Convergence analysis of a two-step method for the nonlinear least squares problem with decomposition of operator

Stepan Shakhno, Roman Iakymchuk, Halyna Yarmola

Журнал обчислювальної та прикладної математики | 2018, № – (–) с. 1–14 | Journal of Computational & Applied Mathematics

UDC 519.6

CONVERGENCE OF A TWO-STEP METHOD FOR THE NONLINEAR LEAST SQUARES PROBLEM WITH DECOMPOSITION OF OPERATOR

S. M. Shakhno, R. P. Iakymchuk, H. P. Yarmola

Резюме. У роботі запропоновано та досліджено збіжність двокрокового методу для розв'язування нелінійної задачі найменших квадратів з декомпозицією оператора за класичних умов Ліпшиця для похідних першого і другого порядків диференційовної частини та поділених різниць першого порядку недиференційовної частини декомпозиції. Встановлено порядок і радіус збіжності методу, а також область єдиності розв'язку нелінійної задачі про найменші квадрати. Проведено чисельні експерименти на ряді тестових задачах.

Abstract. In this article, we propose a two-step method for the nonlinear least squares problem with the decomposition of the operator. We investigate the convergence of this method by applying the classical Lipschitz condition for the first- and second-order derivatives of the differentiable part and for the first-order differences of the non-differentiable part of the decomposition. The convergence order as well as the convergence radius of the method are studied and the uniqueness ball of the solution of the nonlinear least squares problem is examined. Finally, we carry out numerical experiments on a set of test problems.

1. Introduction

Let us consider the nonlinear least squares problem:

$$\min_{x\in\mathbb{R}^n} \frac{1}{2}F(x)^T F(x), \tag{1}$$

where F is a Fréchet differentiable operator defined on $\mathbb{R}^n$ with its values on $\mathbb{R}^m$, $m \geq n$. The best known method for finding an approximate solution of the problem (1) is the Gauss-Newton method, which is defined as

$$x_{k+1} = x_k - [F'(x_k)^T F'(x_k)]^{-1} F'(x_k)^T F(x_k),\ k = 0, 1, 2, ... \tag{2}$$

The convergence analysis of the method (2) under various conditions was conducted in [6, 7, 8]. In paper [17], three free-derivative iterative methods were investigated under the classical Lipschitz conditions. The radius of the convergence ball and the convergence order of these methods were determined. The study of these methods was conducted in the case of both zero and nonzero residuals.

Key words. Nonlinear least squares problem, two-step method, Gauss-Newton method, decomposition of operator, Lipschitz conditions, radius of convergence, uniqueness ball.

In particular, Shakhno [17] proposed the Secant-type method, which was later also studied by Ren and Argyros in [12], as follows

$$x_{k+1} = x_k - [F(x_k, x_{k-1})^T F(x_k, x_{k-1})]^{-1} F(x_k, x_{k-1})^T F(x_k), \quad k = 0, 1, 2, ... \tag{3}$$

This study [17] also determines the convergence order of the method (3) in case of zero residual, which equals to $\dfrac{1+\sqrt{5}}{2} = 1,618....$

In [2, 4, 10, 11], there was considered a two-step modification of the Gauss-Newton method for solving the problem (1)

$$\begin{cases} x_{k+1} = x_k - [F'(z_k)^T F'(z_k)]^{-1} F'(z_k)^T F(x_k), \\ y_{k+1} = x_{k+1} - [F'(z_k)^T F'(z_k)]^{-1} F'(z_k)^T F(x_{k+1}), \ k = 0, 1, 2, ..., \end{cases} \tag{4}$$

where $z_k = (x_k + y_k)/2$; x_0 and y_0 are given. In case when $m = n$, this method is equivalent to the methods proposed by Bartish [3] and Werner [23]. On each iteration, the method (4) computes the inversion of the matrix $[F'(z_k)^T F'(z_k)]^{-1}$ only once.

In [16], we proposed the differential variant of the method (4) that uses divided differences instead of derivatives as follows

$$\begin{cases} x_{k+1} = x_k - [F(x_k, y_k)^T F(x_k, y_k)]^{-1} F(x_k, y_k)^T F(x_k), \\ y_{k+1} = x_{k+1} - [F(x_k, y_k)^T F(x_k, y_k)]^{-1} F(x_k, y_k)^T F(x_{k+1}), \ k = 0, 1, 2, ... \end{cases} \tag{5}$$

This method is built on top of the Secant-type method [12, 17] (3) for solving the nonlinear least squares problem. This method can also be applied to problems with non-differentiable operators.

However, for some problems the nonlinear function in (1) is composed of the differentiable and non-differentiable parts. In this case, the problem (1) can be written as

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} (F(x) + G(x))^T (F(x) + G(x)), \tag{6}$$

where the residual function $F + G$ is defined on $\mathbb{R}^n$ with its values on $\mathbb{R}^m$ and it is nonlinear by x; F is a continuously differentiable function; G is a continuous function, differentiability of which, in general, is not required. To solve the problem (6), we proposed in [13, 18] a method that takes into account the specific features of both F and G as

$$x_{k+1} = x_k - [A_k^T A_k]^{-1} A_k^T (F(x_k) + G(x_k)), \quad k = 0, 1, ..., \tag{7}$$

where $A_k = F'(x_k) + G(x_k, x_{k-1})$; $F'(x_k)$ is a Fréchet derivative of $F(x)$; $G(x_k, x_{k-1})$ is the divided difference of the first-order of the function $G(x)$ at points x_k, x_{k-1}; x_0, x_{-1} are given starting points. This method has the convergence order of $\dfrac{1+\sqrt{5}}{2}$ for solving the problem (6) with zero residual. In case when $m = n$, the method (7) reassembles the well-know Newton-Secant method for nonlinear equations [1, 5, 14].

In this article, we propose a two-step iterative method, for solving the problem (6), which considers the decomposition of the nonlinear operator, as follows

$$\begin{cases} x_{k+1} = x_k - [A_k^T A_k]^{-1} A_k^T (F(x_k) + G(x_k)), \\ y_{k+1} = x_{k+1} - [A_k^T A_k]^{-1} A_k^T (F(x_{k+1}) + G(x_{k+1})), \quad k = 0, 1, ..., \end{cases} \tag{8}$$

where $A_k = F'(\frac{x_k + y_k}{2}) + G(x_k, y_k)$. The main goal of this paper is to analyze the local convergence of the method (8) for the problem (6) with zero as well as non-zero residuals. Additionally, we study both the order and the radius of the convergence of the method (8) as well as the uniqueness ball of the solution of the problem (6). To note, this method as well as the method (5) have the same convergence order of $1 + \sqrt{2}$ in case of zero residual.

In case of $m = n$, the problem (6) converges to solving a system of n nonlinear equations with n unknown and the method (8) to the method [15, 19, 20].

2. Preliminaries

Let us denote $B(x_*, r) = \{x \in D \subseteq \mathbb{R}^n : \|x - x_*\| \leq r\}$ as a closed ball with the radius r $(r > 0)$ at x_*, D is an open convex subset of $\mathbb{R}^n$.

Let $\mathbb{R}^{m \times n}$, $m \geq n$, denote a set of all $m \times n$ matrices. Then, for a full rank matrix $A \in \mathbb{R}^{m \times n}$, its Moore-Penrose pseudo-inverse [8] is defined as $A^\dagger = (A^T A)^{-1} A^T$.

Lemma 1 ([21, 22]). *Let $A, E \in \mathbb{R}^{m \times n}$. Assume that $C = A + E$, $\|A^\dagger\| \|E\| < 1$, and $rank(A) = rank(C)$. Then,*

$$\|C^\dagger\| \leq \frac{\|A^\dagger\|}{1 - \|A^\dagger\| \|E\|}.$$

If $rank(A) = rank(C) = \min(m, n)$, we can obtain

$$\|C^\dagger - A^\dagger\| \leq \frac{\sqrt{2} \|A^\dagger\|^2 \|E\|}{1 - \|A^\dagger\| \|E\|}.$$

Lemma 2 ([6]). *Let $A, E \in \mathbb{R}^{m \times n}$. Assume that $C = A + E$, $\|EA^\dagger\| < 1$, and $rank(A) = n$, then $rank(C) = n$.*

3. Local Convergence Analysis of the Method (8)

In this section, we investigate the convergence of the method (8) and determine its convergence radius.

Theorem 1. *Let $F + G : \mathbb{R}^n \to \mathbb{R}^m$, $m \geq n$, be continuous, where F is a twice Fréchet differentiable operator and G is a continuous operator on a subset $D \subseteq \mathbb{R}^n$. Assume that the problem (6) has a solution $x_* \in D$ and an operator $F'(x_*) + G(x_*, x_*)$ has full rank. Suppose that Fréchet derivatives $F'(x)$ and $F''(x)$ satisfy the Lipschitz conditions on D*

$$\|F'(x) - F'(y)\| \leq L\|x - y\|, \tag{9}$$

$$\|F''(x) - F''(y)\| \leq N\|x - y\|, \tag{10}$$

and the function G has the first order divided difference $G(x,y)$ and

$$\|G(x,y)-G(u,v)\| \leq M(\|x-u\|+\|y-v\|) \tag{11}$$

for all $x,y,u,v \in D$; L, N, and M are non-negative numbers.

Also, the radius $r>0$ is a root of the equation

$$\beta N p^2 + 120\beta T p + 48\sqrt{2}\alpha\beta^2 T - 24 = 0, \tag{12}$$

where

$$2\sqrt{2}\alpha\beta^2 T < 1. \tag{13}$$

Then, for all $x_0, y_0 \in B(x_*, r) \subset D$ the sequences $\{x_k\}$ and $\{y_k\}$, which are generated by the method (8), are well defined, remain in $B(x_*, r)$ for all $k \geq 0$, and converge to x_* such that

$$\begin{aligned}
\rho(x_{k+1}) &\leq \frac{\beta}{1-\beta T\tau_k}\big((N/24)\rho(x_k)^3 + T\rho(x_k)\rho(y_k) + \sqrt{2}\alpha\beta T\tau_k\big), && (14)\\
\rho(y_{k+1}) &\leq \frac{\beta}{1-\beta T\tau_k}\big((N/24)\rho(x_k)^3 + T(\rho(x_{k+1})+\rho(x_k)+\rho(y_k))\rho(x_{k+1}) + \\
&\quad + \sqrt{2}\alpha\beta T\tau_k\big), && (15)\\
r_{k+1} &= \max\{\rho(x_{k+1}), \rho(y_{k+1})\} \leq q r_k \leq \cdots \leq q^{k+1} r_0, && (16)
\end{aligned}$$

where

$$0 < q = \frac{\beta\big((N/24)\rho(x_0)^2 + T(2\rho(x_0)+\rho(y_0)) + \sqrt{2}\alpha\beta T\tau_0/r_0\big)}{1-\beta T\tau_0} < 1, \tag{17}$$

$\rho(x) = \|x - x_*\|$, $\tau_k = \tau(x_k, y_k) = \|x_k - x_*\| + \|y_k - x_*\|$, $r_0 = \max\{\rho(x_0), \rho(y_0)\}$, $z_k = (x_k + y_k)/2$, $\alpha = \|F(x_*) + G(x_*)\|$, $\beta = \|(A_*^T A_*)^{-1} A_*^T\|$, $A_* = F'(x_*) + G(x_*, x_*)$, $T = \dfrac{L+2M}{2}$, $\beta T\tau_0 < 1$.

Proof. From (13) it follows that (12) has the unique positive root, which we annotate as r.

Let choose arbitrary $x_0, y_0 \in B(x_*, r)$ and denote $A_n = F'(\frac{x_n+y_n}{2}) + G(x_n, y_n)$. For $n = 0$, we have the following estimate

$$\begin{aligned}
\|A_0 - A_*\| &= \left\|F'\Big(\frac{x_0+y_0}{2}\Big) + G(x_0,y_0) - (F'(x_*) + G(x_*,x_*))\right\| = \\
&= \left\|F'\Big(\frac{x_0+y_0}{2}\Big) - F'(x_*) + G(x_0,y_0) - G(x_*,x_*)\right\| \leq \\
&\leq \left\|F'\Big(\frac{x_0+y_0}{2}\Big) - F'(x_*)\right\| + \|G(x_0,y_0) - G(x_*,x_*)\| \leq \\
&\leq \frac{L}{2}(\|x_0 - x_*\| + \|y_0 - x_*\|) + M(\|x_0 - x_*\| + \|y_0 - x_*\|) \leq \\
&\leq \frac{L+2M}{2}(\|x_0 - x_*\| + \|y_0 - x_*\|) = T(\|x_0 - x_*\| + \|y_0 - x_*\|)
\end{aligned}$$

and

$$\|(A_*^T A_*)^{-1} A_*^T [A_0 - A_*]\| \leq \beta T(\|x_0 - x_*\| + \|y_0 - x_*\|) = \beta T\tau_0 < 1.$$

According to Lemma 1

$$\|(A_0^T A_0)^{-1} A_0^T\| \le \frac{\beta}{1-\beta T(\|x_0-x_*\|+\|y_0-x_*\|)} = \frac{\beta}{1-\beta T\tau_0},$$

and to Lemma 2

$$\|(A_0^T A_0)^{-1} A_0^T-(A_*^T A_*)^{-1} A_*^T\| \le \frac{\sqrt{2}\beta^2 T(\|x_0-x_*\|+\|y_0-x_*\|)}{1-\beta T(\|x_0-x_*\|+\|y_0-x_*\|)} = \frac{\sqrt{2}\beta^2 T\tau_0}{1-\beta T\tau_0}.$$

For x_1, y_1 that are generated by (8), we have

$$\begin{aligned}
x_1 - x_* &= x_0 - x_* - \left[A_0^T A_0\right]^{-1} A_0^T(F(x_0)+G(x_0)) = \\
&= \left[A_0^T A_0\right]^{-1} A_0^T\left[A_0(x_0-x_*)-(F(x_0)+G(x_0))+(F(x_*)+G(x_*))\right] + \\
&\quad + \left[A_*^T A_*\right]^{-1} A_*^T(F(x_*)+G(x_*)) - \left[A_0^T A_0\right]^{-1} A_0^T(F(x_*)+G(x_*)) = \\
&= \left[A_0^T A_0\right]^{-1} A_0^T\left[F'\left(\frac{x_0+x_*}{2}\right)(x_0-x_*) - F(x_0)+F(x_*)+\right. \\
&\quad + G(x_0,x_*)(x_0-x_*) - G(x_0)+G(x_*) + \\
&\quad \left. + \left(A_0 - F'\left(\frac{x_0+x_*}{2}\right) - G(x_0,x_*)\right)(x_0-x_*)\right] + \\
&\quad + \left[A_*^T A_*\right]^{-1} A_*^T(F(x_*)+G(x_*)) - \left[A_0^T A_0\right]^{-1} A_0^T(F(x_*)+G(x_*));
\end{aligned}$$

$$\begin{aligned}
y_1 - x_* &= x_1 - x_* - \left[A_0^T A_0\right]^{-1} A_0^T(F(x_1)+G(x_1)) = \\
&= \left[A_0^T A_0\right]^{-1} A_0^T\left[A_0(x_1-x_*)-(F(x_1)+G(x_1))+(F(x_*)+G(x_*))\right] + \\
&\quad + \left[A_*^T A_*\right]^{-1} A_*^T(F(x_*)+G(x_*)) - \left[A_0^T A_0\right]^{-1} A_0^T(F(x_*)+G(x_*)) = \\
&= \left[A_0^T A_0\right]^{-1} A_0^T\left[F'\left(\frac{x_1+x_*}{2}\right)(x_1-x_*) - F(x_1)+F(x_*) +\right. \\
&\quad + G(x_1,x_*)(x_1-x_*) - G(x_1)+G(x_*) + \\
&\quad \left. + \left(A_0 - F'\left(\frac{x_1+x_*}{2}\right) - G(x_1,x_*)\right)(x_1-x_*)\right] + \\
&\quad + \left[A_*^T A_*\right]^{-1} A_*^T(F(x_*)+G(x_*)) - \left[A_0^T A_0\right]^{-1} A_0^T(F(x_*)+G(x_*)).
\end{aligned}$$

According to Lemma 1 from [23] with the value $\omega = 1/2$ we can write

$$F(x) - F(y) - F'\left(\frac{x+y}{2}\right)(x-y) =$$
$$= \frac{1}{4}\int_0^1 (1-t)\left[F''\left(\frac{x+y}{2}+\frac{t}{2}(x-y)\right) - F''\left(\frac{x+y}{2}+\frac{t}{2}(y-x)\right)\right](x-y)^2 dt.$$

By setting $x = x_*$ and $y = x_0$ in the equation above, we receive

$$\left\|F(x_*) - F(x_0) - F'\left(\frac{x_0+x_*}{2}\right)(x_*-x_0)\right\| =$$
$$= \frac{1}{4}\left\|\int_0^1 (1-t)\left[F''\left(\frac{x_0+x_*}{2}+\frac{t}{2}(x_*-x_0)\right) - \right.\right.$$

$$- F''\left(\frac{x_0+x_*}{2}+\frac{t}{2}(x_0-x_*)\right)\Big]\,(x_*-x_0)^2dt\Big\| \leq$$
$$\leq \frac{1}{4}\int_0^1 t(1-t)N\|x_0-x_*\|^3dt = \frac{1}{24}N\rho(x_0)^3.$$

Using to the Lipschitz conditions (9), (10) and (11), we get the following estimates

$$\begin{aligned} \Big\|A_0 - F'\Big(\frac{x_0+x_*}{2}\Big) - G(x_0,x_*)\Big\| &= \Big\|F'\Big(\frac{x_0+y_0}{2}\Big) - F'(\frac{x_0+x_*}{2}) + \\ &\quad +G(x_0,y_0) - G(x_0,x_*)\Big\| \leq T\|y_0-x_*\|, \end{aligned}$$

$$\begin{aligned} \Big\|A_0 - F'\Big(\frac{x_1+x_*}{2}\Big) - G(x_1,x_*)\Big\| &= \Big\|F'\Big(\frac{x_0+y_0}{2}\Big) - F'\Big(\frac{x_1+x_*}{2}\Big) + \\ &\quad +G(x_0,y_0) - G(x_1,x_*)\Big\| \leq \\ &\leq L\Big\|\frac{x_0-x_1}{2}+\frac{y_0-x_*}{2}\Big\| + \\ &\quad +M(\|x_1-x_0\|+\|y_0-x_*\|) = \\ &= \frac{L+2M}{2}\|x_0-x_1\| + \frac{L+2M}{2}\|y_0-x_*\| \leq \\ &\leq T(\|x_0-x_*\|+\|x_1-x_*\|+\|y_0-x_*\|). \end{aligned}$$

Hence, from (12) it follows that

$$\begin{aligned} 0<q &= \frac{\beta\big((N/24)\rho(x_0)^2 + T(\rho(x_0)+\rho(x_1)+\rho(y_0)) + \sqrt{2}\alpha\beta T\tau_0/r\big)}{1-\beta T\tau_0/r_0} < \\ &< \frac{\beta\big((N/24)r^2 + 3Tr + 2\sqrt{2}\alpha\beta T\big)}{1-2\beta Tr} = 1. \end{aligned}$$

Thus, by Lemmas 1, 2, conditions (9), (10) and (11), we obtain

$$\|x_1-x_*\| \leq \frac{\beta\big((N/24)\rho(x_0)^3 + T\rho(x_0)\rho(y_0) + \sqrt{2}\alpha\beta T\tau_0\big)}{1-\beta T\tau_0} < qr_0 < r.$$

Similarly,

$$\begin{aligned} \|y_1-x_*\| &\leq \frac{\beta\big((N/24)\rho(x_1)^3 + T(\rho(x_0)+\rho(x_1)+\rho(y_0))\rho(x_1)\big)}{1-\beta T\tau_0} + \\ &\quad + \frac{\sqrt{2}\alpha\beta^2T\tau_0}{1-\beta T\tau_0} < qr_0 < r. \end{aligned}$$

Therefore, $x_1, y_1 \in B(x_*, r)$ and both (14) and (15) follow. Also, (16) is satisfied

$$r_1 = \max\{\|x_1-x_*\|, \|y_1-x_*\|\} \leq qr_0.$$

Using mathematical induction, assume that $x_k, y_k \in B(x_*, r)$ and (16) holds for $k > 0$. Then, for $k+1$ from (8) we obtain that

$$\begin{aligned}\|x_{k+1} - x_*\| &\le \frac{\beta\big((N/24)\rho(x_k)^3 + T\rho(x_k)\rho(y_k) + \sqrt{2}\alpha\beta T\tau_k\big)}{1-\beta T\tau_k} \le \\ &\le \frac{\beta\big((N/24)\rho(x_0)^2 + T\rho(x_0) + 2\sqrt{2}\alpha\beta T\big)r_k}{1-\beta T\tau_0} < qr_k < r\end{aligned}$$

and

$$\begin{aligned}\|y_{k+1} - x_*\| &\le \frac{\beta\big((N/24)\rho(x_{k+1})^3 + T(\rho(k) + \rho(x_{k+1}) + \rho(y_k))\rho(x_{k+1})\big)}{1-\beta T\tau_k} + \\ &+ \frac{\sqrt{2}\alpha\beta^2 T\tau_k}{1-\beta T\tau_k} \le \frac{\beta\big((N/24)\rho(x_0)^2 + 3T\rho(x_0) + 2\sqrt{2}\alpha\beta T\big)r_k}{1-\beta T\tau_0} < \\ &< qr_k < r.\end{aligned}$$

According to (17) and both inequalities (14) and (15), we receive

$$r_{k+1} = \max\{\|x_{k+1} - x_*\|, \|y_{k+1} - x_*\|\} \le qr_k \le q^2 r_{k-1} \le \cdots \le q^{k+1} r_0.$$

Thus, $x_{k+1}, y_{k+1} \in B(x_*, r)$ as well as (14), (15) and (16) hold. □

From (12) it follows that the convergence radius of the method (8) is

$$r = \frac{2(1 - 2\sqrt{2}\alpha\beta^2 T)}{5\beta T + \sqrt{(5\beta T)^2 + \frac{1}{6}\beta N(1 - 2\sqrt{2}\alpha\beta^2 T)}}.$$

Remark 1. *Note that the condition (11) can be replaced with the weaker one*

$$\|G(x,y) - G(u,v)\| \le M_1\|x-u\| + M_2\|y-v\| \tag{18}$$

for all $x, y, u, v \in D$, M_1 and M_2 are positive numbers. This enlarges applicability of the method (8).

For zero residual ($F(x_*) + G(x_*) = 0$), the Theorem 1 can be formulated as

Theorem 2. *Let $F + G : \mathbb{R}^n \to \mathbb{R}^m$, $m \ge n$, is continuous, where F is a twice Fréchet differentiable operator and G is a continuous operator on a subset $D \subseteq \mathbb{R}^n$. Assume that the problem (6) has a solution $x_* \in D$, and the operator $F'(x_*) + G(x_*, x_*)$ has full rank. Suppose that Fréchet derivatives $F'(x)$ and $F''(x)$ on D satisfy the classic Lipschitz conditions as in (9) and (10), respectively; the function G has the first order divided difference $G(x,y)$ that satisfies the Lipschitz conditions as in (11). Moreover, the radius $r > 0$ is a unique positive root of the following equation*

$$\beta N p^2 + 120\beta T p - 24 = 0.$$

Then, the combined method (8) converges to x_* *for all* $x_0, y_0 \in B(x_*, r) \subset D$ *such that*

$$\rho(x_{k+1}) \leq \frac{\beta}{1-\beta T\tau_k}\big((N/24)\rho(x_k)^3 + T\rho(x_k)\rho(y_k)\big), \tag{19}$$

$$\rho(y_{k+1}) \leq \frac{\beta\big((N/24)\rho(x_{k+1})^3 + T(\rho(x_{k+1}) + \rho(x_k) + \rho(y_k))\rho(x_{k+1})\big)}{1-\beta T\tau_k}, \tag{20}$$

$$r_{k+1} = \max\{\rho(x_{k+1}), \rho(y_{k+1})\} \leq qr_k \leq \cdots \leq q^{k+1}r_0,$$

where $\rho(x) = \|x - x_*\|$, $\tau_k = \tau(x_k, y_k) = \|x_k - x_*\| + \|y_k - x_*\|$, $r_0 = \max\{\rho(x_0), \rho(y_0)\}$, $\beta = \|(A_*^T A_*)^{-1}A_*^T\|$, $A_* = F'(x_*) + G(x_*, x_*)$, $\beta T\tau_0 < 1$

$$0 < q = \frac{\beta\big((N/24)\rho(x_0)^2 + T(2\rho(x_0) + \rho(y_0))\big)}{1-\beta T\tau_0} < 1.$$

From Theorem 2, the convergence radius is

$$r = \frac{2}{5\beta T + \sqrt{(5\beta T)^2 + \frac{1}{6}\beta N}} < \frac{1}{5\beta T}.$$

This radius is two times smaller than the convergence radius of the differential method (4) from [11] (a two-step modification of the Gauss-Newton method) and equals to the convergence radius of the difference method (5) from [16].

Corollary 1. *Convergence order of the iterative method (8) in case of zero residual is equal to* $1 + \sqrt{2}$.

Proof. Let us denote $\gamma = \dfrac{\beta N/24}{1-\beta T\tau_0}$, $\eta = \dfrac{\beta T}{1-\beta T\tau_0}$, $a_k = \rho(x_k)$, $b_k = \rho(y_k)$, $k = 0, 1, 2, ...$ Since the residual is zero, i.e. $\alpha = \|F(x_*) + G(x_*)\| = 0$, from the inequalities (19) and (20) we have

$$a_{k+1} \leq a_k(\gamma a_k^2 + \eta b_k), \tag{21}$$

$$\begin{aligned} b_{k+1} &\leq a_{k+1}\left[\gamma a_{k+1}^2 + \eta/3(a_k + a_{k+1} + b_k)\right] \leq \\ &\leq a_{k+1}\left[(\gamma a_k + 2\eta/3)a_k + \eta b_k/3\right] \leq \\ &\leq a_{k+1}a_k\left[\gamma r + \eta\right] = a_{k+1}a_k\phi_1. \end{aligned} \tag{22}$$

From (21) and (22) for large enough k, it follows

$$a_{k+1} \leq a_k(\gamma a_k^2 + \eta b_k) \leq a_k(\gamma a_k^2 + \eta\phi_1 a_k a_{k-1}) \leq a_k^2 a_{k-1}(\gamma + \eta\phi_1) = a_k^2 a_{k-1}\phi_2.$$

From this inequality, we obtain an equation

$$\rho^2 - 2\rho - 1 = 0.$$

The positive root of the latter, which is $\rho_* = 1 + \sqrt{2}$, is the order of convergence of the iterative method (8). □

Under the classic Lipschitz condition a theorem for the uniqueness of the solution can be written as follow

Theorem 3. *Suppose x_* satisfies (6) and $F(x)$ has a continuous derivative $F'(x)$ and $G(x)$ has a divided difference $G(x,y)$ in D. Moreover, $F'(x_*)+G(x_*,x_*)$ has full rank; $F'(x)$ satisfies the Lipschitz condition as in (9); the divided difference $G(x,y)$ satisfies the Lipschitz condition as in (11). Let $r>0$ satisfies*

$$\beta(Lr/2+M)+\alpha\beta_0(L+2M)\leq 1,$$

where $\beta_0=\|(F'(x_*)+G(x_*,x_*))^T(F'(x_*)+G(x_*,x_*))\|$. Then, x_* is a unique solution of the problem (6) in $B(x_*,r)$.

The proof of this theorem is analogous to the one in [6].

To note, in case when $G(x)=0$, we obtain the same results as in Theorem 2 in [11].

5. Numerical experiments

In this section, we give two examples to show the application of our results. We consider method (8) and its partial cases, namely the two-step Gauss-Newton method ($G\equiv 0$) and the two-step Secant method ($F\equiv 0$). We use the norm $\|x\|=\sqrt{\sum\limits_{i=1}^{p}x_i^2}$ for $x\in\mathbb{R}^p$.

Example 1. Consider function $F+G: D=\mathbb{R}\to\mathbb{R}^2$ given by [12]:

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

where $\lambda,\mu\in\mathbb{R}$ are two parameters.

It is known, that $x_*=0$ is the unique solution of the considered problem. Therefore, we can define constants α and β as follows:

$$\alpha=\sqrt{2}|\mu|,\ \beta=\frac{1}{\sqrt{2}}.$$

Let $G(x)=(0,0)^T$. Then

$$F'(x)=\begin{pmatrix} 1 \\ 2\lambda x+1 \end{pmatrix},\quad F''(x)=\begin{pmatrix} 0 \\ 2\lambda \end{pmatrix}$$

and

$$\|F'(x)-F'(y)\|=\left\|\begin{pmatrix} 0 \\ 2\lambda(x-y) \end{pmatrix}\right\|=2|\lambda||x-y|,$$

$$\|F''(x)-F''(y)\|=\left\|\begin{pmatrix} 0 \\ 0 \end{pmatrix}\right\|=0|x-y|.$$

That is, we can set constants $L=2|\lambda|$, $N=0$, $M=0$, $T=\dfrac{L}{2}=\dfrac{2|\lambda|}{2}=|\lambda|$.

Let $F(x)=(0,0)^T$. Then

$$G(x,y)=\begin{pmatrix} \dfrac{x+\mu-y-\mu}{x-y} \\ \dfrac{\lambda x^2+x-\mu-\lambda y^2-y+\mu}{x-y} \end{pmatrix}=\begin{pmatrix} 1 \\ \lambda(x+y)+1 \end{pmatrix}$$

and

$$\|G(x,y)-G(u,v)\| = \left\| \begin{pmatrix} 0 \\ \lambda(x-u+y-v) \end{pmatrix} \right\| \le |\lambda|(|x-y|+|u-v|).$$

That is, we can set constants $L=0$, $N=0$, $M=|\lambda|$, $T=M=|\lambda|$.

Then equation (12) for both methods has form

$$5\sqrt{2}|\lambda|r + 4|\lambda\mu| - 2 = 0.$$

It has unique positive solution

$$r = \frac{\sqrt{2} - 2\sqrt{2}|\lambda\mu|}{5|\lambda|}$$

if parameters λ and μ satisfy

$$\lambda \ne 0,\ |\lambda\mu| < \frac{1}{2}.$$

Let $x_0 = 0.2$, $y_0 = 0.2001$. For this problem $A_k = \begin{pmatrix} 1 \\ \lambda(x_k+y_k)+1 \end{pmatrix}$ in both cases. Therefore, we get the same result by the two-step Gauss-Newton method and the two-step Secant method.

Tabl. 1. The results for $\lambda = 1$, $\mu = 0$.

k	$\rho(x_{k+1})$	The right side of (14)	$\rho(y_{k+1})$	The right side of (15)
0	1.893e-002	3.946e-002	3.412e-003	7.821e-003
1	3.229e-005	4.640e-005	3.600e-007	5.190e-007
2	5.812e-012	8.220e-012	9.487e-017	1.342e-016
3	0	3.899e-028	0	0

Tabl. 2. The results for $\lambda = 0.5$, $\mu = 0.2$.

k	$\rho(x_{k+1})$	The right side of (14)	$\rho(y_{k+1})$	The right side of (15)
0	2.624e-002	6.308e-002	1.881e-002	5.121e-002
1	2.326e-003	4.755e-003	2.230e-003	4.617e-003
2	2.284e-004	4.578e-004	2.274e-004	4.564e-004
3	2.280e-005	4.560e-005	2.279e-005	4.559e-005
4	2.279e-006	4.558e-006	2.279e-006	4.558e-006
5	2.279e-007	4.558e-007	2.279e-007	4.558e-007
6	2.279e-008	4.558e-008	2.279e-008	4.558e-008
7	2.279e-009	4.558e-009	2.279e-009	4.558e-009
8	2.279e-010	4.558e-010	2.279e-010	4.558e-010

If $\lambda = 1$, $\mu = 0$ we obtain $2\sqrt{2}\alpha\beta^2T = 0 < 1$, $\beta T\tau_0 = 0.28291342315 < 1$, $r = 0.2828427124746190$ and $B(x_*, r) \subset D$. If $\lambda = 0.5$, $\mu = 0.2$ we obtain $2\sqrt{2}\alpha\beta^2T = 0.2 < 1$, $\beta T\tau_0 = 0.14145671157 < 1$, $r = 0.4525483399593903$ and $B(x_*, r) \subset D$. From Tables 1, 2, we can see that sequences $\{x_k\}$ and $\{y_k\}$ converges to the solution x_* and error estimates (14) and (15) are true for all $k \ge 0$.

Example 2. Consider function $F+G : D \subseteq \mathbb{R} \to \mathbb{R}^2$ given by:

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

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

where $\lambda, \mu \in \mathbb{R}$ are two parameters.

The unique solution of this problem is $x_* = 0$. Therefore, we can set constants α and β as follows:

$$\alpha = \sqrt{2}|\mu|, \; \beta = \frac{1}{\sqrt{2}}.$$

Let $D = \{x : |x| < 0.5\}$. Then

$$F'(x) = \begin{pmatrix} 1 \\ 3\lambda x^2 + 1 \end{pmatrix}, \quad F''(x) = \begin{pmatrix} 0 \\ 6\lambda x \end{pmatrix}$$

and

$$\|F'(x) - F'(y)\| = \left\| \begin{pmatrix} 0 \\ 3\lambda(x^2 - y^2) \end{pmatrix} \right\| = 3|\lambda||x+y||x-y| \le 3|\lambda||x-y|,$$

$$\|F''(x) - F''(y)\| = \left\| \begin{pmatrix} 0 \\ 6\lambda(x-y) \end{pmatrix} \right\| = 6|\lambda||x-y|;$$

$$G(x,y) = \begin{pmatrix} 0 \\ 0 \\ \dfrac{\lambda|x^2-1| - \lambda - \lambda|y^2-1| + \lambda}{x-y} \end{pmatrix} =$$

$$= \begin{pmatrix} 0 \\ 0 \\ \dfrac{\lambda(1-x^2-1) - \lambda(1-y^2)}{x-y} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -\lambda(x+y) \end{pmatrix}$$

and

$$\|G(x,y) - G(u,v)\| = \left\| \begin{pmatrix} 0 \\ -\lambda(x-u+y-v) \end{pmatrix} \right\| \le |\lambda|(|x-u| + |y-v|).$$

That is, we can set constants $L = 3|\lambda|$, $N = 6|\lambda|$, $M = |\lambda|$, $T = \dfrac{5|\lambda|}{2}$.

Then equation has form

$$\sqrt{2}|\lambda|r^2 + 50\sqrt{2}|\lambda|r + 40|\lambda\mu| - 8 = 0.$$

It has unique positive solution

$$r = \frac{\sqrt{5000|\lambda|^2 - 4\sqrt{2}\lambda(40|\lambda\mu| - 8)} - 50\sqrt{2}|\lambda|}{2\sqrt{2}|\lambda|}$$

if parameters λ and μ satisfy

$$\lambda \neq 0,\ |\lambda\mu| < \frac{1}{5}.$$

Let $x_0 = 0.2$, $y_0 = 0.2001$.

Tabl. 3. The results for $\lambda = 1$, $\mu = 0$.

k	$\rho(x_{k+1})$	The right side of (14)	$\rho(y_{k+1})$	The right side of (15)
0	1.406e-002	2.465e-001	1.681e-003	3.999e-002
1	1.027e-007	4.348e-005	2.225e-011	5.082e-007
2	1.323e-022	4.039e-018	1.223e-036	1.913e-022
3	0	2.860e-058	0	4.098e-067

Tabl. 4. The results for $\lambda = 0.5$, $\mu = 0.2$.

k	$\rho(x_{k+1})$	The right side of (14)	$\rho(y_{k+1})$	The right side of (15)
0	1.132e-002	2.106e-001	6.085e-003	1.622e-001
1	1.179e-005	4.482e-003	1.136e-005	4.420e-003
2	2.010e-011	5.788e-006	2.010e-011	5.788e-006
3	0	1.005e-011	0	1.005e-011

If $\lambda = 1$, $\mu = 0$ we obtain $2\sqrt{2}\alpha\beta^2 T = 0 < 1$, $\beta T \tau_0 = 0.70728355788 < 1$, $r = 0.1128822370012403$ and $B(x_*, r) \subset D$. If $\lambda = 0.5$, $\mu = 0.2$ we obtain $2\sqrt{2}\alpha\beta^2 T = 0.5 < 1$, $\beta T \tau_0 = 0.35364177894 < 1$, $r = 0.1128822370012403$ and $B(x_*, r) \subset D$.

Therefore, all conditions in Theorem 1 are satisfied for the two-step Gauss-Newton method (8). Hence, Theorem 1 applies.

6. Conclusions

We studied the local convergence of the method (8) for the nonlinear least squares problem with the decomposition of the operator under the classic Lipschitz conditions for the first- and second-order derivatives and for the divided differences of the first order. We determined the convergence order and the radius of the method (8) as well as proved the uniqueness ball of the solution of the nonlinear least squares problem (6). Furthermore, the method (8) has promising characteristics for parallelization, which we plan to utilize for constructing and developing new parallel methods for solving the problem (6).

Bibliography

1. Argyros I. K. Convergence and Applications of Newton-type Iterations / I. K. Argyros. – New York: Springer-Verlag, 2008.
2. Bartish M. Ia. About applications of a modification of the Gauss-Newton method / M. Ia. Bartish, A. I. Chypurko // Visnyk of Lviv Univ. Ser. Appl. Math. and Infor. – 1999. – Vol. 1. – P. 3–7 (in Ukrainian).
3. Bartish M. Ia. About one iterative method for solving functional equations / M. Ia. Bartish // Dopov. AN URSR. Ser. A. – 1968. – Vol. 30, № 5. – P. 387–391 (in Ukrainian).

4. Bartish M. Ia. About one modification of the Gauss-Newton method / M. Ia. Bartish, A. I. Chypurko, S. M. Shakhno // Visnyk of Lviv Univ. Ser. Mech. Math. – 1995. – Vol. 42. – P. 35–38 (in Ukrainian).
5. Catinas E. On some iterative methods for solving nonlinear equations / E. Catinas // Revue d'Analyse Numerique et de Theorie de l'Approximation. – 1994. – Vol. 23, № 1. – P. 47–53.
6. Chen J. Convergence of Gauss-Newton method's and uniqueness of the solution / J. Chen, W. Li // Applied Mathematics and Computation. – 2005. – Vol. 170. – P. 686–705.
7. Chong C. Convergence behavior of Gauss-Newton's method and extensions of the Smale point estimate theory / C. Chong, N. Hu, J. Wang // Journal of Complexity. – 2010. – Vol. 26. – P. 268–295.
8. Dennis J. M. Numerical Methods for Unconstrained Optimization and Nonlinear Equations / J. M. Dennis, R. B. Schnabel. – New York: Prentice-Hall, 1983.
9. Hernandez M. A. The Secant method for nondifferentiable operators / M. A. Hernandez, M. J. Rubio // Appl. Math. Lett. –2002. – Vol. 15. – P. 395–399.
10. Iakymchuk R. On the Convergence Analysis of a Two-Step Modification of the Gauss-Newton Method / R. P. Iakymchuk, S. M. Shakhno // PAMM. – 2014. – Vol. 14. – P. 813–814.
11. Iakymchuk R. Convergence analysis of a two-step modification of the Gauss-Newton method and its Applications / R. P. Iakymchuk, S. M. Shakhno // Journal of Numerical and Applied Mathematics. – 2017. – Vol. 126, №3. – P. 61–74.
12. Ren H. Local convergence of a secant type method for solving least squares problems / H. Ren, I. K. Argyros // AMC (Appl. Math. Comp.). – 2010. – Vol. 217. – P. 3816–3824.
13. Shakhno S. M. An iterative method for solving nonlinear least squares problems with nondifferentiable operator / S. M. Shakhno, R. P. Iakymchuk, H. P Yarmola // Mat. Stud. – 2017. – Vol. 48. – P. 97–107.
14. Shakhno S. M. Convergence analysis of combined method for solving nonlinear equations / S. M. Shakhno, I. V. Mel'nyk, H. P Yarmola // J. Math. Sci. – 2016. – Vol. 212. – P. 16–26.
15. Shakhno S. M. Convergence of the two-step combined method and uniqueness of the solution of nonlinear operator equationsS / S. M. Shakhno // Journal of Computational and Applied Mathematics. – 2014. – **261**. – P. 378–386.
16. Shakhno S. M. On a difference method with superquadratic convergence for solving nonlinear least squares problems / S. M. Shakhno, O. P. Gnatyshyn, R. P. Iakymchuk // Visnyk Lviv. Univ. Ser. Appl. Math. Inform. – 2007. – Vol. 13. – P. 51–58 (in Ukrainian).
17. Shakhno S. M. On an iterative algorithm of order 1.839... for solving the nonlinear least squares problems / S. M. Shakhno, O. P. Gnatyshyn // Applied Mathematics and Computation. – 2005. – Vol. 161. – P. 253–264.
18. Shakhno S. One combined method for solving nonlinear least squares problems / S. Shakhno, Yu. Shunkin // Visnyk Lviv. Univ. Ser. Appl. Math. Inform. – 2017. – Vol. 13. – P. 51–58 (in Ukrainian).
19. Shakhno S. M. On the two-step method for solving nonlinear equations with nondifferentiable operator / S. M. Shakhno, H. P Yarmola // Proc. Appl. Math. Mech. – 2012. – **1**. – P. 617–618.
20. Shakhno S. M. Two-step method for solving nonlinear equations with nondifferentiable operator / S. M. Shakhno, H. P Yarmola // Mat. Stud. – 2017. – Vol. 36, №2. – P. 213–220 (in Ukrainian).
21. Steward G. W. On the continuity of the generalized inverse / G. W. Steward // SIAM J. Appl. Math. – 1960. – Vol. 17, № 1. – P. 33–45.
22. Wedin P.-Å. Perturbation theory for pseudo-inverses / P.-Å. Wedin // BIT Numerical Mathematics. – 1973. – Vol. 13, № 2. – P. 217–232.

23. Werner W. Über ein Verfahren der Ordnung $1 + \sqrt{2}$ zur Nullstellenbestimmung / W. Werner // Numer. Math. – 1979. – Vol. 32. – P. 333–342.

Stepan Shakhno
and Halyna Yarmola
Faculty of Applied Mathematics and Informatics
Ivan Franko National University of Lviv
1, Universytets'ka Str., Lviv, 79000, Ukraine

Roman Iakymchuk
Department of Computational Science and Technology
School of Electrical Engineering and Computer Science
KTH Royal Institute of Technology
Lindstedtsvägen 5, 100 44 Stockholm, Sweden