

Extended convergence of Gauss-Newton's method and uniqueness of the solution

IOANNIS K. ARGYROS, YEOL JE CHO and SANTHOSH GEORGE

ABSTRACT. The aim of this paper is to extend the applicability of the Gauss-Newton's method for solving nonlinear least squares problems using our new idea of restricted convergence domains. The new technique uses tighter Lipschitz functions than in earlier papers leading to a tighter ball convergence analysis.

1. INTRODUCTION

Let us consider the nonlinear least squares problem:

$$(1.1) \quad \min f(x) := \frac{1}{2} F(x)^T F(x),$$

where $F : \mathbb{R}^j \rightarrow \mathbb{R}^i$ is Fréchet-differentiable, $j \geq i$ and $\|\cdot\|$ stands for the 2-norm unless otherwise stated. Gauss-Newton's method (GN) defined for each $n = 0, 1, 2 \dots$ by

$$(1.2) \quad x_{n+1} = x_n - [F'(x_n)^T F'(x_n)]^{-1} F'(x_n)^T F(x_n)$$

where x_0 is an initial point is undoubtedly the most popular method for generating a sequence $\{x_n\}$ approximating a solution p of problem (1.1). There is a plethora of ball convergence results for GN based on Lipschitz-type conditions [1–15]. In the present study, we use our new idea of restricted convergence domains leading to tighter Lipschitz functions and to the advantages (A) over earlier work such as [4–8, 13, 15]:

- (a₁) A larger radius of convergence resulting to a wider choice of initial points.
- (a₂) Tighter error bounds on the distances $\|x_n - p\|$ leading to the computation of fewer iterates to obtain a desired error tolerance.
- (a₃) An at least as precise information on the location of the solution.

The advantages (A) are obtained under the same computational cost as in earlier studies, since in practice the computation of the old Lipschitz functions requires the computation of the new Lipschitz functions as special cases. The study of the ball convergence of iterative methods is also important because it shows the degree of difficulty in obtaining good initial points x_0 . Our technique can be used in an analogous way to improve results for other iterative methods [1, 4–15].

The rest of the paper is structured as follows. Section 2 contains auxiliary results, whereas Section 3 includes the ball convergence of GN.

2. AUXILIARY RESULTS

In order to make the paper as self contained as possible we restate some standard concepts and results.

Received: 08.08.2017. In revised form: 13.03.2018. Accepted: 20.03.2018

2010 *Mathematics Subject Classification.* 65G99, 65H10, 47H17, 49M15.

Key words and phrases. Gauss-Newton's method, least squares problems, Ball convergence, Lipschitz condition.

Corresponding author: Ioannis K. Argyros; iargyros@cameron.edu

Let $\mathbb{R}^{i \times j}$ denote the set of all $i \times j$ matrix M_1 , M_1^\dagger denote the Moore-penrose inverse of matrix M_1 , and if M_1 has full rank (namely: $\text{rank}(M_1) = \min\{m, n\} = n$) then $M_1^\dagger = (M_1^T M_1)^{-1} M_1^T$. We state the auxiliary results [4–8, 12, 15].

Lemma 2.1. *Suppose that $M_1, M_3 \in \mathbb{R}^{i \times j}$, $M_2 = M_1 + M_3$, $\|M_1^\dagger\| \|M_3\| < 1$, $\text{rank}(M_1) = \text{rank}(M_2)$, then*

$$(2.1) \quad \|M_2^\dagger\| \leq \frac{\|M_1^\dagger\|}{1 - \|M_1^\dagger\| \|M_3\|}.$$

Moreover, if $\text{rank}(M_2) = \text{rank}(M_1) = \min\{m, n\}$, then, we have

$$(2.2) \quad \|M_2^\dagger - M_1^\dagger\| \leq \frac{\sqrt{2} \|M_1^\dagger\|^2 \|M_3\|}{1 - \|M_1^\dagger\| \|M_3\|}.$$

Lemma 2.2. *Suppose that $M_1, M_3 \in \mathbb{R}^{i \times j}$, $M_2 = M_1 + M_3$, $\|M_3 M_1^\dagger\| < 1$, $\text{rank}(M_1) = \text{rank}(M_2) = n$, then $\text{rank}(M_2) = n$.*

Lemma 2.3. *Let*

$$(2.3) \quad \varphi(t) = \frac{1}{t^\alpha} \int_0^t L(u) u^{\alpha-1} du, \quad \alpha \geq 1, 0 \leq t \leq \rho,$$

where $L(u)$ is a positive integrable function and nondecreasing monotonically in $[0, \rho]$. Then $\varphi(t)$ is nondecreasing with respect to t .

Lemma 2.4. *Suppose that*

$$(2.4) \quad \psi(t) = \frac{1}{t^\alpha} \int_0^t L(u)(t-u) du,$$

where $L(u)$ is a positive integrable function in $[0, \rho]$. Then $\psi(t)$ is nondecreasing monotonically increasing with respect to t .

3. LOCAL CONVERGENCE

Let $U(w, \xi), \bar{U}(w, \xi)$ stand, respectively for the open and closed balls in \mathbb{R}^i with center $w \in \mathbb{R}^i$ and of radius $\xi > 0$.

We base the local convergence analysis of GN on some scalar Lipschitz functions.

Definition 3.1. Let L_0 be a nondecreasing, positive integrable function defined on the interval $[0, \rho]$. We say that F' satisfies the center-radius Lipschitz condition with L_0 average, if

$$\|F'(x) - F'(p)\| \leq \int_0^{d(x)} L_0(u) du \quad \text{for each } x \in U(p, \rho).$$

Define parameters ρ_0 for some $\beta > 0$ by

$$(3.5) \quad \rho_0 = \sup\{u \in [0, \rho] : \beta L_0(u) < 1\}.$$

Set $U_0 = U(p, \rho_0)$. Notice that $U_0 \subseteq U(p, \rho)$.

Definition 3.2. Let L be a nondecreasing, positive integrable function defined on the interval $[0, \rho_0]$. We say that F' satisfies the restricted-radius Lipschitz condition with L average, if

$$\|F'(x) - F'(x^\theta)\| \leq \int_{\theta d(x)}^{d(x)} L(u) du \quad \text{for each } x \in U(p, \rho_0), 0 \leq \theta \leq 1,$$

where $x^\theta = p + \theta(x - p)$ and $d(x) = \|x - p\|$.

Notice: $L = L(L_0)$. That is the construction of function L depends on L_0 and ρ_0 . In earlier studies [4–8, 14, 15] the following definition was used:

Definition 3.3. Let L_1 be a nondecreasing positive integrable function defined on the interval $[0, \rho]$. We say that F' satisfies the radius Lipschitz condition with L_1 average, if

$$\|F'(x) - F'(x^\theta)\| \leq \int_{\theta d(x)}^{d(x)} L_1(u) du \text{ for each } x \in U(p, \rho), 0 \leq \theta \leq 1.$$

Notice that

$$(3.6) \quad L_0(u) \leq L_1(u),$$

$$(3.7) \quad L(u) \leq L_1(u)$$

hold and $\frac{L_1}{L_0}$ can be arbitrarily large [4–6]. In the earlier studies (with the exception of our works where both L_1 and L_0 are used) only the function L_1 is used in the convergence analysis of the Gauss-Newton method. However, in view of (3.6) and (3.7) the earlier results can be improved, if the more precise function L_0 and L are used instead of L_1 (or L_1 and L_0). If one uses the Banach lemma on invertible operators [11] and L_1 , then

$$(3.8) \quad \|[F'(x)^T F'(x)]^{-1} F'(x)^T\| \leq \frac{\beta}{1 - \beta \int_0^{d(x)} L_1(u) du}$$

is obtained (for $\beta > 0$ to be defined later) instead of the more precise estimate using L_0 :

$$(3.9) \quad \|[F'(x)^T F'(x)]^{-1} F'(x)^T\| \leq \frac{\beta}{1 - \beta \int_0^{d(x)} L_0(u) du}.$$

Similarly, at the numerators of the estimates involved L, L_0 can be used instead of the less precise L_0, L_1 leading to the advantages (A). We can state the main local convergence result for the Gauss-Newton method.

Theorem 3.1. Suppose: vector p satisfies problem (1.1); F has a continuous derivative in $U(p, \rho)$; $F'(p)$ has full rank: F' satisfies the center-radius-Lipschitz condition with L_0 average; F' satisfies the restricted-radius-Lipschitz condition with L average and ρ satisfies

$$(3.10) \quad \lambda(L_0, L, \rho) = \lambda(\rho) = \frac{\beta \int_0^\rho L(u) du}{\rho(1 - \beta \int_0^\rho L_0(u) du)} + \frac{\sqrt{2}c\beta^2 \int_0^\rho L_0(u) du}{\rho(1 - \beta \int_0^\rho L_0(u) du)} \leq 1.$$

Then GN is convergent for all $x_0 \in U(p, \rho)$ and

$$(3.11) \quad \begin{aligned} \|x_{n+1} - p\| &\leq \frac{\beta \int_{d(x_0)} L(u) du}{d(x_0)^2(1 - \beta \int_0^{d(x_0)} L_0(u) du)} \|x_n - p\|^2 \\ &+ \frac{\sqrt{2}c\beta^2 \int_0^{d(x_0)} L_0(u) du}{d(x_0)(1 - \beta \int_0^{d(x_0)} L_0(u) du)} \|x_n - p\|, \end{aligned}$$

where

$$(3.12) \quad c = \|F'(x)\|, \beta = \|[F'(p)^T F'(p)]^{-1} F'(p)^T\|$$

and

$$(3.13) \quad \begin{aligned} q_0(L_0, L) = q_0 &= \frac{\beta \int_0^{d(x_0)} L(u) du}{d(x_0)(1 - \beta \int_0^{d(x_0)} L_0(u) du)} \\ &+ \frac{\sqrt{2}c\beta^2 \int_0^{d(x_0)} L_0(u) du}{d(x_0)(1 - \beta \int_0^{d(x_0)} L_0(u) du)} < 1., \end{aligned}$$

Moreover, if $c = 0$, then

$$(3.14) \quad \|x_n - p\| \leq q_0^{2^n - 1} \|x_0 - p\| \text{ for each } n = 1, 2, \dots$$

Proof. We first show that $q_0 \in [0, 1)$. Using monotonicity of L and Lemma 2.3, we have

$$\begin{aligned} q_0 &= \frac{\beta \int_0^{d(x_0)} L(u) u du}{d(x_0)^2 (1 - \beta \int_0^{d(x_0)} L_0(u) du)} d(x_0) \\ &\quad + \frac{\sqrt{2} c \beta^2 \int_0^{d(x_0)} L_0(u) du}{d(x_0)^2 (1 - \beta \int_0^{d(x_0)} L_0(u) du)} d(x_0) \\ &< \frac{\beta \int_0^\rho L(u) u du}{\rho^2 (1 - \beta \int_0^{d(x_0)} L_0(u) du)} d(x_0) \\ &\quad + \frac{\sqrt{2} c \beta^2 \int_0^{d(x_0)} L_0(u) du}{\rho^2 (1 - \beta \int_0^{d(x_0)} L_0(u) du)} d(x_0) \\ &\leq \frac{\|x_0 - p\|}{\rho} < 1. \end{aligned}$$

and

$$\begin{aligned} \|[F'(p)^T F'(p)]^{-1} \| \|F'(x) - F'(p)\| &\leq \beta \int_0^{d(x)} L_0(u) du \\ &\leq \beta \int_0^\rho L_0(u) du < 1, \text{ for each } x \in U(p, \rho). \end{aligned}$$

By Lemma 2.1 and 2.2, we know that for each $x \in U(p, \rho)$, $F'(x)$ has full rank and

$$\begin{aligned} \|[F'(x)^T F'(x)]^{-1} F'(x)^T\| &\leq \frac{\beta}{1 - \beta \int_0^{d(x)} L_0(u) du}, \\ &\text{for each } x \in U(p, \rho), \end{aligned}$$

$$\begin{aligned} \|[F'(x)^T F'(x)]^{-1} F'(x)^T - [F'(p)^T F'(p)]^{-1} F'(p)^T\| &\leq \frac{\sqrt{2} \beta^2 \int_0^{d(x)} L_0(u) du}{d(x_0) (1 - \beta \int_0^{d(x)} L_0(u) du)}, \\ &\text{for each } x \in U(p, \rho). \end{aligned}$$

If $x_n \in U(p, \rho)$, we have by (1.2)

$$\begin{aligned} x_{n+1} - p &= x_n - p - [F'(x_n)^T F'(x_n)]^{-1} F'(x_n)^T F(x_n) \\ &= [F'(x_n)^T F'(x_n)]^{-1} F'(p)^T [F'(x_n)(x_n - p) - F(x_n) + F(p)] \\ &\quad + [F'(x_n)^T F'(x_n)]^{-1} F'(p)^T F(p) - [F'(x_n)^T F'(x_n)]^{-1} F'(x_n)^T F(p). \end{aligned}$$

That is

$$\begin{aligned} \|x_{n+1} - p\| &\leq \| [F'(x_n)^T F'(x_n)]^{-1} F'(p)^T \| \left\| \int_0^1 [F'(x_n) - F'(p + t(x_n - p))](p - x_n) dt \right\| \\ &\quad + \| [F'(x_n)^T F'(x_n)]^{-1} F'(p)^T F'(p) - [F'(x_n)^T F'(x_n)]^{-1} F'(x_n)^T \| \|F(p)\| \\ &\leq \frac{\beta}{1 - \beta \int_0^{d(x_n)} L_0(u) du} \int_0^1 \int_{td(x_n)}^{d(x_n)} L(u) dud(x_n) dt \\ &\quad + \frac{\sqrt{2}c\beta^2 \int_0^{d(x_n)} L_0(u) du}{1 - \beta \int_0^{d(x_n)} L_0(u) du} \\ &\leq \frac{\beta \int_0^{d(x_n)} L(u) du}{1 - \beta \int_0^{d(x_n)} L_0(u) du} + \frac{\sqrt{2}c\beta^2 \int_0^{d(x_n)} L_0(u) du}{1 - \beta \int_0^{d(x_n)} L_0(u) du}. \end{aligned}$$

Setting $n = 0$ above, we get $\|x_1 - p\| \leq q_0 \|x_0 - p\|$. Hence, $x_1 \in U(p, \rho)$, this shows that (1.1) is well defined and by induction $x_n \in U(p, \rho)$ for all n . Further $d(x) = \|x_n - p\|$ is decreasing monotonically. Therefore, we have

$$\begin{aligned} \|x_{n+1} - p\| &\leq \frac{\beta \int_0^{d(x_n)} L(u) u du}{d(x_n)^2 (1 - \beta \int_0^{d(x_n)} L_0(u) du)} d(x_n)^2 \\ &\quad + \frac{\sqrt{2}c\beta^2 \int_0^{d(x_0)} L_0(u) du}{d(x_n) (1 - \beta \int_0^{d(x_0)} L_0(u) du)} d(x_n) \\ &\leq \frac{\beta \int_0^{d(x_0)} L(u) du}{d(x_0)^2 (1 - \beta \int_0^{d(x_0)} L_0(u) du)} d(x_n)^2 + \frac{\sqrt{2}c\beta^2 \int_0^{d(x_0)} L_0(u) du}{d(x_0) (1 - \beta \int_0^{d(x_0)} L_0(u) du)} d(x_n). \end{aligned}$$

In particular, if $c = 0$, we obtain

$$\|x_{n+1} - p\| \leq \frac{\beta \int_0^{d(x_0)} L(u) du}{d(x_0)^2 (1 - \beta \int_0^{d(x_0)} L_0(u) du)} d(x_n)^2 = \frac{q_0}{d(x_0)} \|x_n - p\|^2.$$

□

Concerning the uniqueness of the solution p we can show:

Proposition 3.1. *Suppose p satisfies (1.1), F has a continuous derivative in $U(p, \rho)$, $F'(p)$ has full rank and F' satisfies the center Lipschitz condition with L_0 average. Let $\rho > 0$ satisfy*

$$(3.15) \quad \mu(L_0) = \frac{\beta}{\rho} \int_0^\rho L_0(u)(\rho - u) du + \frac{c\beta_0}{\rho} \int_0^\rho L_0(u) du \leq 1,$$

where c, β are given in (3.12) and

$$(3.16) \quad \beta_0 = \| [F'(p)^T F'(p)]^{-1} \|.$$

Then, problem (1.1) has a unique solution p in $U(p, \rho)$.

Proof. Let $x_0 \in U(p, \rho)$, $x_0 \neq p$ is also a solution of (1.1). Then we have

$$(3.17) \quad [F'(p)^T F'(p)]^{-1} F'(x_0)^T F(x_0) = 0.$$

Hence,

$$\begin{aligned}
x_0 - p &= x_0 - p - [F'(p)^T F'(p)]^{-1} F'(x_0)^T F(x_0) \\
&= [F'(p)^T F'(p)]^{-1} F'(x_0)^T [F'(p)(x_0 - p) - F(x_0) + F(p)] \\
&\quad + [F'(p)^T F'(p)]^{-1} (F'(x_0)^T - F'(x_0)^T) F(x_0) \\
&= [F'(p)^T F'(p)]^{-1} F'(x_0)^T \int_0^t [F'(p) - F'(p + t(x_0 - p))](x_0 - p) dt \\
&\quad + [F'(p)^T F'(p)]^{-1} (F'(x_0)^T - F'(x_0)^T) F(x_0),
\end{aligned}$$

where $0 \leq t \leq 1$. So, we have

$$\begin{aligned}
\|x_0 - p\| &\leq \| [F'(p)^T F'(p)]^{-1} F'(x_0)^T \| \int_0^t \| [F'(p) - F'(p + t(x_0 - p))] \| \| (x_0 - p) \| dt \\
&\quad + \| [F'(p)^T F'(p)]^{-1} \| \| F'(x_0)^T - F'(x_0)^T \| \| F(x_0) \| \\
&\leq \beta \int_0^1 \int_0^{td(x_0)} L_0(u) du d(x_0) dt + c\beta_0 \int_0^{d(x_0)} L_0(u) du \\
&= \beta \int_0^{d(x_0)} L_0(u) (d(x_0) - u) du + c\beta_0 \int_0^{d(x_0)} L_0(u) du.
\end{aligned}$$

From $L(u) > 0$ and Lemma 2.4, we have $\frac{1}{t} \int_0^t L_0(u) du$ is increasing monotonically with respect to t . Therefore, by (3.15) we obtain

$$\begin{aligned}
\|x_0 - p\| &\leq \beta \int_0^{d(x_0)} L_0(u) (d(x_0) - u) du + c\beta_0 \int_0^{d(x_0)} L_0(u) du \\
&\leq \frac{\beta d(x_0)}{d(x_0)} \int_0^{d(x_0)} L_0(u) (d(x_0) - u) du + \frac{c\beta_0 d(x_0)}{d(x_0)} \int_0^{d(x_0)} L_0(u) du \\
&< \frac{\beta d(x_0)}{\rho} \int_0^\rho L_0(u) (\rho - u) du + \frac{c\beta_0 d(x_0)}{\rho} \int_0^\rho L_0(u) du \\
&\leq \|x_0 - p\|.
\end{aligned}$$

□

The optimality of radius

Theorem 3.2. *Suppose that the equality sign holds in the inequality (3.10) in the Theorem 3.1. Then the value ρ of the convergence ball is the best possible, provided that $L_0 = L$.*

Proof. The value of ρ is specified by equation

$$(3.18) \quad \frac{\beta \int_0^\rho L(u) u du}{\rho(1 - \beta \int_0^\rho L_0(u) du)} + \frac{\sqrt{2}c\beta^2 \int_0^\rho L_0(u) u du}{\rho(1 - \beta \int_0^\rho L_0(u) du)} = 1,$$

there exists F satisfying (3.5) in $U(p, \rho)$ and x_0 on the boundary of the closed ball such that GN fails. In fact, the following is an example on the scaled case:

$$(3.19) \quad F(x) = \begin{cases} p - x + \beta \int_0^{x-p} (x-p-u)L(u) du, & p \leq x < p + \rho' \\ p - x + \beta \int_0^{x-p} (x-p+u)L(u) du, & p - \rho \leq x < p, \end{cases}$$

and $x_0 = p + \rho$, $x_n = p + (-1)^n \rho$.

□

Theorem 3.3. *Suppose that the equality sign holds in the inequality (3.15) in the Proposition 3.1. Then the value of the convergence ball is the best possible.*

Proof. Note that when ρ is given by

$$(3.20) \quad \frac{\beta}{\rho} \int_0^\rho L_0(u)(\rho - u)du + \frac{c\beta_0}{\rho} \int_0^\rho L_0(u)du = 1$$

there exist F satisfying Definition 3.1 in $U(p, \rho)$ and x' on the boundary of the closed ball such that $F'(x') = \min\{\frac{1}{2}F'(x)^T, F'(x)\}$. For example consider (3.19) with $x' = p + \rho$. \square

Remark 3.1. If $L_0 = L = L_1$, we obtain the results in [7] which in turn generalized earlier results in [11, 13, 14] when these functions are constants or not. Moreover, if $L = L_1$, then the results reduce to the ones obtained by us in [4–6]. Otherwise, i.e., if $L_0 < L < L_1$ (or $L < L_0 < L_1$) (see [4–6] for examples), then we obtain a larger radius of convergence, tighter error bounds on the distances $\|x_n - p\|$ and an at least as precise information on the location of the solution p , since

$$\begin{aligned} \lambda(L_0, L) &< \lambda(L_0, L_1) < \lambda(L_1, L_1) \\ q_0(L_0, L) &< q_0(L_0, L_1) < q_0(L_1, L_1) \end{aligned}$$

and

$$\mu(L_0) < \mu(L) < \mu(L_1).$$

These advantages are obtained under the same computational cost, since in practice the computation of function L_1 requires the computation of function L_0 and L as special cases.

REFERENCES

- [1] Amat, S., Hernández, M. A. and Romero, N., *Semilocal convergence of a sixth order iterative method for quadratic equations*, Applied Numerical Mathematics, **62** (2012), 833–841
- [2] Argyros, I. K., Cho, Y. J. and Hilout, S., *Numerical methods for equations and its applications*, CRC Press, Taylor and Francis, New York, 2012
- [3] Argyros, I. K., Cho, Y. J. and George, S., *Local convergence for some third-order iterative methods under weak conditions*, J. Korean Math. Soc., **53** (2016), No. 4, 781–793
- [4] Argyros, I. K. and Magreñán, A. A., *Iterative Algorithms I*, Nova Science Publishers Inc., New York, 2016
- [5] Argyros, I. K. and Magreñán, A. A., *Iterative methods and their dynamics with applications*, CRC Press, New York, 2017
- [6] Argyros, I. K. and Magreñán, A. A., *Iterative Algorithms II*, Nova Science Publishers Inc., New York, 2017
- [7] Chen, J. and Li, W., *Convergence of Gauss-Newton's method and uniqueness of the solution*, Appl. Math. Comput., **170** (2005), 687–705
- [8] Haubler, W. M., *A Kantorovich convergence analysis for the Gauss-Newton method*, Numer. Math., **48** (1986), 119–125
- [9] Magreñán, A. A., *Different anomalies in a Jarratt family of iterative root finding methods*, Appl. Math. Comput., **233** (2014), 29–38
- [10] Magreñán, A. A., *A new tool to study real dynamics: The convergence plane*, Appl. Math. Comput., **248** (2014), 29–38
- [11] Ortega, M. and Rheinboldt, W. C., *Iterative solution of nonlinear equations in several variables*, Academic Press, New York, 1970
- [12] Stewart, G. W., *On the continuity of the generalized inverse*, SIAM J. Math., **17** (1960), 33–45
- [13] Smale, S., *Newton's method estimates from data at one point* in: R Ewing, K. Gross, C. Martin (Eds). *The merging of disciplines: New directions in Pure Applied and Computational Mathematics*, Springer, New York, 1886, 185–196
- [14] Traub, J. F. and Wozniakowski, H., *Convergence and complexity of Newton iteration*, J. Assoc. Comput. Math., **29** (1979), 250–258
- [15] Wang, X., *Convergence of Newton's method and uniqueness of the solution of equations in Banach space*, IMA J. Numer. Anal., **20** (2000), 123–134

DEPARTMENT OF MATHEMATICAL SCIENCES
CAMERON UNIVERSITY
LAWTON, OK 73505, USA
E-mail address: iargyros@cameron.edu

DEPARTMENT OF MATHEMATICS EDUCATION AND THE RINS
GYEONGSANG NATIONAL UNIVERSITY
JINJU 660-701, KOREA
E-mail address: yjcho@gnu.ac.kr

DEPARTMENT OF MATHEMATICAL AND COMPUTATIONAL SCIENCES
NIT KARNATAKA, INDIA-575 025
E-mail address: sgeorge@nitk.ac.in