

Article

Improved Convergence Analysis of Gauss-Newton-Secant Method for Solving Nonlinear Least Squares Problems

Ioannis Argyros¹, Stepan Shakhno^{2,*}  and Yuriy Shunkin²¹ Department of Mathematical Sciences, Cameron University, Lawton, OK 73505, USA; iargyros@cameron.edu² Department of Theory of Optimal Processes, Ivan Franko National University of Lviv, 79000 Lviv, Ukraine; yuriy.shunkin@lnu.edu.ua

* Correspondence: stepan.shakhno@lnu.edu.ua

Received: 20 October 2018; Accepted: 15 January 2019; Published: 18 January 2019



Abstract: We study an iterative differential-difference method for solving nonlinear least squares problems, which uses, instead of the Jacobian, the sum of derivative of differentiable parts of operator and divided difference of nondifferentiable parts. Moreover, we introduce a method that uses the derivative of differentiable parts instead of the Jacobian. Results that establish the conditions of convergence, radius and the convergence order of the proposed methods in earlier work are presented. The numerical examples illustrate the theoretical results.

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

MSC: 65F20; 65G99; 65H10; 49M15

1. Introduction

Nonlinear least squares problems often arise while solving overdetermined systems of nonlinear equations, parameter estimation of physical processes by measurement results, constructing nonlinear regression models for solving engineering problems, etc.

The nonlinear least squares problem has the form

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

where the residual function $F : R^p \rightarrow R^m$ ($m \geq p$) is nonlinear in x ; F is a continuously differentiable function. Effective methods for solving nonlinear least squares problems is the Gauss-Newton method [1–3]

$$x_{n+1} = x_n - [F'(x_n)^T F'(x_n)]^{-1} F'(x_n)^T F(x_n), \quad n = 0, 1, \dots \quad (2)$$

However, in practice, there are often problems with the calculation of derivatives. Hence, one can use the iterative-difference methods. These methods do not require calculation of derivatives. Moreover, they do not perform worse than Gauss-Newton method in terms of the convergence rate and the number of iterations. In some cases, nonlinear functions consist of differentiable and nondifferentiable parts. However, it is possible to use iterative-difference methods [4–7]

$$x_{n+1} = x_n - (A_n^T A_n)^{-1} A_n^T F(x_n), \quad n = 0, 1, \dots, \quad (3)$$

where

$$A_n = F(x_n, x_{n-1}),$$

$$A_n = F(2x_n - x_{n-1}, x_{n-1}),$$

or

$$A_n = F(x_n, x_{n-1}) + F(x_{n-2}, x_n) - F(x_{n-2}, x_{n-1}).$$

It is desirable to build iterative methods that take into account properties of the problem. In particular, we can use only derivative of differentiable part of operator instead of full Jacobian, which in fact, does not exist. The methods obtained using this approach converge slowly. More efficient methods use sum of the derivatives of the differentiable part and divided difference of the nondifferentiable part of the operator instead of the Jacobian. Such an approach shows great results in the case of solving nonlinear equations.

In this work we study a combined method for solving nonlinear least squares problem, based on the Gauss-Newton, secant methods. We also use a method, requiring only derivative from the differentiable part of operator. We prove the local convergence and show efficiency on test cases when comparing with secant type methods [5,6]. The convergence region of iterative methods is small in general. This fact limits the number of initial approximations. It is therefore important to extend this region without requiring additional hypotheses. The new approach [8] leads to larger convergence radius than before [9]. We achieve this goal by locating an at least as small region as before containing the iterates. Then, the new Lipschitz constants are at least as tight as the old Lipschitz constants. Moreover, using more precise estimates on the distances involved, under weaker hypotheses, and under the same computational cost, we provide an analysis of the Gauss-Newton-Secant method with the following advantages over the corresponding results in [9]: larger convergence region; finer error estimates on the distances involved, and an at least as precise information on the location of the solution.

The rest of the paper is given as follows. Section 2 contains the statement of the problem, in Sections 3 and 4, we present the local convergence analysis of the first and second method, respectively. In Section 5, we provide the numerical examples. The article ends with some conclusions.

2. Description of the Problem

Consider the nonlinear least squares problem

$$\min_{x \in \mathbb{R}^p} \frac{1}{2} (F(x) + G(x))^T (F(x) + G(x)), \quad (4)$$

where residual function $F + G : \mathbb{R}^p \rightarrow \mathbb{R}^m$ ($m \geq p$) is nonlinear in x ; F is continuously differentiable function; G is continuous function, differentiability of which, in general, is not required.

We propose a modification of the Gauss-Newton method to find a solution of problem (4):

$$\begin{aligned} x_{n+1} &= x_n - (A_n^T A_n)^{-1} A_n^T (F(x_n) + G(x_n)), \\ A_n &= F'(x_n) + G(x_n, x_{n-1}), n = 0, 1, \dots \end{aligned} \quad (5)$$

Here, $F'(x_n)$ is Fréchet derivative by $F(x)$; $G(x_n, x_{n-1})$ is a divided difference of order one for function $G(x)$ [10], where vectors x_n, x_{n-1} and x_0, x_{-1} are given initial approximations, satisfying $G(x, y)(x - y) = G(x) - G(y)$ for $x \neq y$ and $G(x, x) = G'(x)$, if G is differentiable. Setting $A_n = F'(x_n)$, from method (5) we get Gauss-Newton type iterative method for solving problem (4)

$$x_{n+1} = x_n - (F'(x_n)^T F'(x_n))^{-1} F'(x_n)^T (F(x_n) + G(x_n)), n = 0, 1, \dots \quad (6)$$

In case of $m = p$, problem (4) turns into a system of nonlinear equations

$$F(x) + G(x) = 0. \quad (7)$$

Then, it is well known ([3], p. 267) that techniques for minimizing problem (4) are techniques for finding a solution x^* of Equation (7). In this case (5) transforms into the Newton-Secant combined method [11,12]

$$x_{n+1} = x_n - (F'(x_n) + G(x_n, x_{n-1}))^{-1}(F(x_n) + G(x_n)), n = 0, 1, \dots, \quad (8)$$

and method (6) into Newton's-type method for solving nonlinear Equation (7) [13]

$$x_{n+1} = x_n - (F'(x_n))^{-1}(F(x_n) + G(x_n)), n = 0, 1, \dots$$

We assume from now on that function G is differentiable at $x = x^*$.

3. Local Convergence Analysis (5)

Sufficient conditions and the convergence order of the iterative process (5) are presented. However first, we need some crucial definitions. They are needed to provide a clear relationship between the Lipschitz constants appearing in the local convergence analysis and the relationship between them.

Definition 1. The Fréchet derivative F' satisfies the center-Lipschitz condition on D , if there exists $L_0 > 0$ such that for each $x \in D$

$$\|F'(x) - F'(x^*)\| \leq L_0 \|x - x^*\|. \quad (9)$$

Definition 2. The divided difference $G(x, y)$ satisfies the center-Lipschitz condition $D \times D$, if there exists $M_0 > 0$ such that for each $x, y \in D$

$$\|G(x, y) - G(x^*, x^*)\| \leq M_0 (\|x - x^*\| + \|y - x^*\|). \quad (10)$$

Let $B > 0$ and $\alpha > 0$. Define function $\varphi : [0, +\infty) \rightarrow [0, +\infty)$ by

$$\varphi(r) = B[2\alpha + (L_0 + 2M_0)r](L_0 + 2M_0)r. \quad (11)$$

Let $U(x^*, r_*) = \{x : \|x - x^*\| \leq r_*, r_* > 0\}$. Suppose that equation $\varphi(r) = 1$ has at least one positive solution. Denote by γ the smallest such solution. Define

$$D_0 = D \cap U(x^*, \gamma). \quad (12)$$

Definition 3. The Fréchet derivative F' satisfies the restricted Lipschitz condition on D_0 , if there exists $L > 0$ such that for each $x, y \in D_0$

$$\|F'(x) - F'(y)\| \leq L \|x - y\|. \quad (13)$$

Definition 4. The first order divided difference $G(x, y)$ satisfies the restricted Lipschitz condition on $D_0 \times D_0$, if there exists $M > 0$ such that for each $x, y, u \in D_0$

$$\|G(x, y) - G(u, x^*)\| \leq M (\|x - u\| + \|y - x^*\|). \quad (14)$$

Next, we also state the definitions given in [9], so we can compare them to preceding ones.

Definition 5. The Fréchet derivative F' satisfies the Lipschitz condition on D , if there exists $L_1 > 0$ such that for each $x, y \in D$

$$\|F'(x) - F'(y)\| \leq L_1 \|x - y\|. \quad (15)$$

Definition 6. The first order divided difference $G(x, y)$ satisfies the Lipschitz condition on $D \times D$, if there exists $M_1 > 0$ such that for each $x, y, u, v \in D$

$$\|G(x, y) - G(u, v)\| \leq M_1(\|x - u\| + \|y - v\|). \quad (16)$$

Remark 1. It follows from the preceding definitions that $L = L(L_0, M_0)$, $M = M(L_0, M_0)$

$$L_0 \leq L_1, \quad (17)$$

$$L \leq L_1, \quad (18)$$

$$M_0 \leq M_1, \quad (19)$$

and

$$M \leq M_1, \quad (20)$$

since $D_0 \subseteq D$. If any of (17)–(20) are strict inequalities, then the following advantages are obtained over the work in [9] using L_1 and M_1 instead of the new constants:

(a₁) At least as large convergence domain leading to at least as many initial choices.

(a₂) At least as tight upper bounds on the distances $\|x_n - x^*\|$, so at most as many iterations are needed to obtain a desired error tolerance.

It is always true that D_0 is at least as small and included in D by (12). Here lies the new idea and the reason for the advantages. Notice that these advantages are obtained under the same computational cost, as in [9], since the new constants L_0, M_0, L and M are special cases of constants L_1 and M_1 . This technique of using the center Lipschitz condition in combination with the restricted convergence region has been used on Newton's, Secant and Newton-like methods [14] and can be used on other methods in order to extend their applicability.

The Euclidean norm, and the corresponding matrix norm are used in this study which has the advantage $\|A^T\| = \|A\|$.

The proof of the next result follows the corresponding one in [9] but there are crucial differences where we use (L_0, L) instead of L_1 and M_0, M instead of M_1 .

Theorem 1. Let $F + G : R^p \rightarrow R^m$ be continuous on set $D \subseteq R^p$, F be continuously differentiable in this set, and $G(\cdot, \cdot) : D \times D \rightarrow \mathcal{L}(R^p, R^m)$ be a divided difference of order one. Suppose, the problem (4) has a solution x^* on set D , and the inverse operator $(A_*^T A_*)^{-1} = [(F'(x^*) + G(x^*, x^*))^T (F'(x^*) + G(x^*, x^*))]^{-1}$ exists, $\|(A_*^T A_*)^{-1}\| \leq B$, (9), (10), (13), (14) hold, and γ defined in (11) exists. Moreover,

$$\|F(x^*) + G(x^*)\| \leq \eta, \quad \|F'(x^*) + G(x^*, x^*)\| \leq \alpha; \quad (21)$$

$$B(L_0 + 2M_0)\eta < 1, \quad (22)$$

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

$$q(r) = B[(\alpha + (L_0 + 2M_0)r)(L + 2M)r/2 + (L_0 + 2M_0)\eta] \\ + B[2\alpha + (L_0 + 2M_0)r](L_0 + 2M_0)r - 1. \quad (23)$$

Then, for $x_0, x_{-1} \in U(x^*, r_*)$ method (5) is well defined and generates the sequence $\{x_n\}$, $n = 0, 1, \dots$, which belongs to set $U(x^*, r_*)$, and converges to the solution x^* . Moreover, the following error bounds hold

$$\|x_{n+1} - x^*\| \leq C_1\|x_{n-1} - x^*\| + C_2\|x_n - x^*\| + C_3\|x_{n-1} - x^*\|\|x_n - x^*\| \\ + C_4\|x_{n-1} - x^*\|^2\|x_n - x^*\| + C_5\|x_n - x^*\|^2 \\ + C_6\|x_{n-1} - x^*\|\|x_n - x^*\|^2 + C_7\|x_n - x^*\|^3, \quad (24)$$

where

$$\begin{aligned} g(r) &= B[1 - B(2\alpha + (L_0 + 2M_0)r)(L_0 + 2M_0)r]^{-1}; \\ C_1 &= g(r_*)M_0\eta; \quad C_2 = g(r_*)(L_0 + M_0)\eta; \quad C_3 = g(r_*)\alpha M; \quad C_4 = g(r_*)M_0M; \\ C_5 &= g(r_*)\alpha \frac{L}{2}; \quad C_6 = g(r_*) \left(L_0M + M_0M + \frac{M_0L}{2} \right); \quad C_7 = g(r_*)\frac{L}{2}(L_0 + M_0). \end{aligned} \quad (25)$$

Proof. According to the intermediate value theorem on $[0, r]$ for sufficiently large r and in view of (22) function q has at least one positive zero. Denote by r_* the least such positive zero. Moreover, we have $q'(r) \geq 0$ for $r \geq 0$. Indeed, this zero is unique on $[0, r]$.

We shall show estimate (24) by first showing that sequence x_n is well defined.

Let $A_n = F'(x_n) + G(x_n, x_{n-1})$, and set $n = 0$. We need to show that linear operator A_0 is invertible. By assuming, $x_0, x_{-1} \in U(x^*, r_*)$, we obtain the following estimation:

$$\begin{aligned} \|I - (A_*^T A_*)^{-1} A_0^T A_0\| &= \|(A_*^T A_*)^{-1} (A_*^T A_* - A_0^T A_0)\| \\ &= \|(A_*^T A_*)^{-1} (A_*^T (A_* - A_0) + (A_*^T - A_0^T)(A_0 - A_*) + (A_*^T - A_0^T)A_*)\| \\ &\leq \|(A_*^T A_*)^{-1}\| (\|A_*^T\| \|A_* - A_0\| + \|A_*^T - A_0^T\| \|A_0 - A_*\| + \|A_*^T - A_0^T\| \|A_*\|) \\ &\leq B(\alpha \|A_* - A_0\| + \|A_*^T - A_0^T\| \|A_0 - A_*\| + \alpha \|A_*^T - A_0^T\|). \end{aligned} \quad (26)$$

By (9) and (10), we have in turn the estimate

$$\begin{aligned} \|A_0 - A_*\| &= \|(F'(x_0) + G(x_0, x_{-1})) - (F'(x^*) + G(x^*, x^*))\| \\ &= \|F'(x_0) - F'(x^*) + G(x_0, x_{-1}) - G(x^*, x^*)\| \\ &\leq \|F'(x_0) - F'(x^*)\| + \|G(x_0, x_{-1}) - G(x^*, x^*)\| \\ &\leq L_0 \|x_0 - x^*\| + M_0 (\|x_0 - x^*\| + \|x_{-1} - x^*\|). \end{aligned} \quad (27)$$

Then from inequality (26), definition of r_* (23), we get

$$\begin{aligned} \|I - (A_*^T A_*)^{-1} A_0^T A_0\| &\leq B[2\alpha + L_0 \|x_0 - x^*\| + M_0 (\|x_0 - x^*\| + \|x_{-1} - x^*\|)] \\ &\quad \times [L_0 \|x_0 - x^*\| + M_0 (\|x_0 - x^*\| + \|x_{-1} - x^*\|)] \\ &\leq B[2\alpha + (L_0 + 2M_0)r_*](L_0 + 2M_0)r_* = \varphi(r_*) < 1. \end{aligned} \quad (28)$$

By the Banach Lemma on invertible operators [3], and (28) $A_0^T A_0^{-1}$ is invertible. Then from (26), (27) and (28), we get in turn that

$$\begin{aligned} \|(A_0^T A_0)^{-1}\| &\leq g_0 = B\{1 - B[2\alpha + (L_0 \|x_0 - x^*\| + M_0 (\|x_0 - x^*\| + \|x_{-1} - x^*\|)] \\ &\quad \times [L_0 \|x_0 - x^*\| + M_0 (\|x_0 - x^*\| + \|x_{-1} - x^*\|)]\}^{-1} \\ &\leq g(r_*) = B\{1 - B[2\alpha + (L_0 + 2M_0)r_*](L_0 + 2M_0)r_*\}^{-1}. \end{aligned}$$

Hence, iterate x_1 is well defined by method (5) for $n = 0$. Next, we will show that $x_1 \in U(x^*, r_*)$. First of all, we get the estimation

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

Moreover, using (9), (10), (13), (14) and (21), we obtain in turn

$$\begin{aligned}
 & \|A_0 - \int_0^1 F'(x^* + t(x_0 - x^*))dt - G(x_0, x^*)\| \\
 &= \|F'(x_0) - \int_0^1 F'(x^* + t(x_0 - x^*))dt + G(x_0, x_{-1}) - G(x_0, x^*)\| \\
 &= \|\int_0^1 (F'(x_0) - F'(x^* + t(x_0 - x^*)))dt + G(x_0, x_{-1}) - G(x_0, x^*)\| \\
 &\leq \frac{1}{2}L\|x_0 - x^*\| + M\|x_{-1} - x^*\| = \frac{1}{2}(L\|x_0 - x^*\| + 2M\|x_{-1} - x^*\|), \\
 &\|A_0\| \leq \|A_*\| + \|A_0 - A_*\| \leq \alpha + L_0\|x_0 - x^*\| + M_0(\|x_0 - x^*\| + \|x_{-1} - x^*\|).
 \end{aligned}$$

Then, by method (5) for $n = 0$ and the preceding estimate, we have in turn that

$$\begin{aligned}
 \|x_1 - x^*\| &\leq B\{(\alpha + L_0\|x_0 - x^*\| + M_0(\|x_0 - x^*\| + \|x_{-1} - x^*\|)) \\
 &\quad \times \frac{1}{2}(L\|x_0 - x^*\| + 2M\|x_{-1} - x^*\|)\|x_0 - x^*\| \\
 &\quad + \eta(L_0\|x_0 - x^*\| + M_0(\|x_0 - x^*\| + \|x_{-1} - x^*\|))\} \\
 &\quad / \{1 - B[2\alpha + L_0\|x_0 - x^*\| + M_0(\|x_0 - x^*\| + \|x_{-1} - x^*\|)] \\
 &\quad \times (L_0\|x_0 - x^*\| + M_0(\|x_0 - x^*\| + \|x_{-1} - x^*\|))\} \\
 &\leq g_0\{(\alpha + L_0\|x_0 - x^*\| + M_0(\|x_0 - x^*\| + \|x_{-1} - x^*\|)) \\
 &\quad \times \frac{1}{2}(L\|x_0 - x^*\| + 2M\|x_{-1} - x^*\|)\|x_0 - x^*\| \\
 &\quad + \eta(L_0\|x_0 - x^*\| + M_0(\|x_0 - x^*\| + \|x_{-1} - x^*\|))\} \\
 &< g(r_*)[(\alpha + (L_0 + 2M_0)r_*)(L + 2M)r_*/2 + (L_0 + 2M_0)\eta]r_* \\
 &= p(r_*)r_* = r_*,
 \end{aligned}$$

where $p(r) = g(r)[(\alpha + (L_0 + 2M_0)r)(L + 2M)r/2 + (L_0 + 2M_0)\eta]$. That is $x_1 \in U(x^*, r_*)$ and estimate (24) holds for $n = 0$.

Suppose that $x_n \in U(x^*, r_*)$ for $n = 0, 1, \dots, k$ and estimate (24) holds for $n = 0, 1, \dots, k - 1$, where $k \geq 1$ is integer. We shall show that $x_{n+1} \in U$ and estimate (24) holds for $n = k$.

As in the derivation of (28), using (9), (21) and the definition of function φ , we get in turn that

$$\begin{aligned}
 & \|I - (A_*^T A_*^T)^{-1} A_k^T A_k\| = \|(A_*^T A_*)^{-1} (A_*^T A_* - A_k^T A_k)\| \\
 &= \|(A_*^T A_*)^{-1} (A_*^T (A_* - A_k) + (A_*^T - A_k^T) (A_k - A_*) + (A_*^T - A_k^T) A_*)\| \\
 &\leq \|(A_*^T A_*)^{-1}\|(\|A_*^T\|\|A_* - A_k\| + \|A_*^T - A_k^T\|\|A_k - A_*\| + \|A_*^T - A_k^T\|\|A_*\|) \\
 &\leq B(\alpha\|A_* - A_k\| + \|A_*^T - A_k^T\|\|A_k - A_*\| + \alpha\|A_*^T - A_k^T\|) \\
 &\leq B[2\alpha + L_0\|x_k - x^*\| + M_0(\|x_k - x^*\| + \|x_{k-1} - x^*\|)] \\
 &\quad \times [L_0\|x_k - x^*\| + M_0(\|x_k - x^*\| + \|x_{k-1} - x^*\|)] \\
 &\leq B[2\alpha + (L_0 + 2M_0)r_*](L_0 + 2M_0)r_* < 1.
 \end{aligned}$$

Hence, $(A_k^T A_k)^{-1}$ exists and

$$\begin{aligned}
 & \|(A_{k+1}^T A_{k+1})^{-1}\| \leq g_k = B\{1 - B[2\alpha + (L_0\|x_k - x^*\| \\
 &\quad + M_0(\|x_k - x^*\| + \|x_{k-1} - x^*\|))\} \\
 &\quad \times (L_0\|x_k - x^*\| + M_0(\|x_k - x^*\| + \|x_{k-1} - x^*\|))\}^{-1} \leq g(r_*).
 \end{aligned}$$

Therefore, iteration x_{k+1} is well defined, and the following estimate holds

$$\begin{aligned}
 \|x_{k+1} - x^*\| &= \|x_k - x^* - (A_k^T A_k)^{-1} (A_k^T (F(x_k) + G(x_k)) - A_*^T (F(x^*) + G(x^*)))\| \\
 &\leq \| - (A_k^T A_k)^{-1} \| \| (-A_k^T (A_k - \int_0^1 F'(x^* + t(x_k - x^*)) dt \\
 &\quad - G(x_k, x^*)) (x_k - x^*) + (A_k^T - A_*^T) (F(x^*) + G(x^*)) \| \\
 &\leq g_k \{ (\alpha + L_0 \|x_k - x^*\| + M_0 (\|x_k - x^*\| + \|x_{k-1} - x^*\|)) \\
 &\quad \times \frac{1}{2} (L \|x_k - x^*\| + 2M \|x_{k-1} - x^*\|) \|x_k - x^*\| \\
 &\quad + \eta (L_0 \|x_k - x^*\| + M_0 (\|x_k - x^*\| + \|x_{k-1} - x^*\|)) \} \\
 &\leq g(r_*) \{ (\alpha + L_0 \|x_k - x^*\| + M_0 (\|x_k - x^*\| + \|x_{k-1} - x^*\|)) \\
 &\quad \times \frac{1}{2} (L \|x_k - x^*\| + 2M \|x_{k-1} - x^*\|) \|x_k - x^*\| \\
 &\quad + \eta (L_0 \|x_k - x^*\| + M_0 (\|x_k - x^*\| + \|x_{k-1} - x^*\|)) \} \\
 &< p(r_*) r_* = r_*.
 \end{aligned}$$

That proves $x_{k+1} \in U(x^*, r_*)$ and estimate (24) for $n = k$.

Thus, method (5) is well defined, $x_n \in U(x^*, r_*)$ for all $n \geq 0$ and estimate (24) holds for all $n \geq 0$. It remains to prove that $x_n \rightarrow x^*$ for $n \rightarrow \infty$.

Define a and b on $[0, r_*]$ by

$$a(r) = g(r) ((L_0 + M_0)\eta + \alpha Lr/2 + L(L_0 + M_0)r^2/2) \quad (29)$$

and

$$b(r) = g(r) (M_0\eta + \alpha Mr + (2M_0M + L_0M + \frac{M_0L}{2})r^2). \quad (30)$$

According to r_* , we get

$$a(r_*) \geq 0, \quad b(r_*) \geq 0, \quad a(r_*) + b(r_*) = 1. \quad (31)$$

Using estimate (24), the definitions of constants C_i , $i = 1, 2, \dots, 7$, and functions a and b , for $n \geq 0$ we get following

$$\begin{aligned}
 \|x_{n+1} - x^*\| &\leq C_1 \|x_{n-1} - x^*\| + C_2 \|x_n - x^*\| + C_3 r_* \|x_{n-1} - x^*\| \\
 &+ C_4 r_*^2 \|x_{n-1} - x^*\| + C_5 r_* \|x_n - x^*\| + C_6 r_*^2 \|x_{n-1} - x^*\| + C_7 r_*^2 \|x_n - x^*\| \\
 &= a(r_*) \|x_n - x^*\| + b(r_*) \|x_{n-1} - x^*\|.
 \end{aligned} \quad (32)$$

As it was shown in [1], under conditions (29)–(32) sequence $\{x_n\}$ converges to x^* , as $n \rightarrow \infty$. \square

Corollary 1. In case of $\eta = 0$, we have a nonlinear least squares problem with zero residual. Then, $C_1 = 0$ and $C_2 = 0$, and estimate (24) reduces to

$$\|x_{n+1} - x^*\| \leq (C_3 + C_4 r_*) \|x_{n-1} - x^*\| \|x_n - x^*\| + (C_5 + C_6 r_* + C_7 r_*) \|x_n - x^*\|^2.$$

That is method (5) converges with order $\frac{1+\sqrt{5}}{2}$.

Let $G(x) \equiv 0$ in (4), corresponding to the residual functions being differentiable. Then, from Theorem 1, we obtain the following corollary.

Corollary 2. If $G(x) \equiv 0$, then in the conditions of theorem, we set, $M = 0$, $C_3 = 0$, $C_4 = 0$, and estimate (24) reduces to:

$$\|x_{n+1} - x^*\| \leq (C_5 + C_6 r_* + C_7 r_*) \|x_n - x^*\|^2.$$

Hence method (5) has a convergence order two.

Remark 2. If $L_0 = L = L_1$ and $M_0 = M = M_1$ our results specialize to the corresponding ones in [9]. Otherwise, they constitute an improvement as already noted in the Remark 1. As an example let the $a_1, q_1, C_1^1, C_2^1, C_3^1, C_4^1, r_*^1$ denote the functions and parameter 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 (17)–(20) that

$$q(r) \leq q_1(r),$$

$$g(r) \leq g_1(r),$$

$$C_1 \leq C_1^1,$$

$$C_2 \leq C_2^1,$$

$$C_3 \leq C_3^1,$$

$$C_4 \leq C_4^1,$$

so

$$r_*^1 \leq r_*,$$

$$B(L_1 + 2M_1)\eta < 1 \Rightarrow B(L_0 + 2M_0)\eta < 1.$$

Consequently, the new sufficient convergence criteria are weaker than the ones in [9], unless, if $L_0 = L_1$ and $M_0 = M_1$. And moreover, the new error bounds are tighter than the corresponding ones in [9] and the rest of the advantages already mentioned in Remark 1 hold true.

The results can be improved even further, if (10) and (14) are replaced by

$$\|G(x, y) - G(x^*, x^*)\| \leq K_0 \|x - x^*\| + \overline{K}_0 \|y - x^*\|, \quad (33)$$

and

$$\|G(x, y) - G(u, x^*)\| \leq N_0 \|x - u\| + \overline{N}_0 \|y - x^*\|, \quad (34)$$

respectively, since $K_0 \leq M_0$, $\overline{K}_0 \leq M_0$, $N_0 \leq M$ and $\overline{N}_0 \leq M$. We leave the details to the motivated reader.

4. Local Convergence Analysis (6)

Sufficient conditions and the rate of local converges of method (6) are defined in the following theorem.

Theorem 2. Let $F + G : R^p \rightarrow R^m$ be continuous on set $D \subseteq R^p$, F be continuously differentiable in this set, and G be a function on D . Suppose, the problem (4) has a solution x^* on set D , $F(x^*) + G(x^*) = 0$ and the inverse operator $(A_*^T A_*)^{-1} = [F'(x^*)^T F'(x^*)]^{-1}$ exists and $\|(A_*^T A_*)^{-1}\| \leq B$. Fréchet derivative F' and function G satisfy Lipschitz conditions on set D_0

$$\|F'(x) - F'(x^*)\| \leq L_0 \|x - x^*\|, \quad (35)$$

$$\|F'(x) - F'(y)\| \leq L \|x - y\|, \quad (36)$$

$$\|G(x) - G(x^*)\| \leq M_0 \|x - x^*\|. \quad (37)$$

Moreover,

$$\|F'(x^*)\| \leq \alpha; \quad (38)$$

$$BM_0\alpha < 1 \quad (39)$$

and $U(x^*, r_*) \subseteq D$, where r_* is unique positive zero of function q , defined by

$$q(r) = B[(\alpha + L_0r)(Lr + 2M_0)/2] + B[2\alpha + L_0r]L_0r - 1. \quad (40)$$

Then, for $x_0 \in U(x^*, r_*)$ method (6) is well defined and generates the sequence $\{x_n\}$, $n = 0, 1, \dots$ which belongs to set $U(x^*, r_*)$, and converges to the solution x^* . Moreover, the following error bounds hold

$$\|x_{n+1} - x^*\| \leq C_1\|x_n - x^*\| + C_2\|x_n - x^*\|^2 + C_3\|x_n - x^*\|^3, \quad (41)$$

where

$$g(r) = B[1 - B(2\alpha + L_0r)L_0r]^{-1};$$

$$C_1 = g(r_*)M_0\alpha; \quad C_2 = g(r_*)\left(L_0M_0 + \frac{\alpha L}{2}\right); \quad C_3 = g(r_*)\frac{L_0L}{2}. \quad (42)$$

Proof. According to intermediate value theorem on $[0, r]$ for sufficiently large r and in view of (39) function q has a least positive zero, denoted by r_* , and $q'(r) \geq 0$ for $r \geq 0$. Indeed, this zero is unique on $[0, r]$. The proof analogous to the one given in Theorem 1.

Let $A_n = F'(x_n)$, and set $n = 0$. By assuming $x_0, x_{-1} \in U(x^*, r_*)$. By analogy to (26) in Theorem 1, we get

$$\|I - (A_*^T A_*)^{-1} A_0^T A_0\| \leq B(\alpha \|A_* - A_0\| + \|A_*^T - A_0^T\| \|A_0 - A_*\| + \alpha \|A_*^T - A_0^T\|). \quad (43)$$

Taking into account, that

$$\|A_0 - A_*\| = \|F'(x_0) - F'(x^*)\| \leq L_0\|x_0 - x^*\|, \quad (44)$$

from inequality (43), definition of r_* given in (40), we get

$$\begin{aligned} \|I - (A_*^T A_*)^{-1} A_0^T A_0\| &\leq B[2\alpha + L_0\|x_0 - x^*\|]L_0\|x_0 - x^*\| \\ &\leq B[2\alpha + L_0r_*]L_0r_* = \varphi(r_*) < 1. \end{aligned} \quad (45)$$

From the Banach Lemma on invertible operators [3], and (45) $A_0^T A_0$ is invertible. Then, from (43)–(45), we get

$$\begin{aligned} \|(A_0^T A_0)^{-1}\| &\leq g_0 = B\{1 - B[2\alpha + L_0\|x_0 - x^*\|](L_0\|x_0 - x^*\|)\}^{-1} \\ &\leq g(r_*) = B\{1 - B[2\alpha + L_0r_*]L_0r_*\}^{-1}. \end{aligned}$$

Hence, iteration x_1 is well defined.

Next, we will show that $x_1 \in U(x^*, r_*)$. We have the estimate

$$\begin{aligned} \|x_1 - x^*\| &= \|x_0 - x^* - (A_0^T A_0)^{-1}(A_0^T(F(x_0) + G(x_0)) - A_*^T(F(x^*) + G(x^*)))\| \\ &\leq \|(A_0^T A_0)^{-1}\| \left\| -A_0^T(A_0 - \int_0^1 F'(x^* + t(x_0 - x^*))dt(x_0 - x^*) \right. \\ &\quad \left. - (G(x_0) - G(x^*))) + (A_0^T - A_*^T)(F(x^*) + G(x^*)) \right\|. \end{aligned}$$

In view of the estimates

$$\begin{aligned} & ||(A_0 - \int_0^1 F'(x^* + t(x_0 - x^*))dt)(x_0 - x^*) - (G(x_0) - G(x^*))|| \\ &= ||\int_0^1 (F'(x_0) - F'(x^* + t(x_0 - x^*)))dt(x_0 - x^*) - (G(x_0) - G(x^*))|| \\ &\leq \left(\frac{1}{2}L||x_0 - x^*|| + M_0\right) ||x_0 - x^*|| = \frac{1}{2}(L||x_0 - x^*|| + 2M_0) ||x_0 - x^*||, \\ &||A_0|| \leq ||A_*|| + ||A_0 - A_*|| \leq \alpha + L_0||x_0 - x^*||, \end{aligned}$$

we obtain in turn that

$$\begin{aligned} ||x_1 - x^*|| &\leq B\{(\alpha + L_0||x_0 - x^*||) \frac{1}{2}(L||x_0 - x^*|| + 2M_0) ||x_0 - x^*||\} \\ &\quad / \{1 - B[2\alpha + L_0||x_0 - x^*||]L_0||x_0 - x^*||\} \\ &\leq g_0\{(\alpha + L_0||x_0 - x^*||) \frac{1}{2}(L||x_0 - x^*|| + 2M_0) ||x_0 - x^*||\} \\ &\leq g(r_*)\{(\alpha + L_0||x_0 - x^*||) \frac{1}{2}(L||x_0 - x^*|| + 2M_0) ||x_0 - x^*||\} \\ &< g(r_*)[(\alpha + L_0r_*)(Lr_* + 2M_0)/2]r_* = r_*. \end{aligned}$$

Hence, $x_1 \in U(x^*, r_*)$ and inequality (41) holds for $n = 0$.

Suppose $x_n \in U(x^*, r_*)$ for $n = 0, 1, \dots, k$ and estimate (41) holds for $n = 0, 1, \dots, k-1$, where $k \geq 1$ is integer. Next, we show that $x_{n+1} \in U(x^*, r_*)$ and estimate (41) holds for $n = k$.

Then, we obtain

$$\begin{aligned} &||I - (A_*^T A_*^T)^{-1} A_k^T A_k|| = ||(A_*^T A_*)^{-1} (A_*^T A_* - A_k^T A_k)|| \\ &= ||(A_*^T A_*)^{-1} (A_*^T (A_* - A_k) + (A_*^T - A_k^T)(A_k - A_*) + (A_*^T - A_k^T)A_*)|| \\ &\leq ||(A_*^T A_*)^{-1}|| (||A_*^T|| ||A_* - A_k|| + ||A_*^T - A_k^T|| ||A_k - A_*|| + ||A_*^T - A_k^T|| ||A_*||) \\ &\leq B(\alpha ||A_* - A_k|| + ||A_*^T - A_k^T|| ||A_k - A_*|| + \alpha ||A_*^T - A_k^T||) \\ &\leq B[2\alpha + L_0||x_k - x^*||]L_0||x_k - x^*|| \leq B[2\alpha + L_0r_*]L_0r_* < 1. \end{aligned}$$

Hence, $(A_k^T A_k)^{-1}$ exists and

$$||(A_{k+1}^T A_{k+1})^{-1}|| \leq g_k = B\{1 - B[2\alpha + L_0||x_k - x^*||]L_0||x_k - x^*||\}^{-1} \leq g(r_*).$$

Therefore iteration x_{k+1} is well defined, and we get in turn that

$$\begin{aligned} ||x_{k+1} - x^*|| &= ||x_k - x^* - (A_k^T A_k)^{-1} (A_k^T (F(x_k) + G(x_k)) - A_*^T (F(x^*) + G(x^*)))|| \\ &\leq ||-(A_k^T A_k)^{-1}|| ||[-A_k^T (A_k - \int_0^1 F'(x^* + t(x_k - x^*))dt \\ &\quad - G(x_k, x^*))(x_k - x^*) + (A_k^T - A_*^T)(F(x^*) + G(x^*))]]|| \\ &\leq g_k\{(\alpha + L_0||x_k - x^*||) \frac{1}{2}(L||x_k - x^*|| + 2M_0) ||x_k - x^*||\} \\ &\leq g(r_*)\{(\alpha + L_0||x_k - x^*||) \frac{1}{2}(L||x_k - x^*|| + 2M_0) ||x_k - x^*||\} < r_*. \end{aligned}$$

That proves $x_k \in U(x^*, r_*)$, and estimate (41) for $n = k$.

Thus, iterative process (6) is well defined, $x_n \in U(x^*, r_*)$ for all $n \geq 0$ and estimate (41) holds for all $n \geq 0$.

Define function a on $[0, r_*]$

$$a(r) = g(r)(M_0\alpha + (\alpha L/2 + L_0M_0)r + L_0Lr^2/2). \quad (46)$$

Using estimate (41), the definitions of constants C_i , $i = 1, 2, 3$ and function a , for $n \geq 0$, we get the following

$$\begin{aligned} \|x_{n+1} - x^*\| &\leq C_1\|x_n - x^*\| + C_2r_*\|x_n - x^*\| + C_3r_*^2\|x_n - x^*\| \\ &= a(r_*)\|x_n - x^*\|. \end{aligned} \quad (47)$$

For any $r_* > 0$ and initial point $x_0 \in U(x^*, r_*)$, r' exists and $0 < r' < r_*$ such that $x_0 \in U(x^*, r')$. Similarly to the proof that all iterates stay in $U(x^*, r_*)$, we show that all iterates stay in $U(x^*, r')$. So, estimation (47) holds, if r_* is replaced by r' . In particular, from (47) for $n \geq 0$, we get

$$\|x_{n+1} - x^*\| \leq a\|x_n - x^*\|,$$

where $a = a(r')$. Obviously $a \geq 0$, $a < a(r_*) = 1$. Therefore, we obtain

$$\|x_{n+1} - x^*\| \leq a\|x_n - x^*\| \leq \dots \leq a^{n+1}\|x_0 - x^*\|.$$

However, $a^{n+1} \rightarrow 0$ for $n \rightarrow \infty$. Hence, sequence $\{x_n\}$ converges to x^* as $n \rightarrow \infty$, with a rate of geometric progression. \square

The same type of improvements as in Theorem 1 are obtained for Theorem 2 (see Remark 2).

Remark 3. As we can see from estimations (41) and (42), convergence of method (6) depends on α , L_0 , L and M . For problems with weak nonlinearity (α , L_0 , L and M_0 —“small”) convergence rate of iterative process is linear. In case of strongly nonlinear problems (α , L_0 , L and/or M_0 —“large”) method (6) may not converge at all.

5. Numerical Experiments

Let us compare the convergence rate of combined method (5), Gauss-Newton type method (6) Secant-type method for solving nonlinear least squares problem [5,6] on some test cases with

$$\begin{aligned} x_{n+1} &= x_n - (A_n^T A_n)^{-1} A_n^T (F(x_n) + G(x_n)), \\ A_n &= F(x_n, x_{n-1}) + G(x_n, x_{n-1}), \quad n = 0, 1, \dots \end{aligned} \quad (48)$$

Testing is carried out on nonlinear systems with a nondifferentiable operator with zero and non-zero residual. Classic Gauss-Newton and Newton methods can not be used for solving such a problem. Results are searched with an accuracy $\varepsilon = 10^{-8}$. Calculations are performed until the following conditions are satisfied

$$\|x_{n+1} - x_n\| \leq \varepsilon \quad \text{and} \quad \|A_n^T (F(x_n) + G(x_n))\| \leq \varepsilon,$$

in this case $f(x) = \min_{x \in R^n} \frac{1}{2} (F(x) + G(x))^T (F(x) + G(x))$.

Example 1. [11,12].

$$\begin{cases} 3x^2y + y^2 - 1 + |x - 1| = 0, \\ x^4 + xy^3 - 1 + |y| = 0, \end{cases}$$

$$(x^*, y^*) \approx (0.89465537, 0.32782652), \quad f(x^*) = 0.$$

Example 2. $n = 2$, $m = 3$;

$$\begin{cases} 3x^2y + y^2 - 1 + |x - 1| = 0, \\ x^4 + xy^3 - 1 + |y| = 0, \\ |x^2 - y| = 0, \end{cases}$$

$$(x^*, y^*) \approx (0.74862800, 0.43039151), f(x^*) \approx 4.0469349 \cdot 10^{-2}.$$

Remark 4. The results of the numerical experiments are shown in the Table 1. In particular, we compare studied methods with respect to the number of iterations needed to find the solution with given accuracy. In Example 1, all methods converge to one solution. In Example 2 Gauss-Newton type method (6) converges to point $(x^*, y^*) \approx (0.89465537, 0.32782652)$ with residual $f(x^*) \approx 1.11666739 \cdot 10^{-1}$, with the same number of iterations. Such iterations are marked with * symbol in the table. Other methods find the point $(x^*, y^*) \approx (0.74862800, 0.43039151)$ with smaller residual $f(x^*) \approx 4.0469349 \cdot 10^{-2}$. Additional initial approximation (x_{-1}, y_{-1}) is chosen as:

$$(x_{-1}, y_{-1}) = (x_0 - 10^{-4}, y_0 - 10^{-4}).$$

Table 1. Number of iteration made to solve test problem.

Example	(x_0, y_0)	Gauss-Newton Type (6)	Secant Type (48)	Combined Method (5)
1	(1, 0)	19	7	7
	(3, 1)	22	11	10
	(0.5, 0.5)	21	18	10
2	(1, 0)	19*	22	12
	(3, 1)	22*	25	15
	(0.5, 0.5)	21*	19	13

6. Conclusions

Based on the theoretical studies, the numerical experiments, and the comparison of obtained results, we can argue that the combined differential-difference method (5) converges faster than Gauss-Newton type method (6) and Secant type method (48). Moreover, the method has high convergence order $(1 + \sqrt{5})/2$ in case of zero residual and does not require calculation of derivatives of the nondifferentiable part of operator. Therefore, the proposed method (5) solves the problem efficiently and fast.

Author Contributions: All authors contributed equally and significantly to the writing of this article. All authors read and approved the final manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors would like to express their sincere gratitude to the referees for their valuable comments which have significantly improved the presentation of this paper.

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

References

- Argyros, I.K. *Convergence and Applications of Newton-Type Iterations*; Springer: New York, NY, USA, 2008; 506p.
- Dennis, J.E.; Schnabel, R.B. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*; SIAM: Philadelphia, PA, USA, 1996.
- Ortega, J.M.; Rheinboldt, W.C. *Iterative Solution of Nonlinear Equations in Several Variables*; Academic Press: New York, NY, USA, 1970.
- Argyros, I.K.; Ren, H. A derivative free iterative method method for solving least squares problems. *Numer. Algorithms* **2011**, *58*, 555–571.

5. Ren, H.; Argyros, I.K. Local convergence of a secant type method for solving least squares problems. *Appl. Math. Comput.* **2010**, *217*, 3816–3824. [\[CrossRef\]](#)
6. 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. [\[CrossRef\]](#)
7. Shakhno, S.M.; Gnatyshyn, O.P. Iterative-difference methods for solving nonlinear least-squares problem. In *Progress in Industrial Mathematics at ECMI 98*; Vieweg + Teubner Verlag: Stuttgart, Germany, 1999; pp. 287–294.
8. Argyros, I.K.; Hilout S. On an improved convergence analysis of Newton's method. *Appl. Math. Comput.* **2013**, *225*, 372–386. [\[CrossRef\]](#)
9. Shakhno, S.M.; Shunkin, Y.V. One combined method for solving nonlinear least squares problems. *Visnyk Lviv Univ. Ser. Appl. Math. Inform.* **2017**, *25*, 38–48. (In Ukrainian)
10. Ulm, S. On generalized divided differences. *Proc. Acad. Sci. Estonian SSR. Phys. Mathe.* **1967**, *16*, 13–26. (In Russian)
11. Cătinăș, E. On some iterative methods for solving nonlinear equations. *Revue d'Analyse Numérique et de Theorie de l'Approximation* **1994**, *23*, 47–53.
12. Shakhno, S.M.; Mel'nyk, I.V.; Yarmola, H.P. Analysis of the Convergence of a Combined Method for the Solution of Nonlinear Equations. *J. Math. Sci.* **2014**, *201*, 32–43. [\[CrossRef\]](#)
13. 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. [\[CrossRef\]](#)
14. Argyros, I.K.; Magreñán, Á.A. *A Contemporary Study of Iterative Methods: Convergence, Dynamics and Applications*; Academic Press: London, UK, 2018.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).