J. Nonlinear Funct. Anal. 2016 (2016), Article ID 5

MATHRES
Journal of Nonlinear Functional Analysis
Available online at http://jnfa.mathres.org

ON THE GAUSS-NEWTON METHOD FOR CONVEX OPTIMIZATION USING RESTRICTED CONVERGENCE DOMAINS

IOANNIS K. ARGYROS[1]*, SANTHOSH GEORGE[2]

[1]Department of Mathematical Sciences, Cameron University, Lawton, OK 73505, USA

[2]Department of Mathematical and Computational Sciences, NIT Karnataka, 575025, India

Abstract. Using our new idea of restricted convergence domain, we present a semi-local convergence analysis of the Gauss-Newton method for solving convex composite optimization problems [11,13,17,18,22,23,25]. The convergence analysis is based on a combination of a center-majorant and majorant function. The new approach has the advantage of larger convergence domain, tighter error bounds on the distances involved and the information on the location of the solution is at least as precise.

Keywords. Gauss-Newton method; convex composite optimization problem; semi-local convergence; majorant function; center-majorant function.

1. Introduction

It is well known (see, e.g., [8, 11, 12, 14, 17, 20, 25, 27]) that a lot of problems in Mathematical Programming such as convex inclusion, minimax problems, penalization methods, goal programming, constrained optimization and other problems can be formulated like composite optimization problem.

In this study, using the idea of restricted domain, we present a semi-local convergence analysis with tighter error estimates on the distances involved and the

*Corresponding author.

E-mail addresses: iargyros@cameron.edu (I.K. Argyros), sgeorge@nitk.ac.in (S. George)

Received January 25, 2016

information on the location of the solution is at least as precise. These advantages were obtained (under the same computational cost) using same or weaker sufficient convergence hypotheses.

The study is organized as follows: Section 2 contains the definition of Gauss-Newton Algorithm(GNA) and the semi-local convergence analysis of GNA. Numerical examples and applications of our theoretical results and favorable comparisons to earlier studies (see, e.g., [13, 17, 19, 22, 23]) are presented in Section 2.3.

2. The Gauss-Newton algorithm, regularity and majorant condition

2.1. **Gauss-Newton algorithm.** Let $h:\mathbb{R}^m \longrightarrow \mathbb{R}$ is convex, $F:\mathbb{R}^n \longrightarrow \mathbb{R}^m$ is Fréchet-differentiable operator and $m,n\in\mathbb{N}^\star$. We consider the convex composite optimization problem

$$\min_{x\in\mathbb{R}^n} P(x) := h(F(x)). \tag{2.1}$$

The problem (2.1) is very important because the study of (2.1) provides a unified framework for the development and analysis of algorithmic method and it is a powerfull tool for the study of first and second-order optimality conditions in constrained optimization (see, e.g., [2, 11, 13, 20, 22, 23, 25, 27]). We assume that the minimum h_{min} of the function h is attained. Problem (2.1) is related to

$$F(x)\in\mathscr{C}, \tag{2.2}$$

where

$$\mathscr{C} = \operatorname{argmin} h \tag{2.3}$$

is the set of all minimum points of h.

Let $\xi\in[1,\infty)$, $\Delta\in(0,\infty]$ and for each $x\in\mathbb{R}^n$, define $\mathscr{D}_\Delta(x)$ by

$$\begin{aligned}\mathscr{D}_\Delta(x) \;=\; &\{d\in\mathbb{R}^n : \| d \|\le\Delta,\ h(F(x)+F'(x)d)\le h(F(x)+F'(x)d')\\ &\text{for all } d'\in\mathbb{R}^n \text{ with } \| d' \|\le\Delta\}.\end{aligned} \tag{2.4}$$

Let $x_0 \in \mathbb{R}^n$ be an initial point. The Gauss-Newton algorithm (GNA) associated with (ξ, Δ, x_0) as defined in [13] (see also [11, 17]) is as follows:

Algorithm GNA : (ξ, Δ, x_0)

INICIALIZATION. Take $\xi \in [1, \infty)$, $\Delta \in (0, \infty]$ and $x_0 \in \mathbb{R}^n$, set $k = 0$.

STOP CRITERION. Compute $D_\Delta(x_k)$. If $0 \in \mathscr{D}_\Delta(x_k)$, STOP. Otherwise.

ITERATIVE STEP. Compute d_k satisfying $d_k \in D_\Delta(x_k)$, $\|d_k\| \leq \xi d(0, D\Delta(x_k))$,

Then, set $x_{k+1} = x_k + d_k$, $k = k+1$ and GO TO STOP CRITERION.

Here, $d(x, W)$ denotes the distance from x to W in the finite dimensional Banach space containing W. Note that the set $\mathscr{D}_\Delta(x)$ $(x \in \mathbb{R}^n)$ is nonempty and is the solution of the following convex optimization problem

$$\min_{d \in \mathbb{R}^n, \|d\| \leq \Delta} h(F(x) + F'(x)\,d), \tag{2.5}$$

which can be solved by well known methods such as the subgradient or cutting plane or bundle methods (see, e.g., [13, 20, 25–27]).

Let $U(x, r)$ denote the open ball in $\mathbb{R}^n$ (or $\mathbb{R}^m$) centered at x and of radius $r > 0$. By $\overline{U}(x, r)$ we denote its closure. Let W be a closed convex subset of $\mathbb{R}^n$ (or $\mathbb{R}^m$). The negative polar of W denoted by $W^\ominus$ is defined as

$$W^\ominus = \{z : < z, w > \leq 0 \quad \text{for each} \quad w \in W\}. \tag{2.6}$$

2.2. **Regularity.** In order for us to make the study as self contained as possible, we mention some concepts and results on regularities which can be found in [12] (see also, e.g., [11, 17, 22, 23, 25]). For a set-valued mapping $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and for a set A in $\mathbb{R}^n$ or $\mathbb{R}^m$, we denote by

$$D(T) = \{x \in \mathbb{R}^n : Tx \neq \emptyset\}, \quad R(T) = \bigcup_{x \in D(T)} Tx, \tag{2.7}$$

$$T^{-1}y = \{x \in \mathbb{R}^n : y \in Tx\} \quad \text{and} \quad \| A \| = \inf_{a \in A} \| a \|.$$

Consider the inclusion

$$F(x) \in C, \tag{2.8}$$

where C is a closed convex set in $\mathbb{R}^m$. Let $x \in \mathbb{R}^n$ and

$$\mathscr{D}(x) = \{d \in \mathbb{R}^n : F(x) + F'(x)d \in C\}. \tag{2.9}$$

Definition 2.1. Let $x_0 \in \mathbb{R}^n$.

(a) x_0 is quasi-regular point of (2.8) if there exist $R \in (0, +\infty)$ and an increasing positive function β on $[0, R)$ such that

$$\mathscr{D}(x) \neq \emptyset \text{ and } d(0, \mathscr{D}(x)) \leq \beta(\| x - x_0 \|)\, d(F(x), C) \text{ for all } x \in U(x_0, R). \tag{2.10}$$

$\beta(\| x - x_0 \|)$ is an "error bound" in determining how for the origin is away from the solution set of (2.8).

(b) x_0 is a regular point of (2.8) if

$$ker(F'(x_0)^T) \cap (C - F(x_0))^{\ominus} = \{0\}. \tag{2.11}$$

Proposition 2.2. (see, e.g., [13, 17, 22, 25]) *Let x_0 be a regular point of (2.8). Then, there are constants $R > 0$ and $\beta > 0$ such that (2.10) holds for R and $\beta(\cdot) = \beta$. Therefore, x_0 is a quasi-regular point with the quasi-regular radius $R_{x_0} \geq R$ and the quasi-regular bound function $\beta_{x_0} \leq \beta$ on $[0, R]$.*

Remark 2.3. (a) $\mathscr{D}(x)$ can be considered as the solution set of the linearized problem associated to (2.8)

$$F(x) + F'(x)d \in C. \tag{2.12}$$

(b) If C defined in (2.8) is the set of all minimum points of h and if there exists $d_0 \in \mathscr{D}(x)$ with $\| d_0 \| \leq \Delta$, then $d_0 \in \mathscr{D}_\Delta(x)$ and for each $d \in \mathbb{R}^n$, we have the following equivalence

$$d \in \mathscr{D}_\Delta(x) \Longleftrightarrow d \in \mathscr{D}(x) \Longleftrightarrow d \in \mathscr{D}_\infty(x). \tag{2.13}$$

(c) Let R_{x_0} denote the supremum of R such that (2.10) holds for some function β defined in Definition 2.1. Let $R \in [0, R_{x_0}]$ and $\mathscr{B}_R(x_0)$ denotes the set of function β defined on $[0, R)$ such that (2.10) holds. Define

$$\beta_{x_0}(t) = \inf\{\beta(t) : \beta \in \mathscr{B}_{R_{x_0}}(x_0)\} \quad \text{for each} \quad t \in [0, R_{x_0}). \tag{2.14}$$

All function $\beta \in \mathscr{B}_R(x_0)$ with $\lim_{t\to R^-} \beta(t) < +\infty$ can be extended to an element of $\mathscr{B}_{R_{x_0}}(x_0)$ and we have that

$$\beta_{x_0}(t) = \inf\{\beta(t) : \beta \in \mathscr{B}_R(x_0)\} \quad \text{for each} \quad t \in [0, R). \tag{2.15}$$

R_{x_0} and β_{x_0} are called the quasi-regular radius and the quasi-regular function of the quasi-regular point x_0, respectively.

We denote by r_{x_0} the supremum of r such that (2.10) holds for some increasing positive valued function β on $[0, r)$. That is,

$$r_{x_0} := \sup\{r : \exists\beta : [0, r) \to (0, \infty) \text{ satisfing } (2.10)\}. \tag{2.16}$$

Definition 2.4. *([11,13,17,25]) Let $C \subset \mathbb{R}^m$ be non-empty, closed and convex cone, $F : \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable and $x \in \mathbb{R}^n$. Define the multifunction $T_x : \mathbb{R}^n \to P(\mathbb{R}^m)$ by*

$$T_x d = F'(x)d - C.$$

The domain, norm and inverse of T_x are defined, respectively, by

$$D(T_x) := \{d \in \mathbb{R}^n : T_x d \neq \emptyset\},$$

$$\|T_x\| := \sup\{\|T_x d\| : x \in D(T_x), \|d\| < 1\},$$

$$T_x^{-1} y := \{d \in \mathbb{R}^n : F'(x)d \in y + C\}, y \in \mathbb{R}^m,$$

where $\|T_x d\| := \inf\{\|v\| : v \in T_x d\}$. Then, the point $x_0 \in \mathbb{R}^n$ satisfies the Robinson condition if the multifunction T_{x_0} carries $\mathbb{R}^n$ onto $\mathbb{R}^m$. That is, for each $y \in \mathbb{R}^m$ there exist $d \in \mathbb{R}^n$, $c \in C$ such that $y = F'(x_0)d - C$.

We need the following Banach-type perturbation result.

Lemma 2.5. *([2, 11, 17, 21, 25]) Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable and C a nonempty closed convex cone. Suppose that $x_0 \in \mathbb{R}$ satisfies the Robinson condition. Then*

$$\|T_{x_0}^{-1}\| < \infty.$$

Moreover, if S is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ such that $\|T_{x_0}^{-1}\|\|S\| < 1$, then the convex process $\bar{T}$, defined by $\bar{T} := T_{x_0} + S$, carries $\mathbb{R}^n$ onto $\mathbb{R}^m$, $\|T_{x_0}^{-1}\| < \infty$ and

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

2.3. **Majorant Condition.** We need the definition of the majorant function and the definition of the center-majorant function for F in order to study the semi-local convergence of GNA.

Definition 2.6. *Let $R > 0$, $x_0 \in \mathbb{R}^n$ and $F : \mathbb{R}^n \to \mathbb{R}^m$ be continuously Fréchet-differentiable. A twice-differentiable function $f_0 : [0,R) \to \mathbb{R}$ is called a center-majorant function for F on $U(x_0,R)$, if for each $x \in U(x_0,R)$,*

- (h_0^0)**:** $\|F'(x) - F'(x_0)\| \leq f_0'(\|x - x_0\|) - f_0'(0)$;
- (h_1^0)**:** $f_0(0) = 0$, $f_0'(0) = -1$;

  and
- (h_2^0)**:** f_0' *is convex and strictly increasing.*
- (h_3)**:** $f_0(t) \leq f(t)$ *for each* $t \in [0,R_0)$.

Suppose that $R_0 < R$. If $R_0 \geq R$, we do not need to introduce Definition 2.7.

Definition 2.7. *[6, 9, 11, 17] Let $x_0 \in \mathbb{R}^n$ and $F : \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable. Let $R_0 = \sup\{t \in [0,R) : f_0'(t) < 0\}$. A twice-differentiable function $f : [0,R_0) \to \mathbb{R}$ is called a majorant function for F on $U(x_0,R_0)$, if for each $x, y \in U(x_0,R_0)$, $\|x - x_0\| + \|y - x\| < R_0$,*

- (h_0)**:** $\|F'(y) - F'(x)\| \leq f'(\|y - x\| + \|x - x_0\|) - f'(\|x - x_0\|)$;
- (h_1)**:** $f(0) = 0$, $f'(0) = -1$;

  and
- (h_2)**:** f' *is convex and strictly increasing.*

In Section 4, we present examples where hypotheses (h_3) is satisfied. Let $\xi > 0$ and $\alpha > 0$ be fixed and define auxiliary functions $\varphi : [0,R_0) \to \mathbb{R}$ by

$$\varphi(t) = \xi + (\alpha - 1)t + \alpha f(t). \tag{2.17}$$

We shall use the following hypotheses:

(h_4)**:** there exists $s^* \in (0,R_0)$ such that for each $t \in (0,s^*)$, $\varphi(t) > 0$ and $\varphi(s^*) = 0$;

(h_5)**:** $\varphi(t^*) < 0$.

From now on we assume the hypotheses $(h_0)-(h_4)$ and $(h_0^0)-(h_2^0)$ which will be called the hypotheses (H). Hypothesis (h_5) shall be considered to hold only when explicitly stated.

3. Semi-local convergence

We present the semi-local convergence of (GNA) in this section. First, we need an auxiliary result.

Lemma 3.1. *[11] Let $F:\mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable and let $h:\mathbb{R}^m \to \mathbb{R}$ be real-valued and convex with a nonempty minimizer C. Suppose that $x_0 \in \mathbb{R}^n$ satisfies the Robinson condition. Then, x_0 is a regular point of the inclusion $F(x) \in C$. In particular, x_0 is a quasi-regular point of the inclusion $F(x) \in C$. Moreover, suppose C is a cone, $R_0 > 0$ and $f_0 : [0,R_0) \to \mathbb{R}$ is a center majorant function for F on $U(x_0,R)$. Let $\xi > 0$, $\beta_0 = \|T_{x_0}^{-1}\|$, the auxiliary function $f_{0,\xi,\beta_0} : [0,R_0) \to \mathbb{R}$ defined by*

$$f_{0,\xi,\beta_0}(t) := \xi + (\beta_0 - 1)t + \beta_0 f_0(t)$$

and

$$r_{\beta_0} := \sup\{t \in [0,R_0) : f'_{0,\xi,\beta_0}(t) < 0\}.$$

If r_{x_0} is the quasi-regular radius and $\beta_{x_0}(.)$ is the quasi-regular bound function for the quasi-regular point x_0, then

$$r_{x_0} \geq r_{\beta_0} \text{ and } \beta_{x_0}(t) \leq \frac{\beta_0}{1-\beta_0(1+f_0'(t))} \text{ for each } t \in [0,r_{\beta_0}).$$

Remark 3.2. *If $f_0 = f$, then Lemma 3.1 reduces to the corresponding one in [17, Lemma 25]. Otherwise, i.e., if $f_0'(t) < f'(t)$ for each $t \in [0,R_0)$, then our result constitutes an improvement obtained under less computational cost, since in practice* *the computation of function f requires the computation of f_0 as a special case. Let $f_{\xi,\beta_0} : [0,R_0) \to \mathbb{R}$ be defined by*

$$f_{\xi,\beta_0}(t) := \xi + (\beta_0 - 1)t + \beta_0 f(t) \text{ for each } t \in [0,R_0)$$

$$\bar{r}_{\beta_0} := \sup\{t \in [0,R_0) : f'_{\xi,\beta_0}(t) < 0\}$$

and let $\bar{\beta}_{x_0}(.)$ be the quasi-regular bound function for the quasi-regular point x_0. Then, we have that

$$f_{0,\xi,\beta_0}(t) < f_{\xi,\beta_0}(t)$$

$$r_{\beta_0} \le \bar{r}_{\beta_0}$$

and

$$\beta_{x_0}(t) \le \bar{\beta} x_0(t) \text{ for each } t \in [0, r_{\beta_0}).$$

Hence, our result is an improvement (under less computational cost). We need a semi-local convergence result for GNA.

Theorem 3.3. *[11,17, Theorem 23] Let $F : \mathbb{R}^n \to \mathbb{R}^m$ be continuously differentiable. Suppose that $R > 0$, $x_0 \in \mathbb{R}^n$ and $f : [0,R_0) \to \mathbb{R}$ is a majorant function for F on $U(x_0,R_0)$. Take the constant $\alpha > 0$ and $\xi > 0$ and consider the auxiliary function $f_{\xi,\alpha} : [0,R_0) \to \mathbb{R}$,*

$$f_{\xi,\alpha}(t) = \xi + (\alpha - 1)t + \alpha f(t).$$

If t^* is the smallest zero of $f_{\xi,\alpha}$, then the sequence generated by Newton's Method for solving $f_{\xi,\alpha}(t) = 0$, with initial point $t_0 = 0$,

$$t_{k+1} = t_k - f'_{\xi,\alpha}(t_k)^{-1} f_{\xi,\alpha}(t_k),\ k = 0,1,\cdots,$$

is well defined, $\{t_k\}$ is strictly increasing, remains in $[0,t^*)$, and converges $Q-$linearly to t^*. Let $\eta \in [1,\infty)$, $\Delta \in (0,\infty]$ and $h : \mathbb{R}^m \to \mathbb{R}$ be real-valued and convex with a nonempty minimizer set C. Suppose that C is a cone and $x_0 \in \mathbb{R}^n$ satisfies the Robinson condition. Let $\beta_0 = \|T_{x_0}^{-1}\|$. If $d(F(x_0),C) > 0$, $t^* \le r_{\beta_0} := \{t \in [0,R) : \beta_0 - 1 + \beta_0 f'(t) < 0\}$,

$$\Delta \ge \xi \ge \eta \beta_0 d(F(x_0),C),\ \alpha \ge \frac{\eta \beta_0}{1 + (\eta - 1)\beta_0 [f'(\xi) + 1]},$$

then sequence $\{x_k\}$ *generated by GNA remains in* $U(x_0,t^*)$, *satisfies the inequality*

$$\|x_{k+1}-x_k\| \leq t_{k+1}-t_k,$$

for $k=0,1,\cdots$, *converging to a point* $x^* \in \overline{U}(x_0,t^*)$ *such that* $F(x^*) \in C$,

$$\|x^*-x_k\| \leq t^*-t_k,\ k=0,1,\cdots$$

and the convergence is $R-$*linear. If, additionally,* $f_{\xi,\alpha}$ *satisfies* (h_4) *then the sequences* $\{t_k\}$ *and* $\{x_k\}$ *converge* $Q-$*quadratically and* $R-$ *quadratically to* t^* *and* x^*, *respectively.*

It is convenient for the semi-local convergence analysis that follows to introduce scalar sequences $\{r_k\}$, $\{s_k\}$ by

$$\begin{aligned} r_0 &= 0, r_1=\xi, r_2=r_1-\frac{f_{0,\xi,\alpha}(r_1)-f_{0,\xi,\alpha}(r_0)-f'_{0,\xi,\alpha}(r_0)(r_1-r_0)}{f'_{0,\xi,\alpha}(r_1)} \\ r_{k+2} &= r_{k+1}-\frac{f_{\xi,\alpha}(r_{k+1})-f_{\xi,\alpha}(r_k)-f'_{\xi,\alpha}(r_k)(r_{k+1}-r_k)}{f'_{0,\xi,\alpha}(r_{k+1})} \text{ for each } k=1,2,\cdots \end{aligned}$$

and

$$\begin{aligned} s_0 &= 0, s_1=\xi, \\ s_{k+2} &= s_{k+1}-\frac{f_{\xi,\alpha}(s_{k+1})-f_{\xi,\alpha}(s_k)-f'_{\xi,\alpha}(s_k)(s_{k+1}-s_k)}{f'_{0,\xi,\alpha}(s_{k+1})} \text{ for each } k=0,1,2,\cdots \end{aligned}$$

Notice also that by the definition of Newton's sequence $\{t_k\}$ we have that this sequence can be written as

$$\begin{aligned} t_0 &= 0, t_1=\xi, \\ t_{k+2} &= t_{k+1}-\frac{f_{\xi,\alpha}(t_{k+1})-f_{\xi,\alpha}(t_k)-f'_{\xi,\alpha}(t_k)(t_{k+1}-t_k)}{f'_{\xi,\alpha}(t_{k+1})} \text{ for each } k=0,1,2,\cdots \end{aligned}$$

Next, we compare scalar sequences $\{r_k\}$, $\{s_k\}$, $\{t_k\}$.

Lemma 3.4. *Suppose that*

$$-\frac{(f'_{\xi,\alpha}(u+\theta(v-u))-f'_{\xi,\alpha}(u))(v-u)}{f_{0,\xi,\alpha}(v)} \leq -\frac{(f'_{\xi,\alpha}(\bar{u}+\theta(\bar{v}-\bar{u}))-f'_{\xi,\alpha}(\bar{u}))(\bar{v}-\bar{u})}{f_{\xi,\alpha}(\bar{v})} \tag{3.1}$$

for each $u,v,\bar{u},\bar{v}\in[0,R_0)$, $\theta\in[0,1]$ *with* $u\le v$, $v\le\bar{v}$, $u\le\bar{u}$ *and* $\bar{u}\le\bar{v}$. *Then, sequences* $\{r_k\}$, $\{s_k\}$, $\{t_k\}$ *are increasingly convergent to their unique least upper bounds denoted by* r^*, s^* *and* t^*, *respectively. Moreover, the following estimates hold*

$$r_k \;\le\; s_k\le t_k \tag{3.2}$$

$$r_{k+1}-r_k \;\le\; s_{k+1}-s_k\le t_{k+1}-t_k \tag{3.3}$$

and

$$r^*\le s^*\le t^*. \tag{3.4}$$

Moreover, if (3.1) is a strict inequality, then (3.2) is a strict inequality for $k=2,3,\cdots$ *and (3.3) is a strict inequality for* $k=1,2,\cdots$.

Proof. The monotonicity of sequence $\{t_k\}$ is shown in [17, Corollary 8]. Similarly the monotonicity of sequences $\{r_k\}$, $\{s_k\}$ is shown. Then, using a simple inductive argument, the definition of sequences $\{r_k\}$, $\{s_k\}$ and $\{t_k\}$ and (3.1), we show (3.2) and (3.3). Finally (3.4) follows from (3.2).

□

If one simply follows the proof of Theorem 2.3 in [17] (i.e. essentially the proof of Theorem 12 in [17]) it is straight forward to show that $\{r_k\}$ and $\{s_k\}$ can replace $\{t_k\}$ in the proof of this theorems.

Notice also that in the computation of the upper bound on $\|x_2-x_1\|$ only the center-majorant function f_0 can be used. That justifies the definition of sequence $\{r_k\}$. If the less precise majorant function f is used for the computation of the upper bound on $\|x_2-x_1\|$, then one can obtain the definition of sequence $\{s_k\}$. Sequences $\{r_k\}$, $\{s_k\}$, $\{t_k\}$ converge in general under different convergence criteria (see next section). However, in the next result we show that $\{r_k\}$ and $\{s_k\}$ are majorizing sequences for $\{x_k\}$ that converge under the same convergence criteria as $\{t_k\}$.

Theorem 3.5. *Let* $F:\mathbb{R}^n\to\mathbb{R}^m$ *be continuously differentiable. Suppose that* $R>0$, $x_0\in\mathbb{R}^n$, $f_0:[0,R_0)\to\mathbb{R}$ *is a center-majorant function for* F *on* $[0,R_0)$ *and* $f:$

$[0,R_0) \to \mathbb{R}$ *is a majorant function for* F *on* $[0,R_0)$. *Let* $\alpha > 0$ *and* $\xi > 0$. *Define auxiliary functions*

$$f_{0,\xi,\alpha} : [0,R_0) \to \mathbb{R}$$

$$f_{\xi,\alpha} : [0,R_0) \to \mathbb{R}$$

by

$$f_{0,\xi,\alpha}(t) := \xi + (\alpha - 1)t + f_0(t)$$

and

$$f_{\xi,\alpha}(t) := \xi + (\alpha - 1)t + f(t).$$

If t^* *is the smallest zero of* $f_{\xi,\alpha}$, *then sequences* $\{r_k\}$, $\{s_k\}$, $\{t_k\}$ *are increasingly convergent to* r^*, s^*, t^*, *and converge* $Q-$*linearly provided that (3.1) holds.*

Let $\eta \in [1,\infty)$, $\Delta \in (0,\infty)$ *and* $h : \mathbb{R}^m \to \mathbb{R}$ *be real-valued convex function with a nonempty minimizer set* C. *Suppose that* C *is a cone and* $x_0 \in \mathbb{R}^n$ *satisfies the Robinson condition. Let* $\beta_0 = \|T_{x_0}^{-1}\|$. *Suppose that* $d(F(x_0),C) > 0$,

$$t^* \leq r_{\beta_0} := \{t \in [0,R) : \beta_0 - 1 + \beta_0 f_0'(t) < 0\},$$

$$\Delta \geq \xi \geq \eta \beta_0 d(F(x_0),C),$$

$$\alpha \geq \frac{\eta \beta_0}{1 + (\eta - 1)\beta_0(1 + f_0'(\xi))}.$$

Then, sequence $\{x_k\}$ *generated by (GNA) is well defined, remains in* $U(x_0,r^*)$,

$$F(x_k) + F'(x_k)(x_{k+1} - x_k) \in C \text{ for each } k = 0,1,\cdots$$

and converges to some $x^* \in \overline{U}(x_0,r^*)$ *such that* $F(x^*) \in C$. *Moreover, the following estimates hold for each* $k = 0,1,\cdots$

$$\|x_{k+1} - x_k\| \leq r_{k+1} - r_k \leq s_{k+1} - s_k \leq t_{k+1} - t_k,$$

$$\|x_k - x^*\| \leq r^* - r_k,$$

$$\|x_k - x^*\| \leq s^* - s_k$$

and

$$\|x_k - x^*\| \leq t^* - t_k.$$

If, additionally, $f_{\xi,\alpha}$ satisfies (h_4), then sequences $\{r_k\}$, $\{s_k\}$, $\{t_k\}$, $\{x_k\}$ converge $Q-$quadratically and $R-$quadratically to r^*, s^*, t^*, x^*, respectively.

Proof. The point $x_0 \in \mathbb{R}^n$ satisfies the Robinson condition. We have by Lemma 3.1 that x_0 is a quasi-regular point of the inclusion $F(x) \in C$ with the quasi-regular radius $r_{x_0} \geq r_{\beta_0}$. Then, using the assumption $t^* \leq r_{\beta_0}$ we have that

$$t^* < r_{x_0}.$$

Moreover, Lemma 3.1 also implies that the quasi-regular bound function $\beta_{x_0}(.)$ satisfies

$$\beta_{x_0}(t) \leq \frac{\beta_0}{1-\beta_0[f_0'(t)+1]}, \text{ for each } t \in [0, r_{\beta_0}). \tag{3.5}$$

In view of $\Delta \geq \xi \geq \eta \beta_0 d(F(x_0), C)$ and the preceding inequality implies that $\beta_{x_0}(0) \leq \beta_0$, we get that

$$\Delta \geq \xi \geq \eta \beta_{x_0}(0) d(F(x_0), C).$$

Using $0 < \xi$ and $t^* \leq r_{\beta_0}$ with the first statement in [17, Proposition 10] we conclude that $0 < \xi < t^* \leq r_{\beta_0}$. So, using (3.5), $f_0'(0) = -1$, f_0' as strictly increasing and $\eta \geq 1$; after simple algebraic manipulation we obtain

$$\eta[f_0'(t)+1] + \frac{1}{\beta_{x_0}(t)} \geq \frac{1}{\beta_0} + (\eta-1)[f_0'(t)+1] \geq \frac{1}{\beta_0} + (\eta-1)[f_0'(\xi)+1]. \tag{3.6}$$

It follows from (3.6) that

$$\frac{\eta\beta_0}{1+(\eta-1)\beta_0[f_0'(\xi)+1]} \geq \frac{\eta\beta_{x_0}(t)}{\eta\beta_{x_0}(t)[f_0'(t)+1]+1} \text{ for each } t \in [\xi, t^*).$$

Hence, the assumption $\alpha \geq \eta\beta_0/[1+(\eta-1)\beta_0(f_0'(\xi)+1)]$ and the last inequality imply that

$$\alpha \geq \sup\left\{\frac{\eta\beta_{x_0}(t)}{\eta\beta_{x_0}(t)[f_0'(t)+1]+1} : \xi \leq t \leq t^*\right\}.$$

The result now follows from Lemma 3.1, Theorem 3.3 and Lemma 3.4.

□

Remark 3.6. *If $f_0 = f$ then $t_k = s_k = r_k$ and Theorem 3.5 reduces to Theorem 12 in [17]. Otherwise, (i.e., if $f_0'(t) < f'(t)$ for each $t \in [0, R_0)$) it constitutes an improvement (see also Remark 3.2). So, far these improvement were obtained under the same sufficient convergence conditions. At this point, we are wondering if even the sufficient convergence conditions can be weakened. We present such cases in the next section.*

4. Special cases and applications

We present different sufficient convergence conditions for sequences $\{r_k\}$, $\{s_k\}$, $\{t_k\}$ (i.e., of $\{x_k\}$) in some interesting cases. We shall set

$$f_0(t) = \frac{\alpha L_0}{2}t^2 - t \text{ and } f(t) = \frac{\alpha L}{2}t^2 - t$$

for some $L_0 > 0$ and $L > 0$.

Next, we present the following specialization of Theorem 3.5 under the Robinson condition.

Theorem 4.1. *Let $F : U(x_0, R) \to \mathbb{R}^m$ be continuously differentiable. Suppose: there exist positive constants L_0 and L such that*

$$\|F'(x) - F'(x_0)\| \leq L_0\|x - x_0\| \text{ for each } x \in U(x_0, R)$$

$$\|F'(x) - F'(y)\| \leq L\|x - y\| \text{ for each } x, y \in U(x_0, R_0)$$

$$\delta_0 = \alpha l_0 \xi \leq \frac{1}{2} \tag{4.1}$$

and

$$\alpha \geq \frac{\eta \beta_0}{1 + (\eta - 1)L_0\beta_0\xi}$$

where $l_0 = \frac{1}{8}(4L + \sqrt{L_0 L + 8L^2} + \sqrt{L_0 L})$. Then

(a) *Scalar sequence* $\{r_k\}$ *defined by*

$$\begin{aligned} r_0 &= 0, r_1 = \xi, r_2 = r_1 - \frac{\alpha L_0 (r_1 - r_0)^2}{2(1 - \alpha L_0 r_1)} \\ r_{k+1} &= r_k - \frac{\alpha L (r_k - r_{k-1})^2}{2(1 - \alpha L_0 r_k)} \text{ for each } k = 2, 3, \cdots \end{aligned}$$

is increasingly convergent to its unique least upper bound r^*.

(b) *Sequence* $\{x_k\}$ *generated by (GNA) is well defined, remains in* $U(x_0, r^*)$ *for each* $k = 0, 1, 2 \cdots$ *and converges to a limit point* $x^* \in \overline{U}(x_0, r^*)$ *satisfying* $F(x^*) \in C$. *Moreover, the following estimates hold for each* $k = 0, 1, \cdots$

$$\|x_{k+1} - x_k\| \leq r_{k+1} - r_k$$

and

$$\|x_k - x^*\| \leq r^* - r_k.$$

Notice that the convergence condition (4.1) was given by us in [9].

Remark 4.2. (a) *In particular, if* $C = \{0\}$ *and* $n = m$, *the Robinson condition is equivalent to the condition that* $F'(x_0)^{-1}$ *is non-singular. Hence, for* $\eta = 1$ *we obtain the semi-local convergence for Newton's method defined by*

$$x_{k+1} = x_k - F'(x_k)^{-1} F(x_k) \text{ for each } k = 0, 1, \cdots$$

under the Lipschitz condition [5, 11, 27]. However, the convergence condition in [3], [14]- [18], [20]- [24] is given by

$$\delta = \alpha L \xi \leq \frac{1}{2}. \tag{4.2}$$

Notice again that

$$l_0 \leq L$$

holds in general and $\frac{L}{l_0}$ *can be arbitrarily large [9]. Moreover, the corresponding majorizing sequence sequence* $\{t_k\}$ *is defined by*

$$t_0 = 0, t_1 = \xi, t_{k+1} = t_k - \frac{\alpha L (t_k - t_{k-1})^2}{2(1 - \alpha L t_k)} \text{ for each } k = 1, 2, \cdots.$$

Then, we have for $l_0 < L$ *(i.e.,* $L_0 < L$*) that*

$$r_k < t_k \text{ for each } k = 1, 2, \cdots$$

$$r_{k+1} - r_k < t_{k+1} - t_k \text{ for each } k = 1, 2, \cdots$$

and

$$r^* \leq t^*.$$

Moreover, notice that

$$\delta \leq \frac{1}{2} \Rightarrow \delta_0 \leq \frac{1}{2} \tag{4.3}$$

(but not necessarily vice versa unless if $L_0 = L$*).*

(b) If $n \neq m$, *notice also that if* $L_0 < L$ *the* α *given in the preceding result is larger than the old one using* L *instead of* L_0. *Clearly, the rest of the advantages stated in (a) also hold in this setting.*

(c) If we consider the sequence $\{s_k\}$, *we have that*

$$s_0 = 0,\ s_1 = \xi,\ s_2 = s_1 - \frac{L_0(s_1 - s_0)^2}{2(1 - L_0 s_1)},$$

$$s_{k+2} = s_{k+1} - \frac{L_1(s_{k+1} - s_k)^2}{2(1 - L_0 s_{k+1})} \text{ for each } k = 0, 1, 2, \ldots,$$

where L_1 *is the Lipschitz constant on* $U(x_0, R)$. *Notice that* $U(x_0, R_0) \subseteq U(x_0, R)$, *so* $L \leq L_1$. *The sufficient convergence condition given by us in [11] is*

$$\delta_1 = \alpha l_1 \xi \leq \frac{1}{2}, \tag{4.4}$$

where $l_1 = \frac{1}{8}(4L_1 + L_0 + \sqrt{L_1^2 + 8L_0L_1})$. *Notice again that*

$$\delta \leq \frac{1}{2} \Rightarrow \delta_1 \leq \frac{1}{2}$$

and

$$\delta_1 \leq \frac{1}{2} \Rightarrow \delta_0 \leq \frac{1}{2}$$

but not vice versa unless if $L_0 = L_1$.

Finally, f_0 and f as defined above Theorem 4.1 are center-majorant and majorant functions for F, respectively. Then, with the above choices estimate (3.1) is satisfied. Therefore, the conclusions of Lemma 3.4 for sequences $\{r_k\}, \{s_k\}, \{t_k\}$ hold. Hence, the applicability of the Newton's method or (GNA) under the Robinson condition is expanded under the same computational cost, since in practice the computation of constant L requires the computation of L_0 as a special case.

5. Numerical examples

We present a numerical example for $\alpha = 1, f_0(t) = \frac{L_0}{2}t^2 - t$, $f(t) = \frac{L}{2}t^2 - t$, using Newton's method $f_1(t) = \frac{L_1}{2}t^2 - t$ in this section to show that our results can apply to solve equations in cases the ones in [15] cannot.

EXAMPLE 5.1. *Let $\mathscr{B}_1 = \mathscr{B}_2 = \mathbb{R}, x_0 = 1, D = U(1, 1-q)$, $q \in [0, \frac{1}{2})$ and define function F on D by*

$$F(x) = x^3 - q. \tag{5.1}$$

Then, we have that $\beta = \frac{1}{3}(1-q), L_0 = 3-q, L_1 = 2(2-q)$ and $L = 2(1+\frac{1}{K_0})$. The Newton-Kantorovich condition (4.2) is not satisfied, since

$$\delta > \frac{1}{2} \text{ for each } q \in [0, \frac{1}{2}). \tag{5.2}$$

Hence, there is no guarantee by the Newton-Kantorovich Theorem that Newton's method (2.1) converges to a zero of operator F. Let us see what old condition (4.4) [11] gives:

$$\delta_0 \leq \frac{1}{2}, \text{ if } 0.429999999 < q < \frac{1}{2} \tag{5.3}$$

Theorem 4.1(i.e., (4.1)) gives:

$$\delta_0 \leq \frac{1}{2}, \text{ if } 0.40783335 < q < \frac{1}{2}. \tag{5.4}$$

Hence, we have demonstrated the improvements of Theorem 4.1 using this example.

6. Conclusion

We expanded the applicability of (GNA) under the Robinson condition in order to approximate a solution of a convex composite optimization problem. The advantages denoted by $(\mathscr{A})$ of our analysis over earlier works such as [11, 17, 19, 22, 23, 25] are also shown under the same computational cost.

References

[1] S. Amat, S. Busquier, J.M. Gutiérrez, Geometric constructions of iterative functions to solve nonlinear equations, J. Comput. Appl. Math. 157 (2003), 197–205.

[2] I.K. Argyros, Computational theory of iterative methods, Studies in Computational Mathematics, 15, Editors: K. Chui and L. Wuytack. Elsevier, 2007, New York, U.S.A.

[3] I.K. Argyros, Concerning the convergence of Newton's method and quadratic majorants, J. Appl. Math. Comput. 29 (2009), 391–400.

[4] I.K. Argyros, S. Hilout, On the Gauss–Newton method, J. Appl. Math. 2010 (2010), 1–14.

[5] I.K. Argyros, S. Hilout, Extending the applicability of the Gauss–Newton method under average Lipschitz-conditions, Numer. Alg. 58 (2011), 23–52.

[6] I.K. Argyros, A semilocal convergence analysis for directional Newton methods, Math. Comput. AMS, 80 (2011), 327–343.

[7] I.K. Argyros, Y.J. Cho, S. Hilout, Numerical methods for equations and its applications, CRC Press/Taylor and Francis Group, New-York, 2012.

[8] I.K. Argyros, S. Hilout, Improved local convergence of Newton's method under weaker majorant condition, J. Comput. Appl. Math. 236 (2012), 1892–1902.

[9] I.K. Argyros, S. Hilout, Weaker conditions for the convergence of Newton's method, J. Complexity 28 (2012), 364–387 .

[10] I.K. Argyros, S. Hilout, Computational Methods in Nonlinear Analysis, World Scientific Publ. Comp., New Jersey, 2013.

[11] I. K. Argyros, S. George, Expanding the applicability of the Gauss-Newton method for convex optimization under a regularity condition, Commun. Appl. Nonlinear Anal. 21 (2014), 29–44.

[12] A. Ben-Israel, T.N.E. Greville, Generalized inverses, CMS Books in Mathematics/Ouvrages de Mathematiques de la SMC, 15. Springer-Verlag, New York, second edition, Theory and Applications, 2003.

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

[14] J.E. Dennis, R.B. Schnabel, Numerical methods for unconstrained optimization and nonlinear equations, Classics in Appl. Math., 16, SIAM, Philadelphia, PA, 1996.

[15] O.P. Ferreira, M.L.N. Gonçalves, P.R. Oliveira, Local convergence analysis of the Gauss–Newton method under a majorant condition, J. Complexity, 27 (2011), 111–125.

[16] O.P. Ferreira, B.F. Svaiter, Kantorovich's theorem on Newton's method in Riemmanian manifolds, J. Complexity, 18 (2002), 304–329.

[17] O.P. Ferreira, B.F. Svaiter, Kantorovich's majorants principle for Newton's method, Comput. Optim. Appl. 42 (2009), 213–229.

[18] O.P. Ferreira, M.L.N. Gonçalves, P.R. Oliveira, Convergence of the Gauss- Newton method for convex composite optimization under a majorant condition, SIAM J. Optim. 23 (2013), 1757–1783.

[19] W.M. Häussler, A Kantorovich-type convergence analysis for the Gauss-Newton method. Numer. Math. 48 (1986), 119–125.

[20] J.B. Hiriart-Urruty, C. Lemaréchal, Convex analysis and minimization algorithms (two volumes). I. Fundamentals. II. Advanced theory and bundle methods, 305 and 306, Springer–Verlag, Berlin, 1993.

[21] L.V. Kantorovich, G.P. Akilov, Functional Analysis, Pergamon Press, Oxford (1982)

[22] C. Li, K.F. Ng, Majorizing functions and convergence of the Gauss-Newton method for convex composite optimization, SIAM J. Optim. 18, (2007), 613–692.

[23] C. Li, X.H. Wang, On convergence of the Gauss-Newton method for convex composite optimization, Math. Program. Ser A. 91 (2002), 349–356

[24] A. A. Magreñan, A new tool to study real dynamics: The convergence plane, Appl. Math. Comput. 248 (2014), 215–225.

[25] A. A. Magreñan, I. K. Argyros, Improved convergence analysis for Newton-like methods, Numer. Algo. (2015), 1–23.

[26] Potra, F.A.,Pták, V.: Nondiscrete induction and iterative processes, Pitman, (1994).

[27] S.M. Robinson, Extension of Newton's method to nonlinear functions with values in a cone, Numer. Math. 19 (1972), 341–347

[28] R.T. Rockafellar, Convex analysis, Princeton Mathematical Series, 28, Princeton University Press, Princeton, N.J. (1970)

[29] S. Smale, Newton's method estimates from data at one point. The merging of disciplines: new directions in pure, applied, and computational mathematics (Laramie, Wyo., 1985), 185-196, Springer, New York, 1986.