ARTICLE IN PRESS YJMAA:16385

J. Math. Anal. Appl. ••• (••••) •••–•••

Contents lists available at SciVerse ScienceDirect

Journal of Mathematical Analysis and Applications

www.elsevier.com/locate/jmaa

Convergence analysis of the Gauss–Newton method for convex inclusion and convex-composite optimization problems

C. Li [a,b,*], K.F. Ng [c]

[a] Department of Mathematics, Zhejiang University, Hangzhou 310027, China
[b] Department of Mathematics, College of Sciences, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia
[c] Department of Mathematics (and IMS), Chinese University of Hong Kong, Hong Kong, China

ARTICLE INFO

Article history:
Received 20 July 2011
Available online xxxx
Submitted by A. Dontchev

Keywords:
Gauss–Newton method
Convex process
Convex-composite optimization
Weak-Robinson condition
Majorizing function
Convergence criterion

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 smooth 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 and have a remarkable property being affine-invariant) ensuring the convergence of the iterative sequence with initial point x_0. 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 smooth map).

1. Introduction

Let X and Y be Banach spaces. Let F be a smooth map from X to Y and let C be a closed convex set in Y. There are two interesting and closely related problems associated to F and C. One is known as the convex inclusion problem

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

The other to be considered is the convex-composite optimization problem

$$\min_{x\in X}(h \circ F)(x), \tag{1.2}$$

where h is a real-valued convex function on Y and F is as in problem (1.1). If $h(\cdot) := \mathrm{d}(\cdot, C)$, 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 (1.2); see [2,4,14,15,23,29,30] for many such examples. Problem (1.1) has been studied extensively and many problems in optimization such as linear semi-infinite optimization and conic programming can be recast into the form (1.1), see for example [5–7,10–12,17,18]. In [25], Robinson proposed the following algorithm (which is called the extended Newton method) for solving (1.1) (assuming that C is a closed (convex) cone and X is a reflexive space) with starting point x_0:

* Corresponding author at: Department of Mathematics, Zhejiang University, Hangzhou 310027, China. Fax: +86 571 87953794.
E-mail address: cli@zju.edu.cn (C. Li).

0022-247X/$ – see front matter
doi:10.1016/j.jmaa.2011.11.062

Algorithm 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

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

Since $\mathfrak{D}_\infty(x)$ may be empty for some $x \in X$, 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 [25]. One is

$$\mathrm{Range}(T_{x_0}) = Y, \tag{1.4}$$

where T_{x_0} is the convex process defined by

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

The second assumption is that F' is Lipschitz continuous (say with the modulus K). Under these assumptions (so in particular, $T_{x_0}^{-1}$ is normed: $\|T_{x_0}^{-1}\| < \infty$), it was proved in [25] 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:

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

In the present paper, we will prove the same result with a sharper convergence criterion and under weaker assumptions. 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 [3,19,22,30] and in a recent work [21] of ours. Our approach covers both cases when $\|T_{x_0}^{-1}\|$ is finite or otherwise. Even for the finite case our results are shaper than the earlier results. To the best of our knowledge, in all the works regarding the Gauss–Newton methods by the earlier researchers, the convergence criteria that have been put forward for the convergence of a sequence generated by their algorithms do not share the so-called affine-invariant property, an important property enjoyed by the classical Kantorovich convergence criterion for Newton method for nonsingular system (cf. [8,9,20]), which means that it is independent of the decompositions of f as $f = h \circ F$ or $f = \tilde{h} \circ \tilde{F}$, 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 introduce the new notion of the weak-Robinson condition for convex processes and prove some related results for use of the proof of our main results, which are given in Section 3; particularly the convergence criteria given in Theorems 3.1 and 3.2 are affine-invariant. Further comments and examples about the comparison of the results of the present paper with the known ones are given in Section 4.

2. Convex process and the weak-Robinson condition

We always assume that X, Y, Z are Banach spaces. Let $\mathbf{B}(x, r)$ stand for the open ball in X or Y with center x and radius r. Let S be a closed convex subset of X or Y. We use $\mathrm{d}(x, S)$ to denote the distance from x to S. The concept of convex process (which was introduced by Rockafellar [27,28] for convexity problems) plays a key role in the study of this paper.

Definition 2.1. A set-valued map $T : X \to 2^Y$ is called a convex process from X to Y if it satisfies

(a) $T(x + y) \supseteq Tx + Ty$ for all $x, y \in X$;
(b) $T(\lambda x) = \lambda Tx$ for all $\lambda > 0$, $x \in X$;
(c) $0 \in T0$.

Thus $T : X \to 2^Y$ is a convex process if and only if its graph $\mathrm{Gr}(T)$ is a convex cone in $X \times Y$. 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.,

$$\mathrm{D}(T) = \{x \in X \colon Tx \neq \emptyset\},$$

$$\mathrm{R}(T) = \bigcup \{Tx \colon x \in \mathrm{D}(T)\}$$

and

$$T^{-1}y = \{x \in X \colon y \in Tx\} \quad \text{for each } y \in Y.$$

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

$$\|A\| = \inf\{\|a\| \colon a \in A\}, \tag{2.1}$$

with the convention that $\|\emptyset\| = +\infty$. We also make the convention that $A + \emptyset = \emptyset$ for each set A.

Definition 2.2. Suppose that T is a convex process. The norm of T is defined by

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

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

Let $T, S : X \to 2^Y$ and $Q : Y \to 2^Z$ be convex processes. Recall that $T \subseteq S$ means that $\mathrm{Gr}(T) \subseteq \mathrm{Gr}(S)$, that is, $Tx \subseteq Sx$ for each $x \in \mathrm{D}(T)$. By definition, one can verify easily that $\|T\| \geqslant \|S\|$ 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 QS and multiple λT (with $0 \neq \lambda \in \mathbb{R}$) are processes defined respectively by

$$(T+S)(x) = Tx + Sx \quad \text{for each } x \in X,$$

$$QS(x) = Q\big(S(x)\big) = \bigcup_{y \in S(x)} Q(y) \quad \text{for each } x \in X$$

and

$$(\lambda T)(x) = \lambda(Tx) \quad \text{for each } x \in X.$$

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

$$\|T + S\| \leqslant \|T\| + \|S\|, \qquad \|QS\| \leqslant \|Q\|\|S\| \quad \text{and} \quad \|\lambda T\| = |\lambda|\|T\|. \tag{2.2}$$

We also require two propositions below: they can be found in [26].

Proposition 2.1. *Let $T : X \to 2^Y$ be a convex process with $\mathrm{D}(T) = X$. Then T is normed. Consequently, T^{-1} is normed if $\mathrm{R}(T) = Y$.*

Proposition 2.2. *Let $S_1, S_2 : X \to 2^Y$ be convex processes with $\mathrm{D}(S_1) = \mathrm{D}(S_2) = X$ and $\mathrm{R}(S_1) = Y$. Suppose that $\|S_1^{-1}\|\|S_2\| < 1$ and that $(S_1 + S_2)(x)$ is closed for each $x \in X$. Then $\mathrm{R}(S_1 + S_2) = Y$ and $\|(S_1 + S_2)^{-1}\| \leqslant \frac{\|S_1^{-1}\|}{1 - \|S_1^{-1}\|\|S_2\|}$.*

The following definition is a modified version of the classic Lipschitz condition. Let L be a positive constant and let $L(X, Y)$ denote the Banach space of all continuous linear operators from X to Y. Let $x_0 \in X$ and $r \in (0, +\infty]$.

Definition 2.3. Let $T : Y \to 2^Z$ be a convex process and $H : X \to L(X, Y)$ be a map. The pair (T, H) is said to be Lipschitz continuous on $\mathbf{B}(x_0, r)$ with modulus L if

$$\big\|T\big(H(x) - H(y)\big)\big\| \leqslant L\|x - y\| \quad \text{for all } x, y \in \mathbf{B}(x_0, r). \tag{2.3}$$

Clearly, if T is normed and H is Lipschitz continuous on $\mathbf{B}(x_0, r)$, then the pair (T, H) is Lipschitz continuous on $\mathbf{B}(x_0, r)$. In fact, it is not difficult to verify that if H is Lipschitz continuous on $\mathbf{B}(x_0, r)$, then the pair (T, H) is Lipschitz continuous on $\mathbf{B}(x_0, r)$ if and only if

$$\sup_{x, y \in \mathbf{B}(x_0, r)} \|T\|_{V_{xy}} < \infty,$$

where $V_{xy} := R(H(x) - H(y))$ for any $x, y \in X$, and $\|T\|_V$ denotes the norm of T restricted on the linear space V defined by

$$\|T\|_V := \sup\{\|T(v)\|: v \in V,\ \|v\| \leqslant 1\}.$$

For a given continuous vector-valued function $G : [a, b] \to Y$, let $\int_a^b G(\tau)\,\mathrm{d}\tau$ denote the usual Riemann integral of G on $[a, b]$, that is, it is the limit of the corresponding Riemann sums (see, for instance, [20, Chapter 17]).

Lemma 2.1. *Let $g : [0, 1] \to \mathbb{R}$ and $G : [0, 1] \to Y$ be continuous. Let Z be a reflexive Banach space. Suppose that $T : Y \to 2^Z$ is a convex process with closed graph such that $\mathrm{D}(T) \supseteq \mathrm{R}(G)$ and that*

$$\big\|TG(t)\big\| \leqslant g(t) \quad \text{for each } t \in [0, 1]. \tag{2.4}$$

Then $T \int_0^1 G(\tau)\,\mathrm{d}\tau \neq \emptyset$ and

$$\left\|T \int_0^1 G(\tau)\,\mathrm{d}\tau\right\| \leqslant \int_0^1 g(\tau)\,\mathrm{d}\tau. \tag{2.5}$$

Please cite this article in press as: C. Li, K.F. Ng, Convergence analysis of the Gauss–Newton method for convex inclusion and convex-composite optimization problems, J. Math. Anal. Appl. (2011), doi:10.1016/j.jmaa.2011.11.062

Proof. Let $k = 1, 2, \ldots$. Set

$$y_k := \frac{1}{k}\sum_{i=1}^{k} G\left(\frac{i}{k}\right) \quad \text{and} \quad y_0 := \int_0^1 G(\tau)\,\mathrm{d}\tau.$$

Then $\lim_k y_k = y_0$. Furthermore, $Ty_k \supseteq \frac{1}{k}\sum_{i=1}^{k} TG(\frac{i}{k})$ and so

$$\|Ty_k\| \leqslant \frac{1}{k}\sum_{i=1}^{k}\left\|TG\left(\frac{i}{k}\right)\right\| \leqslant \frac{1}{k}\sum_{i=1}^{k} g\left(\frac{i}{k}\right).$$

Thus there exists $z_k \in Ty_k$ such that

$$\|z_k\| \leqslant \frac{1}{k}\sum_{i=1}^{k} g\left(\frac{i}{k}\right) + \frac{1}{k}. \tag{2.6}$$

Since

$$\lim_k \frac{1}{k}\sum_{i=1}^{k} g\left(\frac{i}{k}\right) = \int_0^1 g(\tau)\,\mathrm{d}\tau, \tag{2.7}$$

it follows that $\{z_k\}$ is bounded. Since Z is reflexive, by the Eberlein–Smulian Theorem in Functional Analysis (cf. [31]), we may assume that, without loss of generality (using a subsequence if necessary), $\{z_k\}$ converges weakly to one point in Z, say z_0. Consequently, it follows from the Mazur Theorem in Functional Analysis (cf. [31]) that there exists a sequence $\{\tilde{z}_k\}$ with the expression

$$\tilde{z}_k = \sum_{i=1}^{n_k} \alpha_i^k z_{k_i} \quad \text{for each } k = 1, 2, \ldots,$$

where $\{\alpha_i^k\} \subseteq [0, 1]$ satisfies $\sum_{i=1}^{n_k} \alpha_i^k = 1$ for each k, such that $\lim_k \tilde{z}_k = z_0$ and the corresponding sequence $\{\tilde{y}_k\}$ generated by the convex combinations of $\{y_k\}$ converges to y_0, that is,

$$\tilde{y}_k := \sum_{i=1}^{n_k} \alpha_i^k y_{k_i} \to y_0.$$

Since T is a convex process, it follows that

$$\tilde{z}_k = \sum_{i=1}^{n_k} \alpha_i^k z_{k_i} \in \sum_{i=1}^{n_k} \alpha_i^k T y_{k_i} \subseteq T\left(\sum_{i=1}^{n_k} \alpha_i^k y_{k_i}\right) = T\tilde{y}_k \quad \text{for each } k = 1, 2, \ldots.$$

Since $\mathrm{Gr}(T)$ is closed by assumption, one has that

$$z_0 \in Ty_0 = T\int_0^1 G(\tau)\,\mathrm{d}\tau. \tag{2.8}$$

Thus $T\int_0^1 G(\tau)\,\mathrm{d}\tau \neq \emptyset$. Since $\{z_k\}$ converges weakly to z_0, it follows from (2.6), (2.7) and (2.8) that

$$\left\|T\int_0^1 G(\tau)\,\mathrm{d}\tau\right\| \leqslant \|z_0\| \leqslant \lim_k \|z_k\| \leqslant \int_0^1 g(\tau)\,\mathrm{d}\tau.$$

The proof is complete. □

For the remainder of the present paper, we shall always assume that C is a nonempty closed cone in Y, and that $F : X \to Y$ is a smooth map, that is, its Fréchet derivative F' is continuous. Let $x \in X$. We define a convex process T_x by

$$T_x d = F'(x)d - C \quad \text{for each } d \in X. \tag{2.9}$$

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

$$T_x^{-1} y = \{d \in X\colon F'(x)d \in y + C\} \quad \text{for each } y \in Y. \tag{2.10}$$

Please cite this article in press as: C. Li, K.F. Ng, Convergence analysis of the Gauss–Newton method for convex inclusion and convex-composite optimization problems, J. Math. Anal. Appl. (2011), doi:10.1016/j.jmaa.2011.11.062

Since $F'(x)$ is continuous and C is closed, it is easy to verify that T_x and T_x^{-1} are of closed graphs. Moreover,

$$\mathfrak{D}_\infty(x) = T_x^{-1}\big(-F(x)\big) = T_x^{-1}\big(-F(x) + C\big) \tag{2.11}$$

(since $C + C = C$).

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 [21]). In light of the preceding lemma, we put forward the following definition giving a condition weaker than the Robinson condition. Recall that $x_0 \in X$ and $r \in (0, +\infty]$.

Definition 2.4. The inclusion (1.1) is said to satisfy the weak-Robinson condition at x_0 on $\mathbf{B}(x_0, r)$ if

$$-F(x_0) \in \mathrm{R}(T_{x_0}) \quad \text{and} \quad \mathrm{R}\big(F'(x)\big) \subseteq \mathrm{R}(T_{x_0}) \quad \text{for each } x \in \mathbf{B}(x_0, r). \tag{2.12}$$

Clearly, the following implication holds for the inclusion (1.1) when C is a closed cone:

$$\text{Robinson condition at } x_0 \quad \Longrightarrow \quad \text{weak-Robinson condition at } x_0 \text{ on } X.$$

Lemma 2.2. *Let $x_0, x, x' \in X$ be such that $\mathrm{R}(F'(z)) \subseteq \mathrm{R}(T_{x_0})$ for each z in the line-segment $[x', x]$. Suppose that X is reflexive and*

$$\big\|T_{x_0}^{-1}\big(F'(z) - F'(x')\big)\big\| \leqslant L\|z - x'\| \quad \text{for each } z \in \big[x', x\big]. \tag{2.13}$$

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

$$\left\| T_{x_0}^{-1} \int_0^1 \big(F'\big(x' + \tau\big(x - x'\big)\big) - F'\big(x'\big)\big)\big(x' - x\big)\,\mathrm{d}\tau \right\| \leqslant \frac{L}{2}\big\|x - x'\big\|^2. \tag{2.14}$$

Proof. Define G and g respectively by

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

and

$$g(t) := L\big\|x - x'\big\|^2 t \quad \text{for each } t \in [0, 1].$$

Then, G and g are continuous on $[0, 1]$, and it is easy to verify from (2.13) that

$$\big\|T_{x_0}^{-1} G(t)\big\| \leqslant L\big\|x - x'\big\|^2 t = g(t) \quad \text{for each } t \in [0, 1].$$

(Thus (2.4) holds with T replaced by $T_{x_0}^{-1}$.) Moreover, $T_{x_0}^{-1}$ is of closed graph (as we noted before), and $\mathrm{D}(T_{x_0}^{-1}) = \mathrm{R}(T_{x_0}) \supseteq \mathrm{R}(G)$ thanks to the given assumptions. Therefore, Lemma 2.1 is applicable to getting $T_{x_0}^{-1}\int_0^1 (F'(x' + \tau(x - x')) - F'(x')) \times (x' - x)\,\mathrm{d}\tau \neq \emptyset$ and

$$\left\| T_{x_0}^{-1} \int_0^1 \big(F'\big(x' + \tau\big(x - x'\big)\big) - F'\big(x'\big)\big)\big(x' - x\big)\,\mathrm{d}\,\tau \right\| \leqslant \int_0^1 L\big\|x - x'\big\|^2 \tau\,\mathrm{d}\tau = \frac{L}{2}\big\|x - x'\big\|^2.$$

The proof is complete. □

The following proposition provides a stability result for the weak-Robinson condition (2.12) and a solvability result for the approximated inclusion problem (2.15) (with $x \in X$ near to x_0) below:

$$F(x) + F'(x)d \in C. \tag{2.15}$$

Proposition 2.3. *Let $x_0 \in X$, $L \geqslant 0$ and $0 < r \leqslant \frac{1}{L}$. Suppose that (1.1) satisfies the weak-Robinson condition at x_0 on $\mathbf{B}(x_0, r)$ and that $(T_{x_0}^{-1}, F')$ is Lipschitz continuous on $\mathbf{B}(x_0, r)$ with modulus L. Let $x \in \mathbf{B}(x_0, r)$. Then*

$$\mathrm{R}(T_{x_0}) = \mathrm{R}(T_x), \qquad \mathrm{D}\big(T_x^{-1} F'(x_0)\big) = X \tag{2.16}$$

and

$$\big\|T_x^{-1} F'(x_0)\big\| \leqslant \frac{1}{1 - L\|x - x_0\|}. \tag{2.17}$$

Furthermore, if X is additionally reflexive, we have that

$$\mathfrak{D}_\infty(x) \neq \emptyset. \tag{2.18}$$

Please cite this article in press as: C. Li, K.F. Ng, Convergence analysis of the Gauss–Newton method for convex inclusion and convex-composite optimization problems, J. Math. Anal. Appl. (2011), doi:10.1016/j.jmaa.2011.11.062

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

$$\|S_2\| = \|T_{x_0}^{-1}(F'(x) - F'(x_0))\| \leqslant L\|x - x_0\| < 1$$

as $x \in \mathbf{B}(x_0, r)$. Thus, by Proposition 2.2, $\mathrm{R}(\mathbf{I} + S_2) = X$, and

$$\|(\mathbf{I} + S_2)^{-1}\| \leqslant \frac{\|\mathbf{I}^{-1}\|}{1 - \|\mathbf{I}^{-1}\|\|S_2\|} \leqslant \frac{1}{1 - L\|x - x_0\|}. \tag{2.19}$$

Further, since

$$T_{x_0}^{-1}F'(x_0) \supseteq F'(x_0)^{-1}F'(x_0) \supseteq \mathbf{I} \quad \text{and} \quad T_{x_0}^{-1}F'(x) \supseteq T_{x_0}^{-1}(F'(x) - F'(x_0)) + T_{x_0}^{-1}F'(x_0),$$

it follows that

$$T_{x_0}^{-1}F'(x) \supseteq S_2 + \mathbf{I}. \tag{2.20}$$

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

$$\|-(T_{x_0}^{-1}F'(x))^{-1}\| = \|(T_{x_0}^{-1}F'(x))^{-1}\| \leqslant \|(\mathbf{I} + S_2)^{-1}\| \leqslant \frac{1}{1 - L\|x - x_0\|}. \tag{2.21}$$

Moreover, for any $y, z \in X$, the following equivalences are valid:

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

Then $T_x^{-1}(-F'(x_0)) = -(T_{x_0}^{-1}F'(x))^{-1}$. Hence $\mathrm{D}(T_x^{-1}F'(x_0)) = \mathrm{R}(T_{x_0}^{-1}F'(x)) = X$, and (2.21) implies that

$$\|T_x^{-1}(-F'(x_0))\| \leqslant \frac{1}{1 - L\|x - x_0\|}; \tag{2.22}$$

thus (2.17) and the second equality in (2.16) hold (since $F'(x_0)$ is linear it is evident that $\|T_x^{-1}(-F'(x_0))\| = \|T_x^{-1}F'(x_0)\|$).

To prove $\mathrm{R}(T_{x_0}) = \mathrm{R}(T_x)$, it suffices to show the inclusion $\mathrm{R}(T_{x_0}) \subseteq \mathrm{R}(T_x)$ as the converse inclusion is clear by assumed weak-Robinson condition. To do this, let $y \in F'(x_0)u - C$ for some $u \in X$. Then, by what we have already proved, there exists $w \in X$ 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 cone, it follows that $y \in F'(x_0)u - C \subseteq F'(x)(-w) - C \subseteq \mathrm{R}(T_x)$. This proves that $\mathrm{R}(T_{x_0}) \subseteq \mathrm{R}(T_x)$.

Finally, suppose that X is additionally reflexive. Then thanks to the given assumptions, Lemma 2.2 is applicable to $[x_0, x]$ in place of $[x', x]$. Hence,

$$T_{x_0}^{-1}\int_0^1 \big(F'(x_0 + t(x - x_0)) - F'(x_0)\big)(x_0 - x)\,\mathrm{d}t \neq \emptyset. \tag{2.23}$$

Then (2.16) implies that

$$T_x^{-1}F'(x_0)T_{x_0}^{-1}\int_0^1 \big(F'(x_0 + t(x - x_0)) - F'(x_0)\big)(x_0 - x)\,\mathrm{d}t \neq \emptyset \tag{2.24}$$

and

$$T_x^{-1}F'(x_0)(x_0 - x) \neq \emptyset. \tag{2.25}$$

Note that

$$T_x^{-1}F'(x_0)T_{x_0}^{-1} \subseteq T_x^{-1} \tag{2.26}$$

(which can be checked easily by making use of the fact that $C + C = C$), and that

Please cite this article in press as: C. Li, K.F. Ng, Convergence analysis of the Gauss–Newton method for convex inclusion and convex-composite optimization problems, J. Math. Anal. Appl. (2011), doi:10.1016/j.jmaa.2011.11.062

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

Since T_x^{-1} is a convex process, it follows from (2.26) that

$$\begin{aligned} T_x^{-1}\big(F(x_0) - F(x)\big) &\supseteq T_x^{-1}F'(x_0)T_{x_0}^{-1}\left(\int_0^1 \big(F'\big(x_0 + t(x - x_0)\big) - F'(x_0)\big)(x_0 - x)\,\mathrm{d}t\right) \\ &\quad + \big(T_x^{-1}F'(x_0)\big)(x_0 - x) \\ &\neq \emptyset, \end{aligned} \tag{2.27}$$

where the nonemptiness assertion holds by (2.25) and (2.24). Similarly, by (2.16), (2.12) and (2.26) again, we have that

$$\emptyset \neq T_x^{-1}F'(x_0)T_{x_0}^{-1}\big(-F(x_0)\big) \subseteq T_x^{-1}\big(-F(x_0)\big). \tag{2.28}$$

From the convex process property,

$$T_x^{-1}\big(-F(x)\big) \supseteq T_x^{-1}\big(-F(x_0)\big) + T_x^{-1}\big(F(x_0) - F(x)\big), \tag{2.29}$$

we make use of (2.27) and (2.28) to conclude that $T_x^{-1}(-F(x)) \neq \emptyset$, that is, (2.18) holds (because of (2.11)). The proof is complete. □

3. Gauss–Newton method and convergence criteria

This section is devoted to establishing two of our main convergence results in the Gauss–Newton method. The first regards Robinson's Algorithm $\mathbf{A}(x_0)$ (explained in Section 1 for problem (1.1)); while the second regards Algorithm $\mathbf{A}(\eta, \Delta, x_0)$ for problem (1.2) which has already been studied by many researches (see [3,19,22,30] for the case when the underlying spaces are finite-dimensional).

As in the earlier sections we assume always that X and Y are Banach spaces, $F : X \to Y$ is a smooth map, and C is a closed cone in Y. For the remainder of this paper, we assume in addition that X is reflexive. Moreover, whenever the problem (1.2) or Algorithm $\mathbf{A}(\eta, \Delta, x_0)$ is considered, we will assume implicitly that $h : Y \to \mathbb{R}$ is a (continuous) convex function, and the cone C is the set of its minimum points:

$$C := \operatorname{argmin} h. \tag{3.1}$$

For any $\Delta \in (0, +\infty]$ and $x \in X$, let $D_\Delta(x)$ denote the set of all $d \in X$ satisfying $\|d\| \leqslant \Delta$ and

$$h\big(F(x) + F'(x)d\big) = \min\big\{h\big(F(x) + F'(x)d'\big) \colon d' \in X,\ \|d'\| \leqslant \Delta\big\}. \tag{3.2}$$

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

$$\min\big\{h\big(F(x) + F'(x)d'\big) \colon d' \in X,\ \big\|d'\big\| \leqslant \Delta\big\}. \tag{3.3}$$

Let

$$\mathfrak{D}_\Delta(x) = \big\{d \in X \colon \|d\| \leqslant \Delta,\ F(x) + F'(x)d \in C\big\}. \tag{3.4}$$

Note that $\mathfrak{D}_\Delta(x) \subseteq D_\Delta(x)$ for each $x \in X$.

Remark 3.1.

(a) If $\Delta < +\infty$, then $D_\Delta(x) \neq \emptyset$ for each $x \in X$.
(b) If $F(x^*) \in C$ then x^* solves (1.2).
(c) Suppose that $\mathfrak{D}_\Delta(x) \neq \emptyset$. Then for each $d \in X$ with $\|d\| \leqslant \Delta$, the following equivalences hold.

$$d \in D_\Delta(x) \iff d \in \mathfrak{D}_\Delta(x) \iff d \in \mathfrak{D}_\infty(x) \iff d \in D_\infty(x). \tag{3.5}$$

Let $\eta \in [1, +\infty)$, $\Delta \in (0, +\infty]$ and $x_0 \in X$. Recall the following algorithm (the Gauss–Newton method) for solving (1.2).

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

If $h(F(x_k)) = \min\{h(F(x_k) + F'(x_k)d)\colon d \in X, \ \|d\| \leqslant \Delta\}$, then stop; otherwise, choose $d_k \in D_\Delta(x_k)$ to satisfy $\|d_k\| \leqslant \eta \mathrm{d}(0, D_\Delta(x_k))$, and set $x_{k+1} = x_k + d_k$.

We shall base our analysis on a majorizing function technique. In what follows, we make the following blanket arrangement on notations. Fix a point $x_0 \in X$ and constants $L \in (0, +\infty)$, $\eta \in [1, +\infty)$, $\Delta \in (0, +\infty]$, and we assume that $-F(x_0) \in \mathrm{R}(T_{x_0})$. Define ξ and α by

$$\xi := \eta \left\| T_{x_0}^{-1}\big(-F(x_0)\big)\right\| \quad \text{and} \quad \alpha := \frac{\eta}{1 + (\eta - 1)L\xi}. \tag{3.6}$$

Define the quadratic "majorizing function" ϕ_η by

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

Then

$$\phi'_\eta(t) = -(1 - \alpha L t) \quad \text{for each } t \geqslant 0. \tag{3.7}$$

Let $\{t_{\eta,n}\}$ denote the sequence generated by Newton's method for ϕ_η with initial point $t_{\eta,0} = 0$:

$$t_{\eta,n+1} = t_{\eta,n} - \phi'_\eta(t_{\eta,n})^{-1}\phi_\eta(t_{\eta,n}) \quad \text{for each } n = 0, 1, \ldots. \tag{3.8}$$

In particular,

$$t_{\eta,1} = \xi. \tag{3.9}$$

Lemma 3.1. *Suppose that*

$$\xi \leqslant \frac{1}{L(\eta + 1)}. \tag{3.10}$$

Then the zeros of ϕ_η are given by

$$\left.\begin{matrix} r_\eta^* \\ r_\eta^{**} \end{matrix}\right\} = \frac{1 + (\eta - 1)L\xi \mp \sqrt{1 - 2L\xi - (\eta^2 - 1)(L\xi)^2}}{L\eta}. \tag{3.11}$$

Moreover $\{t_{\eta,n}\}$ is increasingly convergent to r_η^* and has the closed form

$$t_{\eta,n} = \frac{\sum_{i=0}^{2^n-2} {q_\eta}^i}{\sum_{i=0}^{2^n-1} {q_\eta}^i} r_\eta^* \quad \textit{for each } n = 1, 2, \ldots, \tag{3.12}$$

where

$$q_\eta = \frac{1 - L\xi - \sqrt{1 - 2L\xi - (\eta^2 - 1)(L\xi)^2}}{L\eta\xi}. \tag{3.13}$$

Proof. Note that

$$\xi \leqslant \frac{1}{L(\eta + 1)} \quad \Longleftrightarrow \quad \xi \leqslant \frac{1}{2\alpha L}.$$

Thus the zeros of ϕ_η are

$$\left.\begin{matrix} r_\eta^* \\ r_\eta^{**} \end{matrix}\right\} = \frac{1 \mp \sqrt{1 - 2\alpha L\xi}}{\alpha L}. \tag{3.14}$$

It is also known (see for example [16,24]) that

$$t_{\eta,n} = \frac{\sum_{i=0}^{2^n-2} \tilde{q}_\alpha^i}{\sum_{i=0}^{2^n-1} \tilde{q}_\alpha^i} r_\eta^* \quad \text{for each } n = 1, 2, \ldots, \tag{3.15}$$

where

$$\tilde{q}_\alpha := \frac{r_\eta^*}{r_\eta^{**}} = \frac{1 - \sqrt{1 - 2\alpha L\xi}}{1 + \sqrt{1 - 2\alpha L\xi}}. \tag{3.16}$$

Substituting $\alpha = \frac{\eta}{1+(\eta-1)L\xi}$ into (3.14)–(3.16), one sees that the conclusions of this lemma hold. $\square$

Please cite this article in press as: C. Li, K.F. Ng, Convergence analysis of the Gauss–Newton method for convex inclusion and convex-composite optimization problems, J. Math. Anal. Appl. (2011), doi:10.1016/j.jmaa.2011.11.062

Theorem 3.1. *Let $L \in (0, +\infty)$. Suppose that the inclusion (1.1) satisfies the weak-Robinson condition at x_0 on $\mathbf{B}(x_0, r_\eta^*)$, and that $(T_{x_0}^{-1}, F')$ is Lipschitz continuous on $\mathbf{B}(x_0, r_\eta^*)$ with modulus L. Assume that*

$$\xi \leqslant \min\left\{\frac{1}{L(\eta+1)}, \Delta\right\}. \tag{3.17}$$

Then Algorithm $\mathbf{A}(\eta, \Delta, x_0)$ is well defined (even if $\Delta = +\infty$) and any sequence $\{x_n\}$ so generated converges to some x^* with $F(x^*) \in C$ and the following assertions hold for each $n = 1, 2, \ldots$:

$$\|x_n - x_{n-1}\| \leqslant t_{\eta,n} - t_{\eta,n-1}, \tag{3.18}$$

$$F(x_{n-1}) + F'(x_{n-1})(x_n - x_{n-1}) \in C \tag{3.19}$$

and, for each $n = 0, 1, \ldots,$

$$\left\|x_n - x^*\right\| \leqslant \frac{{q_\eta}^{2^n-1}}{\sum_{i=0}^{2^n-1} {q_\eta}^i} r_\eta^*. \tag{3.20}$$

Proof. Let us first note that, for each $n \geqslant 1$,

$$\xi \leqslant t_{\eta,n} < r_\eta^* \leqslant \frac{1}{L}. \tag{3.21}$$

In fact, the first two inequalities hold because $\xi = t_{\eta,1} \leqslant r_\eta^*$ by (3.9) and Lemma 3.1. Moreover note from (3.17) that $L\xi < 1$ and so $\frac{1+(\eta-1)L\xi}{L\eta} \leqslant \frac{1}{L}$; thus $r_\eta^* \leqslant \frac{1}{L}$ by (3.11), and (3.21) is proved. Below we shall use mathematical induction to verify (3.18) and (3.19). For this end, let $k \geqslant 1$ and use $\overline{1,k}$ to denote the set of all integers n satisfying $1 \leqslant n \leqslant k$.

By the weak-Robinson condition assumption, we have from (2.11) together with (3.5) that

$$D_\infty(x_0) = \mathfrak{D}_\infty(x_0) = T_{x_0}^{-1}\bigl(-F(x_0)\bigr) \neq \emptyset. \tag{3.22}$$

Hence, by (3.6), (3.9) and (3.17),

$$\eta \mathrm{d}\bigl(0, D_\infty(x_0)\bigr) = \eta\left\|T_{x_0}^{-1}\bigl(-F(x_0)\bigr)\right\| = \xi = t_{\eta,1} - t_{\eta,0} \leqslant \Delta. \tag{3.23}$$

Since $\eta \geqslant 1$ and X is reflexive, it follows from (3.22) that there exists $d \in D_\infty(x_0)$ such that $\|d\| \leqslant \xi \leqslant \Delta$. Thus, $\mathrm{d}(0, D_\Delta(x_0)) = \mathrm{d}(0, D_\infty(x_0))$ and, by Remark 3.1,

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

In particular, x_1 is well defined and $F(x_0) + F'(x_0)d_0 \in C$; hence (3.19) holds for $n = 1$. Furthermore, by (3.23) and Algorithm $\mathbf{A}(\eta, \Delta, x_0)$ one has that $\|d_0\| \leqslant \eta \mathrm{d}(0, D_\infty(x_0)) \leqslant t_{\eta,1} - t_{\eta,0}$, i.e., $\|x_1 - x_0\| \leqslant t_{\eta,1} - t_{\eta,0}$. This shows that (3.18) holds for $n = 1$.

Now assume that (3.18) and (3.19) hold for all $n \in \overline{1,k}$. Write

$$x_k^\tau = \tau x_k + (1-\tau)x_{k-1} \quad \text{for each } \tau \in [0,1]. \tag{3.24}$$

Note that

$$\|x_k - x_0\| \leqslant \sum_{i=1}^k \|x_i - x_{i-1}\| \leqslant \sum_{i=1}^k (t_{\eta,i} - t_{\eta,i-1}) = t_{\eta,k} \tag{3.25}$$

and

$$\|x_{k-1} - x_0\| \leqslant t_{\eta,k-1} \leqslant t_{\eta,k}. \tag{3.26}$$

It follows from (3.24) and (3.21) that $x_k^\tau \in \mathbf{B}(x_0, r_\eta^*) \subseteq \mathbf{B}(x_0, 1/L)$ for each $\tau \in [0,1]$. Note in particular that, by Remark 3.1 and (2.18) in Proposition 2.3 (applied to r_η^* in place of r),

$$D_\infty(x_k) = \mathfrak{D}_\infty(x_k) \neq \emptyset \tag{3.27}$$

(where $D_\infty(x_k)$ and $\mathfrak{D}_\infty(x_k)$ are defined by (3.2) and (1.3) respectively). In particular, x_{k+1} is well defined and (3.19) holds for $n = k+1$. Letting $\tau = 1$, we further note that $\|x_k - x_0\| < r_\eta^* \leqslant \frac{1}{L}$ and it follows from the Lipschitz continuity assumption that

$$\left\|T_{x_0}^{-1}\bigl(F'(x_k) - F'(x_0)\bigr)\right\| \leqslant L\|x_k - x_0\| < Lr_\eta^* \leqslant 1. \tag{3.28}$$

Please cite this article in press as: C. Li, K.F. Ng, Convergence analysis of the Gauss–Newton method for convex inclusion and convex-composite optimization problems, J. Math. Anal. Appl. (2011), doi:10.1016/j.jmaa.2011.11.062

Together with the assumed weak-Robinson condition, we see that Lemma 4.1 is applicable to x_k in place of x. Hence

$$\mathrm{D}\big(T_{x_k}^{-1}F'(x_0)\big) = X \tag{3.29}$$

and

$$\big\|T_{x_k}^{-1}F'(x_0)\big\| \leqslant \big(1 - L\|x_k - x_0\|\big)^{-1} \leqslant (1 - Lt_{\eta,k})^{-1} \tag{3.30}$$

thanks to (2.17), (3.25) and (3.28). By (3.25) and the weak-Robinson condition, we apply Lemma 2.2 (applied to $[x_{k-1}, x_k]$ in place of $[x', x]$) to get

$$T_{x_0}^{-1}\int_0^1 \big(F'\big(x_k^\tau\big) - F'(x_{k-1})\big)(x_{k-1} - x_k)\,d\tau \neq \emptyset$$

and

$$\left\|T_{x_0}^{-1}\int_0^1 \big(F'\big(x_k^\tau\big) - F'(x_{k-1})\big)(x_{k-1} - x_k)\,d\tau\right\| \leqslant \frac{L}{2}\|x_k - x_{k-1}\|^2 \leqslant \frac{L}{2}(t_{\eta,k} - t_{\eta,k-1})^2$$

(thanks to (3.18)). Since

$$-F(x_k) + F(x_{k-1}) = \int_0^1 F'\big(x_{k-1} + \tau(x_k - x_{k-1})\big)(x_{k-1} - x_k)\,\mathrm{d}\tau, \tag{3.31}$$

it follows that

$$T_{x_0}^{-1}\big[-F(x_k) + F(x_{k-1}) - F'(x_{k-1})(x_{k-1} - x_k)\big] \neq \emptyset \tag{3.32}$$

and

$$\big\|T_{x_0}^{-1}\big[-F(x_k) + F(x_{k-1}) - F'(x_{k-1})(x_{k-1} - x_k)\big]\big\| \leqslant \frac{L}{2}(t_{\eta,k} - t_{\eta,k-1})^2. \tag{3.33}$$

Similar but using (3.8), we have that

$$\phi_\eta(t_{\eta,k}) = \frac{\alpha L}{2}(t_{\eta,k} - t_{\eta,k-1})^2,$$

and it follows from (3.33) that

$$\big\|T_{x_0}^{-1}\big[-F(x_k) + F(x_{k-1}) + F'(x_{k-1})(x_k - x_{k-1})\big]\big\| \leqslant \frac{\phi_\eta(t_{\eta,k})}{\alpha}. \tag{3.34}$$

We claim that

$$\emptyset \neq \big(T_{x_k}^{-1}F'(x_0)\big)T_{x_0}^{-1}\big[-F(x_k) + F(x_{k-1}) + F'(x_{k-1})(x_k - x_{k-1})\big] \subseteq D_\infty(x_k). \tag{3.35}$$

In fact, the above nonemptiness assertion follows from (3.29) and (3.32). To show the inclusion in (3.35), 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 cone. Since (3.19) 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, (3.35) is valid and it follows that

$$\mathrm{d}\big(0, D_\infty(x_k)\big) \leqslant \big\|\big(T_{x_k}^{-1}F'(x_0)\big)T_{x_0}^{-1}\big[-F(x_k) + F(x_{k-1}) + F'(x_{k-1})(x_k - x_{k-1})\big]\big\|. \tag{3.36}$$

Combining this with (3.30) and (3.34) we arrive at

$$\begin{aligned}\mathrm{d}\big(0, D_\infty(x_k)\big) &\leqslant \big\|T_{x_k}^{-1}F'(x_0)\big\|\big\|T_{x_0}^{-1}\big[-F(x_k) + F(x_{k-1}) + F'(x_{k-1})(x_k - x_{k-1})\big]\big\| \\ &\leqslant \frac{\phi_\eta(t_{\eta,k})}{\alpha(1 - Lt_{\eta,k})} \\ &\leqslant -\frac{1}{\eta}\frac{\phi_\eta(t_{\eta,k})}{\phi'_\eta(t_{\eta,k})}\end{aligned} \tag{3.37}$$

Please cite this article in press as: C. Li, K.F. Ng, Convergence analysis of the Gauss–Newton method for convex inclusion and convex-composite optimization problems, J. Math. Anal. Appl. (2011), doi:10.1016/j.jmaa.2011.11.062

where the last inequality holds because, by (3.6) and (3.21),

$$1 \leqslant \eta = \alpha\big(1 + (\eta - 1)L\xi\big) \leqslant \alpha\big(1 + (\eta - 1)Lt_{\eta,k}\big),$$

and so

$$\frac{\eta}{\alpha}(1 - Lt_{\eta,k})^{-1} \leqslant (1 - \alpha Lt_{\eta,k})^{-1} = -\phi'_\eta(t_{\eta,k})^{-1}. \tag{3.38}$$

By (3.37) and (3.8), we have

$$\eta \mathrm{d}\big(0, D_\infty(x_k)\big) \leqslant -\frac{\phi_\eta(t_{\eta,k})}{\phi'_\eta(t_{\eta,k})} = (t_{\eta,k+1} - t_{\eta,k}). \tag{3.39}$$

Noting that the real-valued function $t \mapsto -\phi'_\eta(t)^{-1}\phi_\eta(t)$ is decreasing on $(0, r^*_\eta)$, we have that

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

It follows from (3.39) that $\mathrm{d}(0, D_\infty(x_k)) \leqslant \eta\mathrm{d}(0, D_\infty(x_k)) \leqslant \Delta$, which together with (3.27) implies that there exists $d_0 \in X$ with $\|d_0\| \leqslant \Delta$ such that $F(x_k) + F'(x_k)d_0 \in C$. Consequently, by Remark 3.1,

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

and

$$\mathrm{d}\big(0, D_\Delta(x_k)\big) = \mathrm{d}\big(0, D_\infty(x_k)\big). \tag{3.40}$$

Thus it follows from (3.39) that

$$\|x_{k+1} - x_k\| = \|d_k\| \leqslant \eta\mathrm{d}\big(0, D_\Delta(x_k)\big) \leqslant (t_{\eta,k+1} - t_{\eta,k}),$$

that is (3.18) holds for $n = k + 1$. Hence, by induction, (3.18) and (3.19) hold for all $n = 1, 2, \ldots$. Consequently, $\{x_n\}$ is a Cauchy sequence and so converges to some x^* with $F(x^*) \in C$ by (3.19). Moreover, by (3.18),

$$\big\|x_n - x^*\big\| \leqslant r^*_\eta - t_{\eta,n} \quad \text{for each } n = 0, 1, 2, \ldots.$$

This together with (3.12) implies that (3.20) holds for each $n = 0, 1, \ldots$. □

Theorem 3.2. *Let $L \in (0, +\infty)$ and $\xi = \|T^{-1}_{x_0}(-F(x_0))\|$. Suppose that the inclusion (1.1) satisfies the weak-Robinson condition at x_0 on $\mathbf{B}(x_0, r^*_1)$ and that $(T^{-1}_{x_0}, F')$ is Lipschitz continuous on $\mathbf{B}(x_0, r^*_1)$ with modulus L, where r^*_1 is defined by (3.11) with $\eta = 1$. Assume that*

$$\xi \leqslant \frac{1}{2L}. \tag{3.41}$$

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

$$\big\|x_n - x^*\big\| \leqslant \frac{{q_1}^{2^n-1}}{\sum_{i=0}^{2^n-1} {q_1}^i} r^*_1 \quad \textit{for each } n = 0, 1, \ldots, \tag{3.42}$$

where

$$q_1 = \frac{1 - L\xi - \sqrt{1 - 2L\xi}}{L\xi}. \tag{3.43}$$

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

$$h(y) := \mathrm{d}(y, C) = \inf_{z\in C} \|y - z\| \quad \text{for each } y \in Y. \tag{3.44}$$

Let $\Delta = +\infty$ and $\eta = 1$ (so (3.17) and (3.41) are identical). Since $\mathfrak{D}_\infty(x_0) \neq \emptyset$ by the weak-Robinson condition, there exists $x'_1 \in X$ such that $d_0 := x'_1 - x_0 \in \mathfrak{D}_\infty(x_0)$ and $\|d_0\| = \mathrm{d}(0, \mathfrak{D}_\infty(x_0))$. Noting that $\mathfrak{D}_\infty(x_0) = D_\infty(x_0)$ by Remark 3.1(c), x'_1 can be regarded as a point obtained by Algorithm $\mathbf{A}(\eta, \Delta, x_0)$ at its first iteration. Then Theorem 3.1 is applicable; it follows from (3.19) and Remark 3.1 that there exists $x'_2 \in X$ such that $d_1 := x'_2 - x'_1 \in \mathfrak{D}_\infty(x'_1) = D_\infty(x'_1)$ 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 3.1 and the proof is complete. □

Please cite this article in press as: C. Li, K.F. Ng, Convergence analysis of the Gauss–Newton method for convex inclusion and convex-composite optimization problems, J. Math. Anal. Appl. (2011), doi:10.1016/j.jmaa.2011.11.062

Corollary 3.1. *(See Robinson [25].) Suppose that T_{x_0} is surjective and that F' is Lipschitz continuous on $\mathbf{B}(x_0, \hat{R})$ with modulus $K > 0$:*

$$\|F'(x) - F'(y)\| \leqslant K\|x - y\| \quad \text{for all } x, y \in \mathbf{B}(x_0, \hat{R}), \tag{3.45}$$

where

$$\hat{R} := \frac{1 - \sqrt{1 - 2K\|T_{x_0}^{-1}\|\xi}}{K\|T_{x_0}^{-1}\|} \quad \text{and} \quad \xi = \|T_{x_0}^{-1}(-F(x_0))\|. \tag{3.46}$$

Assume that

$$\|x_1 - x_0\| \leqslant \frac{1}{2K\|T_{x_0}^{-1}\|}. \tag{3.47}$$

Then the conclusions of Theorem 3.2 hold with $r_1^* = \hat{R}$ and

$$q_1 = \frac{1 - K\|T_{x_0}^{-1}\|\xi - \sqrt{1 - 2K\|T_{x_0}^{-1}\|\xi}}{K\|T_{x_0}^{-1}\|\xi}. \tag{3.48}$$

Proof. Since T_{x_0} is surjective, it follows from Proposition 2.1 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, (3.43) and (3.48) are consistent. Likewise, r_1^* given in (3.11) equals $\hat{R}$. Furthermore, by the assumed Lipschitz continuity (3.45), one has that

$$\|T_{x_0}^{-1}(F'(x) - F'(y))\| \leqslant \|T_{x_0}^{-1}\|\|F'(x) - F'(y)\| \leqslant L\|x - y\| \quad \text{for all } x, y \in \mathbf{B}(x_0, r_1^*).$$

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 (2.11), we see that (3.41) and (3.47) are the same. Therefore, the result follows from Theorem 3.2. □

Remark 3.2. The convergence criteria given in Theorems 3.1 and 3.2 are affine-invariant in the sense described below. Let A be an invertible continuous linear operator from Y to itself. Define functions $\tilde{h} := h \circ A^{-1}$ and $\tilde{F} := A \circ F$ and define $\tilde{C} = A(C)$. Then $\operatorname{argmin} \tilde{h} = \tilde{C}$ and $h \circ F = \tilde{h} \circ \tilde{F}$. Hence the minimization problem (1.2) and the corresponding inclusion problem (1.1) can be rewritten respectively as

$$\min_{x \in X} (\tilde{h} \circ \tilde{F})(x) \tag{3.49}$$

and

$$\tilde{F}(x) \in \tilde{C}. \tag{3.50}$$

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 (3.50)) defined by

$$\tilde{T}_{x_0} d := \tilde{F}'(x_0)d - \tilde{C}. \tag{3.51}$$

Then the weak-Robinson condition assumed in Theorem 3.1 for (1.1) is equivalent to the corresponding one for (3.50). Likewise, the Lipschitz continuity 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 3.1 and 3.2 for (1.2) and (1.1) coincide respectively with the corresponding ones for (3.49) and (3.50), that is to say, such convergence criteria are affine-invariant. Note that the convergence criteria given in [21, Theorem 4.1] and [25, Theorem 2] do not have such property.

Remark 3.3. We exclude the trivial case when $L = 0$ in our study because, in this trivial case, if $(T_{x_0}^{-1}, F')$ is Lipschitz continuous with the modulus L 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 3.1 and 3.2, the Gauss–Newton method stops at the first step, that is, $F(x_1) \in C$.

Remark 3.4. Let $\mathcal{F}$ be a set-valued mapping from X to Y. Consider the following generalized equation:

$$0 \in F(x) + \mathcal{F}(x). \tag{3.52}$$

The Newton method for solving the above generalized equation and its local and/or stable local convergence have been explored extensively recently; see for example [1], the monographs [13] (in particular, Chapter 6) and the references therein.

The main tool there to analyze the local convergence is the notion of the metric regularity put on the set-valued mapping $F + \mathcal{F}$. Clearly, problem (1.1) is a special example by taking the set-valued mapping $\mathcal{F} := -C$. According to the Robinson extension of the open mapping theorem to convex processes (see [24]), if T_{x_0} is surjective then the mapping $F - C$ is metric regular around x_0, and so the results in [1] or [13] may apply. However, our approach here for (1.1) has several advantages: a) Our mapping $F(\cdot) - C$ is not required to have the metric regularity (as T_{x_0} is not necessarily surjective); b) the existence of the solution is not initially assumed; and c) our convergence result given in Theorem 3.1 concerned with any sequence provided by Algorithm $\mathbf{A}(x_0)$ (which, in general, is different from the Newton method considered in [1] and [13] for solving Eq. (3.52) with $\mathcal{F} = -C$).

4. Conclusion and examples

Under the assumptions that C is a closed cone, the inclusion (1.1) satisfies the weak-Robinson condition at x_0, and $(T_{x_0}^{-1}, F')$ is Lipschitz continuous with modulus L, we have established a convergence criterion ensuring the convergence of the Gauss–Newton method for solving convex-composite optimization problems. In particular, we obtain the convergence criterion for the extended Newton method for solving the inclusion problem considered by Robinson in [25]. In general, the norm of $\|T_{x_0}^{-1}\|$ is not necessarily finite. Even in the special case when X, Y are finite-dimensional and T_{x_0} is surjective (so $\|T_{x_0}^{-1}\| < +\infty$), our result is sharper than that in [25] as shown in Example 4.1 below.

Example 4.1. Let $X = \mathbb{R}$, $Y = \mathbb{R}^2$ and $\lambda > 0$. Define F by

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

Thus

$$F'(x) = \begin{bmatrix} \sin x + 1 \\ x + 1 \end{bmatrix} \quad \text{for each } x \in \mathbb{R} \tag{4.1}$$

and

$$\|F'(x) - F'(x')\| \leqslant \sqrt{2}|x - x'| \quad \text{for all } x, x' \in \mathbb{R}$$

($K := \sqrt{2}$ is in fact the optimal Lipschitz constant). Take $x_0 = 0$ and $C = \{(t_1, t_2)^T \in \mathbb{R}^2\colon t_1 \leqslant 0,\ t_2 \leqslant 0\}$. Then

$$F(x_0) = (\lambda, \lambda)^T, \qquad F'(x_0) = (1, 1)^T$$

and

$$T_{x_0}^{-1} y = \big(-\infty, \min\{y_1, y_2\}\big] \quad \text{for each } y = (y_1, y_2)^T \in \mathbb{R}^2.$$

Hence,

$$\|T_{x_0}^{-1}\| = 1, \qquad \|T_{x_0}^{-1}\big(-F(x_0)\big)\| = \lambda \tag{4.2}$$

and

$$\|T_{x_0}^{-1}\big(F'(x) - F'(x')\big)\| \leqslant \max\big\{|\sin x - \sin x'|, |x - x'|\big\} \leqslant |x - x'| \quad \text{for all } x, x' \in \mathbb{R}. \tag{4.3}$$

Thus the modulus L in Theorem 3.2 is equal to 1, and (3.41) means that $\lambda \leqslant \frac{1}{2}$ while the corresponding sufficient condition given in [25, Theorem 2] is $\lambda \leqslant \frac{\sqrt{2}}{4}$ $(= \frac{1}{2K\|T_{x_0}^{-1}\|})$. Therefore if $\lambda \in (\frac{\sqrt{2}}{4}, \frac{1}{2}]$ then the corresponding F provides an example showing that Theorem 3.2 properly extends the earlier results.

As the following example shows, it can happen that $T_{x_0}^{-1}$ is not normed even though the weak-Robinson condition at x_0 is satisfied. Clearly, in this case, [25, Theorem 2] is not applicable. For our examples, it would be convenient to recall the following fact:

Fact 4.1. If X is reflexive and $T_{x_0}^{-1}$ is normed, then $\mathrm{R}(T_{x_0})$ is closed, that is $\mathrm{D}(T_{x_0}^{-1})$ is closed.

In fact, let $\{z_n\} \subseteq \mathrm{D}(T_{x_0}^{-1})$ be such that $z_n \to z$. Without loss of generality, we may assume that $\|z_n\| = \|z\| = 1$. Since $\|T_{x_0}^{-1}\| < \infty$, we can take $\{y_n\} \subseteq X$ such that $\{y_n\}$ is bounded and $y_n \in T_{x_0}^{-1} z_n$ for each n. As in the proof of Lemma 2.1, we may assume that without loss of generality, $y_n \to y$ weakly for some $y \in X$. Then there exists a sequence $\{\tilde{y}_n\}$, with each $\tilde{y}_n = \sum_{i=1}^{k_n} \alpha_i^n y_{n_i}$, where $k_n \geqslant 1$ and $\{\alpha_i^n\} \subseteq [0,1]$ satisfies $\sum_{i=1}^{k_n} \alpha_i^n = 1$, such that $\tilde{y}_n \to y$ and the correspond convex combination $\tilde{z}_n := \sum_{i=1}^{k_n} \alpha_i^n z_{n_i} \to z$. Thus

$$F'(x_0)\tilde{y}_n \in \tilde{z}_n + C \quad \text{for each } n = 1, 2, \ldots$$

and taking limits, we have that $F'(x_0)y \in z + C$. Therefore $y \in T_{x_0}^{-1} z$ and so $z \in \mathrm{D}(T_{x_0}^{-1})$. This proves that $\mathrm{D}(T_{x_0}^{-1})$ is closed and Fact 4.1 is established.

Example 4.2. Let $X = \mathbb{R}^2$, $Y = \mathbb{R}^3$ and let the cone $C \subseteq \mathbb{R}^3$ be given by

$$C := \big\{(t_1, t_2, t_3)^T : t_1^2 + (t_3 - t_2)^2 \leqslant t_2^2 \text{ and } t_2 \leqslant 0\big\},$$

that is, C is the cone generated by the origin and the plane disk $\{(t_1, -1, t_3)^T : t_1^2 + (t_3+1)^2 \leqslant 1\}$. Let $x_0 = 0$ and $\lambda \in (0, \frac{5}{6}]$. Define F by

$$F(x) = \begin{pmatrix} 0 \\ t_1 + t_2^2 + \frac{\lambda}{5} \\ -\frac{\lambda}{5} \end{pmatrix} \quad \text{for each } x = (t_1, t_2)^T \in \mathbb{R}^2.$$

Then

$$F'(x) = \begin{pmatrix} 0 & 0 \\ 1 & 2t_2 \\ 0 & 0 \end{pmatrix} \quad \text{for each } x = (t_1, t_2)^T \in \mathbb{R}^2.$$

In particular, $F'(x_0) = \left(\begin{smallmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)$, and so

$$\mathrm{R}(T_{x_0}) = \big\{(t_1, t_2, t_3)^T \in \mathbb{R}^3 : t_3 > 0\big\} \cup \big\{(t_1, t_2, t_3)^T \in \mathbb{R}^3 : t_1 = t_3 = 0\big\}.$$

Hence,

$$-F(x_0) \in \mathrm{R}(T_{x_0}) \quad \text{and} \quad \mathrm{R}\big(F'(x)\big) \subseteq \mathrm{R}(T_{x_0}) \quad \text{for each } x \in \mathbb{R}^2.$$

Therefore, the inclusion (1.1) satisfies the weak-Robinson condition at x_0 with $r = +\infty$. Since

$$\big(F'(x) - F'(x_0)\big)y = (0, 2t_2 y_2, 0)^T \quad \text{for all } x = (t_1, t_2)^T \in \mathbb{R}^2 \text{ and } y = (y_1, y_2)^T \in \mathbb{R}^2,$$

it is easy to verify that $(T_{x_0}^{-1}(F'(x) - F'(x_0)))y = \{(z_1, z_2)^T \in \mathbb{R}^2 : z_1 \leqslant 2t_2 y_2\}$. Therefore,

$$\big\|T_{x_0}^{-1}\big(F'(x) - F'(x_0)\big)\big\| = 2|t_2| \leqslant 2\|x - x_0\| \quad \text{for each } x \in \mathbb{R}^2.$$

Then the (best) modulus L in Theorem 3.2 is equal to 2. Noting that $\mathrm{R}(T_{x_0})$ is not closed, we see from Fact 4.1 that $T_{x_0}^{-1}$ is not normed. Since $F(x_0) = (0, \frac{\lambda}{5}, -\frac{\lambda}{5})$, it follows that

$$T_{x_0}^{-1}\big(-F(x_0)\big) = \left\{(z_1, z_2) \in \mathbb{R}^2 : \left(0, z_1 + \frac{\lambda}{5}, -\frac{\lambda}{5}\right) \in C\right\} = \left\{(z_1, z_2) \in \mathbb{R}^2 : z_1 \leqslant -\frac{3\lambda}{10}\right\}$$

and so $\|T_{x_0}^{-1}(-F(x_0))\| = \frac{3\lambda}{10} \leqslant \frac{1}{2L}$ because $\lambda \leqslant \frac{5}{6}$. Thus Theorem 3.2 is applicable and we can conclude that Algorithm $\mathbf{A}(x_0)$ is well defined and any sequence $\{x_k\}$ so generated converges to a solution of the inclusion problem (1.1).

Below we make some comparison of our results in Section 3 with that reported in [21]. Recall from [21] that $x_0 \in X$ is called 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

$$\mathfrak{D}_\infty(x) \neq \emptyset \quad \text{and} \quad \mathrm{d}\big(0, \mathfrak{D}_\infty(x)\big) \leqslant \beta\big(\|x - x_0\|\big)\mathrm{d}\big(F(x), C\big) \quad \text{for all } x \in \mathrm{B}(x_0, r). \tag{4.4}$$

Furthermore, let $\mathbf{r}_{x_0}$ denote the supremum $\mathbf{r}_{x_0}$ of r such that (4.4) holds for some increasing positive-valued function β on $[0, r)$, and β_{x_0} the infimum of β such that (4.4) 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.

In general, the quasi-regularity at point x_0 doesn't imply the weak-Robinson condition at x_0 even in the case when $\|T_{x_0}^{-1}\| < +\infty$ and $(T_{x_0}^{-1}, F')$ is Lipschitz continuous, see [21, Example 6.1]. The following proposition establishes a relationship between the weak-Robinson condition and the quasi-regularity.

Proposition 4.1. *Suppose that X is reflexive. Let $x_0 \in X$, $L \geqslant 0$ and $0 < r \leqslant \frac{1}{L}$. Suppose that* (1.1) *satisfies the weak-Robinson condition at x_0 on $\mathbf{B}(x_0, r)$ and that $(T_{x_0}^{-1}, F')$ is Lipschitz continuous on $\mathbf{B}(x_0, r)$ with modulus L. Then the following assertions hold*:

(i) *If $F(x_0) \notin C$, then x_0 is a quasi-regular point.*
(ii) *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 $\mathbf{r}_{x_0} \geqslant r$ and*

$$\beta_{x_0}(t) \leqslant \frac{\|T_{x_0}^{-1}\|}{1 - Lt} \quad \textit{for each } t \in [0, r).$$

Proof. (i). Assume that $F(x_0) \notin C$ and set $\rho := \mathrm{d}(F(x_0), C)$. Then $\rho > 0$. By the continuity, there exists $\bar{r} \in (0, r)$ such that

$$\inf\bigl\{\mathrm{d}\bigl(F(x), C\bigr) \colon x \in \mathbf{B}(x_0, \bar{r})\bigr\} \geqslant \frac{\rho}{2}.$$

By (2.18) in Proposition 2.3, $\mathfrak{D}_\infty(x) \neq \emptyset$ for each $x \in \mathbf{B}(x_0, \bar{r})$. Below we will show that there exists a constant $\theta > 0$ such that

$$\mathrm{d}\bigl(0, \mathfrak{D}_\infty(x)\bigr) \leqslant \theta \quad \text{for each } x \in \mathbf{B}(x_0, \bar{r}). \tag{4.5}$$

Granting this, one sees that

$$\mathrm{d}\bigl(0, \mathfrak{D}_\infty(x)\bigr) \leqslant \frac{2\theta}{\rho}\mathrm{d}\bigl(F(x), C\bigr) \quad \text{for each } x \in \mathbf{B}(x_0, \bar{r}),$$

and so x_0 is a quasi-regular point. To verify (4.5), let $x \in \mathbf{B}(x_0, \bar{r})$. By (2.27),

$$\begin{aligned} \bigl\|T_x^{-1}\bigl(F(x_0) - F(x)\bigr)\bigr\| &\leqslant \bigl\|T_x^{-1}F'(x_0)\bigr\| \left(\left\| T_{x_0}^{-1} \int_0^1 \bigl(F'\bigl(x_0 + t(x - x_0)\bigr) - F'(x_0)\bigr)(x_0 - x)\,\mathrm{d}t \right\| + \bar{r} \right) \\ &\leqslant \bigl\|T_x^{-1}F'(x_0)\bigr\| \left(\frac{L\bar{r}^2}{2} + \bar{r} \right), \end{aligned} \tag{4.6}$$

where the last inequality holds because, by (2.14) (applied to $[x_0, x]$ in place of $[x', x]$),

$$\left\| T_{x_0}^{-1} \int_0^1 \bigl(F'\bigl(x_0 + t(x - x_0)\bigr) - F'(x_0)\bigr)(x_0 - x)\,\mathrm{d}t \right\| \leqslant \frac{L\bar{r}^2}{2}. \tag{4.7}$$

Further, by (2.28),

$$\bigl\|T_x^{-1}\bigl(-F(x_0)\bigr)\bigr\| \leqslant \bigl\|T_x^{-1}F'(x_0)\bigr\| \bigl\|T_{x_0}^{-1}\bigl(-F(x_0)\bigr)\bigr\|. \tag{4.8}$$

By (2.29), (4.6) and (4.8), we have that

$$\bigl\|T_x^{-1}\bigl(-F(x)\bigr)\bigr\| \leqslant \theta. \tag{4.9}$$

where

$$\theta := \bigl\|T_x^{-1}F'(x_0)\bigr\| \left(\bigl\|T_{x_0}^{-1}\bigl(-F(x_0)\bigr)\bigr\| + \frac{L\bar{r}^2}{2} + \bar{r} \right).$$

Note that $\theta < +\infty$ by (2.17) and (2.12). By (2.11), (4.9) means that $\mathrm{d}(0, \mathfrak{D}_\infty(x)) \leqslant \theta$ and so (4.5) is shown.

(ii). Assume that $T_{x_0}^{-1}$ is normed. Thus it follows from (2.17), (2.26) and the given Lipschitz continuity assumption, we have that, for each $x \in \mathrm{B}(x_0, r)$,

$$\bigl\|T_x^{-1}\bigr\| \leqslant \bigl\|T_x^{-1}F'(x_0)T_{x_0}^{-1}\bigr\| \leqslant \frac{\|T_{x_0}^{-1}\|}{1 - L\|x - x_0\|}$$

and so (2.11) entails that

$$\mathrm{d}\bigl(0, \mathfrak{D}_\infty(x)\bigr) = \bigl\|T_x^{-1}\bigl(C - F(x)\bigr)\bigr\| \leqslant \bigl\|T_x^{-1}\bigr\|\mathrm{d}\bigl(F(x), C\bigr) \leqslant \frac{\|T_{x_0}^{-1}\|}{1 - L\|x - x_0\|}\mathrm{d}\bigl(F(x), C\bigr).$$

Recalling the definition of β, we complete the proof. □

In spite of Proposition 4.1, below we give an example to show that Theorem 3.1 is applicable but not [21, Corollary 4.3] (note in particular that the strict inequalities in (4.11) below hold in this example). For discussion, we continue to use ξ to denote $\eta\|T_{x_0}^{-1}(-F(x_0))\|$ as in Section 3. Recall that in the discussion of the main results of [21] (see Corollary 4.3 there), it was assumed that $\|T_{x_0}^{-1}\| < +\infty$ and that the quantity $\eta\|T_{x_0}^{-1}\| \mathrm{d}(F(x_0), C)$ (to be denoted by $\hat{\xi}$) played an important role in the convergence criterion in [21]. Noting the obvious inclusion

$$T_{x_0}^{-1}\big(C - F(x_0)\big) \subseteq T_{x_0}^{-1}\big(-F(x_0)\big), \tag{4.10}$$

one has that

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

that is,

$$\xi \leqslant \hat{\xi}. \tag{4.11}$$

Example 4.3. Let $X = \mathbb{R}$, $Y = \mathbb{R}^2$ and $\lambda > 0$. Let F be as in Example 4.1. Define h by

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

Thus

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

Clearly, $C = \{(t_1, t_2) \in \mathbb{R}^2\colon t_1 \leqslant 0,\ t_2 \leqslant 0\}$. Take $x_0 = 0$. Then, by (4.3), $(T_{x_0}^{-1}, F')$ is Lipschitz continuous on $\mathbb{R}$ with modulus $L = 1$. Let $\eta = 1$ and $\Delta = +\infty$ (and so α, defined in [21, Corollary 4.3], is equal to $\|T_{x_0}^{-1}\| = 1$). Then by (4.2) $\xi = \lambda$. Let $\lambda \in (\frac{1}{4}, \frac{1}{2}]$. Hence (3.41) is satisfied and so Theorem 3.2 is applicable. Below we shall show that [21, Corollary 4.3] is not applicable. In fact, otherwise, there exist $\Lambda \geqslant r > 0$ and a positive-valued increasing absolutely continuous function $\hat{L}_r$ defined on $[0, \Lambda)$ with $\int_0^{\Lambda} \hat{L}_r(t)\,\mathrm{d}t = +\infty$ such that F' satisfies the $\hat{L}_r$-average Lipschitz condition on $\mathbf{B}(x_0, r)$ in the sense defined in [21, Definition 2.5] and

$$\hat{\xi} \leqslant \hat{b}_1, \qquad \hat{r}_1^* \leqslant r \tag{4.12}$$

where $\hat{b}_1, \hat{r}_1^*$ are the corresponding b_α, r_α^* defined for $L = \hat{L}_r$ in [21, Section 2] with $\alpha = 1$. Then, by (4.1) and the assumed $\hat{L}_r$-average Lipschitz condition, we have

$$\big\|F'\big(x'\big) - F'(x)\big\| = \sqrt{\big(\sin x' - \sin x\big)^2 + \big(x' - x\big)^2} \leqslant \int_{|x|}^{|x'-x|+|x|} \hat{L}_r(\tau)\,\mathrm{d}\tau \quad \text{for all } x', x \in (-r, r).$$

In particular,

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

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

$$\hat{L}_r(t) \geqslant \hat{L}_r(0) \geqslant \sqrt{2} \quad \text{for each } t \in [0, \Lambda) \tag{4.13}$$

because $\hat{L}_r$ is increasing. Let $\hat{\phi}_1$ (resp. $\bar{\phi}_1$) denote the function ϕ_1 defined in [21, Section 2] with $\alpha = 1$, $\xi = \hat{\xi}$ but with L replaced by $\hat{L}_r$ (resp. $\sqrt{2}$), namely,

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

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 $\bar{\phi}_1 \leqslant \hat{\phi}_1$ by (4.13), and hence $\bar{\phi}_1(\hat{r}_1^*) \leqslant \hat{\phi}_1(\hat{r}_1^*) = 0$ with $\hat{r}_1^* \leqslant r < \Lambda$ (see (4.12)). 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

$$\hat{\xi} \leqslant \frac{1}{2\sqrt{2}}. \tag{4.14}$$

Noting that $\mathrm{d}(F(x_0), C) = \sqrt{2}\lambda$, it follows that $\hat{\xi} = \eta\|T_{x_0}^{-1}\| \mathrm{d}(F(x_0), C) = \sqrt{2}\lambda$. This together with (4.14) implies that $\sqrt{2}\lambda \leqslant \frac{1}{2\sqrt{2}}$, which contradicts that $\lambda > \frac{1}{4}$.

Acknowledgments

The authors are indebted to the anonymous referees for their helpful remarks, which allowed us to improve the original presentation. The first author was supported in part by the National Natural Science Foundation of China (grant 11171300) and by Zhejiang Provincial Natural Science Foundation of China (grant Y6110006) and the second author was supported by an Earmarked Grant (GRF) from the Research Grant Council of Hong Kong.

References

[1] F.J. Aragon Artacho, A.L. Dontchev, M. Gaydu, M.H. Geoffroy, V.M. Veliov, Metric regularity of Newton's iteration, SIAM J. Control Optim. 49 (2011) 339–362.
[2] J.V. Burke, Descent methods for composite nondifferentiable optimization problems, Math. Program. 33 (1985) 260–279.
[3] J.V. Burke, M.C. Ferris, A Gauss–Newton method for convex composite optimization, Math. Program. 71 (1995) 179–194.
[4] J.V. Burke, R.A. Poliquin, Optimality conditions for non-finite valued convex composite function, Math. Program. 57 (1992) 103–120.
[5] M.J. Cánovas, M.A. López, B. Mordukhovich, J. Parra, Variational analysis in semi-infinite and infinite programming, I: Stability of linear inequality systems of feasible solutions, SIAM J. Optim. 20 (2009) 1504–1526.
[6] M.J. Cánovas, M.A. López, B. Mordukhovich, J. Parra, Variational analysis in semi-infinite and infinite programming, II: Necessary optimality conditions, SIAM J. Optim. 20 (2010) 2788–2806.
[7] J.P. Dedieu, Approximate solutions of analytic inequality systems, SIAM J. Optim. 11 (2000) 411–425.
[8] P. Deuflhard, G. Heindl, Affine invariant convergence theorems for Newton's method and extensions to related methods, SIAM J. Numer. Anal. 16 (1979) 1–10.
[9] P. Deuflhard, Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms, Springer-Verlag, Berlin, Heidelberg, 2004.
[10] N. Dinh, M.A. Goberna, M.A. López, On the stability of the feasible set in optimization problems, SIAM J. Optim. 20 (2010) 2254–2280.
[11] N. Dinh, B. Mordukhovich, T.T.A. Nghia, Subdifferentials of value functions and optimality conditions for DC and bilevel infinite and semi-infinite programs, Math. Program. 123 (2010) 101–138.
[12] N. Dinh, G. Vallet, T.T.A. Nghia, Farkas-type results and duality for DC programs with convex constraints, J. Convex Anal. 2 (2008) 235–262.
[13] A.L. Dontchev, R.T. Rockafellar, Implicit Functions and Solution Mappings: A View from Variational Analysis, Springer Monogr. Math., Springer, Dordrecht, 2009.
[14] R. Flecher, Second order correction for nondifferentiable optimization, in: G.A. Watson (Ed.), Numerical Analalysis, in: Lect. Notes in Math., vol. 912, Springer, Berlin, 1982, pp. 85–114.
[15] R. Flecher, Practical Methods of Optimization, 2nd ed., Wiley, New York, 1987.
[16] W.B. Gragg, R.A. Tapai, Optimal error bounds for the Newton–Kantorovich theorems, SIAM J. Numer. Anal. 11 (1974) 10–13.
[17] V. Jeyakumar, G.M. Lee, Complete characterization of stable Farkas' lemma and cone-convex programming duality, Math. Program. 114 (2008) 335–347.
[18] V. Jeyakumar, A.M. Rubinov, B.M. Glover, Y. Ishizuka, Inequality systems and global optimization, J. Math. Anal. Appl. 202 (1996) 900–919.
[19] K. Jittorntrum, M.R. Osborne, Strong uniqueness and second order convergence in nonlinear discrete approximation, Numer. Math. 34 (1980) 439–455.
[20] L.V. Kantorovich, G.P. Akilov, Functional Analysis, second ed., Pergamon Press, Oxford, 1982 (translated from Russian by Howard L. Silcock).
[21] C. Li, K.F. Ng, Majorizing functions and convergence of the Gauss–Newton method for convex composite optimization, SIAM J. Optim. 18 (2007) 613–642.
[22] C. Li, X.H. Wang, On convergence of the Gauss–Newton method for convex composite optimization, Math. Program. 91 (2002) 349–356.
[23] K. Madsen, Minimization of nonlinear approximation function, PhD Thesis, Institute of Numerical Analysis, Technical University of Denmark, Lyngby, 1985.
[24] A.M. Ostrowski, Solutions of Equations in Euclidean and Banach Spaces, Academic Press, New York, 1973.
[25] S.M. Robinson, Extension of Newton's method to nonlinear function with values in a cone, Numer. Math. 19 (1972) 341–347.
[26] S.M. Robinson, Normed convex process, Trans. Amer. Math. Soc. 174 (1972) 127–140.
[27] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, NJ, 1970.
[28] R.T. Rockafellar, Monotone Processes of Convex and Concave Type, Mem. Amer. Math. Soc., vol. 77, 1967.
[29] R.T. Rockafellar, First- and second-order epi-differentiability in nonlinear programming, Trans. Amer. Math. Soc. 307 (1988) 75–108.
[30] R.S. Womersley, Local properties of algorithms for minimizing nonsmooth composite function, Math. Program. 32 (1985) 69–89.
[31] K. Yosida, Functional Analysis, Springer-Verlag, New York, 1968.

Please cite this article in press as: C. Li, K.F. Ng, Convergence analysis of the Gauss–Newton method for convex inclusion and convex-composite optimization problems, J. Math. Anal. Appl. (2011), doi:10.1016/j.jmaa.2011.11.062