm $\mathbf{A}(\eta, \Delta, x_0)$. Thus, the conclusion follows from Theorem 4.1 and the proof is complete. ☐

Remark 4.3. The convergence criteria given in Theorems 4.1 and 4.2 are affine invariant in the sense described below. Let A be a $m \times m$ nonsingular matrix. Define functions $\tilde{h} := h \circ A^{-1}$ and $\tilde{F} := A \circ F$ and define $\tilde{C} = A(C)$. Then argmin $\tilde{h} = \tilde{C}$ and $h \circ F = \tilde{h} \circ \tilde{F}$. Hence the minimization problem (1.2) and so the corresponding inclusion problem (1.1) can be rewritten respectively as

$$\min_{x \in \mathbb{R}^n} (\tilde{h} \circ \tilde{F})(x) \tag{4.26}$$

and

$$\tilde{F}(x) \in \tilde{C}. \tag{4.27}$$

Moreover $\tilde{T}_{x_0} = A \circ T_{x_0}$ and $\tilde{T}_{x_0}^{-1} = T_{x_0}^{-1} \circ A^{-1}$, where $\tilde{T}_{x_0}$ denotes the convex process (associated with (4.27)) defined by

$$\tilde{T}_{x_0} d := \tilde{F}'(x_0)\, d - \tilde{C}. \tag{4.28}$$

Then the weak-Robinson condition assumed in Theorem 4.1 for (1.1) is equivalent to the corresponding one for (4.27). Likewise, the L-average Lipschitz condition for $(T_{x_0}^{-1}, F')$ is equivalent to that for $(\tilde{T}_{x_0}^{-1}, \tilde{F}')$. Moreover, $\xi = \eta \|T_{x_0}^{-1}(-F(x_0))\| = \eta \|\tilde{T}_{x_0}^{-1}(-\tilde{F}(x_0))\|$. Therefore, the convergence criteria given in Theorems 4.1 and 4.2 for (1.2) and (1.1) coincide respectively with the corresponding ones for (4.26) and (4.27), that is to say, such convergence criteria are affine invariant. Note that the convergence criteria given in [11, Theorem 4.1] and [19, Theorem 2] do not have such property.

Remark 4.4. We exclude the trivial case when $L \equiv 0$ in our study because, in this trivial case, if $(T_{x_0}^{-1}, F')$ satisfies the weak L-average Lipschitz condition on $\mathbf{B}(x_0, r)$, then

$$F(x) - F(x_0) - F'(x_0)(x - x_0) \in C \quad \text{for each } x \in \mathbf{B}(x_0, r),$$

and therefore, under the assumption made in Theorems 4.1 and 4.2, the Gauss-Newton method stops at the first step, that is, $F(x_1) \in C$.

5. Applications

This section is divided into three subsections: for the first two we consider applications of our main results specializing respectively in Kantorovich's type and in the type of the weak γ-condition studied by Wang and Han in [28]. The last subsection is devoted to a similar study of the famous Smale point estimate theory (for analytic equations) for the inclusion problem (1.1) with F assumed to be analytic. We introduce a new notion of the weak-Smale condition for (1.1), and show, under a mild and reasonable assumption, that the weak-Smale condition implies the weak γ-condition and the weak-Robinson condition. Recall the blanket assumption made at the beginning of section 4; in particular, $x_0 \in \mathbb{R}^v$, $\eta \in [1, +\infty)$, and $\Delta \in (0, +\infty)$. Moreover, unless explicitly mentioned otherwise, ξ and α are defined by (4.1).

5.1. **Kantorovich's type condition.** Throughout this subsection, we assume that that L is a positive constant function on $[0, +\infty)$. Then, by (2.5) and (2.7), we have for all $\alpha > 0$ that

$$r_\alpha = \frac{1}{\alpha L}, \quad b_\alpha = \frac{1}{2\alpha L} \tag{5.1}$$

and

$$\phi_\alpha(t) = \xi - t + \frac{\alpha L}{2} t^2 \quad \text{for each } t \geq 0.$$

Moreover, the zeros of ϕ_α are given by

$$\left.\begin{array}{c} r_\alpha^* \\ r_\alpha^{**} \end{array}\right\} = \frac{1 \mp \sqrt{1 - 2\alpha L \xi}}{\alpha L}, \tag{5.2}$$

provided that $\xi \leq \frac{1}{2\alpha L}$. It is also known (see for example [9, 18, 27]) that $\{t_{\alpha,n}\}$ has the closed form

$$t_{\alpha,n} = \frac{1 - q_\alpha^{2^n - 1}}{1 - q_\alpha^{2^n}} r_\alpha^* \quad \text{for each } n = 0, 1, \ldots, \tag{5.3}$$

where

$$q_\alpha := \frac{r_\alpha^*}{r_\alpha^{**}} = \frac{1 - \sqrt{1 - 2\alpha L \xi}}{1 + \sqrt{1 - 2\alpha L \xi}}. \tag{5.4}$$

The following lemma is direct by definition.

Lemma 5.1. *Let α be defined by* (4.1)*, that is,*

$$\alpha = \frac{\eta}{1 + (\eta - 1) L \xi}. \tag{5.5}$$

Then b_α, r^*_α and q_α defined at the beginning of this subsection are given by

$$b_\alpha = \frac{1+(\eta-1)L\xi}{2L\eta}, \tag{5.6}$$

$$r^*_\alpha = \frac{1+(\eta-1)L\xi - \sqrt{1-2L\xi-(\eta^2-1)(L\xi)^2}}{L\eta} \tag{5.7}$$

and

$$q_\alpha = \frac{1-L\xi - \sqrt{1-2L\xi-(\eta^2-1)(L\xi)^2}}{L\eta\xi}. \tag{5.8}$$

In particular, in the case when $\eta = 1$ (and so $\alpha = 1$),

$$r^*_1 = \frac{1-\sqrt{1-2L\xi}}{L} \tag{5.9}$$

$$q_1 = \frac{1-L\xi-\sqrt{1-2L\xi}}{L\xi}. \tag{5.10}$$

Theorem 5.2. *Let $\{x_n\}$ be a sequence generated by Algorithm $\mathbf{A}(\eta, \Delta, x_0)$. Let $L \in (0, +\infty)$ and let α be defined by (5.5) (so r^*_α and q_α are given in (5.7) and (5.8)). Suppose that the inclusion (1.1) satisfies the weak-Robinson condition at x_0 on $\mathbf{B}(x_0, r^*_\alpha)$, and that $(T^{-1}_{x_0}, F')$ is Lipschitz continuous on $\mathbf{B}(x_0, r^*_\alpha)$ with modulus L. Assume that*

$$\xi \le \min\left\{\frac{1}{L(\eta+1)}, \Delta\right\}. \tag{5.11}$$

Then $\{x_n\}$ converges to some x^* with $F(x^*) \in C$ and

$$\|x_n - x^*\| \le \frac{{q_\alpha}^{2^n-1}}{\sum_{i=0}^{2^n-1} {q_\alpha}^i} r^*_\alpha \quad \textit{for each } n = 0, 1, \ldots. \tag{5.12}$$

Proof. By (5.6), the following equivalences hold:

$$\xi \le b_\alpha \Longleftrightarrow \xi \le \frac{1+(\eta-1)L\xi}{2L\eta} \Longleftrightarrow \xi \le \frac{1}{L(1+\eta)}$$

(where the second equivalence holds by elementary verification). Thus (4.2) and (5.11) are the same and so the conclusions in Theorem 4.1 hold. Moreover, by (5.3), we have that

$$r^*_\alpha - t_{\alpha,n} = r^*_\alpha - \frac{1-q_\alpha^{2^n-1}}{1-q_\alpha^{2^n}} r^*_\alpha = r^*_\alpha \left(\frac{1-q_\alpha}{1-q_\alpha^{2^n}}\right) q_\alpha^{2^n-1}$$

and so (5.12) holds from (4.6). □

Letting $\Delta = +\infty$, and $\eta = 1$ (so $\alpha = 1$ and $b_1 = \frac{1}{2L}$ in (5.5) and (5.6)), the following result follows immediately from (5.3), Theorems 4.2 and 5.2.

Theorem 5.3. *Let $L \in (0, +\infty)$, $\xi = \|T_{x_0}^{-1}(-F(x_0))\|$, and let r_1^*, q_1 be defined by* (5.9)*,* (5.10)*. Suppose that the inclusion* (1.1) *satisfies the weak-Robinson condition at x_0 on $\mathbf{B}(x_0, r_1^*)$ and that $(T_{x_0}^{-1}, F')$ is Lipschitz continuous on $\mathbf{B}(x_0, r_1^*)$ with modulus L. Assume that*

$$\xi \le \frac{1}{2L}. \tag{5.13}$$

Then Algorithm $\mathbf{A}(x_0)$ is well-defined and any sequence $\{x_n\}$ so generated converges to a solution x^* of (1.1) *satisfying* (5.12) *with $\alpha = 1$.*

Corollary 5.4. *(Robinson*[19]*) Suppose that T_{x_0} is surjective and that F' is Lipschitz continuous on $\mathbf{B}(x_0, \hat{R})$ with modulus $K > 0$:*

$$\|F'(x) - F'(y)\| \le K\|x - y\| \quad \textit{for all } x, y \in \mathbf{B}(x_0, \hat{R}), \tag{5.14}$$

where

$$\hat{R} = \frac{1 - \sqrt{1 - 2K\|T_{x_0}^{-1}\|\xi}}{K\|T_{x_0}^{-1}\|} \quad \textit{and} \quad \xi = \|T_{x_0}^{-1}(-F(x_0))\|. \tag{5.15}$$

Assume that

$$\|x_1 - x_0\| \le \frac{1}{2K\|T_{x_0}^{-1}\|}. \tag{5.16}$$

Then the conclusions of Theorem 5.3 hold with $r_1^* = \hat{R}$ and

$$q_1 = \frac{1 - K\|T_{x_0}^{-1}\|\xi - \sqrt{1 - 2K\|T_{x_0}^{-1}\|\xi}}{K\|T_{x_0}^{-1}\|\xi}. \tag{5.17}$$

Proof. Since T_{x_0} is surjective, it follows that $\|T_{x_0}^{-1}\| < +\infty$ and the inclusion (1.1) satisfies the weak-Robinson condition at x_0 on $\mathbf{B}(x_0, +\infty)$. Let $L := K\|T_{x_0}^{-1}\|$. Then, (5.10) and (5.17) are consistent. Likewise, r_1^* given in (5.9) equals $\hat{R}$. Furthermore, by the assumed Lipschitz continuity (5.14), one has that

$$\|T_{x_0}^{-1}(F'(x) - F'(y))\| \le \|T_{x_0}^{-1}\|\|F'(x) - F'(y)\| \le L\|x - y\|$$

for all $x, y \in \mathbf{B}(x_0, r_1^*)$. This means that $(T_{x_0}^{-1}, F')$ is Lipschitz continuous on $\mathbf{B}(x_0, r_1^*)$ with modulus L. Since $\xi = \|T_{x_0}^{-1}(-F(x_0))\| = \|x_1 - x_0\|$ by Algorithm $\mathbf{A}(x_0)$ and (3.7), we see that (5.13) and (5.16) are the same. Therefore, the result follows from Theorem 5.3. □

5.2. **Weak γ-condition.** Throughout this subsection, γ denotes an arbitrary but fixed positive constant. The notion of the γ-condition for operators in Banach spaces was introduced in [28] by Wang and Han to study the Smale point estimate theory, which is recently extended in [15] to suit the setting of vector fields or mappings on Riemannian manifolds. Below we give an analogue of this notion to suit the setting of inclusion problems.

Let $k \ge 1$ and assume that F is C^k (kth continuously differentiable) on $\mathbb{R}^v$ (or on a neighbouthbood of x_0). Fix $x \in \mathbb{R}^v$. The kth derivative $F^{(k)}(x)$ at x is a k-multilinear operator from $(\mathbb{R}^v)^k$ to $\mathbb{R}^m$. It follows that, for any $k - 1$ points

$z_1, z_2, \ldots, z_{k-1} \in \mathbb{R}^v$, $T_{x_0}^{-1}(F^{(k)}(x)(z_1, z_2, \ldots, z_{k-1}))$ is a convex process from $\mathbb{R}^v$ to $\mathbb{R}^m$. Define

$$(5.18)\quad \|T_{x_0}^{-1}F^{(k)}(x)\| := \sup\{\|T_{x_0}^{-1}(F^{(k)}(x)(z_1, z_2, \ldots, z_{k-1}))\| : \{z_i\}_{i=1}^{k-1} \subset \mathbf{B}_{\mathbb{R}^v}\}.$$

In particular, for each $j \le k$,

$$(5.19)\qquad \|T_{x_0}^{-1}F^{(k)}(x)z^j\| \le \|T_{x_0}^{-1}F^{(k)}(x)\|\|z\|^j \quad \text{for each } z \in \mathbb{R}^v,$$

where and in the sequel, the z^j denotes, as usual, $(z, \ldots, z) \in (\mathbb{R}^v)^j = \mathbb{R}^v \times \cdots \times \mathbb{R}^v$ for each $z \in \mathbb{R}^v$; moreover, if $z_1, \ldots, z_l \in \mathbb{R}^v$, then $z^j z_1 \ldots z_l$ denotes the corresponding element in $\mathbb{R}^{j+l}$.

Definition 5.5. Let $0 < r \le \frac{1}{\gamma}$. Suppose that F is of the continuous second derivative F'' on $\mathbf{B}(x_0, r)$. We say that $(T_{x_0}^{-1}, F)$ satisfies the weak γ-condition at x_0 on $\mathbf{B}(x_0, r)$ if

$$(5.20)\qquad \|T_{x_0}^{-1}F''(x)\| \le \frac{2\gamma}{(1 - \gamma\|x - x_0\|)^3} \quad \text{for each } x \in \mathbf{B}(x_0, r).$$

Assume, for the remainder of this subsection, that L is the function defined by

$$(5.21)\qquad L(t) = \frac{2\gamma}{(1 - \gamma t)^3} \quad \text{for each } t \text{ with } 0 \le t < \frac{1}{\gamma}.$$

Note that $\int_0^{\frac{1}{\gamma}} L(t)\mathrm{d}t = +\infty$; also, by (2.5), (2.7) and elementary calculation (cf. [26]), one has that for all $\alpha > 0$,

$$(5.22)\qquad r_\alpha = \left(1 - \sqrt{\frac{\alpha}{1+\alpha}}\right)\frac{1}{\gamma}, \quad b_\alpha = \left(1 + 2\alpha - 2\sqrt{\alpha(1+\alpha)}\right)\frac{1}{\gamma}$$

and

$$(5.23)\qquad \phi_\alpha(t) = \xi - t + \frac{\alpha\gamma t^2}{1 - \gamma t} \quad \text{for each } t \text{ with } 0 \le t < \frac{1}{\gamma}.$$

Hence,

$$(5.24)\qquad \xi \le b_\alpha \Longleftrightarrow \xi \le \frac{1 + 2\alpha - 2\sqrt{\alpha(1+\alpha)}}{\gamma}.$$

Moreover, it is known in [26] that, if $\xi \le \frac{1+2\alpha-2\sqrt{\alpha(1+\alpha)}}{\gamma}$, the zeros of ϕ_α are given by

$$(5.25)\qquad \left.\begin{array}{c} r_\alpha^* \\ r_\alpha^{**} \end{array}\right\} = \frac{1 + \gamma\xi \mp \sqrt{(1+\gamma\xi)^2 - 4(1+\alpha)\gamma\xi}}{2(1+\alpha)\gamma}$$

and the sequence $\{t_{\alpha,n}\}$ has the closed form:

$$(5.26)\qquad t_{\alpha,n} = \frac{1 - q_\alpha^{2^n-1}}{1 - q_\alpha^{2^n-1}p_\alpha} r_\alpha^* \quad \text{for each } n = 0, 1, \ldots,$$

where

$$(5.27)\qquad q_\alpha = \frac{1 - \gamma\xi - \sqrt{(1+\gamma\xi)^2 - 4(1+\alpha)\gamma\xi}}{1 - \gamma\xi + \sqrt{(1+\gamma\xi)^2 - 4(1+\alpha)\gamma\xi}}$$

and

$$p_\alpha := \frac{1+\gamma\xi - \sqrt{(1+\gamma\xi)^2 - 4(1+\alpha)\gamma\xi}}{1+\gamma\xi + \sqrt{(1+\gamma\xi)^2 - 4(1+\alpha)\gamma\xi}}. \tag{5.28}$$

Elementarily, it follows that, if $\xi \le b_\alpha$,

$$q_\alpha \le \frac{1}{2} \Longleftrightarrow \gamma\xi \le \frac{4+9\alpha - 3\sqrt{\alpha(9\alpha+8)}}{4}, \tag{5.29}$$

$$q_\alpha = \frac{1}{2} \Longleftrightarrow \gamma\xi = \frac{4+9\alpha - 3\sqrt{\alpha(9\alpha+8)}}{4}, \tag{5.30}$$

and

$$\frac{t_{\alpha,n+1} - t_{\alpha,n}}{t_{\alpha,n} - t_{\alpha,n-1}} \le {q_\alpha}^{2^n - 1} \quad \text{for each } n = 0, 1, \ldots \tag{5.31}$$

(see [26] for example).

Proposition 5.6. *Let $0 < r \le \frac{1}{\gamma}$. Suppose that the inclusion (1.1) satisfies the second weak-Robinson condition at x_0 on $\mathbf{B}(x_0, r)$ and that $(T_{x_0}^{-1}, F)$ satisfies the weak γ-condition at x_0 on $\mathbf{B}(x_0, r)$. Then $(T_{x_0}^{-1}, F')$ satisfies the L-average Lipschitz condition on $\mathbf{B}(x_0, r)$.*

Proof. Let $x, x' \in \mathbf{B}(x_0, r)$ be such that $\|x - x'\| + \|x' - x_0\| < r$. Let $u \in \mathbb{R}^v$ with $\|u\| \le 1$. We have to show that

$$\|T_{x_0}^{-1}(F'(x) - F'(x'))u\| \le \int_{\|x'-x_0\|}^{\|x-x'\|+\|x'-x_0\|} L(\tau)\mathrm{d}\tau. \tag{5.32}$$

By the assumed weak γ-condition,

$$\|T_{x_0}^{-1} F''(x' + t(x - x'))(x - x')u\| \le \frac{2\gamma\|x - x'\|}{(1 - \gamma(\|x' - x_0\| + t\|x - x'\|))^3}.$$

By Lemma 3.6, it follows that

$$\begin{aligned} \|T_{x_0}^{-1}(F'(x) - F'(x'))u\| &= \left\| T_{x_0}^{-1} \int_0^1 F''(x' + t(x - x'))(x - x')u \,\mathrm{d}\,t \right\| \\ &\le \int_0^1 \frac{2\gamma\|x-x'\|}{(1-\gamma(\|x'-x_0\|+t\|x-x'\|))^3} \mathrm{d}\,t \\ &= \int_{\|x'-x_0\|}^{\|x-x'\|+\|x'-x_0\|} L(\tau)\mathrm{d}\tau. \end{aligned}$$

This proves (5.32) and completes the proof. □

As mentioned at the beginning of this section, we assume, unless explicitly mentioned, that $x_0 \in \mathbb{R}^v$, $\eta \in [1, +\infty)$, $\Delta \in (0, +\infty)$ and $\xi = \eta\|T_{x_0}^{-1}(-F(x_0))\|$. Let

$$\alpha = \frac{\eta(1-\gamma\xi)^2}{(\eta - 1) + (2 - \eta)(1 - \gamma\xi)^2}. \tag{5.33}$$

Using the fact that $\eta \in [1, +\infty)$ and $(1-\gamma\xi)^2 \le 1$, one checks that $0 \le \frac{(1-\gamma\xi)^2}{(\eta-1)+(2-\eta)(1-\gamma\xi)^2} \le 1$, and that

$$0 \le \alpha \le \eta. \tag{5.34}$$

Recall that γ denotes any fixed positive constant and that $r_\alpha^* < \frac{1}{\gamma}$ by (2.11) and (5.22).

Theorem 5.7. *Let $\{x_n\}$ be a sequence generated by Algorithm $\mathbf{A}(\eta, \Delta, x_0)$. Let r_α^* and q_α be given respectively by (5.25) and (5.27) with α defined by (5.33). Suppose that the inclusion (1.1) satisfies the second weak-Robinson condition at x_0 on $\mathbf{B}(x_0, r_\alpha^*)$ and that $(T_{x_0}^{-1}, F)$ satisfies the weak γ-condition at x_0 on $\mathbf{B}(x_0, r_\alpha^*)$. Assume that*

$$\xi \leq \min\left\{\frac{1+2\eta-2\sqrt{\eta(1+\eta)}}{\gamma},\ \Delta\right\}. \tag{5.35}$$

Then $\{x_n\}$ converges at a quadratic rate to some x^* with $F(x^*) \in C$ and the following assertions hold:

$$\|x_{n+1} - x_n\| \leq q_\alpha^{2^{n-1}} \|x_n - x_{n-1}\| \quad \textit{for all } n = 1, 2, \ldots, \tag{5.36}$$

and

$$\|x_n - x^*\| \leq {q_\alpha}^{2^n-1} r_\alpha^* \quad \textit{for all } n = 0, 1, \ldots. \tag{5.37}$$

Proof. Let L be defined as in (5.21). Thanks to the given assumptions, Proposition 5.6 implies that $(T_{x_0}^{-1}, F')$ satisfies the L-average condition on $\mathbf{B}(x_0, r_\alpha^*)$. Note further that, by (5.21), $\int_0^\xi L(t)\,\mathrm{d}t = (1-\gamma\xi)^{-2} - 1$; hence α given in (5.33) is consistent with (4.1). Moreover, since the function $t \mapsto 1 + 2t - 2\sqrt{t(1+t)}$ is monotonically decreasing (because $1 + 2t - 2\sqrt{t(1+t)} = (1 + 2t + 2\sqrt{t(1+t)})^{-1}$ for each $t \geq 0$), it follows from (5.34) that

$$\frac{1+2\eta-2\sqrt{\eta(1+\eta)}}{\gamma} \leq \frac{1+2\alpha-2\sqrt{\alpha(1+\alpha)}}{\gamma};$$

so (4.2) holds by (5.35) and (5.24). Hence the conclusions of Theorem 4.1 hold. By (5.31) and (4.3), we have (5.36). Combining (5.26) and (4.6), we have that

$$\|x_n - x^*\| \leq r_\alpha^* - t_{\alpha,n} = r_\alpha^* q_\alpha^{2^n-1}\left(\frac{1-p_\alpha}{1-q_\alpha^{2^n-1}p_\alpha}\right) \quad \text{for each } n = 0, 1, \ldots$$

and so (5.37) holds. □

Theorem 5.8. *Let $\xi = \|T_{x_0}^{-1}(-F(x_0))\|$ (that is, $\eta = 1$), and let r_1^* and q_1 be defined respectively by (5.25) and (5.27) with $\alpha = 1$, that is,*

$$r_1^* = \frac{1+\gamma\xi-\sqrt{(1+\gamma\xi)^2-8\gamma\xi}}{4\gamma} \quad \textit{and} \quad q_1 = \frac{1-\gamma\xi-\sqrt{(1+\gamma\xi)^2-8)\gamma\xi}}{1-\gamma\xi+\sqrt{(1+\gamma\xi)^2-8\gamma\xi}}.$$

Suppose that the inclusion (1.1) satisfies the second weak-Robinson condition at x_0 on $\mathbf{B}(x_0, r_1^*)$ and that $(T_{x_0}^{-1}, F)$ satisfies the weak γ-condition at x_0 on $\mathbf{B}(x_0, r_1^*)$. Assume that

$$\xi \leq \frac{3-2\sqrt{2}}{\gamma}. \tag{5.38}$$

Then Algorithm $\mathbf{A}(x_0)$ is well-defined and any sequence $\{x_n\}$ so generated converges at a quadratic rate to a solution x^* of the inclusion problem (1.1), and (5.36) and (5.37) hold with $\alpha = 1$.

Proof. Let $\Delta = +\infty$ and $\eta = 1$. Then (5.33) and $\alpha = 1$ are consistent. By (5.38), (5.35) holds. Therefore, Theorem 5.7 is applicable (recall from Theorem 4.2 that $\{x_n\}$ is a sequence generated by Algorithm $\mathbf{A}(1, +\infty, x_0)$). □

As in [11], $x_0 \in \mathbb{R}^v$ is called an (η, Δ)-approximate solution of (1.2) if any sequence $\{x_n\}$ generated by Algorithm $\mathbf{A}(\eta, \Delta, x_0)$ converges to a limit x^* solving (1.2) and satisfies Smale's condition:

$$\|x_{n+1} - x_n\| \le \left(\frac{1}{2}\right)^{2^{n-1}} \|x_n - x_{n-1}\| \quad \text{for each } n = 1, 2, \ldots. \tag{5.39}$$

Similarly, with respect to Algorithm $\mathbf{A}(x_0)$, one can define the notion of an approximate solution of (1.1). For the following theorem, we note that

$$\frac{4 + 9\eta - 3\sqrt{\eta(9\eta + 8)}}{4\gamma} < \frac{1 + 2\eta - 2\sqrt{\eta(1 + \eta)}}{\gamma} \tag{5.40}$$

(thanks to $\eta + 8\sqrt{\eta(1 + \eta)} < 3\sqrt{\eta(9\eta + 8)}$, which in turn follows from $2\sqrt{\eta(1 + \eta)} < 1 + 2\eta$ by squaring on both sides).

Theorem 5.9. *Let r_η be defined as in* (5.22)*, that is,*

$$r_\eta = \left(1 - \sqrt{\frac{\eta}{1 + \eta}}\right) \frac{1}{\gamma}.$$

Suppose that the inclusion (1.1) *satisfies the second weak-Robinson condition at x_0 on $\mathbf{B}(x_0, r_\eta)$ and that $(T_{x_0}^{-1}, F)$ satisfies the weak γ-condition at x_0 on $\mathbf{B}(x_0, r_\eta)$. Assume that*

$$\xi \le \min\left\{\frac{4 + 9\eta - 3\sqrt{\eta(9\eta + 8)}}{4\gamma},\ \Delta\right\}. \tag{5.41}$$

Then, x_0 is an (η, Δ)-approximate solution of (1.2). In fact, any sequence $\{x_n\}$ generated by Algorithm $\mathbf{A}(\eta, \Delta, x_0)$ converges to a limit x^* with $F(x^*) \in C$ such that (5.39) *holds.*

Proof. Let $\{x_n\}$ be any sequence generated by Algorithm $\mathbf{A}(\eta, \Delta, x_0)$. Let α be as in Theorem 5.7. An elementary calculation shows that the function $t \mapsto r_t^*$ is monotonically increasing and it follows from (5.34) that $r_\alpha^* \le r_\eta^*$. Consequently, one has that $r_\alpha^* \le r_\eta^* \le r_\eta$. Thus, by the assumption, $(T_{x_0}^{-1}, F)$ satisfies the weak γ-condition at x_0 on $\mathbf{B}(x_0, r_\alpha^*)$. Moreover, one sees from (5.40) that (5.41) implies (5.35). Therefore, one can apply Theorem 5.7 to conclude that the sequence $\{x_n\}$ converges to a solution x^* of (1.2) and (5.36) holds. For (5.39), we need only to show that $q_\alpha \le \frac{1}{2}$. To do this, we need to emphasize the dependence on the parameters and so we write $q(\alpha, \xi)$ for q_α defined by (5.27). Then one checks that $q(\alpha, \xi)$ is monotonically increasing with respect to each of its variables, and sees by (5.29) and (5.30) that $q\left(\alpha, \frac{4+9\alpha-3\sqrt{\alpha(9\alpha+8)}}{4\gamma}\right) = \frac{1}{2}$ for each $0 < \alpha \le \eta$. It follows that

$$q(\alpha, \xi) \le q(\eta, \xi) \le q\left(\eta, \frac{4 + 9\eta - 3\sqrt{\eta(9\eta + 8)}}{4\gamma}\right) = \frac{1}{2},$$

completing the proof. □

Theorem 5.10. *Let $x_0 \in \mathbb{R}^v$ and let $\xi = \|T_{x_0}^{-1}(-F(x_0))\|$ (that is, $\eta = 1$). Suppose that the inclusion* (1.1) *satisfies the second weak-Robinson condition at x_0 on $\mathbf{B}(x_0, (1-\frac{\sqrt{2}}{2})\frac{1}{\gamma})$ and that $(T_{x_0}^{-1}, F)$ satisfies the weak γ-condition at x_0 on $\mathbf{B}(x_0, (1-\frac{\sqrt{2}}{2})\frac{1}{\gamma})$. Assume that*

$$\xi \leq \frac{13 - 3\sqrt{17}}{4\gamma}. \tag{5.42}$$

Then, x_0 is an approximate solution of (1.1).

Proof. Let $\Delta = \infty$, $\eta = \alpha = 1$. Then $\gamma_\eta = (1 - \frac{\sqrt{2}}{2})\frac{1}{\gamma}$, and (5.41)$\Longleftrightarrow$(5.42). Therefore, the result follows from Theorem 5.9 (as in the proof of Theorem 5.8). □

5.3. **Weak-Smale condition.** In his fundamental work on point estimate theory regarding Newton's method for solving the nonlinear analytic equation $F = 0$, where F is an analytic function from a Banach space to another, Smale (cf. [2, 24, 25]) made an important use of his assumption (on initial point x_0) that

$$F'(x_0) \text{ is surjective.} \tag{5.43}$$

In the course of his study, the quantity $\gamma_F(x_0) \in \mathbb{R}$ defined by

$$\gamma_F(x_0) := \sup_{k\geq 2} \left\| \frac{F'(x_0)^\dagger F^{(k)}(x_0)}{k!} \right\|^{\frac{1}{k-1}}, \tag{5.44}$$

also plays a key role, where $F'(x_0)^\dagger$ stands for the Moore-Penrose inverse of $F'(x_0)$. The present subsection is devoted to an attempt to address similar issues for the inclusion problem (1.1). Here and for the whole subsection, F is assumed to be analytic as in Smale's theory. Recall from [13, Page 653] that the inclusion (1.1) satisfies the weak-Smale condition at x_0 if

$$-F(x_0) \in \mathrm{R}(T_{x_0}), \quad \mathrm{R}(F^{(k)}(x_0)) \subseteq \mathrm{R}(T_{x_0}) \quad \text{for each } k = 2, 3, \ldots \tag{5.45}$$

and

$$\gamma_{(F,C)}(x_0) := \sup_{k\geq 2} \left\| \frac{T_{x_0}^{-1} F^{(k)}(x_0)}{k!} \right\|^{\frac{1}{k-1}} < \infty, \tag{5.46}$$

where $\|T_{x_0}^{-1} F^{(k)}(x_0)\|$ is defined as in (5.18).

Remark 5.11. (a) By definition, one sees that if $\|T_{x_0}^{-1}\| < \infty$ and (5.45) holds, then the inclusion (1.1) satisfies the weak-Smale condition.

(b) By Lemma 3.9, we have that if $\mathrm{R}(T_{x_0})$ is closed and that the inclusion (1.1) satisfies the weak-Robinson condition on $\mathbf{B}(x_0, \delta)$ for some $\delta > 0$, then the inclusion (1.1) satisfies (5.45).

(c) As explained in [13, Remark 4.1], in the case when $C = \{0\}$, one has that $\gamma_{(F,C)}(x_0) = \gamma_F(x_0)$ if (5.43) holds.

Without loss of generality, we assume for the remainder that $\gamma_{(F,C)}(x_0) > 0$ (see Remark 4.4). The assertion (a) in following proposition holds trivially by definition while (b) is known in [13, Proposition 4.1].

Proposition 5.12. *Let $\gamma := \gamma_{(F,C)}(x_0)$ (see (5.46)) and suppose that the inclusion (1.1) satisfies the weak-Smale condition at x_0. Then*

(a) *the inclusion (1.1) satisfies the second weak-Robinson condition at x_0 on $\mathbf{B}(x_0, \frac{1}{\gamma})$;*

(b) *$(T_{x_0}^{-1}, F)$ satisfies the weak γ-condition at x_0 on $\mathbf{B}(x_0, \frac{1}{\gamma})$.*

We end this paper with some comparison of our results in section 4 with that reported in [11]. To do this, we continue to use ξ to denote $\eta\|T_{x_0}^{-1}(-F(x_0))\|$ as in sections 4 and 5. Recall that in the discussion of the main results of [11] (see Corollary 4.3 there), it was assumed that $\|T_{x_0}^{-1}\| < +\infty$ and that the quantity $\eta\|T_{x_0}^{-1}\|d(F(x_0), C)$ (to be denoted by $\hat{\xi}$) played an important role in the convergence criterion in [11]. Noting the obvious inclusion

$$T_{x_0}^{-1}(C - F(x_0)) \subseteq T_{x_0}^{-1}(-F(x_0)), \tag{5.47}$$

one has that

$$\|T_{x_0}^{-1}(-F(x_0))\| \le \|T_{x_0}^{-1}(C - F(x_0))\| \le \|T_{x_0}^{-1}\|d(F(x_0), C),$$

that is,

$$\xi \le \hat{\xi}. \tag{5.48}$$

Note also that

$$\|T_{x_0}^{-1}(F'(x) - F'(y))\| \le \|T_{x_0}^{-1}\|\|F'(x) - F'(y)\| \quad \text{for all } x, y \in \mathbb{R}. \tag{5.49}$$

Therefore, if F' satisfies the (weak) $\hat{L}$-average condition on $\mathbf{B}(x_0, r)$ for some positive-valued increasing absolutely continuous function $\hat{L}$, then $(T_{x_0}^{-1}, F')$ satisfies the (weak) L-average condition on $\mathbf{B}(x_0, r)$ with

$$L(\tau) \le \|T_{x_0}^{-1}\|\hat{L}(\tau) \quad \text{for each } 0 < \tau < r. \tag{5.50}$$

Thus Theorem 4.1 extends [11, Corollary 4.3]. The example below shows that the extension is proper and it points to the situation that Theorem 4.1 is applicable but not [11, Corollary 4.3] (note in particular that the strict inequalities in (5.48) and (5.50) hold in this example).

Example 5.13. Let $v = 1$, $m = 2$ and $\lambda \in (\frac{1}{4}, \frac{1}{2}]$, and take $x_0 = 0$. Let F and h be defined by

$$F(x) = \begin{bmatrix} x - \cos x + 1 + \lambda \\ \frac{1}{2}x^2 + x + \lambda \end{bmatrix} \quad \text{for each } x \in \mathbb{R}$$

and

$$h(y_1, y_2) = \max\{y_1,\ 0\} + \max\{0,\ y_2\}| \quad \text{for each } y = (y_1, y_2)^T \in \mathbb{R}^2,$$

respectively. Thus, $C = \{(t_1, t_2) \in \mathbb{R}^2 :\ t_1 \le 0, t_2 \le 0\}$, and

$$(h \circ F)(x) = \max\{x - \cos x + 1 + \lambda,\ 0\} + \max\{0,\ \frac{1}{2}x^2 + x + \lambda\} \quad \text{for each } x \in \mathbb{R}.$$

Note that $F'(x) = \begin{bmatrix} \sin x + 1 \\ x + 1 \end{bmatrix}$ (for each $x \in \mathbb{R}$), and that

$$\|T_{x_0}^{-1}\| = 1, \qquad \|T_{x_0}^{-1}(-F(x_0))\| = \lambda. \tag{5.51}$$

It follows that, for all $x, x' \in \mathbb{R}$,

$$\|F'(x) - F'(x')\| = \sqrt{(\sin x' - \sin x)^2 + (x' - x)^2} \leq \sqrt{2}\, |x - x'|. \tag{5.52}$$

Note by definition that $T_{x_0}^{-1} y = (-\infty, \min\{y_1, y_2\}]$ for each $y = (y_1, y_2)^T \in \mathbb{R}^2$. Hence, for all $x, x' \in \mathbb{R}$,

$$\|T_{x_0}^{-1}(F'(x) - F'(x'))\| \leq \max\{|\sin x - \sin x'|, |x - x'|\} = |x - x'|. \tag{5.53}$$

Hence, $(T_{x_0}^{-1}, F')$ is Lipschitz continuous on $\mathbb{R}$ with modulus $L = 1$. Let $\eta = 1$ and $\Delta = +\infty$ (and so α, defined in [11, Corollary 4.3], is equal to $\|T_{x_0}^{-1}\| = 1$). Then by (5.51) $\xi = \lambda \leq \frac{1}{2} = \frac{1}{2L}$. Hence (5.13) is satisfied and so Theorem 5.3 is applicable. Below we shall show that [11, Corollary 4.3] is not applicable with the initial point x_0. In fact, otherwise, there exist some positive constants Λ, r and a positive-valued increasing absolutely continuous function $\hat{L}$ defined on $[0, \Lambda)$ with $\int_0^\Lambda \hat{L}(t)\, \mathrm{d}t = +\infty$ and $1 < r \leq \Lambda$ such that F' satisfies the $\hat{L}$-average Lipschitz condition on $\mathbf{B}(x_0, r)$ in the sense defined in [11, Definition 2.5] and

$$\hat{\xi} \leq \hat{b}_1, \quad \hat{r}_1^* \leq r \tag{5.54}$$

where $\hat{\xi} := \|T_{x_0}^{-1}\| d(F(x_0), C)$ is defined as in [11, (4.1)] with $\eta = 1$ and $\|T_{x_0}^{-1}\|$ in place of β_{x_0}, and $\hat{b}_1$, $\hat{r}_1^*$ are the corresponding b_α, r_α^* defined as in (2.5) with $\alpha = 1$ and $\hat{L}$ in place of L, i.e.,

$$\int_0^{r_1} \hat{L}(\tau)\ \mathrm{d}\tau = 1 \quad \text{and} \quad b_1 = \int_0^{r_1} \hat{L}(\tau)\tau\ \mathrm{d}\tau.$$

Then, by (5.52) and the assumed $\hat{L}$-average Lipschitz condition, we have

$$\|F'(x') - F'(x)\| = \sqrt{(\sin x' - \sin x)^2 + (x' - x)^2} \leq \int_{|x|}^{|x'-x|+|x|} \hat{L}(\tau)\, \mathrm{d}\tau$$

holds for all x', $x \in (-r, r)$ with $|x' - x| + |x| < r$. In particular,

$$\sqrt{\sin^2 t + t^2} \leq \int_0^t \hat{L}(\tau)\, \mathrm{d}\tau \quad \text{for all } t \in [0, r),$$

where the equality holds when $t = 0$. Differentiating on both sides at $t = 0$, it follows that $\hat{L}(0) \geq \sqrt{2}$. Hence

$$\hat{L}(t) \geq \hat{L}(0) \geq \sqrt{2} \quad \text{for each } t \in [0, \Lambda) \tag{5.55}$$

because $\hat{L}$ is increasing. Let $\hat{\phi}_1$ (resp. $\bar{\phi}_1$) denote the function ϕ_1 defined in (2.7) with $\alpha = 1$, $\xi = \hat{\xi}$ but with L replaced by $\hat{L}$ (resp. $\sqrt{2}$), namely,

$$\hat{\phi}_1(t) = \hat{\xi} - t + \int_0^t \hat{L}(\tau)\ (t - \tau)\ \mathrm{d}\tau \quad \text{for each } t \in [0, \Lambda)$$

and

$$\bar{\phi}_1(t) = \hat{\xi} - t + \int_0^t \sqrt{2}(t - \tau)\ \mathrm{d}\tau = \hat{\xi} - t + \frac{\sqrt{2}}{2} t^2 \quad \text{for each } t \in [0, \Lambda).$$

Then $\bar{\phi}_1 \leq \hat{\phi}_1$ by (5.55), and hence $\bar{\phi}_1(\hat{r}_1^*) \leq \hat{\phi}_1(\hat{r}_1^*) = 0$ with $\hat{r}_1^* \leq r < \Lambda$ (see (5.54)). This means that $\bar{\phi}_1$ has a zero in $(0, \Lambda)$. Noting that $\bar{\phi}_1$ is a quadratic function with real zeros, we have that

$$\hat{\xi} \leq \frac{1}{2\sqrt{2}}. \tag{5.56}$$

Noting that $d(F(x_0), C) = \sqrt{2}\lambda$, it follows that $\hat{\xi} = \|T_{x_0}^{-1}\| d(F(x_0), C) = \sqrt{2}\lambda$. This, together with (5.56), implies that $\sqrt{2}\lambda \leq \frac{1}{2\sqrt{2}}$, which contradicts that $\lambda > \frac{1}{4}$.

Acknowledgments Research of the first author was supported in part by the National Natural Science Foundation of China(grant number 11971429) and Zhejiang Provincial Natural Science Foundation of China (grant numbers LY18A010004). Research of the second author was supported in part by an Earmarked Grants (GRF) from the Research Grant Council of Hong Kong (Project CUHK 14302516).

References

[1] H. H. Bauschke and J. M. Borwein, *On projection algorithms for solving convex feasibility problems*, SIAM Rev. **38** (1996), 367–426.

[2] L. Blum, F. Cucker, M. Shub and S. Smale, *Complexity and Real Computation*, New York: Springer-Verlag, 1997.

[3] J. V. Burke, *Descent methods for composite nondifferentiable optimization problems*, Math. programming **33** (1985), 260–279.

[4] J. V. Burke, *An exact penalization viewpoint of constrained optimization*, SIAM J. Control Optim. **29** (1991), 968–998.

[5] J. V. Burke and M. C. Ferris, *A Gauss-Newton method for convex composite optimization*, Math. Programming **71** (1995), 179–194.

[6] J. V. Burke and R. A. Poliquin, *Optimality conditions for non-finite valued convex composite function*, Math. Programming **57** (1992), 103–120.

[7] R. Flecher, *Second order correction for nondifferentiable optimization*, in: Numerical Analysis, Lecture Notes in Mathematics, G.A. Watson (ed.), vol. 912, Spring, Berlin, 1982, pp. 85–114.

[8] R. Flecher, *Practical Methods of Optimization*, Wiley, New York, 2nd ed. 1987.

[9] W. B. Gragg and R. A. Tapai, *Optimal error bounds for the Newton-Kantorovich theorems*, SIAM J. Numer. Anal. **11** (1974), 10–13.

[10] K. Jittorntrum and M. R. Osborne, *Strong uniqueness and second order convergence in nonlinear discrete approximation*, Numer. Math. **34** (1980), 439–455.

[11] C. Li and K. F. Ng, *Majorizing functions and convergence of the Gauss-Newton method for convex composite optimization*, SIAM J. Optim. **18** (2007), 613–642.

[12] C. Li and K. F. Ng, *Convergence analysis of the Gauss-Newton method for convex inclusion and convex-composite optimization problems*, J. Math. Anal. Appl. **389** (2012), 469–485.

[13] C. Li and K. F. Ng, *Extended Newton methods for conic inequalities: Approximate solutions and the extended Smale α-theory*, J. Math. Anal. Appl. **440** (2016), 636–660.

[14] C. Li and K. F. Ng, *Quantitative analysis for perturbed abstract inequality systems in Banach spaces*, SIAM J. Optim. **28** (2018), 2872–2901.

[15] C. Li and J. H. Wang, *Newton's method on Riemannian manifolds: Smale's point estimate theory under the γ-condition*, IMA J. Numer. Anal. **26** (2006), 228–251.

[16] C. Li and X. H. Wang, *On convergence of the Gauss-Newton method for convex composite optimization*, Math. Programming **91** (2002), 349–356.

[17] K. Madsen, *Minimization of nonlinear approximation function*, Ph.D.Thesis, Institute of Numerical Anakysis, Technical University of Denmark, Lyngby, 1985.

[18] A. M. Ostrowski, *Solutions of Equations in Euclidean and Banach Spaces*, Academic Press, New York, 1973.

[19] S. M. Robinson, *Extension of Newton's method to nonlinear function with values in a cone*, Numer. Math. **19** (1972), 341–347.
[20] S. M. Robinson, *Normed convex process*, Trans. Amer. Math. Soc. **174** (1972), 127–140.
[21] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ. 1970.
[22] R. T. Rockafellar, *Monotone processes of convex and concave type*, vol. 77, Amer. Math. Soc. Memoir, 1967.
[23] R. T. Rockafellar, *First- and second-order epi-differentiability in nonlinear programming*, Trans. Amer. Math. Soc. **307**(1988), 75–108.
[24] M. Shub and S. Smale, *Complexity of Bézout's theorem IV: Probability of success; extensions*, SIAM J. Numer. Anal. **33** (1996), 128–148.
[25] S. Smale, *Newton's method estimates from data at one point*, The Merging of Disciplines: New Directions in Pure, Applied and Computational Mathematics (R.Ewing, K.Gross and C. Martin, eds), New York: Springer, 1986, pp. 185–196.
[26] X. H. Wang, *Convergence of Newton's method and inverse function theorem in Banach space*, Math. Comp. **68** (1999), 169–186.
[27] X. H. Wang, *Convergence of an iteration process*, KeXue Tongbao **20** (1975), 558–559.
[28] X. H. Wang and D. F. Han, *Criterion α and Newton's method*, Chinese J. Numer. Appl. Math. **19** (1997), 96–105.
[29] R. S. Womersley, *Local properties of algorithms for minimizing nonsmooth composite function*, Math. Programming **32** (1985), 69–89

Manuscript received April 30 2019
revised October 14 2019

C. Li
School of Mathematical Sciences, Zhejiang University, Hangzhou 310027, P. R. China
E-mail address: `cli@zju.edu.cn`

K. F. Ng
Department of Mathematics and IMS, Chinese University of Hong Kong, Hong Kong, P. R. China
E-mail address: `kfng@math.cuhk.edu.hk`