

MAJORIZING FUNCTIONS AND CONVERGENCE OF THE GAUSS-NEWTON METHOD FOR CONVEX COMPOSITE OPTIMIZATION

CHONG LI* AND K. F. NG†

Abstract. We introduce a notion of quasi-regularity for points with respect to the inclusion $F(x) \in C$ where F is a nonlinear Frechét differentiable function from \mathbb{R}^v to \mathbb{R}^m . When C is the set of minimum points of a convex real-valued function h on \mathbb{R}^m and F' satisfies the L -average Lipschitz condition of Wang, we use the majorizing function technique to establish the semi-local linear/quadratic convergence of sequences generated by the Gauss-Newton method (with quasi-regular initial points) for the convex composite function $h \circ F$. Results are new even when the initial point is regular and F' is Lipschitz.

Key words. The Gauss-Newton method, convex composite optimization, majorizing function, convergence.

AMS subject classifications. 47J15 65H10 Secondary, 41A29

1. Introduction . The convex composite optimization to be considered is as follows.

$$\min_{x \in \mathbb{R}^v} f(x) := h(F(x)), \quad (1.1)$$

where h is a real-valued convex function on \mathbb{R}^m and F is a nonlinear Frechét differentiable map from \mathbb{R}^v to \mathbb{R}^m (with norm $\|\cdot\|$). We assume throughout that h attains its minimum h_{\min} .

This problem has recently received a great deal of attention. As observed by Burke and Ferris in their seminal paper [4], a wide variety of its applications can be found throughout the mathematical programming literature especially in convex inclusion, minimax problems, penalization methods and goal programming, see also [2, 6, 7, 15, 22]; the study of (1.1) not only provides a unifying framework for the development and analysis of algorithmic for solutions but also a convenient tool for the study of first- and second-order optimality conditions in constrained optimization [3, 5, 7, 22]. As in [4, 13], the study of (1.1) naturally relates to the convex inclusion problem

$$F(x) \in C, \quad (1.2)$$

where

$$C := \operatorname{argmin} h, \quad (1.3)$$

the set of all minimum points of h . Of course it is meaningful to study (1.2) by its own right for a general closed convex set (cf. [14, 16]). In section 3, we introduce a new notion of quasi-regularity for $x_0 \in \mathbb{R}^v$ with respect to the inclusion (1.2). This new notion covers the case of regularity studied by Burke and Ferris [4] as well as the case when $F'(x_0) - C$ is surjective employed by Robinson [18]. More importantly, we introduce notions of the quasi-regular radius r_{x_0} and of the quasi-regular bound function β_{x_0} attached to

*Department of Mathematics, Zhejiang University, Hangzhou 310027, P. R. China, (cli@zju.edu.cn). This author was supported in part by the National Natural Science Foundation of China (grant 10671175) and Program for New Century Excellent Talents in University.

†Department of Mathematics, Chinese University of Hong Kong, Hong Kong, P. R. China, (kfng@math.cuhk.edu.hk). This author was supported by a direct grant (CUHK) and an Earmarked Grant from the Research Grant Council of Hong Kong.

each quasi-regular point x_0 . For the general case this pair (r_{x_0}, β_{x_0}) together with a suitable Lipschitz type assumption on F' enables us to address the issue of convergence of Gauss-Newton sequence provided by the following well known algorithm (cf. [4, 10, 13, 31]).

Algorithm A (η, Δ, x_0) . Let $\eta \in [1, +\infty)$, $\Delta \in (0, +\infty]$ and for each $x \in \mathbb{R}^v$ we define $D_\Delta(x)$ by

$$D_\Delta(x) = \{d \in \mathbb{R}^v : \|d\| \leq \Delta, h(F(x) + F'(x)d) \leq h(F(x) + F'(x)d') \quad \forall d' \in \mathbb{R}^v \text{ with } \|d'\| \leq \Delta\}. \quad (1.4)$$

Let $x_0 \in \mathbb{R}^v$ be given. For $k = 0, 1, \dots$, having x_0, x_1, \dots, x_k , determine x_{k+1} as follows.

If $0 \in D_\Delta(x_k)$ then stop; if $0 \notin D_\Delta(x_k)$, choose d_k such that $d_k \in D_\Delta(x_k)$ and

$$\|d_k\| \leq \eta d(0, D_\Delta(x_k)), \quad (1.5)$$

and set $x_{k+1} = x_k + d_k$. Here $d(x, W)$ denotes the distance from x to W in the finite-dimensional Banach space containing W .

Note that $D_\Delta(x)$ is nonempty and is the solution set of the following convex optimization problem

$$\min_{d \in \mathbb{R}^v, \|d\| \leq \Delta} h(F(x) + F'(x)d). \quad (1.6)$$

which can be solved by standard methods such as the subgradient method, the cutting plane method, the bundle method etc (cf. [9]).

If the initial point x_0 is a quasi-regular point with (r_{x_0}, β_{x_0}) and if F' satisfies a Lipschitz type condition (introduced by Wang [28]) with an absolutely continuous function L satisfying a suitable property in relation to (r_{x_0}, β_{x_0}) , our main results presented in section 4 show that the Gauss-Newton sequence $\{x_n\}$ provided by Algorithm **A** (η, Δ, x_0) converges at a quadratic rate to some x^* with $F(x^*) \in C$ (in particular, x^* solves (1.1)). Even in the special case when x_0 is regular and F' is Lipschitz the advantage of allowing β_{x_0} and L to be functions (rather than constants) provides results which are new even for the above special case. Examples are given in section 6 to show there are situations where our results are applicable but not the earlier results in the literature; in particular, Example 6.1 is a simple example to demonstrate a quasi-regular point which is not regular. We shall show that the Gauss-Newton sequence $\{x_n\}$ is “majorized” by the corresponding numerical sequence $\{t_n\}$ generated by the classical Newton method with initial point $t_0 = 0$ for a “majorizing” function of the following type (again introduced by Wang [28])

$$\phi_\alpha(t) = \xi - t + \alpha \int_0^t L(u)(t-u) \, du \quad \text{for each } t \geq 0, \quad (1.7)$$

where ξ, α are positive constants and L is a positive-valued increasing (more precisely, nondecreasing) absolutely continuous function on $[0, +\infty)$. In the case when L is a constant function, (1.7) reduces to

$$\phi_\alpha(t) = \xi - t + \frac{\alpha L}{2} t^2 \quad \text{for each } t \geq 0, \quad (1.8)$$

the majorizing function used by Kantorovich [11, 12]. In the case when

$$L(u) = \frac{2\gamma}{(1-\gamma u)^3}, \quad (1.9)$$

(1.7) reduces to

$$\phi_\alpha(t) = \xi - t + \frac{\alpha\gamma t^2}{1-\gamma t}, \quad (1.10)$$

the majorizing function that Wang made use of in his work [28] on approximate zeros of Smale (cf. [26]). Motivated by this and as an application of our results in section 4, we provide a sufficient condition ensuring for a point $x_0 \in \mathbb{R}^v$ to be an ‘‘approximate solution’’ of (1.1) in the sense that the Gauss-Newton sequence $\{x_n\}$ generated by Algorithm $\mathbf{A}(\eta, \Delta, x_0)$ converges to a solution of (1.1) and satisfies the condition:

$$\|x_{n+1} - x_n\| \leq \left(\frac{1}{2}\right)^{2^{n-1}} \|x_n - x_{n-1}\| \quad \text{for each } n = 1, 2, \dots, \quad (1.11)$$

(the last condition was used by Smale [26] in his study of approximate zeros for Newton’s method).

2. Preliminaries. Let $\mathbf{B}(x, r)$ stand for the open ball in \mathbb{R}^v or \mathbb{R}^m with center x radius r while the corresponding closed ball is denoted by $\overline{\mathbf{B}}(x, r)$. Let W be a closed convex subset of \mathbb{R}^v or \mathbb{R}^m . The negative polar of W is denoted by W^\ominus and defined by

$$W^\ominus = \{z : \langle z, w \rangle \leq 0 \text{ for each } w \in W\}.$$

Let L be a positive-valued increasing absolutely continuous function on $[0, +\infty)$, and let $\alpha > 0$. Let $r_\alpha > 0$ and $b_\alpha > 0$ such that

$$\alpha \int_0^{r_\alpha} L(u) \, du = 1 \quad \text{and} \quad b_\alpha = \alpha \int_0^{r_\alpha} L(u)u \, du \quad (2.1)$$

(thus $b_\alpha < r_\alpha$). Let $\xi \geq 0$ and define

$$\phi_\alpha(t) = \xi - t + \alpha \int_0^t L(u)(t - u) \, du \quad \text{for each } t \geq 0. \quad (2.2)$$

Thus

$$\phi'_\alpha(t) = -1 + \alpha \int_0^t L(u) \, du, \quad \phi''_\alpha(t) = \alpha L(t) \quad \text{for each } t \geq 0 \quad (2.3)$$

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 $t_{\alpha,0} = 0$:

$$t_{\alpha,n+1} = t_{\alpha,n} - \phi'_\alpha(t_{\alpha,n})^{-1} \phi_\alpha(t_{\alpha,n}) \quad n = 0, 1, \dots \quad (2.4)$$

In particular, by (2.2) and (2.3),

$$t_{\alpha,1} = \xi. \quad (2.5)$$

Below we list a series of useful lemmas for our purpose. They are either known or can be verified easily by elementary methods (such as by differential calculus). In particular Lemma 2.3 and Lemma 2.1(i) are taken from [28], while Lemma 2.1 (ii) and (iii) are well known. Here we shall give a proof of Lemma 2.4 as an illustration.

LEMMA 2.1. *Suppose that $0 < \xi \leq b_\alpha$. Then $b_\alpha < r_\alpha$ and the following assertions hold.*

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

$$\phi_\alpha(\xi) > 0, \quad \phi_\alpha(r_\alpha) = \xi - b_\alpha \leq 0, \quad \phi_\alpha(+\infty) \geq \xi > 0. \quad (2.6)$$

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

$$\xi < r_\alpha^* < \frac{r_\alpha}{b_\alpha} \xi < r_\alpha < r_\alpha^{**}, \quad (2.7)$$

and, if $\xi = b_\alpha$, ϕ_α has a unique zero r_α^* in $(\xi, +\infty)$ (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 $\xi = b_\alpha$.

LEMMA 2.2. Let r_α , b_α and ϕ_α be defined by (2.1) and (2.2). 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 $(0, +\infty)$.

(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 $\xi \leq b_\alpha$.

LEMMA 2.3. Let $0 \leq c < +\infty$. Define

$$\chi(t) = \frac{1}{t^2} \int_0^t L(c+u)(t-u) \, du, \quad 0 \leq t < +\infty. \quad (2.8)$$

Then χ is increasing on $[0, +\infty)$.

LEMMA 2.4. Define

$$\omega_\alpha(t) = \phi'_\alpha(t)^{-1} \phi_\alpha(t) \quad t \in [0, r_\alpha^*].$$

Suppose that $0 < \xi \leq b_\alpha$. Then ω_α is increasing on $[0, r_\alpha^*]$.

Proof. Since

$$\omega'_\alpha(t) = \frac{\phi'_\alpha(t)^2 - \phi_\alpha(t)\phi''_\alpha(t)}{\phi'_\alpha(t)^2} \quad \text{for each } t \in [0, r_\alpha^*],$$

it suffices to show that

$$\zeta_\alpha(t) := \phi'_\alpha(t)^2 - \phi_\alpha(t)\phi''_\alpha(t) \geq 0 \quad \text{for each } t \in [0, r_\alpha^*].$$

Since $\zeta_\alpha(r_\alpha^*) = \phi'_\alpha(r_\alpha^*)^2 \geq 0$, it remains to show that ζ_α is decreasing on $[0, r_\alpha^*]$. To do this, note that by (2.3) ζ_α is absolutely continuous and so the derivative of ζ_α exists almost everywhere on $[0, r_\alpha^*]$ with

$$\zeta'_\alpha(t) = \phi'_\alpha(t)\phi''_\alpha(t) - \phi_\alpha(t)\phi'''_\alpha(t) \leq 0 \quad \text{a.e. } t \in [0, r_\alpha^*].$$

because $\phi'_\alpha \leq 0$ while $\phi_\alpha, \phi''_\alpha, \phi'''_\alpha \geq 0$ a.e. on $[0, r_\alpha^*]$. Therefore, ζ_α is decreasing on $[0, r_\alpha^*]$ and the proof is complete. \square

The following conditions were introduced by Wang in [28] but using the terminologies of “the center Lipschitz condition with the L average” and “the center Lipschitz condition in the inscribed sphere with the L average” respectively for (a) and (b).

DEFINITION 2.5. Let Y be a Banach space and let $x_0 \in \mathbb{R}^v$. Let G be a mapping from \mathbb{R}^v to Y . Then G is said to satisfy

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

$$\|G(x) - G(x_0)\| \leq \int_0^{\|x-x_0\|} L(u) du \quad \text{for each } x \in \mathbf{B}(x_0, r); \quad (2.9)$$

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

$$\|G(x) - G(x')\| \leq \int_{\|x'-x_0\|}^{\|x-x'\|+\|x'-x_0\|} L(u) du \quad \text{for all } x, x' \in \mathbf{B}(x_0, r) \text{ with } \|x-x'\| + \|x'-x_0\| \leq r. \quad (2.10)$$

3. Regularities. Let C be a closed convex set in \mathbb{R}^m . Consider the inclusion

$$F(x) \in C. \quad (3.1)$$

Let $x \in \mathbb{R}^v$ and

$$\mathcal{D}(x) = \{d \in \mathbb{R}^v : F(x) + F'(x)d \in C\}. \quad (3.2)$$

REMARK 3.1. In the case when C is the set of all minimum points of h and if there exists $d_0 \in \mathbb{R}^v$ with $\|d_0\| \leq \Delta$ such that $d_0 \in \mathcal{D}(x)$, then $d_0 \in D_\Delta(x)$ and for each $d \in \mathbb{R}^v$ with $\|d\| \leq \Delta$ one has

$$d \in D_\Delta(x) \iff d \in \mathcal{D}(x) \iff d \in D_\infty(x). \quad (3.3)$$

REMARK 3.2. The set $\mathcal{D}(x)$ defined in (3.2) can be viewed as the solution set of the following "linearized" problem associated with (3.1):

$$(P_x) : \quad F(x) + F'(x)d \in C. \quad (3.4)$$

Thus $\beta(\|x - x_0\|)$ in (3.5) is an "error bound" in determining how far the origin is away from the solution set of (P_x) .

DEFINITION 3.1. A point $x_0 \in \mathbb{R}^v$ is called a quasi-regular point of the inclusion (3.1) if there exist $r \in (0, +\infty)$ and an increasing positive-valued function β on $[0, r)$ such that

$$\mathcal{D}(x) \neq \emptyset \quad \text{and} \quad d(0, \mathcal{D}(x)) \leq \beta(\|x - x_0\|) d(F(x), C) \quad \text{for all } x \in \mathbf{B}(x_0, r). \quad (3.5)$$

Let \mathbf{r}_{x_0} denote the supremum of r such that (3.5) holds for some increasing positive-valued function β on $[0, r)$. Let $r \in [0, \mathbf{r}_{x_0}]$ and let $\mathcal{B}_r(x_0)$ denote the set of all increasing positive-valued function β on $[0, r)$ such that (3.5) holds. Define

$$\beta_{x_0}(t) = \inf\{\beta(t) : \beta \in \mathcal{B}_{\mathbf{r}_{x_0}}(x_0)\} \quad \text{for each } t \in [0, \mathbf{r}_{x_0}). \quad (3.6)$$

Note that each $\beta \in \mathcal{B}_r(x_0)$ with $\lim_{t \rightarrow r^-} \beta(t) < +\infty$ can be extended to an element of $\mathcal{B}_{\mathbf{r}_{x_0}}(x_0)$. From this we can verify that

$$\beta_{x_0}(t) = \inf\{\beta(t) : \beta \in \mathcal{B}_r(x_0)\} \quad \text{for each } t \in [0, r]. \quad (3.7)$$

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 .

DEFINITION 3.2. *A point $x_0 \in \mathbb{R}^v$ is a regular point of the inclusion (3.1) if*

$$\ker(F'(x_0)^T) \cap (C - F(x_0))^\ominus = \{0\}. \quad (3.8)$$

The notion of regularity relates to some other notions of regularity that can be found in the papers [1, 3, 20, 21, 25], which have played an important role in the study of nonsmooth optimizations. Some equivalent conditions on the regular points for (3.1) are given in [4]. In the following proposition the existence of constants r and β is due to Burke-Ferris [4] and the second assertion then follows from a remark after Definition 3.1

PROPOSITION 3.3. *Let x_0 be a regular point of (3.1). Then there are constants $r > 0$ and $\beta > 0$ such that (3.5) holds for r and $\beta(\cdot) = \beta$; consequently, x_0 is a quasi-regular point with the quasi-regular radius $\mathbf{r}_{x_0} \geq r$ and the quasi-regular bound function $\beta_{x_0} \leq \beta$ on $[0, r]$.*

Another important link of the present study relates to Robinson's condition [18, 19] that the convex process $d \mapsto F'(x)d - C$ is onto \mathbb{R}^m . To see this, let us first recall the concept of convex process which was introduced by Rockafeller [23, 24] for convexity problems (see also Robinson [19]).

DEFINITION 3.4. *A set-valued mapping $T : \mathbb{R}^v \rightarrow 2^{\mathbb{R}^m}$ 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 $T : \mathbb{R}^v \rightarrow 2^{\mathbb{R}^m}$ 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 $D(T)$, $R(T)$, T^{-1} ; i.e.,

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

$$R(T) = \cup\{Tx : x \in D(T)\},$$

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

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

$$\|A\| = \inf\{\|a\| : a \in A\}. \quad (3.9)$$

DEFINITION 3.5. *Suppose that T is a convex process. The norm of T is defined by*

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

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

For two convex processes T and S from \mathbb{R}^v to \mathbb{R}^m , the addition and multiplication are defined respectively as follows.

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

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

Let C be a closed convex set in \mathbb{R}^m and let $x \in \mathbb{R}^v$. We define T_x by

$$T_x d = F'(x)d - C \quad \text{for each } d \in \mathbb{R}^v. \quad (3.10)$$

Then its inverse is

$$T_x^{-1}y = \{d \in \mathbb{R}^v : F'(x)d \in y + C\} \quad \text{for each } y \in \mathbb{R}^m. \quad (3.11)$$

Note that T_x is a convex process in the case when C is a cone. Note also that $D(T_x) = \mathbb{R}^v$ for each $x \in \mathbb{R}^v$ and $D(T_{x_0}^{-1}) = \mathbb{R}^m$ if $x_0 \in \mathbb{R}^v$ is such that the following condition of Robinson is satisfied:

$$T_{x_0} \text{ carries } \mathbb{R}^v \text{ onto } \mathbb{R}^m. \quad (3.12)$$

Proposition 3.7 below shows that condition of Robinson (3.12) implies that x_0 is a regular point of (3.1) and an estimate of the quasi-regular bound function is provided. For its proof we need the following lemma which is known in [18].

LEMMA 3.6. *Let C be a closed convex cone in \mathbb{R}^m and let $x_0 \in \mathbb{R}^v$ be such that the condition of Robinson (3.12) is satisfied. Then the following assertions hold:*

(i) $T_{x_0}^{-1}$ is normed.

(ii) If S is a linear transformation from \mathbb{R}^v to \mathbb{R}^m such that $\|T_{x_0}^{-1}\|\|S\| < 1$, then the convex process \tilde{T} defined by

$$\tilde{T} = T_{x_0} + S$$

carries \mathbb{R}^v onto \mathbb{R}^m . Furthermore, \tilde{T}^{-1} is normed and

$$\|\tilde{T}^{-1}\| \leq \frac{\|T_{x_0}^{-1}\|}{1 - \|T_{x_0}^{-1}\|\|S\|}.$$

PROPOSITION 3.7. *Let $x_0 \in \mathbb{R}^v$ and let T_{x_0} be defined as in (3.10). Suppose that the condition of Robinson (3.12) is satisfied. Then the following assertion hold.*

(i) x_0 is a regular point of (3.1).

(ii) Suppose further that C is a closed convex cone in \mathbb{R}^m and F' satisfies the weak L -average Lipschitz condition on $\mathbf{B}(x_0, r)$ for some $r > 0$. Let $\beta_0 = \|T_{x_0}^{-1}\|$ and let r_{β_0} be defined by

$$\beta_0 \int_0^{r_{\beta_0}} L(u) \, du = 1 \quad (3.13)$$

(cf. (2.1)). Then the quasi-regular radius \mathbf{r}_{x_0} and the quasi-regular bound function β_{x_0} satisfy $\mathbf{r}_{x_0} \geq \min\{r, r_{\beta_0}\}$ and

$$\beta_{x_0}(t) \leq \frac{\beta_0}{1 - \beta_0 \int_0^t L(u) \, du} \quad \text{for each } t \text{ with } 0 \leq t < \min\{r, r_{\beta_0}\}. \quad (3.14)$$

Proof. Suppose that the condition (3.12) is satisfied and let y belong to the set of the intersection in (3.8). Then, in view of the definition of T_{x_0} , there exist $u \in \mathbb{R}^v$ and $c \in C$ such that $-y - F(x_0) = F'(x_0)u - c$. Hence

$$\langle y, F'(x_0)u \rangle = \langle F'(x_0)^T y, u \rangle = 0 \quad \text{and} \quad \langle y, c - F(x_0) \rangle \leq 0. \quad (3.15)$$

It follows that

$$\langle y, y \rangle = \langle y, c - F(x_0) - F'(x_0)u \rangle = \langle y, c - F(x_0) \rangle \leq 0 \quad (3.16)$$

and hence $y = 0$. This shows that (3.8) holds and so x_0 is a regular point of the inclusion (3.1).

Now let $r > 0$ and suppose that F' satisfies the weak L -average Lipschitz condition on $\mathbf{B}(x_0, r)$. Let $x \in \mathbb{R}^v$ such that $\|x - x_0\| < \min\{r, r_{\beta_0}\}$. Then

$$\|F'(x) - F'(x_0)\| \leq \int_0^{\|x-x_0\|} L(u) \, du < \int_0^{r_{\beta_0}} L(u) \, du;$$

hence, by (3.13),

$$\|T_{x_0}^{-1}\| \|F'(x) - F'(x_0)\| < \|T_{x_0}^{-1}\| \int_0^{r_{\beta_0}} L(u) \, du = 1.$$

This with Lemma 3.6 implies that the convex process defined by

$$T_x d = F'(x) d - C = T_{x_0} d + [F'(x) - F'(x_0)] d \quad \text{for each } d \in \mathbb{R}^v$$

carries \mathbb{R}^v onto \mathbb{R}^m and

$$\|T_x^{-1}\| \leq \frac{\|T_{x_0}^{-1}\|}{1 - \|T_{x_0}^{-1}\| \|F'(x) - F'(x_0)\|} \leq \frac{\|T_{x_0}^{-1}\|}{1 - \|T_{x_0}^{-1}\| \int_0^{\|x-x_0\|} L(u) \, du}. \quad (3.17)$$

Since T_x is surjective, we have that $\mathcal{D}(x)$ is nonempty; in particular, for each $c \in C$,

$$T_x^{-1}(c - F(x)) \subseteq \mathcal{D}(x) \quad (3.18)$$

To see this, let $d \in T_x^{-1}(c - F(x))$. Then, by (3.11), one has that $F'(x) d \in c - F(x) + C \subseteq C - F(x)$ and so $F'(x) d + F(x) \in C$, that is, $d \in \mathcal{D}(x)$. Hence (3.18) is true. Consequently,

$$d(0, \mathcal{D}(x)) \leq \|T_x^{-1}(c - F(x))\| \leq \|T_x^{-1}\| \|c - F(x)\|.$$

Since this is valid for each $c \in C$, it is seen that

$$d(0, \mathcal{D}(x)) \leq \|T_x^{-1}\|d(F(x), C).$$

Combining this with (3.17) and (3.6) gives the desired result (3.14), and the proof is complete. \square

4. Convergence criterion. We assume throughout the remainder of this paper that C is the set of all minimum points of h . Let $x_0 \in \mathbb{R}^v$ be a quasi-regular point of the inclusion (3.1) with the quasi-regular radius \mathbf{r}_{x_0} and the quasi-regular bound function β_{x_0} . Let $\eta \in [1, +\infty)$ and let

$$\xi := \eta\beta_{x_0}(0)d(F(x_0), C). \quad (4.1)$$

For all $\mathbf{r} \in (0, \mathbf{r}_{x_0}]$, we define

$$\alpha_0(\mathbf{r}) := \sup \left\{ \frac{\eta\beta_{x_0}(t)}{\eta\beta_{x_0}(t) \int_0^t L(u) du + 1} : \xi \leq t < \mathbf{r} \right\} \quad (4.2)$$

with the usual convention that $\sup \emptyset = -\infty$.

THEOREM 4.1. *Let $\eta \in [1, \infty)$ and $\Delta \in (0, \infty]$. Let $x_0 \in \mathbb{R}^v$ be a quasi-regular point of the inclusion (3.1) with the quasi-regular radius \mathbf{r}_{x_0} and the quasi-regular bound function β_{x_0} . Let $\xi > 0$, $0 < \mathbf{r} \leq \mathbf{r}_{x_0}$ and $\alpha_0(\mathbf{r})$ be as described above. Let $\alpha \geq \alpha_0(\mathbf{r})$ be a positive constant and let b_α, r_α be defined by (2.1). Let r_α^* denote the smaller zero of the function ϕ_α defined by (2.2). Suppose that F' satisfies the L -average Lipschitz condition on $\mathbf{B}(x_0, r_\alpha^*)$, and that*

$$\xi \leq \min\{b_\alpha, \Delta\} \quad \text{and} \quad r_\alpha^* \leq \mathbf{r}; \quad (4.3)$$

(for example, (4.3) is satisfied if

$$\xi \leq \min\left\{b_\alpha, \frac{b_\alpha}{r_\alpha}\mathbf{r}, \Delta\right\} \quad (4.4)$$

holds). Let $\{x_n\}$ denote the sequence generated by Algorithm **A**(η, Δ, x_0). Then, $\{x_n\}$ converges to some x^* such that $F(x^*) \in C$, and the following assertions hold for each $n = 1, 2, \dots$:

$$\|x_n - x_{n-1}\| \leq t_{\alpha,n} - t_{\alpha,n-1}, \quad (4.5)$$

$$\|x_{n+1} - x_n\| \leq (t_{\alpha,n+1} - t_{\alpha,n}) \left(\frac{\|x_n - x_{n-1}\|}{t_{\alpha,n} - t_{\alpha,n-1}} \right)^2, \quad (4.6)$$

$$F(x_n) + F'(x_n)(x_{n+1} - x_n) \in C \quad (4.7)$$

and

$$\|x_{n-1} - x^*\| \leq r_\alpha^* - t_{\alpha,n-1}. \quad (4.8)$$

Proof. By (2.7) and (4.4)

$$r_\alpha^* \leq \frac{r_\alpha}{b_\alpha} \xi \leq \mathbf{r}. \quad (4.9)$$

Hence (4.4) \implies (4.3). Thus it suffices to prove the theorem for the case when (4.3) is assumed. By (4.3), (2.5) and Lemma 2.1, one has that, for each n ,

$$\xi \leq t_{\alpha,n} < r_\alpha^* \leq \mathbf{r} \leq \mathbf{r}_{x_0}. \quad (4.10)$$

By the quasi-regularity assumption, it follows that

$$\mathcal{D}(x) \neq \emptyset \quad \text{and} \quad d(0, \mathcal{D}(x)) \leq \beta_{x_0}(\|x - x_0\|) d(F(x), C) \quad \text{for each } x \in \mathbf{B}(x_0, \mathbf{r}). \quad (4.11)$$

Let $k \geq 1$. We use $\overline{1, k}$ to denote the set of all integers n satisfying $1 \leq n \leq k$. Below we will verify the following implication:

$$(4.5) \text{ hold for all } n \in \overline{1, k} \text{ and } (4.7) \text{ holds for } n = k - 1 \implies (4.6) \text{ and } (4.7) \text{ hold for } n = k. \quad (4.12)$$

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

$$x_k^\tau = \tau x_k + (1 - \tau)x_{k-1} \quad \tau \in [0, 1]. \quad (4.13)$$

Note that

$$\|x_k - x_0\| \leq \sum_{i=1}^k \|x_i - x_{i-1}\| \leq \sum_{i=1}^k (t_{\alpha,i} - t_{\alpha,i-1}) = t_{\alpha,k} \quad (4.14)$$

and

$$\|x_{k-1} - x_0\| \leq t_{\alpha,k-1} \leq t_{\alpha,k}. \quad (4.15)$$

It follows from (4.13) and (4.10) that $x_k^\tau \in \mathbf{B}(x_0, r_\alpha^*) \subseteq \mathbf{B}(x_0, \mathbf{r})$ for each $\tau \in [0, 1]$. Hence (4.11) holds for $x = x_k$, namely,

$$\mathcal{D}(x_k) \neq \emptyset \quad \text{and} \quad d(0, \mathcal{D}(x_k)) \leq \beta_{x_0}(\|x_k - x_0\|) d(F(x_k), C). \quad (4.16)$$

We claim that

$$\eta d(0, \mathcal{D}(x_k)) \leq (t_{\alpha,k+1} - t_{\alpha,k}) \left(\frac{\|x_k - x_{k-1}\|}{t_{\alpha,k} - t_{\alpha,k-1}} \right)^2 \leq t_{\alpha,k+1} - t_{\alpha,k}, \quad (4.17)$$

(the second inequality need no proof by the assumption of (4.12)). To show the first inequality, using (4.7) for $n = k - 1$ and the fact that F' satisfies L -average Lipschitz condition on $\mathbf{B}(x_0, r_\alpha^*)$, together with the elementary identity

$$\int_0^1 \int_A^{A+\tau B} L(u) \, du \, d\tau = \int_0^B L(A+u) \left(1 - \frac{u}{B}\right) \, du \quad \text{for all } A, B > 0, \quad (4.18)$$

we have by (4.16) that

$$\begin{aligned}
\eta d(0, \mathcal{D}(x_k)) &\leq \eta \beta_{x_0}(\|x_k - x_0\|) d(F(x_k), C) \\
&\leq \eta \beta_{x_0}(\|x_k - x_0\|) \|F(x_k) - F(x_{k-1}) - F'(x_{k-1})(x_k - x_{k-1})\| \\
&\leq \eta \beta_{x_0}(\|x_k - x_0\|) \left\| \int_0^1 (F'(x_k^\tau) - F'(x_{k-1}))(x_k - x_{k-1}) \, d\tau \right\| \\
&\leq \eta \beta_{x_0}(\|x_k - x_0\|) \int_0^1 \left(\int_{\|x_{k-1} - x_0\|}^{\tau \|x_k - x_{k-1}\| + \|x_{k-1} - x_0\|} L(u) \, du \right) \|x_k - x_{k-1}\| \, d\tau \\
&= \eta \beta_{x_0}(\|x_k - x_0\|) \left(\int_0^{\|x_k - x_{k-1}\|} L(\|x_{k-1} - x_0\| + u)(\|x_k - x_{k-1}\| - u) \, du \right) \\
&\leq \eta \beta_{x_0}(t_{\alpha, k}) \left(\int_0^{\|x_k - x_{k-1}\|} L(t_{\alpha, k-1} + u)(\|x_k - x_{k-1}\| - u) \, du \right),
\end{aligned}$$

where the last inequality is valid because L and β_{x_0} are increasing and thanks to (4.14) and (4.15). Since (4.5) holds for $n = k$, Lemma 2.3 implies that

$$\frac{\int_0^{\|x_k - x_{k-1}\|} L(t_{\alpha, k-1} + u)(\|x_k - x_{k-1}\| - u) \, du}{\|x_k - x_{k-1}\|^2} \leq \frac{\int_0^{t_{\alpha, k} - t_{\alpha, k-1}} L(t_{\alpha, k-1} + u)(t_{\alpha, k} - t_{\alpha, k-1} - u) \, du}{(t_{\alpha, k} - t_{\alpha, k-1})^2}$$

and it follows from the earlier estimate that

$$\eta d(0, \mathcal{D}(x_k)) \leq \eta \beta_{x_0}(t_{\alpha, k}) \left(\int_0^{t_{\alpha, k} - t_{\alpha, k-1}} L(t_{\alpha, k-1} + u)(t_{\alpha, k} - t_{\alpha, k-1} - u) \, du \right) \left(\frac{\|x_k - x_{k-1}\|}{t_{\alpha, k} - t_{\alpha, k-1}} \right)^2. \quad (4.19)$$

Similarly by (2.2), (2.3), (2.4) and (4.18), we have

$$\begin{aligned}
\phi_\alpha(t_{\alpha, k}) &= \phi_\alpha(t_{\alpha, k}) - \phi_\alpha(t_{\alpha, k-1}) - \phi'_\alpha(t_{\alpha, k-1})(t_{\alpha, k} - t_{\alpha, k-1}) \\
&= \left(\int_0^1 [\phi'_\alpha(t_{\alpha, k-1} + \tau(t_{\alpha, k} - t_{\alpha, k-1})) - \phi'_\alpha(t_{\alpha, k-1})] \, d\tau \right) (t_{\alpha, k} - t_{\alpha, k-1}) \\
&= \left(\alpha \int_0^1 \int_{t_{\alpha, k-1}}^{\tau(t_{\alpha, k} - t_{\alpha, k-1}) + t_{\alpha, k-1}} L(u) \, du \, d\tau \right) (t_{\alpha, k} - t_{\alpha, k-1}) \\
&= \alpha \int_0^{t_{\alpha, k} - t_{\alpha, k-1}} L(t_{\alpha, k-1} + u)(t_{\alpha, k} - t_{\alpha, k-1} - u) \, du.
\end{aligned} \quad (4.20)$$

On the other hand, by (4.10) and (4.2),

$$\frac{\eta \beta_{x_0}(t_{\alpha, k})}{\alpha_0(\mathbf{r})} \leq \left(1 - \alpha_0(\mathbf{r}) \int_0^{t_{\alpha, k}} L(u) \, du \right)^{-1}.$$

Since $\alpha \geq \alpha_0(\mathbf{r})$ and by (2.3), it follows that

$$\frac{\eta \beta_{x_0}(t_{\alpha, k})}{\alpha} \leq \left(1 - \alpha \int_0^{t_{\alpha, k}} L(u) \, du \right)^{-1} = -(\phi'_\alpha(t_{\alpha, k}))^{-1}. \quad (4.21)$$

Combining (4.19)-(4.21) together with (2.4), the first inequality in (4.17) is seen to hold. Moreover by Lemma 2.4 and (4.3), we have

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

so (4.17) implies that $d(0, \mathcal{D}(x_k)) \leq \Delta$. Hence there exists $d_0 \in \mathbb{R}^v$ with $\|d_0\| \leq \Delta$ such that $F(x_k) + F'(x_k)d_0 \in C$. Consequently, by Remark 3.1,

$$D_\Delta(x_k) = \{d \in \mathbb{R}^v : \|d\| \leq \Delta, F(x_k) + F'(x_k)d \in C\}$$

and

$$d(0, D_\Delta(x_k)) = d(0, \mathcal{D}(x_k)).$$

Since $d_k = x_{k+1} - x_k \in D_\Delta(x_k)$ by Algorithm **A**(η, Δ, x_0), it follows that (4.7) holds for $n = k$. Furthermore, one has that

$$\|x_{k+1} - x_k\| \leq \eta d(0, D_\Delta(x_k)) = \eta d(0, \mathcal{D}(x_k)).$$

This with (4.17) yields that (4.6) holds for $n = k$ and hence implication (4.12) is proved.

Clearly, if (4.5) holds for each $n = 1, 2, \dots$, then $\{x_n\}$ is a Cauchy sequence by the monotonicity of $\{t_n\}$ and hence converges to some x^* . Thus (4.8) is clear. Therefore, to prove the theorem, we only need to prove that (4.5), (4.6) and (4.7) hold for each $n = 1, 2, \dots$. We will proceed by the mathematical induction. First, by (4.1), (4.3) and (4.11), $\mathcal{D}(x_0) \neq \emptyset$ and

$$\eta d(0, \mathcal{D}(x_0)) \leq \eta \beta_{x_0}(\|x_0 - x_0\|) d(F(x_0), C) = \eta \beta_{x_0}(0) d(F(x_0), C) = \xi \leq \Delta.$$

Then using the same arguments just used above, we have that (4.7) holds for $n = 0$ and

$$\|x_1 - x_0\| = \|d_0\| \leq \eta d(0, D_\Delta(x_0)) \leq \eta \beta_{x_0}(0) d(F(x_0), C) = \xi = t_{\alpha,1} - t_{\alpha,0};$$

that is, (4.5) holds for $n = 1$. Thus, by (4.12), (4.6) and (4.7) hold for $n = 1$. Furthermore, assume that (4.5), (4.6) and (4.7) hold for all $1 \leq n \leq k$. Then

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

This shows that (4.5) holds for $n = k + 1$ and hence (4.5) holds for all n with $1 \leq n \leq k + 1$. Thus, (4.12) implies that (4.6) and (4.7) hold for $n = k + 1$. Therefore, (4.5), (4.6) and (4.7) hold for each $n = 1, 2, \dots$. The proof is complete. \square

REMARK 4.1. (a) In Theorem 4.1 if one assumes in addition that either **(i)** $\xi < b_\alpha$ or **(ii)** $\alpha > \alpha_0 := \alpha_0(\mathbf{r})$, then the convergence of the sequence $\{x_n\}$ is of quadratic rate. For the case of **(i)**, this remark follows from immediately from Lemma 2.1 **(iii)** thanks to (4.5) and (4.8). If **(ii)** is assumed then, by Lemma 2.2, $\xi \leq b_\alpha < b_{\alpha_0}$ and $r_{\alpha_0}^* \leq r_\alpha^* \leq \mathbf{r}$. Hence (4.3) holds with α_0 in place of α . Since $\xi < b_{\alpha_0}$, we are now in the case **(i)** if α is replaced by α_0 and hence our remark here is established.

(b) Refinements for results presented in the remainder of this paper can also be established in the similar manner as in **(a)** above.

REMARK 4.2. Suppose that there exists a pair $(\bar{\alpha}, \bar{\mathbf{r}})$ such that

$$\begin{cases} \bar{\alpha} = \alpha_0(\bar{\mathbf{r}}) \\ r_{\bar{\alpha}}^* = \bar{\mathbf{r}}. \end{cases} \quad (4.22)$$

Note that the functions $\alpha \mapsto b_\alpha$ is decreasing by Lemma 2.2. Then, if (4.3) holds for some (α, \mathbf{r}) with $\alpha \geq \bar{\alpha}$ and $r \geq \bar{r}$, (4.3) does for $(\alpha, \mathbf{r}) = (\bar{\alpha}, \bar{\mathbf{r}})$ (hence Theorem 4.1 is applicable).

Recall from Proposition 3.3 that the assumption for the existence of r, β in the following corollary is automatically satisfied when $x_0 \in \mathbb{R}^v$ is a regular point of the inclusion (3.1). This remark also applies to Theorems 5.1, 5.6 and Corollary 5.2.

COROLLARY 4.2. *Let $x_0 \in \mathbb{R}^v$ be a regular point of the inclusion (3.1) with $r > 0$ and $\beta > 0$ such that*

$$\mathcal{D}(x) \neq \emptyset \quad \text{and} \quad d(0, \mathcal{D}(x)) \leq \beta d(F(x), C) \quad \text{for all } x \in \mathbf{B}(x_0, r). \quad (4.23)$$

Let $\eta \in [1, \infty)$, $\Delta \in (0, \infty]$, $\xi = \eta\beta d(F(x_0), C)$,

$$\alpha = \frac{\eta\beta}{1 + \eta\beta \int_0^\xi L(u) du} \quad (4.24)$$

and let b_α, r_α be defined by (2.1). Let r_α^* denote the smaller zero of the function ϕ_α defined by (2.2). Suppose that F' satisfies the L -average Lipschitz condition on $\mathbf{B}(x_0, r_\alpha^*)$ and that

$$\xi \leq \min\{b_\alpha, \Delta\} \quad \text{and} \quad r_\alpha^* \leq r; \quad (4.25)$$

(for example, (4.25) is satisfied if

$$\xi \leq \min\{b_\alpha, \frac{b_\alpha}{r_\alpha} r, \Delta\} \quad (4.26)$$

holds). Then the conclusions of Theorem 4.1 hold.

Proof. Note that $\xi < r_\alpha^* \leq r$ by Lemma 2.1 and (4.25). By (3.6) and (3.7), it is clear that $\mathbf{r}_{x_0} \geq r$ and $\beta_{x_0}(\cdot) \leq \beta$ on $[0, r)$. Let $\mathbf{r} := r$ and let $\alpha_0(\mathbf{r})$ be defined by (4.2) as in Theorem 4.1. Then, by (4.24), we have that

$$\alpha \geq \frac{\eta\beta_{x_0}(t)}{1 + \eta\beta_{x_0}(t) \int_0^t L(u) du} \quad \text{for each } t \in [\xi, r).$$

Hence $\alpha \geq \alpha_0(\mathbf{r})$ by (4.2). Note that (4.26) (resp. (4.25)) is identical to (4.4) (resp. (4.3)), Theorem 4.1 is applicable and the proof is complete. \square

COROLLARY 4.3. *Let $\eta \in [1, +\infty)$, $\Delta \in (0, +\infty]$ and let C be a cone. Let $x_0 \in \mathbb{R}^v$ be such that T_{x_0} carries \mathbb{R}^v onto \mathbb{R}^m . Let*

$$\xi = \eta \|T_{x_0}^{-1}\| d(F(x_0), C), \quad (4.27)$$

$$\alpha = \frac{\eta \|T_{x_0}^{-1}\|}{1 + (\eta - 1) \|T_{x_0}^{-1}\| \int_0^\xi L(u) du} \quad (4.28)$$

and let b_α, r_α be defined by (2.1). Let r_α^* denote the smaller zero of the function ϕ_α defined by (2.2). Suppose that F' satisfies the L -average Lipschitz condition on $\mathbf{B}(x_0, r_\alpha^*)$ and that

$$\xi \leq \min\{b_\alpha, \Delta\}. \quad (4.29)$$

Then the conclusions of Theorem 4.1 hold.

Proof. Let $\beta_0 = \|T_{x_0}^{-1}\|$ and let r_{β_0} be defined by (3.13). Then, by Proposition 3.7 (ii), we know that x_0 is a quasi-regular point with the quasi-regular radius

$$\mathbf{r}_{x_0} \geq \min\{r_\alpha^*, r_{\beta_0}\} \quad (4.30)$$

and the the quasi-regular bound function

$$\beta_{x_0}(t) \leq \frac{\beta_0}{1 - \beta_0 \int_0^t L(u) \, du} \quad \text{for each } t \text{ with } 0 \leq t < \min\{r_\alpha^*, r_{\beta_0}\}. \quad (4.31)$$

Let $\mathbf{r} := \min\{r_\alpha^*, r_{\beta_0}\}$, and let $\alpha_0(\mathbf{r})$ be defined by (4.2). We claim that

$$\alpha \geq \alpha_0(\mathbf{r}) \quad (4.32)$$

and

$$r_{\beta_0} \geq r_\alpha^*. \quad (4.33)$$

Granting this, the minimum on the right-hand side of (4.30) is simply r_α^* , and so $\mathbf{r} = r_\alpha^* \leq \mathbf{r}_{x_0}$. Moreover we note that $\beta_{x_0}(0) \leq \beta_0$ by (4.31) and so the ξ defined by (4.1) is majorized by that defined by (4.27); thus (4.29) entails that (4.3) holds and Theorem 4.1 is applicable. Therefore we need only to prove our claim. Note by (4.31) that, for each $\xi \leq t < \min\{r_\alpha^*, r_{\beta_0}\} = \mathbf{r}$, we have

$$\eta \int_0^t L(u) \, du + \frac{1}{\beta_{x_0}(t)} \geq \frac{1}{\beta_0} + (\eta - 1) \int_0^t L(u) \, du \geq \frac{1}{\beta_0} + (\eta - 1) \int_0^\xi L(u) \, du, \quad (4.34)$$

that is,

$$\frac{\beta_{x_0}(t)}{1 + \eta \beta_{x_0}(t) \int_0^t L(u) \, du} \leq \frac{\beta_0}{1 + (\eta - 1) \beta_0 \int_0^t L(u) \, du}; \quad (4.35)$$

thus (4.32) follows by definitions of $\alpha_0(\mathbf{r})$ and of α respectively given by (4.2) and (4.28). To verify (4.33), consider the two cases namely: (i) $\alpha \geq \beta_0$ and (ii) $\alpha < \beta_0$. In (i), since by Lemma 2.2 r_α is decreasing with respect to α , we have that $r_\alpha^* \leq r_\alpha \leq r_{\beta_0}$. In (ii), since r_α^* is increasing with respect to α by Lemma 2.2, we have that $r_\alpha^* \leq r_{\beta_0}^* \leq r_{\beta_0}$. Therefore, (4.33) holds in all cases and the proof is complete. \square

REMARK 4.3. (a) *If the strict inequalities in (4.25) (resp. (4.29)) of Corollary 4.2 (resp. Corollary 4.3) holds, then the starting point x_0 of the sequence $\{x_n\}$ can be replaced by a nearby point, that is, there exists a neighbourhood $U(x_0)$ of x_0 such that the sequence $\{x_n\}$ generated by Algorithm **A**(η, Δ, \bar{x}) with initial point \bar{x} from $U(x_0)$ converges to some solution of the inclusion problem (3.1) at a quadratic rate.*

(b) *Refinements for results presented in the remainder of this paper can also be established in the similar manner as in (a) above.*

5. Special cases and applications. This section is devoted to some applications. Firstly we specialize results of the preceding section to two important cases of the function L : $L = \text{constant}$ and $L = \frac{2\gamma}{(1-\gamma u)^3}$. Secondly, mimicking Smale's γ -theory about the approximation zeros for Newton's method in solving nonlinear equations, we do the same for the Gauss-Newton method in solving composite convex optimization.

5.1. Kantorovich type. Throughout this subsection, we assume that the function L is a constant function. Then, by (2.1) and (2.2), we have that, for all $\alpha > 0$,

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

and

$$\phi_\alpha(t) = \xi - t + \frac{\alpha L}{2} t^2.$$

Moreover, if $\xi \leq \frac{1}{2\alpha L}$, then the zeros of ϕ_α are given by

$$\left. \begin{array}{l} r_\alpha^* \\ r_\alpha^{**} \end{array} \right\} = \frac{1 \mp \sqrt{1 - 2\alpha L \xi}}{\alpha L}. \quad (5.2)$$

It is also known (see for example [8, 17, 29]) that $\{t_{\alpha, n}\}$ has the closed form

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

where

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

For the present case (L is a positive constant), a commonly used version of Lipschitz continuity on $\mathbf{B}(x_0, r)$ is of course the following: a function G is Lipschitz continuous with modulus L (Lipschitz constant) if

$$\|G(x_1) - G(x_2)\| \leq L \|x_1 - x_2\| \quad \text{for all } x_1, x_2 \in \mathbf{B}(x_0, r).$$

Clearly, this is a stronger requirement than the corresponding ones given in Definition 2.5. Although the weaker requirement of Definition 2.1 (b) is sufficient for results in this subsection, we prefer to use the Lipschitz continuity in this regard to be in line with the common practice.

THEOREM 5.1. *Let $x_0 \in \mathbb{R}^v$ be a regular point of the inclusion (3.1) with $r > 0$ and $\beta > 0$ such that (4.23) holds. Let $L \in (0, +\infty)$, $\eta \in [1, +\infty)$, $\Delta \in (0, +\infty]$, $\xi = \eta\beta d(F(x_0), C)$,*

$$R^* = \frac{1 + L\eta\beta\xi - \sqrt{1 - (L\eta\beta\xi)^2}}{L\eta\beta} \quad \text{and} \quad Q = \frac{1 - \sqrt{1 - (L\eta\beta\xi)^2}}{L\eta\beta\xi}. \quad (5.5)$$

Assume that F' is Lipschitz continuous on $\mathbf{B}(x_0, R^)$ with modulus L , and that*

$$\xi \leq \min\left\{\frac{1}{L\beta\eta}, \Delta\right\} \quad \text{and} \quad r \geq R^* \quad (5.6)$$

(for example, (5.6) is satisfied if

$$\xi \leq \min\left\{\frac{1}{L\beta\eta}, \frac{1}{2}r, \Delta\right\} \quad (5.7)$$

holds). Let $\{x_n\}$ denote the sequence generated by Algorithm $\mathbf{A}(\eta, \Delta, x_0)$. Then $\{x_n\}$ converges to some x^ with $F(x^*) \in C$ and*

$$\|x_n - x^*\| \leq \frac{Q^{2^n - 1}}{\sum_{i=0}^{2^n - 1} Q^i} R^* \quad \text{for each } n = 0, 1, \dots. \quad (5.8)$$

Proof. Let α be given as in (4.24) namely $\alpha = \frac{\eta\beta}{1+L\eta\beta\xi}$. Moreover, by (5.1)-(5.5), one has that

$$r_\alpha^* = R^*, \quad q_\alpha = Q, \quad r_\alpha = \frac{1+L\eta\beta\xi}{L\eta\beta}, \quad b_\alpha = \frac{1+L\eta\beta\xi}{2L\eta\beta} \quad (5.9)$$

and

$$t_{\alpha,n} = \frac{1-Q^{2^n-1}}{1-Q^{2^n}} R^*. \quad (5.10)$$

Hence condition (4.3) (resp. (4.4)) is equivalent the three inequalities together:

$$\xi \leq \frac{1+L\eta\beta\xi}{2L\eta\beta}, \quad \xi \leq \Delta \quad \text{and} \quad r_\alpha^* \leq r \quad (\text{resp. } \xi \leq \frac{b_\alpha}{r_\alpha}),$$

and is hence, by (5.9), also equivalent to condition (5.6) (resp. (5.7)). Thus we apply Corollary 4.2 to conclude that the sequence $\{x_n\}$ converges to some x^* with $F(x^*) \in C$ and, for each $n = 1, 2, \dots$,

$$\|x_n - x^*\| \leq r_\alpha^* - t_{\alpha,n}.$$

Noting, by (5.9) and (5.10), that

$$r_\alpha^* - t_{\alpha,n} = \left(1 - \frac{1-Q^{2^n-1}}{1-Q^{2^n}}\right) R^* = \frac{Q^{2^n-1}}{\sum_{i=0}^{2^n-1} Q^i} R^*,$$

it follows that (5.8) holds and the proof is complete. \square

The following corollary (which requires no proof by virtue of Theorem 5.1 and Remark 4.3 **(a)**) is a slight extension of [13, Theorem 1] (which, in turn, extends a result of Burke and Ferris [4, Theorem 4.1]) and our conditions such as (5.11) are more direct than the corresponding ones in [13]. In fact, the conditions **(a)** – **(c)** of [13, Theorem 1] clearly imply the condition (5.11) below. Moreover, by **(a)** and **(b)** of [13, Theorem 1], $\bar{\delta} > 4\eta\bar{\beta}d(F(\bar{x}), C) = 4\bar{\xi}$. Since, for each $\xi \leq \frac{1}{L\bar{\beta}\eta}$,

$$\frac{1+L\eta\bar{\beta}\bar{\xi} - \sqrt{1-(L\eta\bar{\beta}\bar{\xi})^2}}{L\eta\bar{\beta}\bar{\xi}} \leq 2,$$

one has that

$$\frac{1+L\eta\bar{\beta}\bar{\xi} - \sqrt{1-(L\eta\bar{\beta}\bar{\xi})^2}}{L\eta\bar{\beta}} = \frac{1+L\eta\bar{\beta}\bar{\xi} - \sqrt{1-(L\eta\bar{\beta}\bar{\xi})^2}}{L\eta\bar{\beta}\bar{\xi}} \bar{\xi} \leq 2\bar{\xi} < \bar{\delta}.$$

Hence, $\bar{r}^* := \bar{\delta}$ satisfies the requirements of Corollary 5.2 below.

COROLLARY 5.2. *Let $\bar{x} \in \mathbb{R}^v$ be a regular point of the inclusion (3.1) with positive constants \bar{r} and $\bar{\beta}$ satisfying (4.23) in places of r and β , respectively. Let $L \in (0, +\infty)$, $\eta \in [1, +\infty)$, $\Delta \in (0, +\infty]$, $\bar{\xi} = \eta\bar{\beta}d(F(\bar{x}), C)$ and let $\bar{r}^* > \frac{1+L\eta\bar{\beta}\bar{\xi} - \sqrt{1-(L\eta\bar{\beta}\bar{\xi})^2}}{L\eta\bar{\beta}}$. Assume that F' is Lipschitz continuous on $\mathbf{B}(\bar{x}, \bar{r}^*)$ with modulus L , and that*

$$\bar{\xi} < \min\left\{\frac{1}{L\bar{\beta}\eta}, \frac{1}{2}\bar{r}, \Delta\right\}. \quad (5.11)$$

Then, there exists a neighbourhood $U(\bar{x})$ of \bar{x} such that the sequence $\{x_n\}$ generated by Algorithm **A**(η, Δ, x_0) with $x_0 \in U(\bar{x})$ converges at a quadratic rate to some x^* with $F(x^*) \in C$ and the estimate (5.8) holds.

THEOREM 5.3. *Let $\eta \in [1, +\infty)$, $\Delta \in (0, +\infty]$ and let C be a cone. Let $x_0 \in \mathbb{R}^v$ be such that T_{x_0} carries \mathbb{R}^v onto \mathbb{R}^m . Let $L \in (0, +\infty)$ and $\xi = \eta \|T_{x_0}^{-1}\| d(F(x_0), C)$. Instead of (5.5), we write*

$$R^* = \frac{1 + (\eta - 1)L \|T_{x_0}^{-1}\| \xi - \sqrt{1 - 2L \|T_{x_0}^{-1}\| \xi - (\eta^2 - 1)(L \|T_{x_0}^{-1}\| \xi)^2}}{L \|T_{x_0}^{-1}\| \eta} \quad (5.12)$$

and

$$Q = \frac{1 - L \|T_{x_0}^{-1}\| \xi - \sqrt{1 - 2L \|T_{x_0}^{-1}\| \xi - (\eta^2 - 1)(L \|T_{x_0}^{-1}\| \xi)^2}}{L \|T_{x_0}^{-1}\| \eta \xi}. \quad (5.13)$$

Suppose that F' is Lipschitz continuous on $\mathbf{B}(x_0, R^*)$ with modulus L , and that

$$\xi \leq \min \left\{ \frac{1}{L \|T_{x_0}^{-1}\| (\eta + 1)}, \Delta \right\}. \quad (5.14)$$

Then, the same conclusions as that in Theorem 5.1 hold.

Proof. Let α be defined as in (4.28), that is, $\alpha = \frac{\eta \|T_{x_0}^{-1}\|}{1 + (\eta - 1)L \|T_{x_0}^{-1}\| \xi}$. Then, by (5.1), the following equivalences holds:

$$\xi \leq b_\alpha \iff 2L\xi\eta \|T_{x_0}^{-1}\| \leq 1 + (\eta - 1)L \|T_{x_0}^{-1}\| \xi \iff \xi L \|T_{x_0}^{-1}\| (1 + \eta) \leq 1;$$

that is, (4.29) and (5.14) are equivalent. Moreover it is easy to verify that R^* and Q defined in (5.12) and (5.13) respectively are equal to r_α^* and q_α defined in (5.2) and (5.4). Therefore, one can complete the proof in the same way as for Theorem 5.1 but using Corollary 4.3 in place of Corollary 4.2. \square

5.2. Smale's type. Let $\gamma > 0$. For the remainder of this paper we assume that L is the function defined by

$$L(u) = \frac{2\gamma}{(1 - \gamma u)^3} \quad \text{for each } u \text{ with } 0 \leq u < \frac{1}{\gamma}. \quad (5.15)$$

Then, by (2.1), (2.2) and elementary calculation (cf. [28]), one has that for all $\alpha > 0$,

$$r_\alpha = \left(1 - \sqrt{\frac{\alpha}{1 + \alpha}}\right) \frac{1}{\gamma}, \quad b_\alpha = \left(1 + 2\alpha - 2\sqrt{\alpha(1 + \alpha)}\right) \frac{1}{\gamma} \quad (5.16)$$

and

$$\phi_\alpha(t) = \xi - t + \frac{\alpha\gamma t^2}{1 - \gamma t} \quad \text{for each } t \text{ with } 0 \leq t < \frac{1}{\gamma}. \quad (5.17)$$

Thus, from [28], we have the following lemma.

LEMMA 5.4. *Let $\alpha > 0$. Assume that $\xi \leq b_\alpha$, namely,*

$$\gamma\xi \leq 1 + 2\alpha - 2\sqrt{\alpha(1 + \alpha)}. \quad (5.18)$$

Then the following assertions hold.

(a) ϕ_α has two zeros given by

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

(b) the sequence $\{t_{\alpha,n}\}$ generated by Newton's method for ϕ_α with initial point $t_{\alpha,0} = 0$ has the closed form:

$$t_{\alpha,n} = \frac{1 - q_\alpha^{2^n - 1}}{1 - q_\alpha^{2^n - 1} p_\alpha} r_\alpha^* \quad \text{for each } n = 0, 1, \dots, \quad (5.19)$$

where

$$q_\alpha := \frac{1 - \gamma\xi - \sqrt{(1 + \gamma\xi)^2 - 4(1 + \alpha)\gamma\xi}}{1 - \gamma\xi + \sqrt{(1 + \gamma\xi)^2 - 4(1 + \alpha)\gamma\xi}} \quad \text{and} \quad p_\alpha := \frac{1 + \gamma\xi - \sqrt{(1 + \gamma\xi)^2 - 4(1 + \alpha)\gamma\xi}}{1 + \gamma\xi + \sqrt{(1 + \gamma\xi)^2 - 4(1 + \alpha)\gamma\xi}}. \quad (5.20)$$

(c)

$$\frac{t_{\alpha,n+1} - t_{\alpha,n}}{t_{\alpha,n} - t_{\alpha,n-1}} = \frac{1 - q_\alpha^{2^n}}{1 - q_\alpha^{2^{n-1}}} \cdot \frac{1 - q_\alpha^{2^{n-1} - 1} p_\alpha}{1 - q_\alpha^{2^{n+1} - 1} p_\alpha} q_\alpha^{2^{n-1}} \leq q_\alpha^{2^n - 1}. \quad (5.21)$$

For the following lemma, we define, for $\xi > 0$,

$$I(\xi) = \{\alpha > 0 : \xi \leq b_\alpha\} = \{\alpha > 0 : \gamma\xi \leq 1 + 2\alpha - 2\sqrt{\alpha(1 + \alpha)}\}. \quad (5.22)$$

Sometimes in order to emphasize the dependence, we write $q(\alpha, \xi)$ for q_α defined by (5.20).

LEMMA 5.5. *The following assertions hold.*

(i) For each $\alpha > 0$, the function $q(\alpha, \cdot)$ is strictly increasing on $(0, b_\alpha]$.

(ii) For each $\xi > 0$, the function $q(\cdot, \xi)$ is strictly increasing on $I(\xi)$.

Proof. We only prove the assertion (i) as (ii) can be proved similarly. Let $\alpha > 0$. Define

$$g_1(t) = (1 + t)^2 - 4(1 + \alpha)t \quad \text{for each } t$$

and

$$g_2(t) = 1 - t + \sqrt{g_1(t)} \quad \text{for each } t \in (0, 1 + 2\alpha - 2\sqrt{\alpha(1 + \alpha)}]$$

Then,

$$g_1'(t) = 2(1 + t) - 4(1 + \alpha) \quad \text{for each } t$$

and

$$g_2'(t) = -1 + \frac{g_1'(t)}{2\sqrt{g_1(t)}} = \frac{g_1'(t) - 2\sqrt{g_1(t)}}{2\sqrt{g_1(t)}} \quad \text{for each } t \in (0, 1 + 2\alpha - 2\sqrt{\alpha(1 + \alpha)}].$$

Define

$$g(t) = 1 - \frac{2\sqrt{g_1(t)}}{g_2(t)} \quad \text{for each } t \in (0, 1 + 2\alpha - 2\sqrt{\alpha(1+\alpha)}].$$

Then, for each $t \in (0, 1 + 2\alpha - 2\sqrt{\alpha(1+\alpha)}]$,

$$g'(t) = -\frac{g_1'(t)g_2(t) + 2g_1(t) - \sqrt{g_1(t)}g_1'(t)}{g_2^2(t)\sqrt{g_1(t)}} = -\frac{(1-t)g_1'(t) + 2g_1(t)}{g_2^2(t)\sqrt{g_1(t)}}$$

Since (as can be verified easily)

$$(1-t)g_1'(t) + 2g_1(t) = -4\alpha(1+t) < 0,$$

it follows that $g' > 0$ on $(0, 1 + 2\alpha - 2\sqrt{\alpha(1+\alpha)}]$ and hence g is increasing on $(0, 1 + 2\alpha - 2\sqrt{\alpha(1+\alpha)}]$. Noting that

$$q(\alpha, \xi) = g(\gamma\xi) \quad \text{for each } \xi \in (0, b_\alpha],$$

the desired conclusion holds. The proof is complete. \square

THEOREM 5.6. *Let $x_0 \in \mathbb{R}^p$ be a regular point of the inclusion (3.1) with $r > 0$ and $\beta > 0$ such that (4.23) holds. Let $\eta \in [1, +\infty)$, $\Delta \in (0, +\infty]$, $\xi = \eta\beta d(F(x_0), C)$ and $\alpha = \frac{\eta\beta(1-\gamma\xi)^2}{\eta\beta+(1-\eta\beta)(1-\gamma\xi)^2}$. Set*

$$r_\alpha^* = \frac{1 + \gamma\xi - \sqrt{(1 + \gamma\xi)^2 - 4(1 + \alpha)\gamma\xi}}{2(1 + \alpha)\gamma} \quad \text{and} \quad q_\alpha = \frac{1 - \gamma\xi - \sqrt{(1 + \gamma\xi)^2 - 4(1 + \alpha)\gamma\xi}}{1 - \gamma\xi + \sqrt{(1 + \gamma\xi)^2 - 4(1 + \alpha)\gamma\xi}}. \quad (5.23)$$

Assume that F' satisfies the L -average Lipschitz condition on $\mathbf{B}(x_0, r_\alpha^*)$ and that

$$\xi \leq \min \left\{ \frac{1 + 2\eta\beta - 2\sqrt{\eta\beta(1 + \eta\beta)}}{\gamma}, \frac{(1 + 2\eta\beta - 2\sqrt{\eta\beta(1 + \eta\beta)})(1 + \eta\beta)}{1 + \eta\beta - \sqrt{\eta\beta(1 + \eta\beta)}} r, \Delta \right\}. \quad (5.24)$$

Let $\{x_n\}$ denote the sequence generated by Algorithm $\mathbf{A}(\eta, \Delta, x_0)$. Then $\{x_n\}$ converges at a quadratic rate to some x^* with $F(x^*) \in C$ and the following assertions hold.

$$\|x_n - x^*\| \leq q_\alpha^{2^n - 1} r_\alpha^* \quad \text{for all } n = 0, 1, \dots \quad (5.25)$$

and

$$\|x_{n+1} - x_n\| \leq q_\alpha^{2^n - 1} \|x_n - x_{n-1}\| \quad \text{for all } n = 1, 2, \dots \quad (5.26)$$

Proof. By (5.15), $\int_0^\xi L(u) du = (1 - \gamma\xi)^{-2} - 1$; hence α given in the statement of the theorem is consistent with (4.24). Set $\alpha' = \eta\beta$. Then, by (5.16),

$$b_{\alpha'} = \frac{1}{\gamma} \left(1 + 2\eta\beta - 2\sqrt{\eta\beta(1 + \eta\beta)} \right) \quad (5.27)$$

and

$$\frac{b_{\alpha'}}{r_{\alpha'}} = \frac{(1 + 2\eta\beta - 2\sqrt{\eta\beta(1 + \eta\beta)})(1 + \eta\beta)}{1 + \eta\beta - \sqrt{\eta\beta(1 + \eta\beta)}}.$$

Thus (5.24) reads

$$\xi \leq \min\{b_{\alpha'}, \frac{b_{\alpha'}}{r_{\alpha'}}r, \Delta\}. \quad (5.28)$$

Since $\gamma\xi < 1$ by (5.24), it is clear from the definition of α that $\alpha < \alpha'$ if $\xi > 0$ and $\alpha = \alpha'$ if $\xi = 0$. Since the function $u \mapsto b_u$ strictly decreasing by Lemma 2.2, it follows that if $\xi > 0$,

$$b_{\alpha'} < b_{\alpha}. \quad (5.29)$$

We claim that

$$\xi < b_{\alpha}, \quad \xi \leq \Delta \quad \text{and} \quad r_{\alpha}^* < r. \quad (5.30)$$

In fact, this claim is trivially true if $\xi = 0$ and so we can assume that $\xi > 0$. Then the first two inequalities follow from (5.28) and (5.29), while the last inequality follows from the fact that $r_{\alpha}^* < r_{\alpha'}^* \leq \frac{r_{\alpha'}}{b_{\alpha'}}\xi \leq r$ thanks to (5.28), Lemma 2.1 (i) and Lemma 2.2 (iii). Therefore, (5.30) is true. Moreover, by (5.30) and (5.16), $\gamma\xi < 1 + 2\alpha - 2\sqrt{\alpha(1+\alpha)}$, the smaller root of the function $t \mapsto (1+t)^2 - 4(1+\alpha)t$. Therefore, $(1+\gamma\xi)^2 - 4(1+\alpha)\gamma\xi > 0$ and so $q_{\alpha} < 1$ by (5.23). Now, by Corollary 4.2, the sequence $\{x_n\}$ converges to some x^* with $F(x^*) \in C$ and the following estimates hold for each n ,

$$\|x_n - x^*\| \leq r_{\alpha}^* - t_{\alpha,n}, \quad (5.31)$$

$$\|x_{n+1} - x_n\| \leq (t_{\alpha,n+1} - t_{\alpha,n}) \left(\frac{\|x_n - x_{n-1}\|}{t_{\alpha,n} - t_{\alpha,n-1}} \right)^2 \leq \left(\frac{t_{\alpha,n+1} - t_{\alpha,n}}{t_{\alpha,n} - t_{\alpha,n-1}} \right) \|x_n - x_{n-1}\|. \quad (5.32)$$

Hence (5.25) and (5.26) are true because, by (5.19) and (5.21), one has

$$r_{\alpha}^* - t_{\alpha,n} = \frac{q_{\alpha}^{2^n-1}(1-p_{\alpha})}{1-q_{\alpha}^{2^n-1}p_{\alpha}} r_{\alpha}^* \leq q_{\alpha}^{2^n-1} r_{\alpha}^* \quad (5.33)$$

and

$$\frac{t_{\alpha,n+1} - t_{\alpha,n}}{t_{\alpha,n} - t_{\alpha,n-1}} \leq q_{\alpha}^{2^n-1}. \quad (5.34)$$

Thus the convergence of $\{x_n\}$ is quadratic and the proof is complete. \square

The following result can be proved similarly but apply Corollary 4.3 in place of Corollary 4.2, and use $\alpha' := \eta\|T_{x_0}^{-1}\|$ (thus $b_{\alpha'} = \frac{1+2\eta\|T_{x_0}^{-1}\| - 2\sqrt{\eta\|T_{x_0}^{-1}\|(1+\eta\|T_{x_0}^{-1}\|)}}{\gamma}$).

THEOREM 5.7. *Let $\eta \in [1, +\infty)$, $\Delta \in (0, +\infty]$ and let C be a cone. Let $x_0 \in \mathbb{R}^v$ such that T_{x_0} carries \mathbb{R}^v onto \mathbb{R}^m . Let $\xi = \eta\|T_{x_0}^{-1}\|d(F(x_0), C)$ and*

$$\alpha = \frac{\eta\|T_{x_0}^{-1}\|(1-\gamma\xi)^2}{(\eta-1)\|T_{x_0}^{-1}\| + (1-(\eta-1)\|T_{x_0}^{-1}\|)(1-\gamma\xi)^2}.$$

Set, as in (5.23),

$$r_{\alpha}^* = \frac{1+\gamma\xi - \sqrt{(1+\gamma\xi)^2 - 4(1+\alpha)\gamma\xi}}{2(1+\alpha)\gamma} \quad \text{and} \quad q_{\alpha} = \frac{1-\gamma\xi - \sqrt{(1+\gamma\xi)^2 - 4(1+\alpha)\gamma\xi}}{1-\gamma\xi + \sqrt{(1+\gamma\xi)^2 - 4(1+\alpha)\gamma\xi}}. \quad (5.35)$$

Suppose that F' satisfies the L -average Lipschitz condition on $\mathbf{B}(x_0, r_{\alpha}^*)$ and that

$$\xi \leq \min \left\{ \frac{1+2\eta\|T_{x_0}^{-1}\| - 2\sqrt{\eta\|T_{x_0}^{-1}\|(1+\eta\|T_{x_0}^{-1}\|)}}{\gamma}, \Delta \right\}, \quad (5.36)$$

Then the conclusions as that in Theorem 5.6 hold.

5.3. Extension of Smale's approximate zeros. The following notion of approximate zeros was introduced in [26] for Newton's method. Let f be an operator from a domain D in a Banach space X to another one Y . Recall that Newton's iteration for f is defined as follows.

$$x_{n+1} = x_n - f'(x_n)^{-1}f(x_n) \quad n = 0, 1, \dots. \quad (5.37)$$

The sequence $\{x_n\}$ is said to satisfy Smale's condition if

$$\|x_{n+1} - x_n\| \leq \left(\frac{1}{2}\right)^{2^{n-1}} \|x_n - x_{n-1}\| \quad \text{for each } n = 1, 2, \dots. \quad (5.38)$$

Note that (5.38) implies that $\{x_n\}$ is a Cauchy sequence and hence converges (with limit denoted by x^*). By (5.37) it follows that x^* is a zero of f .

DEFINITION 5.8. *Suppose that $x_0 \in D$ is such that Newton iteration (5.37) is well-defined for f and $\{x_n\}$ satisfies Smale's condition. Then x_0 is said to be an approximate zero of f .*

Note that if x_0 is an approximate zero of f then Newton iteration (5.37) converges to a zero x^* of f . We now extend the notion of approximate zeros to the Gauss-Newton method for convex composite optimization problem.

DEFINITION 5.9. *Suppose that $x_0 \in D$ is such that the sequence $\{x_n\}$ generated by Algorithm **A**(η, Δ, x_0) converges to a limit x^* solving (1.1) and satisfies Smale's condition. Then x_0 is said to be an (η, Δ) -approximate solution of (1.1).*

Recall that L is defined by (5.15).

THEOREM 5.10. *Let $x_0 \in \mathbb{R}^v$ be a regular point of the inclusion (3.1) with $r > 0$ and $\beta > 0$ such that (4.23) holds. Let $\eta \in [1, +\infty)$, $\Delta \in (0, +\infty]$, $\xi = \eta\beta d(F(x_0), C)$ and*

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

Suppose that F' satisfies the L -average Lipschitz condition on $\mathbf{B}(x_0, \hat{R})$ and that

$$\xi \leq \min \left\{ \frac{4 + 9\eta\beta - 3\sqrt{\eta\beta(9\eta\beta + 8)}}{4\gamma}, \frac{\left(1 + 2\eta\beta - 2\sqrt{\eta\beta(1 + \eta\beta)}\right)(1 + \eta\beta)}{1 + \eta\beta - \sqrt{\eta\beta(1 + \eta\beta)}} r, \Delta \right\}. \quad (5.39)$$

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

Proof. Let α be defined as in Theorem 5.6 and set $\alpha' = \eta\beta$. Then, as in the proof of Theorem 5.6, we have $\alpha \leq \alpha'$ and $r_{\alpha'} = \hat{R}$ by (5.16). By Lemma 2.2 (iii) and (2.7), it follows that

$$r_{\alpha}^* \leq r_{\alpha'}^* \leq r_{\alpha'} = \hat{R}.$$

Thus, by assumptions, F' satisfies the L -average Lipschitz condition on $\mathbf{B}(x_0, r_{\alpha}^*)$. On the other hand, noting that

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

we see that (5.39) implies (5.24). Therefore, one can apply Theorem 5.6 to conclude that the sequence $\{x_n\}$ converges to a solution x^* of (1.1) and

$$\|x_{n+1} - x_n\| \leq q_\alpha^{2^{n-1}} \|x_n - x_{n-1}\| \quad \text{for all } n = 1, 2, \dots \quad (5.41)$$

It remains to show that $q_\alpha \leq \frac{1}{2}$. To do this we need to emphasize the dependence on the parameters and so we write $q(\alpha, \xi)$ for q_α defined by (5.20) as before. Note that, by (5.27) the right-hand side member of the inequality (5.40) is simply $b_{\alpha'}$ while the left-hand side member majorizes ξ by (5.39). It follows from the monotonicity of $q(\cdot, \cdot)$ established in Lemma 5.5 that

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

where the last equality can be verified elementarily. This completes the proof. \square

Similar to the above proof, we can use Theorem 5.7 in place of Theorem 5.6 to verify the following result.

THEOREM 5.11. *Let $\eta \in [1, +\infty)$, $\Delta \in (0, +\infty]$ and let C be a cone. Let $x_0 \in \mathbb{R}^v$ such that T_{x_0} carries \mathbb{R}^v onto \mathbb{R}^m . Let $\xi = \eta\|T_{x_0}^{-1}\|d(F(x_0), C)$ and*

$$\tilde{R} = \left(1 - \sqrt{\frac{\eta\|T_{x_0}^{-1}\|}{1 + \eta\|T_{x_0}^{-1}\|}}\right) \frac{1}{\gamma}.$$

Suppose that F' satisfies the L -average Lipschitz condition on $\mathbf{B}(x_0, \tilde{R})$ and that

$$\xi \leq \min \left\{ \frac{4 + 9\eta\|T_{x_0}^{-1}\| - 3\sqrt{\eta\|T_{x_0}^{-1}\|(9\eta\|T_{x_0}^{-1}\| + 8)}}{4\gamma}, \Delta \right\}. \quad (5.42)$$

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

6. Examples. Let us begin with a simple example demonstrating a quasis-regular point which is not a regular point.

EXAMPLE 6.1. *Consider the operator F from \mathbb{R}^2 to \mathbb{R}^2 defined by*

$$F(x) = \begin{pmatrix} 1 - t_1 + t_2 + t_1^2 \\ 1 - t_1 + t_2 \end{pmatrix} \quad \text{for each } x = (t_1, t_2) \in \mathbb{R}^2,$$

where \mathbb{R}^2 is endowed with the l_1 -norm. Let $x_0 = 0 \in \mathbb{R}^2$ and $C = \{0\} \subseteq \mathbb{R}^2$. Then

$$F'(x) = \begin{pmatrix} -1 + 2t_1 & 1 \\ -1 & 1 \end{pmatrix} \quad \text{for each } x = (t_1, t_2) \in \mathbb{R}^2;$$

in particular, $F(x_0) = (1, 1)$ and

$$F'(x_0) = \begin{pmatrix} -1 & 1 \\ -1 & 1 \end{pmatrix}.$$

Thus x_0 does not satisfy (3.12). Moreover,

$$\ker F'(x_0) \cap (C - F(x_0))^\ominus = \{(t, t) : t \geq 0\} \neq \{0\}$$

and hence x_0 is not a regular point of the inclusion (3.1). In view of definition of $\mathcal{D}(x)$ in (3.2), we have that, for $x = (t_1, t_2) \in \mathbb{R}^2$,

$$\mathcal{D}(x) = \begin{cases} \{(-\frac{t_1}{2}, -1 + \frac{t_1}{2} - t_2)\} & t_1 \neq 0, \\ \{(d_1, d_1 - 1 - t_2) : d_1 \in \mathbb{R}\} & t_1 = 0 \end{cases}$$

(note that $F'(x)$ is of full-rank if and only if $t_1 \neq 0$). Therefore,

$$d(0, \mathcal{D}(x)) \leq 1 + |t_1| + |t_2| = 1 + \|x\| \quad \text{for each } x = (t_1, t_2) \in \mathbb{R}^2$$

and

$$d(F(x), C) = |1 - t_1 + t_2 + t_1^2| + |1 - t_1 + t_2| \geq 1 - \|x\| \quad \text{for each } x = (t_1, t_2) \in \mathbf{B}(x_0, 1).$$

This implies that

$$d(0, \mathcal{D}(x)) \leq \beta(\|x - x_0\|)d(F(x), C) \quad \text{for each } x = (t_1, t_2) \in \mathbf{B}(x_0, 1),$$

where $\beta(t) = \frac{1+t}{1-t}$ for each $t \in [0, 1)$. Thus, x_0 is a quasi-regular point with quasi-regular radius $\mathbf{r}_{x_0} \geq 1$. In fact, $\mathbf{r}_{x_0} = 1$ because

$$\lim_{t_1 \rightarrow 0^+} \frac{d(0, \mathcal{D}(x))}{d(F(x), C)} = \lim_{t_1 \rightarrow 0^+} \frac{1}{t_1} = +\infty$$

as x goes to $(0, -1)$ on the radial $l : 1 - t_1 + t_2 = 0, t_1 \geq 0$.

Next we give a few examples to illustrate some situations where our results are applicable but not the earlier results in the literature. For the following examples, recall that C is defined by (1.3) and we take

$$\eta = 1 \quad \text{and} \quad \Delta = +\infty. \quad (6.1)$$

Regarding the convergence issue of Gauss-Newton methods, the advantage of considering Wang's L -average Lipschitz condition rather than the classical Lipschitz condition is shown in the following example for which Theorem 5.7 is applicable but not Theorem 5.3.

EXAMPLE 6.2. Let $m = n = 1$ and h be defined by

$$h(y) = \begin{cases} 0 & y \leq 0, \\ y & y \geq 0. \end{cases}$$

Let τ be a constant satisfying

$$10\sqrt{2} - 14 < \tau < 3 - 2\sqrt{2} \quad (6.2)$$

and define

$$F(x) = \begin{cases} \tau - x + \frac{x^2}{1-x} & x \leq \frac{1}{2}, \\ \tau - \frac{1}{2} + 2x^2 & x \geq \frac{1}{2}. \end{cases} \quad (6.3)$$

Then $C = (-\infty, 0]$,

$$F'(x) = \begin{cases} -2 + \frac{1}{(1-x)^2} & x \leq \frac{1}{2}, \\ 4x & x \geq \frac{1}{2} \end{cases}$$

and

$$F''(x) = \begin{cases} \frac{2}{(1-x)^3} & x < \frac{1}{2}, \\ 4 & x > \frac{1}{2}. \end{cases} \quad (6.4)$$

Let $\gamma = 1$ and let L be defined as in (5.15), that is

$$L(u) = \frac{2}{(1-u)^3} \quad \text{for each } u \text{ with } 0 \leq u < 1. \quad (6.5)$$

Then

$$L(u) < L(v) \quad \text{whenever } 0 \leq u < v < 1. \quad (6.6)$$

It follows from (6.4) that

$$\sup\{F''(x) : x \in [-r, r] \setminus \{1/2\}\} = \begin{cases} 16 & r \geq \frac{1}{2}, \\ \frac{2}{(1-r)^3} & 0 < r \leq \frac{1}{2} \end{cases} \quad (6.7)$$

and that

$$0 < F''(u) \leq F''(|u|) \leq L(|u|) \quad \text{whenever } 1/2 \neq u < 1. \quad (6.8)$$

Let $x_0 = 0$. Then, for all $x, x' \in \mathbf{B}(x_0, 1)$ with $|x'| + |x - x'| < 1$, it follows from (6.6) and (6.8) that

$$|F'(x) - F'(x')| = |x - x'| \int_0^1 F''(x' + t(x - x')) dt \leq |x - x'| \int_0^1 L(|x'| + t|x - x'|) dt. \quad (6.9)$$

Thus F' satisfies the L -average Lischitz condition on $\mathbf{B}(x_0, 1)$ with L defined by (6.5). Note that T_{x_0} carries \mathbb{R} onto \mathbb{R} and $\|T_{x_0}^{-1}\| = 1$ as $F'(x_0) = -1$. Let ξ be defined as in Theorems 5.3 and 5.7. Since $F(x_0) = \tau$ and by (6.2), we have

$$\xi = \|T_{x_0}^{-1}\| d(F(x_0), C) = \tau < 3 - 2\sqrt{2}. \quad (6.10)$$

Thus (5.36) is satisfied. Recalling the definitions of α and r_α^* in Theorem 5.7, we have that $\alpha = 1$ and

$$r_\alpha^* = \frac{1 + \xi - \sqrt{(1 + \xi)^2 - 8\xi}}{4} \leq \frac{1 + \xi}{4} \leq \frac{1 + 3 - 2\sqrt{2}}{4} < 1.$$

Therefore Theorem 5.7 is applicable with initial point x_0 . We show next that Theorem 5.3 is not applicable here. In fact, by (6.7), one has that for any $r > 0$, F' is also Lipschitz continuous on $\mathbf{B}(x_0, r)$ with the (least) Lipschitz constant L_r given by

$$L_r = \begin{cases} \frac{2}{(1-r)^3} & r \leq \frac{1}{2}, \\ 16 & r \geq \frac{1}{2} \end{cases} \quad (6.11)$$

Suppose there are ξ , L and R^* satisfying the assumptions stated in Theorem 5.3. For simplicity of notations we write r for R^* . Then by the least property of L_r , (5.12), (5.14) and by a similar argument as we did for (6.10), we have

$$L \geq L_r, \quad (6.12)$$

$$r = \frac{1 - \sqrt{1 - 2L\xi}}{L} \quad (6.13)$$

and

$$\tau = \xi \leq \frac{1}{2L} \leq \frac{1}{2L_r}. \quad (6.14)$$

Since $\tau > 10\sqrt{2} - 14 > \frac{1}{32}$, we have from (6.14) that $L_r < 16$ and it follows from (6.11) that $r < 1/2$ and hence $L_r = \frac{2}{(1-r)^3} \geq 2$. Consequently, by (6.12) and (6.13), we have

$$\tau = \xi = r - \frac{Lr^2}{2} \leq r - \frac{L_r r^2}{2} \leq r - r^2. \quad (6.15)$$

Combining this with (6.14) and (6.11), we have that

$$\tau \leq \min\left\{\frac{(1-r)^3}{4}, r - r^2\right\}. \quad (6.16)$$

Note that the function $r \mapsto \frac{(1-r)^3}{4}$ is decreasing and $r \mapsto r - r^2$ increasing on $[0, \frac{1}{2}]$. Hence

$$\tau \leq \min\left\{\frac{(1-r)^3}{4}, r - r^2\right\} = r_0 - r_0^2 = 10\sqrt{2} - 14, \quad (6.17)$$

where $r_0 = 3 - 2\sqrt{2}$ is the least positive root of equation $\frac{(1-r)^3}{4} = r - r^2$. But (6.17) contradicts (6.2) and therefore the Theorem 5.3 is not applicable to x_0 .

Even when initial point x_0 is regular, the advantage to considering the quasi-regularity bound functions rather than a constant β with the property stated in Proposition 3.3 is shown in the following example for which Theorem 5.1 (and hence results in [4, 13]) are not applicable while Theorems 5.3 is applicable, which is based on the quasi-regular bound function β_{x_0} satisfying (3.14) rather than the quasi-regular bound constant β given by Proposition 3.3.

EXAMPLE 6.3. Let $m = n = 1$ and h be defined by

$$h(y) = |y| \quad \text{for each } y \in \mathbb{R}.$$

Then $C = \{0\}$. Let $\frac{\sqrt{3}-1}{4} < \tau \leq \frac{1}{4}$ and define

$$F(x) = \tau - x + x^2 \quad \text{for each } x \in \mathbb{R}.$$

Then

$$F'(x) = -1 + 2x \quad \text{for each } x \in \mathbb{R};$$

hence F' is Lipschitz continuous with the modular $L = 2$. Let $x_0 = 0$. It is clear that T_{x_0} carries \mathbb{R} onto \mathbb{R} and $\|T_{x_0}^{-1}\| = 1$ as $F'(x_0) = -1$. Since

$$\|T_{x_0}^{-1}\|d(F(x_0), C) = \tau \leq \frac{1}{4}, \quad (6.18)$$

Theorem 5.3 is applicable with initial point $x_0 = 0$. Below we shall show that Theorem 5.1 is not applicable. Suppose on the contrary that there exist $r > 0$ and $\beta > 0$ satisfying the assumptions stated in Theorem 5.1 for x_0 . Then

$$\mathcal{D}(x) \text{ is nonempty and } d(0, \mathcal{D}(x)) \leq \beta d(F(x), C) \quad \text{for each } x \in \mathbf{B}(x_0, r), \quad (6.19)$$

$$r \geq \frac{1 + 2\beta\xi - \sqrt{1 - (2\beta\xi)^2}}{2\beta} \quad (6.20)$$

and

$$\beta\tau = \xi \leq \frac{1}{2\beta}. \quad (6.21)$$

By definitions, it is easy to see that, for each $x \in \mathbb{R}$,

$$\mathcal{D}(x) = \begin{cases} \{-F'(x)^{-1}F(x)\} & x \neq \frac{1}{2}, \\ \emptyset & x = \frac{1}{2}. \end{cases} \quad (6.22)$$

and it follows from (6.19) that $r \leq \frac{1}{2}$ and for each $x \in \mathbf{B}(x_0, r)$,

$$d(0, \mathcal{D}(x)) = |F'(x)^{-1}F(x)| = \frac{1}{1-2|x|}|F(x)| = \frac{1}{1-2|x|}d(F(x), C). \quad (6.23)$$

By (6.19) this implies that

$$\frac{1}{1-2|x|} \leq \beta \quad \text{for each } x \in \mathbf{B}(x_0, r). \quad (6.24)$$

Considering $x_0 = 0$, this implies $\frac{1}{1-2r} \leq \beta$, that is,

$$2\beta r \leq \beta - 1. \quad (6.25)$$

It follows from (6.20) that

$$1 + 2\beta\xi - \sqrt{1 - (2\beta\xi)^2} \leq \beta - 1, \quad (6.26)$$

or equivalently,

$$((2\xi - 1)^2 + (2\xi)^2)\beta^2 + 4(2\xi - 1)\beta + 3 \leq 0. \quad (6.27)$$

Hence

$$(4(2\xi - 1))^2 - 4 \cdot 3((2\xi - 1)^2 + (2\xi)^2) \geq 0 \quad (6.28)$$

which implies that $\xi \leq \frac{\sqrt{3}-1}{4}$. This contradicts the assumption that $\tau > \frac{\sqrt{3}-1}{4}$ because $\xi = \beta\tau \geq \tau$ by (6.21) and (6.24).

We remark that, on one hand, the results in section 5 cover (and improve) the cases considered by Burke-Ferris and Robinson (the initial points are regular in [4, 18]), and on the other hand, there are examples of quasi-regular but not regular points x_0 for which Theorem 4.1 is applicable.

EXAMPLE 6.4. Let $m = n = 3$. To ease our computation, let \mathbb{R}^3 be endowed with the l_1 -norm. Let h be defined by

$$h(x) = \chi(t_1) + \chi(t_2) + \left| t_3 - t_1 - t_2 - \frac{1}{8} \right| \quad \text{for each } x = (t_1, t_2, t_3) \in \mathbb{R}^3,$$

where $\chi(t)$ is a real-valued function on \mathbb{R} defined by

$$\chi(t) = \begin{cases} -1 - t & t \leq -1 \\ 0 & -1 \leq t \leq 0 \\ t & t \geq 0. \end{cases}$$

Define

$$A = \{(c_1, c_2, c_3) : c_3 = c_1 + c_2\} \quad (6.29)$$

and let $F : \mathbb{R}^3 \mapsto \mathbb{R}^3$ be defined by

$$F(x) = \begin{pmatrix} \frac{1}{16} - t_1 + t_1^2 + t_2 + t_3 \\ \frac{1}{16} + t_1 - t_2 + t_2^2 + t_3 \\ t_1^2 + t_2^2 + 2t_3 \end{pmatrix} \quad \text{for each } x = (t_1, t_2, t_3) \in \mathbb{R}^3. \quad (6.30)$$

Then

$$C = \{(c_1, c_2, c_3) : c_1, c_2 \in [-1, 0], c_3 = c_1 + c_2 + \frac{1}{8}\}, \quad (6.31)$$

$$C - F(x) \text{ is contained in } A, \quad (6.32)$$

$$F'(x) = \begin{pmatrix} -1 + 2t_1 & 1 & 1 \\ 1 & -1 + 2t_2 & 1 \\ 2t_1 & 2t_2 & 2 \end{pmatrix} \quad \text{for each } x = (t_1, t_2, t_3) \in \mathbb{R}^3. \quad (6.33)$$

and hence

$$C - F(x) \text{ is contained in } A = \{F'(x)d : d \in \mathbb{R}^3\} \quad \text{for each } x \in \mathbb{R}^3. \quad (6.34)$$

In particular, for $x_0 = 0$, we have that

$$F'(x_0) = \begin{pmatrix} -1 & 1 & 1 \\ 1 & -1 & 1 \\ 0 & 0 & 2 \end{pmatrix}, \quad (6.35)$$

and hence

$$\ker F'(x_0) = \{(t, t, 0) : t \in \mathbb{R}\}. \quad (6.36)$$

Since $F(x_0) = (\frac{1}{16}, \frac{1}{16}, 0)$, one has that

$$\ker F'(x_0) \cap (C - F(x_0))^\ominus = \{(t, t, 0) : t \geq 0\}; \quad (6.37)$$

hence x_0 is not a regular point of (3.1) and the condition of Robinson is not satisfied (see Proposition 3.7). Below we shall show that 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 respectively

$$\mathbf{r}_{x_0} \geq \frac{3}{4} \quad \text{and} \quad \beta_{x_0}(t) \leq \frac{2}{3-4t} \quad \text{for each } t \in [0, \frac{3}{4}]. \quad (6.38)$$

To do this, we note first that F' satisfies the L -average Lipschitz condition on \mathbb{R}^3 with $L = 2$:

$$\|F'(x) - F'(y)\| \leq 2\|x - y\| \quad \text{for each } x, y \in \mathbb{R}^3, \quad (6.39)$$

and the rank of $F'(x)$ is given by:

$$\text{rank } F'(x) = 2 \quad \text{for each } x \in \mathbb{R}^3. \quad (6.40)$$

Since

$$F'(x) = F'(x_0) + F'(x) - F'(x_0) \quad \text{and} \quad \|F'(x) - F'(x_0)\| \leq 2\|x - x_0\|,$$

it follows from the perturbation property of matrixes (cf. [27, 30]) that

$$\|F'(x)^\dagger\| \leq \frac{\|F'(x_0)^\dagger\|}{1 - 2\|x - x_0\|\|F'(x_0)^\dagger\|} \quad (6.41)$$

holds for each $x \in \mathbb{R}^3$ with $2\|x - x_0\|\|F'(x_0)^\dagger\| < 1$, where A^\dagger denotes the Moore-Penrose generalized inverse of the matrix A (cf. [27, 30]). By (6.35), one has that

$$F'(x_0)^\dagger = \begin{pmatrix} -\frac{1}{4} & \frac{1}{4} & 0 \\ \frac{1}{4} & -\frac{1}{4} & 0 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{3} \end{pmatrix} \quad \text{and} \quad \|F'(x_0)^\dagger\| = \frac{2}{3}. \quad (6.42)$$

This together with (6.41) implies that

$$\|F'(x)^\dagger\| \leq \frac{2}{3 - 4\|x - x_0\|} \quad \text{for each } x \in \mathbf{B}(x_0, \frac{3}{4}). \quad (6.43)$$

On the other hand, by (6.33) and (3.2), we have

$$\mathcal{D}(x) = F'(x)^\dagger(C - F(x)) \quad \text{for each } x \in \mathbb{R}^3 \quad (6.44)$$

and, consequently, for each $x \in \mathbf{B}(x_0, \frac{3}{4})$,

$$d(0, \mathcal{D}(x)) \leq \|F'(x)^\dagger\|d(F(x), C) \leq \frac{2}{3 - 4\|x - x_0\|}d(F(x), C). \quad (6.45)$$

This shows that 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 (6.38). Let $\mathbf{r} = \frac{3}{4}$. Recalling (4.2) and (6.1), it follows from (6.38) that

$$\alpha_0(\mathbf{r}) \leq \sup \left\{ \frac{\beta(t)}{\beta(t)2t + 1} : \xi \leq t < \mathbf{r} \right\} = \frac{2}{3}, \quad (6.46)$$

where $\beta(t) := \frac{2}{3-4t}$ for each $t \in [0, \frac{3}{4})$. Thus taking $\alpha = \frac{2}{3}$ in (2.1), we get that

$$r_\alpha = \frac{3}{4} \quad \text{and} \quad b_\alpha = \frac{3}{8}. \quad (6.47)$$

By (4.1) and (6.38),

$$\xi = \beta_{x_0}(0)d(F(x_0), C) \leq \beta(0)\|F(x_0) - c_0\| = \frac{1}{6},$$

where $c_0 = (0, 0, \frac{1}{8})$. It follows that (4.4) is satisfied. Hence Theorem 4.1 is applicable with initial point x_0 even though it is not a regular point.

7. Conclusion. In connection with inclusion problem (1.2) and for a given point x_0 we introduce two new notions: **(a)** the L -average Lipschitz condition for F' and **(b)** quasi-regularity (with the associate quasi-regular radius r_{x_0} and quasi-regular bound function β_{x_0}). The notion **(a)** extends the classical Lipschitz condition and Smale's condition and notion **(b)** extends the regularity. When Robinson's condition (3.12) is satisfied, x_0 is shown to be a regular point and the associate quasi-regular radius r_{x_0} as well as the associate quasi-regular bound function β_{x_0} are estimated if in addition F' satisfies **(a)** with suitable L . We provide sufficient conditions for convergence results with a quasi-regular initial point x_0 in the Gauss-Newton method for the convex composition optimization problem (1.1) with C given to be the set of all minimizers of a convex function h . These conditions are given in terms of r_{x_0} , β_{x_0} and L in the above **(a)**. Examples are given to show that the new concept and results are nontrivial extensions of the existing ones. We thank the referees for suggestions which help our presentations.

REFERENCES

- [1] J. M. Borwein, *Stability and regular points of inequality systems*, J. Optim. Theory Appl., 48(1986), pp. 9-52.
- [2] J. V. Burke, *Descent methods for composite nondifferentiable optimization problems*, Math. programming, 33(3)(1985), pp. 260-279.
- [3] J. V. Burke, *An exact penalization viewpoint of constrained optimization*, SIAM J. Control Optim., 29(1991), pp. 968-998.
- [4] J. V. Burke and M. C. Ferris, *A Gauss-Newton method for convex composite optimization*, Math. Programming, 71(1995), pp. 179-194.
- [5] J. V. Burke and R. A. Poliquin, *Optimality conditions for non-finite valued convex composite function*, Math. Programming, 57(1)(1992), pp. 103-120.
- [6] R. Fletcher, *Second order correction for nondifferentiable optimization*, in: G.A. Watson, ed. Numerical Analysis, Lecture Notes in Mathematics, Vol.912 (Spring, Berlin, 1982), pp. 85-114.
- [7] R. Fletcher, *Practical Methods of Optimization*, Wiley, New York, 2nd ed. 1987.
- [8] W. B. Gragg and R. A. Tapai, *Optimal error bounds for the Newton-Kantorovich theorems*, SIAM J. Numer. Anal., 11(1974), pp. 10-13.
- [9] J. Hiriart-Urruty and C. Lemarechal, *Convex Analysis and Minimization Algorithms II*, Vol. 305 of Grundlehren der Mathematischen Wissenschaften, Springer, New York, 1993.
- [10] K. Jittorntrum and M. R. Osborne, *Strong uniqueness and second order convergence in nonlinear discrete approximation*, Numer. Math., 34(1980), pp. 439-455.
- [11] L. V. Kantorovich, *On Newton method for functional equations*, Dokl. Acad. N USSR, 59(1948), pp. 1237-1240.
- [12] L. V. Kantorovich and G. P. Akilov, *Functional Analysis*, New York: Pergamon Press, 1982.
- [13] C. Li and X. H. Wang, *On convergence of the Gauss-Newton method for convex composite optimization*, Math. Programming, 91(2002), pp. 349-356.
- [14] W. Li and I. Singer, *Global error bounds for convex multifunctions and applications*, Math. Oper. Res., 23(1998), pp. 443-462.
- [15] K. Madsen, *Minimization of nonlinear approximation function*, Ph.D.Thesis, Institute of Numerical Analysis, Technical University of Denmark, (Lyngby, 1985).
- [16] K. F. Ng and X. Y. Zheng, *Characterizations of error bounds for convex multifunctions on Banach spaces*, Math. Oper. Res., 29(2004), pp. 45-63.
- [17] A. M. Ostrowski, *Solutions of Equations in Euclidean and Banach Spaces*, Academic Press, New York, 1973.
- [18] S. M. Robinson, *Extension of Newton's method to nonlinear function with values in cone*, Numer. Math., 19(1972), pp. 341-347.
- [19] S. M. Robinson, *Normed convex process*, Trans. Amer. Math. Soc., 174(1972), pp. 127-140.
- [20] S. M. Robinson, *Stability theory for systems of inequalities, Part I: Linear systems*, SIAM J. Nume. Anal., 12(1975), pp. 754-769.
- [21] S. M. Robinson, *Stability theory for systems of inequalities, Part II: Differentiable nonlinear systems*, SIAM J. Nume. Anal., 13(1976), pp. 479-513.
- [22] R. T. Rockafellar, *First and second order epi-differentiability in nonlinear programming*, Trans. Amer. Math. Soc., 307(1988), pp. 75-108.
- [23] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ. 1970.
- [24] R. T. Rockafellar, *Monotone processes of convex and concave type*, Amer. Math. Soc. Memoir, 77(1967).

- [25] R. T. Rockafellar, *First- and second-order epi-differentiability in nonlinear programming*, Trans. Amer. Math. Soc., 307(1988), pp. 75-108.
- [26] 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.
- [27] G. Stewart, *On the continuity of the generalized inverse*, SIAM J. Appl. Math., 17(1960), 33-45.
- [28] X. H. Wang, *Convergence of Newton's method and inverse function theorem in Banach space*, Math. Comp., 68(1999), pp. 169-186.
- [29] X. H. Wang, *Convergence of an iteration process*, KeXue Tongbao, 20(1975), pp. 558-559.
- [30] P. Wedin, *Perturbation theory for pseudo-inverse*, BIT, 13(1973), pp. 217-232.
- [31] R. S. Womersley, *Local properties of algorithms for minimizing nonsmooth composite function*, Math. Programming, 32(1985), pp. 69-89