axioms

MDPI

Article

Gauss–Newton–Secant Method for Solving Nonlinear Least Squares Problems under Generalized Lipschitz Conditions

Ioannis K. Argyros [1], Stepan Shakhno [2,*], Roman Iakymchuk [3,4], Halyna Yarmola [5] and Michael I. Argyros [6]

[1] Department of Mathematical Sciences, Cameron University, Lawton, OK 73505, USA; iargyros@cameron.edu
[2] Department of Theory of Optimal Processes, Ivan Franko National University of Lviv, Universytetska Str. 1, 79000 Lviv, Ukraine
[3] PEQUAN, LIP6, Sorbonne Université, 4 Place Jussieu, 75252 Paris, France; roman.iakymchuk@sorbonne-universite.fr
[4] Fraunhofer ITWM, Fraunhofer-Platz 1, 67663 Kaiserslautern, Germany
[5] Department of Computational Mathematics, Ivan Franko National University of Lviv, Universytetska Str. 1, 79000 Lviv, Ukraine; halyna.yarmola@lnu.edu.ua
[6] Department of Computer Science, University of Oklahoma, Norman, OK 73071, USA; michael.i.argyros-1@ou.edu
* Correspondence: stepan.shakhno@lnu.edu.ua; Tel.: +38-0322-394-791

Abstract: We develop a local convergence of an iterative method for solving nonlinear least squares problems with operator decomposition under the classical and generalized Lipschitz conditions. We consider the case of both zero and nonzero residuals and determine their convergence orders. We use two types of Lipschitz conditions (center and restricted region conditions) to study the convergence of the method. Moreover, we obtain a larger radius of convergence and tighter error estimates than in previous works. Hence, we extend the applicability of this method under the same computational effort.

Keywords: nonlinear least squares problem; differential-difference method; divided differences; radius of convergence; residual; error estimates

MSC: 65J15

Citation: Argyros, I.K.; Shakhno, S.; Iakymchuk, R.; Yarmola H.; Argyros, M.I. Gauss–Newton–Secant Method for Solving Nonlinear Least Squares Problems under Generalized Lipschitz Conditions. *Axioms* **2021**, *10*, 158. https://doi.org/10.3390/axioms10030158

Academic Editors: Jesús Martín Vaquero, Deolinda M. L. Dias Rasteiro, Araceli Queiruga-Dios and Fatih Yilmaz

Received: 22 June 2021
Accepted: 19 July 2021
Published: 21 July 2021

1. Introduction

Nonlinear least squares problems often arise while solving overdetermined systems of nonlinear equations, estimating parameters of physical processes by measurement results, constructing nonlinear regression models for solving engineering problems, etc. The most used method for solving nonlinear least squares problems is the Gauss–Newton method [1]. In the case when the derivative can not be calculated, difference methods are used [2,3].

Some nonlinear functions have a differentiable and a nondifferentiable part. In this case, a good idea is to use a sum of the derivative of the differentiable part of the operator and the divided difference of the nondifferentiable part instead of the Jacobian [4–6]. Numerical study shows that these methods converge faster than Gauss–Newton type's method or difference methods.

In this paper, we study the local convergence of the Gauss–Newton–Secant method under the classical and generalized Lipschitz conditions for first-order Fréchet derivative and divided differences.

Let us consider the nonlinear least squares problem:

$$\min_{x \in R^p} \frac{1}{2}(F(x) + G(x))^T (F(x) + G(x)), \tag{1}$$

where residual function $F + G : R^p \to R^m$ $(m \geq p)$ is nonlinear in x, F is a continuously differentiable function, and G is a continuous function, the differentiability of which, in general, is not required.

We propose the following modification of the Gauss–Newton method combined with the Secant-type method [4,6] for finding the solution to problem (1):

$$x_{n+1} = x_n - (A_n^T A_n)^{-1} A_n^T (F(x_n) + G(x_n)), n = 0, 1, \ldots, \quad (2)$$

where $A_n = F'(x_n) + G(x_n, x_{n-1})$, $F'(x_n)$ is a Fréchet derivative of $F(x)$; $G(x_n, x_{n-1})$ is a divided difference of the first order of function $G(x)$ [7] at points x_n, x_{n-1}; and x_0, x_{-1} are given.

Setting $A_n = F'(x_n)$, for solving problem (1), from (2) we obtain an iterative Gauss–Newton-type method:

$$x_{n+1} = x_n - (F'(x_n)^T F'(x_n))^{-1} F'(x_n)^T (F(x_n) + G(x_n)), \quad n = 0, 1, \ldots. \quad (3)$$

For $m = p$, problem (1) turns into a system of nonlinear equations:

$$F(x) + G(x) = 0. \quad (4)$$

In this case, method (2) is transformed into the combined Newton–Secant method [8–10]:

$$x_{n+1} = x_n - (F'(x_n) + G(x_n, x_{n-1}))^{-1} (F(x_n) + G(x_n)), \; n = 0, 1, \ldots, \quad (5)$$

and method (3) into the Newtons-type method for solving nonlinear equations [11]:

$$x_{n+1} = x_n - (F'(x_n))^{-1} (F(x_n) + G(x_n)), \; n = 0, 1, \ldots. \quad (6)$$

The convergence domain is small (in general), and error estimates are pessimistic. These problems restrict the applicability of these methods. The novelty of our work is in the claim that these problems can be addressed without adding hypotheses. In particular, our idea is to use a center and restricted radius Lipschitz conditions. Such an approach to the study of the convergence of methods allows for extending the convergence ball of the method and improving error estimates.

The remainder of the paper is organized as follows: Section 2 deals with the local convergence analysis. The numerical experiments appear in Section 3. Section 4 contains the concluding remarks and ideas about future works.

2. Local Convergence Analysis

Let us consider, at first, some auxiliary lemmas needed to obtain the main results. Let D be an open subset of R^p.

Lemma 1 ([4]). *Let $e(t) = \int_0^t E(u)du$, where E is an integrable and positive nondecreasing function on $[0, T]$. Then, $e(t)$ is monotonically increasing with respect to t on $[0, T]$.*

Lemma 2 ([1,12]). *Let $h(t) = \frac{1}{t} \int_0^t H(u)du$, where H is an integrable and positive nondecreasing function on $[0, T]$. Then, $h(t)$ is nondecreasing with respect to t on $(0, T]$.*

Additionally, $h(t)$ at $t = 0$ is defined as $h(0) = \lim_{t \to 0} \left(\frac{1}{t} \int_0^t H(u)du \right)$.

Lemma 3 ([13]). *Let $s(t) = \frac{1}{t^2} \int_0^t S(u)u\,du$, where S is an integrable and positive nondecreasing function on $[0, T]$. Then, $s(t)$ is nondecreasing with respect to t on $(0, T]$.*

Definition 1. *The Fréchet derivative F' satisfies the center Lipschitz condition on D with L_0 average if*

$$\|F'(x) - F'(x^*)\| \leq \int_0^{\rho(x)} L_0(u)du, \quad \text{for each} \quad x \in D \subset R^p, \tag{7}$$

where $\rho(x) = \|x - x^*\|$, $x^* \in D$ is a solution of problem (1), and L_0 is an integrable, positive, and nondecreasing function on $[0, T]$.

The functions M_0, L, M, L_1 and M_1 introduced next are as the function L_0: integrable, positive, and nondecreasing functions defined on $[0, 2R]$.

Definition 2. *The first order divided difference $G(x, y)$ satisfies the center Lipschitz condition on $D \times D$ with M_0 average if*

$$\|G(x, y) - G(x^*, x^*)\| \leq \int_0^{\rho(x)+\rho(y)} M_0(u)du, \quad \text{for each} \quad x, y \in D. \tag{8}$$

Let $B > 0$ and $\alpha > 0$. We define function φ on $[0, +\infty)$ by

$$\varphi(t) = B\left[2\alpha + \int_0^t L_0(u)du + \int_0^{2t} M_0(u)du\right]\left[\int_0^t L_0(u)du + \int_0^{2t} M_0(u)du\right]. \tag{9}$$

Suppose that equation

$$\varphi(t) = 1 \tag{10}$$

has at least one positive solution. Denote by γ the minimal such solution. Then, we can define $\Omega_0 = D \cap \Omega(x^*, \gamma)$, where $\Omega(x^*, \gamma) = \{x : \|x - x^*\| < \gamma\}$.

Definition 3. *The Fréchet derivative F' satisfies the restricted radius Lipschitz condition on Ω_0 with L average if*

$$\|F'(x) - F'(x^\tau)\| \leq \int_{\tau\rho(x)}^{\rho(x)} L(u)du, \; x^\tau = x^* + \tau(x - x^*), 0 \leq \tau \leq 1, \text{for each } x \in \Omega_0. \tag{11}$$

Definition 4. *The first order divided difference $G(x, y)$ satisfies the restricted radius Lipschitz condition on Ω_0 with M average if*

$$\|G(x, y) - G(u, v)\| \leq \int_0^{\|x-u\|+\|y-v\|} M(u)du, \quad \text{for each} \quad x, y, u, v \in \Omega_0. \tag{12}$$

Definition 5. *The Fréchet derivative F' satisfies the radius Lipschitz condition on D with L_1 average if*

$$\|F'(x) - F'(x^\tau)\| \leq \int_{\tau\rho(x)}^{\rho(x)} L_1(u)du, \quad \text{for each} \quad x \in D. \tag{13}$$

Definition 6. *The first order divided difference $G(x, y)$ satisfies the radius Lipschitz condition on D with M_1 average if*

$$\|G(x, y) - G(u, v)\| \leq \int_0^{\|x-u\|+\|y-v\|} M_1(u)du, \quad \text{for each} \quad x, y, u, v \in D. \tag{14}$$

Remark 1. *It follows from the preceding definitions that $L = L(L_0, M_0)$, $M = M(L_0, M_0)$, and for each $t \in [0, \gamma]$*

$$L_0(t) \leq L_1(t), \tag{15}$$

$$L(t) \leq L_1(t), \tag{16}$$

$$M(t) \leq M_1(t), \tag{17}$$

since $\Omega_0 \subseteq D$. *By* $L(L_0, M_0)$, *we mean that* L *(or* M*) depends on* L_0 *and* M_0 *by the definition of* Ω_0. *In case any of* (15)–(17) *are strict inequalities, the following benefits are obtained over the work in [4] using* L_1, M_1 *instead of the new functions:*

(a1) An at least as large convergence region leading to at least as many initial choices;
(a2) At least as tight upper bounds on the distances $\|x_n - x^*\|$, *so at least as few iterations are needed to obtain a desired error tolerance.*

These benefits are obtained under the same computational effort as in [4], since the new functions L_0, M_0, L, and M are special cases of the functions L_1 and M_1. This technique of using the center Lipschitz condition in combination with the restricted convergence region has been used by us on Newton's, Secant, Newton-like methods [14,15], and can be used on other methods, too, with the same benefits.

The proof of the next result follows as the corresponding one in [4], but there are crucial differences, where we use (L_0, L) instead of L_1 and (M_0, M) instead of M_1 used in [4].

We use the Euclidean norm. Note that the following equality is satisfied for the Euclidean norm $\|A - B\| = \|A^T - B^T\|$, where $A, B \in R^{m\times p}$.

Theorem 1. *Let* $F + G : R^p \to R^m$ *be continuous on an open convex subset* $D \subset R^p$, F *be a continuously differentiable function, and* G *be a continuous function. Suppose that problem* (1) *has a solution* $x^* \in D$; *the inverse operation*

$$(A_*^T A_*)^{-1} = [(F'(x^*) + G(x^*, x^*))^T (F'(x^*) + G(x^*, x^*))]^{-1} \tag{18}$$

exists, such that $\|(A_*^T A_*)^{-1}\| \leq B$; (7), (8), (11), *and* (12) *hold, and* γ *given in* (10) *exists.*

Furthermore,

$$\|F(x^*) + G(x^*)\| \leq \eta, \qquad \|F'(x^*) + G(x^*, x^*)\| \leq \alpha; \tag{19}$$

$$\frac{B}{R}\Big(\int_0^R L_0(u)du + \int_0^{2R} M_0(u)du\Big)\eta < 1 \tag{20}$$

and $\Omega = \Omega(x^*, r_*) \subseteq D$, *where* r_* *is the unique positive zero of the function* q *given by*

$$\begin{aligned} q(r) \quad = \quad & B\Bigg[\Big(\alpha + \int_0^r L_0(u)du + \int_0^{2r} M_0(u)du\Big)\Big(\int_0^r L(u)udu + \int_0^r M(u)du\Big) \\ & + \Big(2\alpha + \int_0^r L_0(u)du + \int_0^{2r} M_0(u)du\Big)\Big(\int_0^r L_0(u)du + \int_0^{2r} M_0(u)du\Big) \\ & + \Big(\frac{1}{r}\int_0^r L_0(u)du + \frac{1}{r}\int_0^{2r} M_0(u)du\Big)\eta\Bigg] - 1. \end{aligned} \tag{21}$$

Then, for x_0, $x_{-1} \in \Omega$, *the iterative sequence* $\{x_n\}$, $n = 0, 1, \ldots$, *generated by* (2), *is well defined, remains in* Ω, *and converges to* x^*. *Moreover, the following error estimates hold for each* $n = 0, 1, 2, \ldots$:

$$\begin{aligned} \|x_{n+1} - x^*\| \quad \leq \quad & C_1\|x_{n-1} - x^*\| + C_2\|x_n - x^*\| + C_3\|x_{n-1} - x^*\|\|x_n - x^*\| \\ & + C_4\|x_n - x^*\|^2, \end{aligned} \tag{22}$$

where

$$g(r) = \frac{B}{1-\varphi(r)}; \quad C_1 = g(r_*)\frac{1}{2r_*}\int_0^{2r_*} M_0(u)du\,\eta; \tag{23}$$

$$C_2 = g(r_*)\left(\frac{1}{r_*}\int_0^{r_*} L_0(u)du + \frac{1}{2r_*}\int_0^{2r_*} M_0(u)du\right)\eta; \tag{24}$$

$$C_3 = g(r_*)\left(\alpha + \int_0^{r_*} L_0(u)du + \int_0^{2r_*} M_0(u)du\right)\frac{1}{r_*}\int_0^{r_*} M(u)du; \tag{25}$$

$$C_4 = g(r_*)\left(\alpha + \int_0^{r_*} L_0(u)du + \int_0^{2r_*} M_0(u)du\right)\frac{1}{r_*}\int_0^{r_*} L(u)udu. \tag{26}$$

Proof. We obtain

$$\lim_{r\to 0^+}\frac{1}{r}\int_0^r L_0(u)du \le \lim_{r\to 0^+}\frac{L_0(r)r}{r} \le L_0(0), \tag{27}$$

$$\lim_{r\to 0^+}\frac{1}{r}\int_0^{2r} M_0(u)du \le \lim_{r\to 0^+}\frac{M_0(2r)2r}{r} \le 2M_0(0), \tag{28}$$

since L_0 and M_0 are positive and nondecreasing functions on $[0,R]$, and $[0,2R]$, respectively. Taking into account Lemma 1 for a sufficiently small η, $q(0) = B(L_0(0)+2M_0(0))\eta - 1 < 0$. With a sufficiently large R, the inequality $q(R) > 0$ holds. By the intermediate value theorem, the function q has a positive zero on $(0,R)$ denoted by r_*. Moreover, this zero is the only one on $(0,R)$. Indeed, according to Lemma 2, the function $\left(\frac{1}{r}\int_0^r L_0(u)du + \frac{1}{r}\int_0^{2r} M_0(u)du\right)\eta$ is non-decreasing with respect to r on $(0,R]$. By Lemma 1, functions $\int_0^r L(u)du$, $\int_0^r M(u)du$, and $\int_0^{2r} M(u)du$ are monotonically increasing on $[0,R]$. Furthermore, by Lemma 3, the function $\int_0^r L(u)udu = r^2\left(\frac{1}{r^2}\int_0^r L(u)udu\right)$ is monotonically increasing with respect to r on $(0,R]$. Therefore, $q(r)$ is monotonically increasing on $(0,R]$. Thus, the graph of function $q(r)$ crosses the positive r-axis only once on $(0,R)$. Finally, from the monotonicity of q and since $q(\gamma) > 0$, we obtain $r_* < \gamma$, so $\Omega(x^*, r_*) \subset \Omega_0$.

We denote $A_n = F'(x_n) + G(x_n, x_{n-1})$. Let $n = 0$. By the assumption $x_0,\ x_{-1} \in \Omega$, we obtain the following estimation:

$$\begin{aligned}
\left\|I - (A_*^TA_*)^{-1}A_0^TA_0\right\| &= \left\|(A_*^TA_*)^{-1}(A_*^TA_* - A_0^TA_0)\right\| \\
&= \left\|(A_*^TA_*)^{-1}\left[A_*^T(A_* - A_0) + (A_*^T - A_0^T)(A_0 - A_*) + (A_*^T - A_0^T)A_*\right]\right\| \\
&\le \left\|(A_*^TA_*)^{-1}\right\|\left[\|A_*^T\|\|A_* - A_0\| + \|A_*^T - A_0^T\|\|A_0 - A_*\| + \|A_*^T - A_0^T\|\|A_*\|\right] \\
&\le B\left[\alpha\|A_* - A_0\| + \|A_*^T - A_0^T\|\|A_0 - A_*\| + \alpha\|A_*^T - A_0^T\|\right].
\end{aligned} \tag{29}$$

Using conditions (11) and (12), we obtain

$$\begin{aligned}
\|A_0 - A_*\| &= \|(F'(x_0) + G(x_0, x_{-1})) - (F'(x^*) + G(x^*, x^*))\| \\
&= \|F'(x_0) - F'(x_*) + G(x_0, x_{-1}) - G(x^*, x^*)\| \\
&\le \|F'(x_0) - F'(x^*)\| + \|G(x_0, x_{-1}) - G(x^*, x^*)\| \\
&\le \int_0^{\rho_0} L_0(u)du + \int_0^{\rho_0+\rho_{-1}} M_0(u)du,
\end{aligned} \tag{30}$$

where $\rho_k = \rho(x_k)$. Then, from inequality (29) and the equation $q(r) = 0$, we obtain by (10)

$$\|I - (A_*^T A_*)^{-1} A_0^T A_0\| \le B\Big[2\alpha + \int_0^{\rho_0} L_0(u)du + \int_0^{\rho_0+\rho_{-1}} M_0(u)du\Big] \times \Big[\int_0^{\rho_0} L_0(u)du + \int_0^{\rho_0+\rho_{-1}} M_0(u)du\Big] \le B\Big[2\alpha + \int_0^{r_*} L_0(u)du + \int_0^{2r_*} M_0(u)du\Big]\Big[\int_0^{r_*} L_0(u)du + \int_0^{2r_*} M_0(u)du\Big] < 1. \tag{31}$$

Next, from (29)–(31) and the Banach lemma [16], it follows that $(A_0^T A_0)^{-1}$ exists, and

$$\begin{aligned}\left\|(A_0^T A_0)^{-1}\right\| &\le g_0 = B\Big\{1 - B\Big[2\alpha + \int_0^{\rho_0} L_0(u)du + \int_0^{\rho_0+\rho_{-1}} M_0(u)du\Big] \times \Big[\int_0^{\rho_0} L_0(u)du + \int_0^{\rho_0+\rho_{-1}} M_0(u)du\Big]\Big\}^{-1} \\ &\le g(r_*) = B\Big\{1 - B\Big[2\alpha + \int_0^{r_*} L_0(u)du + \int_0^{2r_*} M_0(u)du\Big] \times \Big[\int_0^{r_*} L_0(u)du + \int_0^{2r_*} M_0(u)du\Big]\Big\}^{-1}.\end{aligned} \tag{32}$$

Hence, x_1 is correctly defined. Next, we will show that $x_1 \in \Omega(x^*, r_*)$.
Using the fact

$$A_*^T(F(x^*) + G(x^*)) = (F'(x^*) + G(x^*, x^*))^T(F(x^*) + G(x^*)) = 0, \tag{33}$$

$x_0, x_{-1} \in \Omega(x^*, r_*)$ and the choice of r_*, we obtain the estimate

$$\begin{aligned}\|x_1 - x^*\| &= \left\|x_0 - x^* - (A_0^T A_0)^{-1}\left[A_0^T(F(x_0) + G(x_0)) - A_*^T(F(x^*) + G(x^*))\right]\right\| \\ &\le \left\|-(A_0^T A_0)^{-1}\right\| \Big\| - A_0^T\Big[A_0 - \int_0^1 F'(x_* + t(x_0 - x^*))dt \\ &\quad -G(x_0, x^*)\Big](x_0 - x^*) + (A_0^T - A_*^T)(F(x^*) + G(x^*))\Big\|.\end{aligned} \tag{34}$$

So, considering the inequalities

$$\begin{aligned}&\left\|A_0 - \int_0^1 F'(x^* + t(x_0 - x^*))dt - G(x_0, x^*)\right\| \\ &= \left\|F'(x_0) - \int_0^1 F'(x^* + t(x_0 - x^*))dt + G(x_0, x_{-1}) - G(x_0, x^*)\right\| \\ &= \Big\| \int_0^1 \left[F'(x_0) - F'(x^* + t(x_0 - x^*))\right]dt + G(x_0, x_{-1}) - G(x_0, x^*)\Big\| \\ &= \Big\| \int_0^1 \left[F'(x_0) - F'(x_0^t)\right]dt + G(x_0, x_{-1}) - G(x_0, x^*)\Big\| \\ &\le \int_0^1 \int_{t\rho_0}^{\rho_0} L(u)dudt + \int_0^{\rho_{-1}} M(u)du = \int_0^{\rho_0} L(u)udu + \int_0^{\rho_{-1}} M(u)du \\ &\le \frac{1}{r_*^2}\int_0^{r_*} L(u)udu\,\rho_0^2 + \frac{1}{r_*}\int_0^{r_*} M(u)du\,\rho_{-1},\end{aligned} \tag{35}$$

$$\|A_0\| \le \|A_*\| + \|A_0 - A_*\| \le \alpha + \int_0^{\rho_0} L_0(u)du + \int_0^{\rho_0+\rho_{-1}} M_0(u)du, \tag{36}$$

we obtain

$$\begin{aligned}\|x_1 - x^*\| &\leq g_0\Big\{\Big[\alpha + \int_0^{\rho_0} L_0(u)du + \int_0^{\rho_0+\rho_{-1}} M_0(u)du\Big] \\ &\times\Big[\int_0^{\rho_0} L(u)udu + \int_0^{\rho_{-1}} M(u)du\Big]\|x_0 - x^*\| + \eta\Big[\int_0^{\rho_0} L_0(u)du \\ &+ \int_0^{\rho_0+\rho_{-1}} M_0(u)du\Big]\Big\} \leq g_0\Big\{\Big[\alpha + \int_0^{r_*} L_0(u)du + \int_0^{2r_*} M_0(u)du\Big] \\ &\times\Big[\frac{1}{r_*^2}\int_0^{r_*} L(u)udu\rho_0^2 + \frac{1}{r_*}\int_0^{r_*} M(u)du\,\rho_{-1}\Big]\|x_0 - x^*\| \\ &+\eta\Big[\frac{1}{r_*}\int_0^{r_*} L_0(u)du\rho_0 + \frac{1}{2r_*}\int_0^{2r_*} M_0(u)du(\rho_0+\rho_{-1})\Big]\Big\} \qquad (37) \\ &< g(r_*)\Big\{\Big[\alpha + \int_0^{r_*} L_0(u)du + \int_0^{2r_*} M_0(u)du\Big]\Big[\int_0^{r_*} L(u)udu + \int_0^{r_*} M(u)du\Big] \\ &+\frac{1}{r_*}\Big[\int_0^{r_*} L_0(u)du + \int_0^{2r_*} M_0(u)du\Big]\eta\Big\}r_* = p(r_*)r_* = r_*,\end{aligned}$$

where

$$\begin{aligned}p(r) &= g(r)\Big\{\Big[\alpha + \int_0^{r} L_0(u)du + \int_0^{2r} M_0(u)du\Big]\Big[\int_0^{r} L(u)udu + \int_0^{r} M(u)du\Big] \\ &+\frac{1}{r}\Big[\int_0^{r} L_0(u)du + \int_0^{2r} M_0(u)du\Big]\eta\Big\}. \qquad (38)\end{aligned}$$

Therefore, $x_1 \in \Omega(x^*, r_*)$, and estimate (22) holds for $n = 0$.

Let us assume that $x_n \in \Omega(x^*, r_*)$ for $n = 0, 1, ..., k$ and estimate (22) holds for $n = 0, 1, ..., k-1$, where $k \geq 1$ is an integer. We shall show $x_{n+1} \in \Omega$ and that the estimate (22) holds for $n = k$.

We can write

$$\begin{aligned}&\|I - (A_*^TA_*^T)^{-1}A_k^TA_k\| = \|(A_*^TA_*)^{-1}(A_*^TA_* - A_k^TA_k)\| \\ &= \|(A_*^TA_*)^{-1}(A_*^T(A_* - A_k) + (A_*^T - A_k^T)(A_k - A_*) + (A_*^T - A_k^T)A_*)\| \\ &\leq B\Big(\alpha\|A_* - A_k\| + \|A_*^T - A_k^T\|\|A_k - A_*\| + \alpha\|A_*^T - A_k^T\|\Big) \\ &\leq B\Big[2\alpha + \int_0^{\rho_k} L_0(u)du + \int_0^{\rho_k+\rho_{k-1}} M_0(u)du\Big]\Big[\int_0^{\rho_k} L_0(u)du \qquad (39) \\ &+\int_0^{\rho_k+\rho_{k-1}} M_0(u)du\Big] \leq B\Big[2\alpha + \int_0^{r_*} L_0(u)du + \int_0^{2r_*} M_0(u)du\Big] \\ &\times\Big[\int_0^{r_*} L_0(u)du + \int_0^{2r_*} M_0(u)du\Big] < 1.\end{aligned}$$

Consequently, $\left(A_k^TA_k\right)^{-1}$ exists, and

$$\begin{aligned}\|(A_{k+1}^TA_{k+1})^{-1}\| &\leq g_k = B\Big\{1 - B\Big[2\alpha + \int_0^{\rho_k} L_0(u)du + \int_0^{\rho_k+\rho_{k-1}} M_0(u)du\Big] \\ &\times\Big[\int_0^{\rho_k} L_0(u)du + \int_0^{\rho_k+\rho_{k-1}} M_0(u)du\Big]\Big\}^{-1} \leq g(r_*). \qquad (40)\end{aligned}$$

Therefore, x_{k+1} is correctly defined, and the following estimate holds:

$$\begin{aligned}
\|x_{k+1}-x_*\| &= \|x_k-x^*-(A_k^TA_k)^{-1}[A_k^T(F(x_k)+G(x_k))-A_*^T(F(x^*)\\
&+G(x^*))]\| \le \|-(A_k^TA_k)^{-1}\|\,\Big\|-A_k^T\Big[A_k-\int_0^1 F'(x^*+t(x_k-x^*))dt\\
&-G(x_k,x_*)\Big](x_k-x^*)+(A_k^T-A_*^T)(F(x^*)+G(x^*))\Big\|\\
&\le \|-(A_k^TA_k)^{-1}\|\,\Big\|-A_k^T\Big[A_k-\int_0^1 F'(x^*+t(x_k-x^*))dt\\
&-G(x_k,x_*)\Big](x_k-x^*)+(A_k^T-A_*^T)(F(x^*)+G(x^*))\Big\| \qquad (41)\\
&\le g_k\Big\{\Big[\alpha+\int_0^{\rho_k}L_0(u)du+\int_0^{\rho_k+\rho_{k-1}}M_0(u)du\Big]\Big[\int_0^{\rho_k}L(u)udu\\
&+\int_0^{\rho_{k-1}}M(u)du\Big]\|x_k-x^*\|+\eta\Big[\int_0^{\rho_k}L_0(u)du+\int_0^{\rho_k+\rho_{k-1}}M_0(u)du\Big]\Big\}\\
&\le g(r_*)\Big\{\Big[\alpha+\int_0^{r_*}L_0(u)du+\int_0^{2r_*}M_0(u)du\Big]\\
&\times\Big[\frac{1}{r_*^2}\int_0^{r_*}L(u)udu\rho_k^2+\frac{1}{r_*}\int_0^{r_*}M(u)du\rho_{k-1}\Big]\|x_k-x^*\|\\
&+\eta\Big[\frac{1}{r_*}\int_0^{r_*}L_0(u)du\rho_k+\frac{1}{2r_*}\int_0^{2r_*}M_0(u)du(\rho_k+\rho_{k-1})\Big]\Big\}<p(r_*)r_*=r_*.
\end{aligned}$$

This proves that $x_{k+1}\in\Omega(x^*,r_*)$ and estimate (22) for $n=k$.

Thus, by the induction method, (2) is correctly defined, $x_n\in\Omega(x^*,r_*)$, and estimate (22) holds for each $n=0,1,2,\ldots$.

It remains to be proven that $x_n\to x^*$ for $n\to\infty$.

Let us define functions a and b on $[0,\,r_*]$ as

$$\begin{aligned}
a(r) &= g(r)\Big\{\Big[\alpha+\int_0^r L_0(u)du+\int_0^{2r}M_0(u)du\Big]\Big[\int_0^r L(u)udu+\int_0^r M(u)du\Big]\\
&+\Big[\frac{1}{r}\int_0^r L_0(u)du+\frac{1}{2r}\int_0^{2r}M_0(u)du\Big]\eta\Big\}; \qquad (42)\\
b(r) &= g(r)\frac{1}{2r}\int_0^{2r}M(u)du\,\eta. \qquad (43)
\end{aligned}$$

According to the choice of r_*, we obtain

$$a(r_*)\ge 0,\quad b(r_*)\ge 0,\quad a(r_*)+\,b(r_*)=1. \qquad (44)$$

Using estimate (22), the definition of functions a, b and constants C_i $(i=1,2,3,4)$, we have

$$\begin{aligned}
\|x_{n+1}-x^*\| &\le C_1\|x_{n-1}-x^*\|+(C_2+C_3r_*+C_4r_*)\|x_n-x^*\|\\
&= a(r_*)\|x_n-x^*\|+b(r_*)\|x_{n-1}-x^*\|. \qquad (45)
\end{aligned}$$

According to the proof in [17], under the conditions (42)–(45), the sequence $\{x_n\}$ converges to x^* for $n\to\infty$. □

Corollary 1 ([4]). *The convergence order of method* (2) *for the problem* (1) *with zero residual is equal to* $\dfrac{1+\sqrt{5}}{2}$.

If $\eta = 0$, we have the nonlinear least squares problem with zero residual. Then, the constants $C_1 = 0$ and $C_2 = 0$, and estimate (22) takes the form

$$\|x_{n+1} - x^*\| \leq C_3\|x_{n-1} - x^*\| \, \|x_n - x^*\| + C_4\|x_n - x^*\|^2. \tag{46}$$

This inequality can be written as

$$\|x_{n+1} - x^*\| \leq (C_3 + C_4)\|x_{n-1} - x^*\| \, \|x_n - x^*\|. \tag{47}$$

Then, we can write an equation for determining the convergence order as follows:

$$t^2 - t - 1 = 0. \tag{48}$$

Therefore, the positive root, $t^* = \dfrac{1+\sqrt{5}}{2}$ of the latter equation is the order of convergence of method (2).

In case $G(x) \equiv 0$ in (1), we obtain the following consequences.

Corollary 2 ([4]). *The convergence order of method* (2) *for problem* (1) *with zero residual is quadratic.*

Indeed, if $G(x) \equiv 0$, then $C_3 = 0$, and estimate (22) takes the form

$$\|x_{n+1} - x^*\| \leq C_4\|x_n - x^*\|^2, \tag{49}$$

which indicates the quadratic convergence rate of method (2).

Remark 2. *If $L_0 = L = L_1$ and $M_0 = M = M_1$, our results specialize to the corresponding ones in [4]. Otherwise, they constitute an improvement as already noted in Remark 1. As an example, let $q_1, g_1, C_1^1, C_2^1, C_3^1, C_4^1, r_*^1$ denote the functions and parameters where L_0, L, M_0, M are replaced by L_1, L_1, M_1, M_1, respectively. Then, we have in view of* (15)–(17) *that*

$$q(r) \leq q_1(r), \tag{50}$$
$$g(r) \leq g_1(r), \tag{51}$$
$$C_1 \leq C_1^1, \tag{52}$$
$$C_2 \leq C_2^1, \tag{53}$$
$$C_3 \leq C_3^1, \tag{54}$$

and

$$C_4 \leq C_4^1. \tag{55}$$

Hence, we have

$$r_*^1 \leq r_*, \tag{56}$$

the new error bounds (22) *being tighter than the corresponding (6) in [4], and the rest of the advantages (already mentioned in Remark 1) holding true.*

Next, we study the convergence of method (2) if L_0, L, M_0, M are constants, as a consequence of Theorem 1.

Corollary 3. *Let $F + G : R^p \to R^m$ be continuous on an open convex subset $D \subset R^p$, F be a continuously differentiable, and G be a continuous function on D. Suppose that problem* (1) *has a solution $x^* \in D$, and the inverse operation*

$$(A_*^T A_*)^{-1} = [(F'(x^*) + G(x^*, x^*))^T (F'(x^*) + G(x^*, x^*))]^{-1} \tag{57}$$

exists, such that $\|(A_*^T A_*)^{-1}\| \leq B$.

Suppose that the Fréchet derivative F' *satisfies the classic Lipschitz conditions*

$$\|F'(x) - F'(x^*)\| \leq L_0\|x - x^*\|, \text{ for each } x \in D, \tag{58}$$

$$\|F'(x) - F'(y)\| \leq L\|x - y\|, \text{ for each } x, y \in \Omega_0 \tag{59}$$

and the function G *has a first order divided difference* $G(x, y)$ *that satisfies*

$$\|G(x, y) - G(x^*, x^*)\| \leq M_0(\|x - x^*\| + \|y - x^*\|), \text{ for each } x, y \in D, \tag{60}$$

$$\|G(x, y) - G(u, v)\| \leq M(\|x - u\| + \|y - v\|), \text{ for each } x, y, u, v \in \Omega_0, \tag{61}$$

where $\Omega_0 = D \cap \Omega\left(x^*, \dfrac{\sqrt{B^2\alpha^2 + B} - B\alpha}{B(L_0 + 2M_0)}\right)$.

Furthermore,

$$\|F(x^*) + G(x^*)\| \leq \eta,\ \|F'(x^*) + G(x^*, x^*)\| \leq \alpha,\ B(L_0 + 2M_0)\eta < 1 \tag{62}$$

and $\Omega = \Omega(x^*, r_*) \subseteq D$, *where*

$$r_* = \frac{4(1 - BT_0\eta)}{B\alpha(4T_0 + T) + \sqrt{B^2\alpha^2(4T_0 + T)^2 + 8BT_0(2T_0 + T)(1 - BT_0\eta)}}, \tag{63}$$

$T_0 = L_0 + 2M_0$, $T = L + 2M$. *Then, for each* x_0, $x_{-1} \in \Omega$, *the iterative sequence* $\{x_n\}$, $n = 0, 1, ...$, *generated by* (2) *is well defined, remains in* Ω, *and converges to* x^*, *such that the following error estimate holds for each* $n = 0, 1, 2, \ldots$:

$$\begin{aligned} \|x_{n+1} - x^*\| \leq\ & C_1\|x_{n-1} - x^*\| + C_2\|x_n - x^*\| \\ & + C_3\|x_{n-1} - x^*\|\|x_n - x^*\| + C_4\|x_n - x^*\|^2, \end{aligned} \tag{64}$$

where

$$g(r) = B[1 - B(2\alpha + (L_0 + 2M_0)r)(L_0 + 2M_0)r]^{-1}; \tag{65}$$

$$C_1 = g(r_*)M_0\eta;\ C_2 = g(r_*)(L_0 + M_0)\eta; \tag{66}$$

$$C_3 = g(r_*)(\alpha + (L_0 + 2M_0)r_*)M; \tag{67}$$

$$C_4 = g(r_*)(\alpha + (L_0 + 2M_0)r_*)\frac{L}{2}. \tag{68}$$

The proof of Corollary 3 is analogous to the proof of Theorem 1.

3. Numerical Examples

In this section, we give examples to show the applicability of method (2) and to confirm Remark 2. We use the norm $\|x\| = \sqrt{\sum_{i=1}^{p} x_i^2}$ for $x \in \mathrm{R}^{\mathrm{p}}$.

Example 1. *Let function* $F + G : R^2 \to R^3$ *be defined by*

$$F(x) + G(x) = \begin{pmatrix} 3u^2v + v^2 - 1 + |u^2 - 1| \\ u^4 + uv^3 - 1 + |v| \\ v - 0.3 + |u - 1| \end{pmatrix}, \tag{69}$$

$$F(x) = \begin{pmatrix} 3u^2v + v^2 - 1 \\ u^4 + uv^3 - 1 \\ v - 0.3 \end{pmatrix},\ G(x) = \begin{pmatrix} |u^2 - 1| \\ |v| \\ |u - 1| \end{pmatrix}, \tag{70}$$

where $x = (u, v)$. *The solution of this problem* $x^* \approx (0.917889,\ 0.288314)$ *and* $\eta \approx 0.079411$.

Let us give the number of iterations needed to obtain an approximate solution of this problem. We test method (2) for the different initial points $x_0 = \delta(1.1,\ 0.5)^T$, where $\delta \in R$, and use the stopping criterion $\|x_{n+1} - x_n\| \leq \varepsilon$. The additional point $x_{-1} = x_0 + 10^{-4}$. The numerical results are shown in Table 1.

Table 1. Results for Example 1, $\varepsilon = 10^{-8}$.

	$\delta = 0.1$	$\delta = 1$	$\delta = 5$	$\delta = 10$	$\delta = 100$
Number of iterations	12	8	15	17	25

In Table 2, we give values of x_{n+1}, $\|x_{n+1} - x_n\|$ and the norm of residual at each iteration.

Table 2. Iterative sequence, norm of growth, and residual for Example 1, $x_0 = (0.8,\ 0.2)^T$, $\varepsilon = 10^{-6}$.

n	x_{n+1}	$\|x_{n+1} - x_n\|$	$\|F(x_{n+1}) + G(x_{n+1})\|$
0	(0.937901, 0.312602)	0.178033	0.143759
1	(0.918455, 0.290216)	2.965298×10^{-2}	7.973496×10^{-2}
2	(0.917850, 0.288333)	1.977741×10^{-3}	7.941104×10^{-2}
3	(0.917888, 0.288313)	4.346993×10^{-5}	7.941092×10^{-2}
4	(0.917889, 0.288314)	7.873833×10^{-7}	7.941092×10^{-2}

Example 2. *Let function* $F + G : D \subseteq \mathrm{R} \to \mathrm{R}^3$ *be defined by [5]:*

$$F(x) + G(x) = \begin{pmatrix} x + \mu \\ \lambda x^3 + x - \mu \\ \lambda |x^2 - 1| - \lambda \end{pmatrix}, \tag{71}$$

$$F(x) = \begin{pmatrix} x + \mu \\ \lambda x^3 + x - \mu \\ 0 \end{pmatrix}, \ G(x) = \begin{pmatrix} 0 \\ 0 \\ \lambda |x^2 - 1| - \lambda \end{pmatrix}, \tag{72}$$

where $\lambda, \mu \in \mathrm{R}$ *are two parameters. Here* $x^* = 0$ *and* $\eta = \sqrt{2}|\mu|$. *Thus, if* $\mu = 0$, *then we have a problem with zero residual.*

Let us consider Example 2 and show that $r_*^1 \leq r_*$ and the new error estimates (64) are tighter than the corresponding ones in [4]. We consider the case of the classical Lipschitz conditions (Corollary 3). Error estimates from [4] are as follows:

$$\begin{aligned} \|x_{n+1} - x^*\| \ \leq \ & C_1^1 \|x_{n-1} - x^*\| + C_2^1 \|x_n - x^*\| \\ & + C_3^1 \|x_{n-1} - x^*\| \|x_n - x^*\| + C_4^1 \|x_n - x^*\|^2, \end{aligned} \tag{73}$$

where

$$g^1(r) = B[1 - B(2\alpha + (L_1 + 2M_1)r)(L_1 + 2M_1)r]^{-1}; \tag{74}$$

$$C_1^1 = g^1(r_*^1) M_1 \eta;\ C_2^1 = g^1(r_*^1)(L_1 + M_1)\eta; \tag{75}$$

$$C_3^1 = g^1(r_*^1)(\alpha + (L_1 + 2M_1) r_*^1) M_1; \tag{76}$$

$$C_4^1 = g^1(r_*^1)(\alpha + (L_1 + 2M_1) r_*^1) \frac{L_1}{2}. \tag{77}$$

They can be obtained from (64) by replacing r_*, L_0, L, M_0, M in $g(r)$, C_1, C_2, C_3, C_4 by $r_*^1, L_1, L_1, M_1, M_1$, respectively. Similarly,

$$r_*^1 = \frac{4(1 - BT_1\eta)}{5B\alpha T_1 + \sqrt{25B^2\alpha^2T_1^2 + 24BT_1^2(1 - BT_1\eta)}}, \; T_1 = L_1 + 2M_1. \tag{78}$$

Let us choose $D = (-0.5;\ 0.5)$. Thus, we have $B = 0.5$, $\eta = \sqrt{2}|\mu|$, $\alpha = \sqrt{2}$, $L_0 = \max_{x\in D} 3|\lambda||x|$, $L = \max_{x,y\in\Omega_0} 3|\lambda||x+y|$, $L_1 = \max_{x,y\in D} 3|\lambda||x+y|$, $M_0 = M = M_1 = |\lambda|$. Radii are written in Table 3.

Table 3. Radii of convergence domains.

λ	μ	L_0	L	L_1	M	r_*	r_*^1
0.4	0	0.6	1.004205	1.2	0.4	0.319259	0.235702
0.1	0.2	0.15	0.3	0.3	0.1	1.192633	0.885163

Tables 4 and 5 report the left and right side of error estimates (64) and (73). We obtained these results for $\varepsilon = 10^{-8}$ and starting approximations $x_{-1} = 0.2001$, $x_0 = 0.2$. We see that the new error bounds (64) are tighter than the corresponding (73) from [4].

Table 4. Results for $\lambda = 0.4$, $\mu = 0$.

n	$\lvert x_{n+1} - x^*\rvert$	The Right Side of (64)	The Right Side of (73)
0	4.364164×10^{-3}	0.125318	0.169740
1	1.425535×10^{-5}	1.245455×10^{-3}	1.529729×10^{-3}
2	2.179258×10^{-11}	8.675961×10^{-8}	1.060957×10^{-7}
3	3.542853×10^{-22}	4.314684×10^{-16}	5.272102×10^{-16}

Table 5. Results for $\lambda = 0.1$, $\mu = 0.2$.

n	$\lvert x_{n+1} - x^*\rvert$	The Right Side of (64)	The Right Side of (73)
0	2.063103×10^{-3}	5.909333×10^{-2}	8.484100×10^{-2}
1	5.453349×10^{-7}	9.113893×10^{-3}	1.080560×10^{-2}
2	2.054057×10^{-14}	9.051468×10^{-5}	1.057648×10^{-4}
3	1.447579×10^{-18}	2.390964×10^{-8}	2.792694×10^{-8}

4. Conclusions

We developed an improved local convergence analysis of the Gauss–Newton–Secant method for solving nonlinear least squares problems with nondifferentiable operator. We use a center and restricted radius Lipschitz conditions to study the method. As a consequence, we obtain a larger radius of convergence and tighter error estimates under the same computational effort as in earlier papers. This idea can be used to extend the usage of other methods with inverses, such as Newton-type, Secant-type, single-step, or multi-step, to mention a few. This should be our future work. Finally, it is worth mentioning that except for the methods used in this paper, some of the most representative computational intelligence algorithms can be used to solve the problems, such as monarch butterfly optimization (MBO) [18], the earthworm optimization algorithm (EWA) [19], elephant herding optimization (EHO) [20], the moth search (MS) algorithm [21], the slime mould algorithm (SMA), and Harris hawks optimization (HHO) [22].

Author Contributions: Editing, I.K.A.; Conceptualization S.S.; Investigation I.K.A., S.S., R.I., H.Y. and M.I.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Li, C.; Zhang, W.; Jin, X. Convergence and uniqueness properties of Gauss-Newton's method. *Comput. Math. Appl.* **2004**, *47*, 1057–1067. [CrossRef]
2. Argyros, I.K.; Ren, H. A derivative free iterative method for solving least squares problems. *Numer. Algorithms* **2011**, *58*, 555–571.
3. Shakhno, S.M.; Gnatyshyn, O.P. On an iterative algorithm of order 1.839... for solving the nonlinear least squares problems. *Appl. Math. Comput.* **2005**, *161*, 253–264. [CrossRef]
4. Shakhno, S.M.; Iakymchuk, R.P.; Yarmola, H.P. An iterative method for solving nonlinear least squares problems with nondifferentiable operator. *Mat. Stud.* **2017**, *48*, 97–107. [CrossRef]
5. Shakhno, S.M.; Iakymchuk, R.P.; Yarmola, H.P. Convergence analysis of a two-step method for the nonlinear least squares problem with decomposition of operator. *J. Numer. Appl. Math.* **2018**, *128*, 82–95.
6. Shakhno, S.; Shunkin, Yu. One combined method for solving nonlinear least squares problems. *Visnyk Lviv Univ. Ser. Appl. Math. Comp. Sci.* **2017**, *25*, 38–48. (In Ukrainian)
7. Ulm, S. On generalized divided differences. *Izv. ESSR Ser. Phys. Math.* **1967**, *16*, 13–26. (In Russian)
8. Cătinaş, E. On some iterative methods for solving nonlinear equations. *Rev. Anal. Numér. Théor. Approx.* **1994**, *23*, 47–53.
9. Shakhno, S.M.; Mel'nyk, I.V.; Yarmola, H.P. Convergence analysis of combined method for solving nonlinear equations. *J. Math. Sci.* **2016**, *212*, 16–26. [CrossRef]
10. Shakhno, S.M. Convergence of combined Newton-Secant method and uniqueness of the solution of nonlinear equations. *Sci. J. Tntu* **2013**, *1*, 243–252. (In Ukrainian)
11. Zabrejko, P.P.; Nguen, D.F. The majorant method in the theory of Newton-Kantorovich approximations and the Pták error estimates. *Numer. Funct. Anal. Optim.* **1987**, *9*, 671–686. [CrossRef]
12. Wang, X.; Li, C. Convergence of Newton's method and uniqueness of the solution of equations in Banach space II. *Acta Math. Sin.* **2003**, *19*, 405–412. [CrossRef]
13. Wang, X. Convergence of Newton's method and uniqueness of the solution of equations in Banach space. *IMA J. Numer. Anal.* **2000**, *20*, 123–134. [CrossRef]
14. Argyros, I.K.; Hilout, S. On an improved convergence analysis of Newton's method. *Appl. Math. Comput.* **2013**, *225*, 372–386. [CrossRef]
15. Argyros, I.K.; Magreñán, A.A. *Iterative Methods and Their Dynamics with Applications: A Contemporary Study*; CRC Press: Boca Raton, FL, USA, 2017.
16. Dennis, J.E.; Schnabel, R.B. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*; SIAM: Philadelphia, PA, USA, 1996.
17. Ren, H.; Argyros, I.K. Local convergence of a secant type method for solving least squares problems. *Appl. Math. Comput.* **2010**, *217*, 3816–3824. [CrossRef]
18. Wang, G.G.; Deb, S.; Cui, Z. Monarch butterfly optimization. *Neural Comput. Appl.* **2019**, *31*, 1995–2014. [CrossRef]
19. Wang, G.G.; Deb, S.; Dos, L.; Coelho, L.D.S. Earthworm optimization algorithm: A bio-inspired metaheuristic algorithm for global optimization problems. *Int. J. Bio-Inspired Comput.* **2018**, *12*, 1–22. [CrossRef]
20. Wang, G.G.; Deb, S.; Coelho, L.D.S. Elephant Herding Optimization. In Proceedings of the 3rd International Symposium on Computational and Business Intelligence (ISCBI 2015), Bali, Indonesia, 7–9 December 2015; pp. 1–5.
21. Mirjalili, S. Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm. *Knowl.-Based Syst.* **2015**, *89*, 228–249 [CrossRef]
22. Zhao, J.; Gao, Z.-M. The hybridized Harris hawk optimization and slime mould algorithm. *J. Phys. Conf. Ser.* **2020**, *1682*, 012029. [CrossRef]