MATHEMATICAL MODELING AND COMPUTING, Vol. 7, No. 2, pp. 248–258 (2020)

Local convergence analysis of the Gauss–Newton–Kurchatov method

Argyros I. K.[1], Shakhno S. M.[2], Yarmola H. P.[2]

[1] *Cameron University, Lawton, OK 73505, USA*
[2] *Ivan Franko National University of Lviv, 1 Universitetska Str., 79000, Lviv, Ukraine*

(Received 18 March 2020; Accepted 21 June 2020)

We present a local convergence analysis of the Gauss–Newton–Kurchatov method for solving nonlinear least squares problems with the decomposition of the operator. The method uses the sum of the derivative of the differentiable part of the operator and the divided difference of the nondifferentiable part instead of computing the full Jacobian. A theorem, which establishes the conditions of convergence, radius, and the convergence order of the proposed method, is proved [1]. However, the radius of convergence is small in general limiting the choice of initial points. Using tighter estimates on the distances, under weaker hypotheses [2], we provide an analysis of the Gauss–Newton–Kurchatov method with the following advantages over the corresponding results [1]: extended convergence region; finer error distances, and an at least as precise information on the location of the solution. The numerical examples illustrate the theoretical results.

Keywords: *Gauss–Newton–Kurchatov method, local convergence, Fréchet–derivative, Lipschitz condition, center Lipschitz condition, convergence domain.*

2010 MSC: 65F20, 65G99, 65H10, 49M15 **DOI:** 10.23939/mmc2020.02.248

1. Introduction

Let us consider the problem of finding an approximate solution of the nonlinear least squares problem

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

where the residual function $F : D \subseteq \mathbb{R}^p \to \mathbb{R}^m$, $m \geqslant p$ is nonlinear in x, F is continuously differentiable, and D is an open convex set in $\mathbb{R}^p$.

A large number of problems in applied mathematics and also in engineering are solved by finding the solutions of the problem (1). They are, for example, solving overdetermined systems of nonlinear equations, estimating parameters of physical processes by measurement results, constructing nonlinear regressions models for solving engineering, problems dynamic systems, etc. The used solution methods are iterative — when starting from one or several initial approximations, a sequence is constructed that converges to a solution of the problems (1).

The known methods of the Gauss–Newton type [3–6] are used to solve the problem (1), which have derivatives of function in their iterative formulas. However, in practice, problems with calculations of derivative arise. In this case, we can use iterative-difference methods [3, 7–10] that do not require the calculation of the matrix of derivatives and often are not inferior over the Gauss–Newton method at the order of convergence and the number of iterations. But sometimes the nonlinear function consists of differentiable and non-differentiable parts. Then a nonlinear least squares problem arises

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

where the residual function $F + G\colon D \subseteq \mathbb{R}^p \to \mathbb{R}^m$, $m \geqslant p$, is nonlinear in x, F is continuously differentiable, G is continuous function, differentiability of which, in general, is not assumed, and D is an open convex set in $\mathbb{R}^p$. Although it is possible to apply iterative-difference methods for solving a nonlinear problem (2), but it is also possible to construct iterative methods that take into account the decomposition of the residual function. In this case, when solving nonlinear equations, methods [11–16] were constructed as combinations of the Newton method [4, 5, 17, 18] and iterative-difference methods of chord (secant) and Kurchatov [3–5, 7, 8, 10, 19, 20].

In the paper [1], we proposed a method for solving a nonlinear problem of least squares with a non-differentiable operator (2) constructed on the basis of the Gauss–Newton method method [4, 5] and the Kurchatov type method [8, 10, 15]. We studied its local convergence under Lipschitz conditions and showed its effectiveness in comparison with other methods using test problems.

2. Preliminaries

To find a solution of the problem (2) we consider the Gauss–Newton–Kurchatov method [1]:

$$\begin{aligned} x_{n+1} &= x_n - (A_n^\top A_n)^{-1} A_n^\top \big(F(x_n) + G(x_n)\big), \\ A_n &= F'(x_n) + G(2x_n - x_{n-1}, x_{n-1}), \quad n = 0, 1, \ldots, \end{aligned} \tag{3}$$

where $F'(x_n)$ is matrix of Jacobi of $F(x)$; $G(2x_n - x_{n-1}, x_{n-1})$ is the divided difference of the first order of functions [21], and the points $2x_n - x_{n-1}$, x_{n-1}, x_0, x_{-1} are initial approximations. Method (3) is a combination of the Gauss–Newton method [4, 5] and the Kurchatov type method [8, 10, 15].

If $m = p$, method (3) reduces to the Newton–Kurchatov method for solving the nonlinear equation $F(x) + G(x) = 0$ [12, 13, 16]:

$$\begin{aligned} x_{n+1} &= x_n - A_n^{-1} \big(F(x_n) + G(x_n)\big), \\ A_n &= F'(x_n) + G(2x_n - x_{n-1}, x_{n-1}), \quad n = 0, 1, \ldots \end{aligned} \tag{4}$$

Setting in (3) $A_n = F'(x_n) + G(x_n, x_{n-1})$, we obtain a combination of the Gauss-Newton method [4, 5] and the Secant type method [7, 10] of the form [1]

$$\begin{aligned} x_{n+1} &= x_n - A_n^{-1} (F(x_n) + G(x_n)), \\ A_n &= F'(x_n) + G(x_n, x_{n-1}), \quad n = 0, 1, \ldots \end{aligned} \tag{5}$$

We need the following Lipschitz conditions.

Definition 1. *We say that the Fréchet derivative F' satisfies the center Lipschitz conditions on D, if there exist $L_0 > 0$ such that for each $x \in D$*

$$\left\| F'(x) - F'(x^*) \right\| \leqslant L_0 \left\| x - x^* \right\|, \tag{6}$$

where $x^* \in D$ solves problem (2).

Definition 2. *We say that divided differences $G(\cdot,\cdot)$ and $G(\cdot,\cdot,\cdot)$ satisfy the special Lipschitz conditions on $D \times D$ and $D \times D \times D$, if there exist $M_0 > 0$ and $N_0 > 0$ such that for each $x, y \in D$*

$$\| G(x, y) - G(u, v) \| \leqslant M_0 \big(\|x - u\| + \|y - v\| \big), \tag{7}$$

and

$$\| G(u, x, y) - G(v, x, y) \| \leqslant N_0 \|u - v\|. \tag{8}$$

Let $B > 0$ and $\alpha > 0$. Define function h on $[0, +\infty)$

$$h(t) = B \left[(2\alpha + (L_0 + 2M_0)t + N_0 t^2 \right] \left[(L_0/2 + M_0)t + N_0 t^2 \right]. \tag{9}$$

Suppose that equation $h(t) = 1$ has at least one positive solution. Denote by γ such the smallest solution. Set $D_0 = D \cap \Omega(x^*, \gamma)$.

Definition 3. *We say that the Fréchet derivative F' satisfies the restricted special Lipschitz conditions on D_0, if there exist $L > 0$ such that for each $x, y \in D_0$*

$$\left\|F'(x) - F'(y)\right\| \leqslant L \left\|x - y\right\| \tag{10}$$

Definition 4. *We say that divided differences $G(\cdot,\cdot)$ and $G(\cdot,\cdot,\cdot)$ satisfy the special Lipschitz conditions on $D_0 \times D_0$ and $D_0 \times D_0 \times D_0$, respectively, if there exist $M > 0$ and $N > 0$ such that for each $x, y, u, v \in D_0$*

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

and

$$\|G(u, x, y) - G(v, x, y)\| \leqslant N \|u - v\|. \tag{12}$$

The following condition together with (7) and (8) have been used instead of the preceding ones in the study of such iterative methods [15].

Definition 5. *We say that the Fréchet derivative F' satisfies the Lipschitz conditions on D, if there exist $L_1 > 0$ such that for euch $x, y \in D$*

$$\left\|F'(x) - F'(y)\right\| \leqslant L_1 \left\|x - y\right\| \tag{13}$$

Let $\Omega(x^*, 3r_*) = \{x\colon \|x - x^*\| < 3r_*\}$.

3. Convergence analysis of the iterative process (3)

Further, we improve Theorem 1 [1].

Theorem 1. *Let the function $F + G : \mathbb{R}^p \to \mathbb{R}^m$ be continuous on the open subset $D \subseteq \mathbb{R}^p$, F continuously differentiable in this domain, and let G be a continuous function. Assume that the problem (1) has a solution x^* in the domain and there exists the inverse operator $(A_*^\top A_*)^{-1} = [(F'(x^*)+G(x^*, x^*))^\top (F'(x^*) + G(x^*, x^*))]^{-1}$ and*

$$\|(A_*^\top A_*)^{-1}\| \leqslant B.$$

Estimates (6), (7), (8), (10), (11), (12) hold and γ given by (9) exists,

$$\|F(x^*) + G(x^*)\| \leqslant \eta, \quad \left\|F'(x^*) + G(x^*, x^*)\right\| \leqslant \alpha, \tag{14}$$

$$B(L + 2M)\eta < 1, \tag{15}$$

$$\Omega(x^*, 3r_*) \subseteq D,$$

where r_* is unique positive zero of the function q, given by

$$q(r) = B\left[\left(\alpha + (L + 2M)r + 4Nr^2\right)\left((L/2 + M)r + 4Nr^2\right) + (L + 2M + 4Nr)\eta\right] + B\left[2\alpha + (L_0 + 2M_0)r + 4N_0r^2\right]\left[(L_0 + 2M_0)r + 4N_0r^2\right] - 1. \tag{16}$$

Then for $x_0, x_{-1} \in \Omega(x^*, r_*)$ the iterative process (3) is well defined, the sequence $\{x_n\}$, $n = 0, 1, \ldots$, generated by it, remains in the open subset $\Omega(x^*, r_*)$, and converges to the solution x^*. Moreover, the following error estimates hold for $n = 0, 1, \ldots$

$$\|x_{n+1} - x^*\| \leqslant C_1 \|x_n - x^*\| + C_2 \|x_n - x_{n-1}\|^2 + C_3 \|x_n - x^*\|^2 + C_4 \|x_{n-1} - x^*\|^2 \|x_n - x^*\|, \tag{17}$$

where

$$g(r) = B\left[1 - B\left(2\alpha + (L_0 + 2M_0)r + 4N_0r^2\right)\left((L_0 + 2M_0)r + 4N_0r^2\right)\right]^{-1},$$
$$C_1 = g(r_*)(L + 2M)\eta, \quad C_2 = g(r_*)N\eta,$$
$$C_3 = g(r_*)(L/2 + M)\left(\alpha + (L + 2M)r_* + 4Nr_*^2\right), \tag{18}$$
$$C_4 = g(r_*)N\left(\alpha + (L + 2M)r_* + 4Nr_*^2\right).$$

Proof. According to the intermediate value theorem on $[0, r]$, the function q for a sufficiently large r and by (15) has a positive zero denoted by r_*. But $q'(r) \geqslant 0$ for $r \geqslant 0$. So, this root is the only one on $[0, r]$.

By assumption $x_0, x_{-1} \in \Omega(x^*, r_*)$. Then we have

$$\|2x_0 - x_{-1} - x^*\| \leqslant \|x_0 - x^*\| + \|x_0 - x_{-1}\| \leqslant \|x_0 - x^*\| + \|x_0 - x^*\| + \|x_{-1} - x^*\| < 3r_*.$$

So, $2x_0 - x_{-1} \in \Omega(x^*, 3r_*)$.

Let us denote $A_n = F'(x_n) + G(2x_n - x_{n-1}, x_{n-1})$. Let $n = 0$ and we will get this estimate:

$$\begin{aligned}\left\|I - \left(A_*^\top A_*\right)^{-1}A_0^\top A_0\right\| &= \left\|\left(A_*^\top A_*\right)^{-1}\left(A_*^\top A_* - A_0^\top A_0\right)\right\| \\ &= \left\|\left(A_*^\top A_*\right)^{-1}\left(A_*^\top(A_* - A_0) + (A_*^\top - A_0^\top)(A_0 - A_*) + (A_*^\top - A_0^\top)A_*\right)\right\| \\ &\leqslant \left\|(A_*^\top A_*)^{-1}\right\|\left(\left\|A_*^\top\right\|\left\|A_* - A_0\right\| + \left\|A_*^\top - A_0^\top\right\|\left\|A_0 - A_*\right\| + \left\|A_*^\top - A_0^\top\right\|\left\|A_*\right\|\right) \\ &\leqslant B(\alpha\left\|A_* - A_0\right\| + \left\|A_*^\top - A_0^\top\right\|\left\|A_0 - A_*\right\| + \alpha\left\|A_*^\top - A_0^\top\right\|).\end{aligned} \tag{19}$$

Using (8), we get

$$\begin{aligned}\|G(2x_0 - x_{-1}, x_{-1}) - G(x_0, x_0)\| &= \|G(2x_0 - x_{-1}, x_{-1}) - G(x_0, x_{-1}) + G(x_0, x_{-1}) - G(x_0, x_0)\| \\ &= \|G(2x_0 - x_{-1}, x_{-1}, x_0)(x_0 - x_{-1}) - G(x_0, x_{-1}, x_0)(x_0 - x_{-1})\| \\ &\leqslant N_0\|x_0 - x_{-1}\|^2\end{aligned} \tag{20}$$

and

$$\begin{aligned}\|G(2x_0 - x_{-1}, x_{-1}) - G(x_0, x^*)\| &= \|G(2x_0 - x_{-1}, x_{-1}) - G(x_0, x_0) + G(x_0, x_0) - G(x_0, x^*)\| \\ &\leqslant N_0\|x_0 - x_{-1}\|^2 + M_0\|x_0 - x^*\|.\end{aligned} \tag{21}$$

We use the inequalities (7), (20), (21):

$$\begin{aligned}\|A_0 - A_*\| &= \left\|\left(F'(x_0) + G(2x_0 - x_{-1}, x_{-1})\right) - \left(F'(x^*) + G(x^*, x^*)\right)\right\| \\ &= \left\|F'(x_0) - F'(x^*) + G(2x_0 - x_{-1}, x_{-1}) - G(x_0, x^*) + G(x_0, x^*) - G(x^*, x^*)\right\| \\ &\leqslant L\|x_0 - x^*\| + N\|x_0 - x_{-1}\|^2 + 2M\|x_0 - x^*\| \\ &= (L_0 + 2M_0)\|x_0 - x^*\| + N_0\|x_0 - x_{-1}\|^2.\end{aligned} \tag{22}$$

Then

$$\|A_0\| \leqslant \|A_0\| + \|A_0 - A_*\| \leqslant \alpha + (L_0 + 2M_0)\|x_0 - x^*\| + N_0\|x_0 - x_{-1}\|^2. \tag{23}$$

Then we obtain from the inequality (19) and the definition r_* (16)

$$\begin{aligned}\left\|I - (A_*^\top A_*)^{-1}A_0^\top A_0\right\| &\leqslant B\left[2\alpha + (L_0 + 2M_0)\|x_0 - x^*\| + N_0\|x_0 - x_{-1}\|^2\right] \\ &\quad\times\left[(L_0 + 2M_0)\|x_0 - x^*\| + N_0\|x_0 - x_{-1}\|^2\right] \\ &\leqslant B\left[2\alpha + (L_0 + 2M_0)r_* + 4N_0r_*^2\right]\left[(L_0 + 2M_0)r_* + 4N_0r_*^2\right] \\ &= h(r_*) < 1.\end{aligned} \tag{24}$$

According to the Banach's theorem on the inverse operator [5], there exists $(A_0^\top A_0)^{-1}$ and from (24) we have

$$\begin{aligned}\left\|(A_0^\top A_0)^{-1}\right\| \leqslant g_0 &= B\Big\{1 - B\left[2\alpha + (L_0+2M_0)\left\|x_0-x^*\right\| + N_0\left\|x_0-x_{-1}\right\|^2\right] \\ &\times \left[(L_0+2M_0)\left\|x_0-x^*\right\| + N_0\left\|x_0-x_{-1}\right\|^2\right]\Big\}^{-1} \\ &\leqslant g(r_*) = B\left\{1 - B\left[2\alpha + (L_0+2M_0)r_* + 4N_0r_*^2\right]\left[(L_0+2M_0)r_* + 4N_0r_*^2\right]\right\}^{-1}.\end{aligned}$$

Consequently, the iteration x_1 is well defined.

Then let us show that $x_1 \in \Omega(x^*, r_*)$. Using the equality

$$A_*^\top(F(x^*)+G(x^*)) = 0,$$

we will obtain an estimate

$$\begin{aligned}\|x_1 - x^*\| &= \left\|x_0 - x^* - \left(A_0^\top A_0\right)^{-1}\left(A_0^\top\left(F(x_0)+G(x_0)\right) - A_*^\top\left(F(x^*)+G(x^*)\right)\right\| \\ &\leqslant \left\|-(A_0^\top A_0)^{-1}\right\| \left\|\Big[-A_0^\top\Big(A_0 - \int_0^1 F'\big(x^*+t(x_0-x^*)\big)dt \right. \\ &\left. - G(x_0,x^*)\Big)(x_0-x^*) + (A_0^\top - A_*^\top)\big(F(x^*)+G(x^*)\big)\Big]\right\|.\end{aligned}$$

Hence, taking into account (21), (23) and inequalities

$$\begin{aligned}&\left\|A_0 - \int_0^1 F'\big(x^*+t(x_0-x^*)\big)dt - G(x_0,x^*)\right\| \\ &\qquad = \left\|F'(x_0) - \int_0^1 F'\big(x^*+t(x_0-x^*)\big)dt + G(2x_0-x_{-1},x_{-1}) - G(x_0,x^*)\right\| \\ &\qquad = \left\|\int_0^1 \big(F'(x_0) - F'(x^*+t(x_0-x^*))\big)dt + G(2x_0-x_{-1},x_{-1}) - G(x_0,x^*)\right\| \\ &\qquad \leqslant \frac{1}{2}L\left\|x_0-x^*\right\| + M\left\|x_0-x^*\right\| + N\left\|x_0-x_{-1}\right\|^2 \\ &\qquad \leqslant \frac{1}{2}L\left\|x_0-x^*\right\| + M\left\|x_0-x^*\right\| + N(\|x_0-x^*\| + \|x_{-1}-x^*\|)^2\end{aligned}$$

we will obtain

$$\begin{aligned}\|x_1-x^*\| \leqslant B\Big\{&\left(\alpha + (L+2M)\left\|x_0-x^*\right\| + N\left\|x_0-x_{-1}\right\|^2\right) \\ &\times \left(\frac{1}{2}L\left\|x_0-x^*\right\| + M\left\|x_0-x^*\right\| + N\left\|x_0-x_{-1}\right\|^2\right)\left\|x_0-x^*\right\| \\ &+ \eta\left((L+2M)\left\|x_0-x^*\right\| + N\left\|x_0-x_{-1}\right\|^2\right)\Big\} \\ &\times \Big\{1 - B\left[2\alpha + (L+2M)\left\|x_0-x^*\right\| + N\left\|x_0-x_{-1}\right\|^2\right] \\ &\times \left((L+2M)\left\|x_0-x^*\right\| + N\left\|x_0-x_{-1}\right\|^2\right)\Big\}^{-1} \\ = g_0\Big\{&\left(\alpha + (L+2M)\left\|x_0-x^*\right\| + N\left\|x_0-x_{-1}\right\|^2\right) \\ &\times \left(\frac{1}{2}L\left\|x_0-x^*\right\| + M\left\|x_0-x^*\right\| + N\left\|x_0-x_{-1}\right\|^2\right)\left\|x_0-x^*\right\| \\ &+ \eta\left((L+2M)\left\|x_0-x^*\right\| + N\left\|x_0-x_{-1}\right\|^2\right)\Big\}\end{aligned}$$

$$< g(r_*)\big[\,\big(\alpha + (L+2M)r_* + 4Nr_*^2\big)\,\big((L/2+M)r_* + 4Nr_*^2\big) \\ + (L+2M+4Nr_*)\eta\big]r_* = p(r_*)r_* = r_*,$$

where

$$p(r) = g(r)\,[(\alpha + (L+2M)r + 4Nr^2)((L/2+M)r + 4Nr^2) + (L+2M+4Nr)\eta\,].$$

Hence, $x_1 \in \Omega(x^*, r_*)$ and inequality (16) is true for $n = 0$.

Assume that $x_n \in \Omega(x^*, r_*)$ for $n = 0, 1, \dots, k$, and the estimate (17) for $n = 0, 1, \dots, k-1$, where $k \geqslant 1$ is an integer, holds. Further, we prove that $x_{n+1} \in \Omega(x^*, r_*)$, and the estimate (17) holds for $n = k$.

Define

$$\begin{aligned}\big\|I - (A_*^\top A_*^\top)^{-1}A_k^\top A_k\big\| &= \big\|(A_*^\top A_*)^{-1}(A_*^\top A_* - A_k^\top A_k)\big\| \\ &= \big\|(A_*^\top A_*)^{-1}(A_*^\top(A_* - A_k) + (A_*^\top - A_k^\top)(A_k - A_*) + (A_*^\top - A_k^\top)A_*)\big\| \\ &\leqslant \|(A_*^\top A_*)^{-1}\|\big(\|A_*^\top\|\|A_* - A_k\| + \|A_*^\top - A_k^\top\|\|A_k - A_*\| + \|A_*^\top - A_k^\top\|\|A_*\|\big) \\ &\leqslant B\big(\alpha\|A_* - A_k\| + \|A_*^\top - A_k^\top\|\|A_k - A_*\| + \alpha\|A_*^\top - A_k^\top\|\big) \\ &\leqslant B\left[2\alpha + (L+2M)\|x_k - x^*\| + N\|x_k - x_{k-1}\|^2\right] \\ &\quad \times \left[(L/2+M)\|x_k - x^*\| + N\|x_k - x_{k-1}\|^2\right] \\ &\leqslant B\left[2\alpha + (L+2M)r_* + 4Nr_*^2\right]\left[(L+2M)r_* + 4Nr_*^2\right] = h(r_*) < 1.\end{aligned}$$

Thus, $(A_k^\top A_k)^{-1}$ exists and

$$\begin{aligned}\big\|(A_{k+1}^\top A_{k+1})^{-1}\big\| \leqslant g_k = B\Big\{1 - B\left[2\alpha + (L_0 + 2M_0)\,\|x_k - x^*\| + N_0\,\|x_k - x_{k-1}\|^2\right] \\ \times\left[(L_0/2 + M_0)\,\|x_k - x^*\| + N_0\,\|x_k - x_{k-1}\|^2\right]\Big\}^{-1} \leqslant g(r_*).\end{aligned}$$

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

$$\begin{aligned}\|x_{k+1} - x^*\| &= \Big\|x_k - x^* - (A_k^\top A_k)^{-1}(A_k^\top\big(F(x_k) + G(x_k)\big) - A_*^\top\big(F(x^*) + G(x^*)\big)\Big\| \\ &\leqslant \|-(A_k^\top A_k)^{-1}\|\,\Big\|-A_k^\top\Big(A_k - \int_0^1 F'\big(x^* + t(x_k - x^*)\big)dt - G(x_k, x^*)\Big)(x_k - x^*)\Big\| \\ &\quad + \|-(A_k^\top A_k)^{-1}\|\|(A_k^\top - A_*^\top)\big(F(x^*) + G(x^*)\big)\| \\ &\leqslant g_k\Big\{\left[\alpha + (L+2M)\,\|x_k - x^*\| + N\,\|x_k - x_{k-1}\|^2\right] \\ &\quad \times\left[(L/2+M)\,\|x_k - x^*\| + N\,\|x_k - x_{k-1}\|^2\right]\|x_k - x^*\| \\ &\qquad + \eta\Big((L+2M)\,\|x_k - x^*\| + N\,\|x_k - x_{k-1}\|^2\Big)\Big\} \\ &\leqslant g(r_*)\Big\{\left[\alpha + (L+2M)\,\|x_k - x^*\| + N\,\|x_k - x_{k-1}\|^2\right] \\ &\quad \times\left[(L/2+M)\,\|x_k - x^*\| + N\,\|x_k - x_{k-1}\|^2\right]\|x_k - x^*\| \\ &\qquad + \eta\Big((L+2M)\,\|x_k - x^*\| + N\,\|x_k - x_{k-1}\|^2\Big)\Big\} < p(r_*)r_* = r_*,\end{aligned}$$

i.e. $x_{k+1} \in \Omega(x^*, r_*)$, and estimate (17) holds for $n = k$.

Consequently, the iterative process (3) is well defined, $x_n \in \Omega(x^*, r_*)$ for all $n \geqslant 0$, and estimate (17) holds for all $n \geqslant 0$.

Further, we prove that $x_n \to x^*$ for $n \to \infty$. Define functions a and b on $[0, r_*]$:

$$a(r) = g(r)((L + 2M + 3Nr)\eta + \varphi(r)((L/2 + M)r + 4Nr^2)), \quad b(r) = g(r)Nr\eta, \tag{25}$$

where $\varphi(r) = \alpha + (L + 2M)r + 4Nr^2$.

According to the choice r_*, we have

$$a(r_*) \geqslant 0, \quad b(r_*) \geqslant 0, \quad a(r_*) + b(r_*) = 1. \tag{26}$$

Using the estimate (17), the definition of constants C_i, $i = 1, 2, 3, 4$, as well as the functions a and b, for $n \geqslant 0$, we obtain

$$\begin{aligned} \|x_{n+1} - x^*\| &\leqslant \left(C_1 + C_3 r + 4C_4 r_*^2\right) \|x_n - x^*\| \\ &\quad + C_2 \left(\|x_n - x^*\|^2 + 2\|x_{n-1} - x^*\| \|x_n - x^*\| + \|x_{n-1} - x^*\|^2\right) \\ &< \left(C_1 + 3C_2 r_* + C_3 r_* + 4C_4 r_*^2\right) \|x_n - x^*\| + C_2 r_* \|x_{n-1} - x^*\| \\ &= a(r_*) \|x_n - x^*\| + b(r_*) \|x_{n-1} - x^*\|. \end{aligned} \tag{27}$$

Similarly to [8], we prove that under the conditions (25), (26) the sequence $\{x_n\}$ for $n \to \infty$ converges to x^*.

First of all, for a real number $r_* > 0$ and initial points $x_0, x_{-1} \in \Omega(x^*, r_*)$ there exists a real number r' such that $0 < r' < r_*$, $x_0, x_{-1} \in \Omega(x^*, r')$. Then all the above estimates for the sequence $\{x_n\}$ are valid, if replaced r_* by r'. In particular, from (27) for $n \geqslant 0$, we obtain

$$\|x_{n+1} - x^*\| \leqslant a \|x_n - x^*\| + b \|x_{n-1} - x^*\|, \tag{28}$$

where $a = a(r')$, $b = b(r')$.

Clearly, we also have

$$a \geqslant 0, \quad b \geqslant 0, \quad a + b < a(r_*) \|x_n - x^*\| + b(r_*) \|x_{n-1} - x^*\| < 1.$$

Define sequences $\{\theta_n\}$, $\{\rho_n\}$:

$$\begin{gathered} \theta_n = \frac{\|x_n - x^*\|}{r'}, \quad n = -1, 0, 1, \ldots, \\ \rho_{-1} = \theta_{-1}, \quad \rho_0 = \theta_0, \quad \rho_{n+1} = a\rho_n + b\rho_{n-1}, \quad n = 0, 1, 2, \ldots. \end{gathered} \tag{29}$$

We divide the two parts of inequality (28) into r' and obtain $\theta_{n+1} = a\theta_n + b\theta_{n-1}$, $n = 0, 1, 2, \ldots$.

By definition of the sequence $\{\rho_n\}$, we have

$$0 \leqslant \theta_n \leqslant \rho_n, \quad n = -1, 0, 1, \ldots. \tag{30}$$

For the sequence $\{\rho_n\}$, the explicit formulas are known

$$\rho_n = \omega_1 \lambda_1^n + \omega_2 \lambda_2^n, \quad n = -1, 0, 1, \ldots, \tag{31}$$

where

$$\lambda_1 = \frac{a - \sqrt{a^2 + 4b}}{2}, \quad \lambda_2 = \frac{a + \sqrt{a^2 + 4b}}{2}$$

and

$$\omega_1 = \frac{\lambda_2^{-1}\rho_0 - \rho_{-1}}{\lambda_2^{-1} - \lambda_1^{-1}}, \quad \omega_2 = \frac{\rho_{-1} - \lambda_1^{-1}\rho_0}{\lambda_2^{-1} - \lambda_1^{-1}}.$$

Note that
$$0 \leqslant |\lambda_1| \leqslant |\lambda_2| < \frac{a+\sqrt{a^2+4(1-a)}}{2} = \frac{a+2-a}{2} = 1.$$
Taking into account (30) and (31), we conclude that $\{\theta_n\} \to 0$ as $n \to \infty$. Therefore, we conclude that $x_n \to x^*$ as $n \to \infty$. ■

Remark 1. If $L_0 = L = L_1$, $M_0 = M$ and $N_0 = N$, our results are specialized to the corresponding ones [1]. Otherwise they constitute an improvement. As an example let q_1, g_1, C_1^1, C_2^1, C_3^1, C_4^1, r_*^1 used in [1], denote the functions and parameters, where L_0, L, M, N are replaced by L_1, L_1, M_0, N_0, respectively. Then, since $L_0 \leqslant L_1$, $L \leqslant L_1$, $M \leqslant M_0$, $N \leqslant N_0$ and since $D_0 \subseteq D$, we have $q(r) \leqslant q_1(r)$, $g(r) \leqslant g_1(r)$, $C_1 \leqslant C_1^1$, $C_2 \leqslant C_2^1$, $C_3 \leqslant C_3^1$, $C_4 \leqslant C_4^1$, so $r_*^1 \leqslant r_*$, and the new error bounds are tighter than the corresponding ones (23) [1] .

Moreover, we have
$$B(L_0+2M_0)\eta < 1 \quad \Rightarrow \quad B(L+2M)\eta < 1$$
but not vice versa, unless if $L_0 = L$ and $M_0 = M$.

Hence, the new sufficient convergence criteria for method (3) are weaker. These advantages are obtained under the same computational cost as [1], since in practice the new constants are special cases of the previous ones.

Corollary 1. *In the case of zero residual, the convergence order of the iterative process (3) is quadratic.*

If $\eta = 0$, we have a nonlinear least squares problem with zero residual in the solution. Then the constants $C_1 = 0$ and $C_2 = 0$ and (17) reduces to
$$\|x_{n+1} - x^*\| \leqslant C_3 \|x_n - x^*\|^2 + C_4 \|x_n - x_{n-1}\|^2 \|x_n - x^*\| . \tag{32}$$
It follows from the inequality (32) that the order of convergence (3) is not higher than quadratic. Consequently, there exist a constant $C_5 \geqslant 0$ and a positive integer N such that for all $n \geqslant N$
$$\|x_n - x^*\| \geqslant C_5 \|x_{n-1} - x^*\|^2 .$$
By
$$\|x_n - x^*\| \leqslant \|x_{n-1} - x^*\| ,$$
we have
$$\|x_n - x_{n-1}\|^2 \leqslant \big(\|x_n - x^*\| + \|x_{n-1} - x^*\| \big)^2 \leqslant 4 \|x_{n-1} - x^*\|^2 ,$$
and from (32) we have
$$\begin{aligned} \|x_{n+1} - x^*\| &\leqslant C_3 \|x_n - x^*\|^2 + 4C_4 \|x_{n-1} - x^*\|^2 \|x_n - x^*\| \\ &\leqslant C_3 \|x_n - x^*\|^2 + 4\frac{C_4}{C_5} \|x_n - x^*\|^2 = C_6 \|x_n - x^*\|^2 . \end{aligned}$$

Consequently, the convergence order of the iterative process (3) is quadratic.

As we see from the estimates (17) and (18), the convergence of the iterative process (3) essentially depends on the terms containing the values η, α, L, M and N.

For problems with zero residual in the solution ($\eta = 0$), the quadratic convergence of the iterative process (3) is established.

For problems with a small residual in the solution (η is "small") and with weak nonlinearity (α, L_0, L, M and N are "small"), the convergence of the iterative process is linear. In the case of large residual (η is "large") or for strongly nonlinear problems (α, L_0, L, M and N are "large"), the iterative process (3) cannot converge at all.

4. Results of numerical experiment

On several test cases, we compare the convergence rates of the Gauss-Newton-Kurchatov method (3), the Gauss-Newton-Secant method (5) and the Secant-type difference method [7, 10]

$$\begin{gathered} x_{n+1} = x_n - (A_n^\top A_n)^{-1} A_n^\top \big(F(x_n) + G(x_n)\big), \\ A_n = F(x_n, x_{n-1}) + G(x_n, x_{n-1}), \quad n = 0, 1, \ldots, \end{gathered} \tag{33}$$

and the Kurchatov–type difference method [8, 10]

$$\begin{gathered} x_{n+1} = x_n - (A_n^\top A_n)^{-1} A_n^\top \big(F(x_n) + G(x_n)\big), \\ A_n = F(2x_n - x_{n-1}, x_{n-1}) + G(2x_n - x_{n-1}, x_{n-1}), \quad n = 0, 1, \ldots. \end{gathered} \tag{34}$$

We tested methods on nonlinear systems with a non-differentiable operator with zero and non-zero residual. The classical Gauss-Newton method and the Newton method cannot be applied to solve these problems.

Solution results are of the accuracy $\varepsilon = 10^{-8}$. The additional approximation was chosen as follows: $x_{-1} = x_0 - 10^{-4}$. The calculations were carried out until the conditions were fulfilled

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

with $f(x) = \min\limits_{x \in \mathrm{R}^p} \frac{1}{2}(F(x) + G(x))^\top (F(x) + G(x))$.

Example 1. [3, 11, 14], $p = 2$, $m = 2$:

$$\begin{cases} 3x_1^2 x_2 + x_2^2 - 1 + |x_1 - 1|, \\ x_1^4 + x_1 x_2^3 - 1 + |x_2|, \end{cases}$$

$$(x_1^*, x_2^*) \approx (0.89465537,\ 0.32782652), \quad f(x^*) = 0.$$

Example 2. $p = 2$, $m = 3$:

$$\begin{cases} 3x_1^2 x_2 + x_2^2 - 1 + |x_1 - 1|, \\ x_1^4 + x_1 x_2^3 - 1 + |x_2|, \\ \left|x_1^2 - x_2\right|, \end{cases}$$

$$(x_1^*, x_2^*) \approx (0.74862800,\ 0.43039151), \quad f(x^*) \approx 4.0469349 \cdot 10^{-2}.$$

Example 3. $p = 3$, $m = 10$:

$$\begin{cases} F_i(x) + G_i(x) = e^{-r_i x_1} - e^{-r_i x_2} - x_3(e^{-r_i} - e^{-10 r_i}) - |r_i(x_1 + x_2) - \sqrt{1 - r_i x_3}|, \\ r_i = 0.1i, \quad i = 1, \ldots, m, \end{cases}$$

$$(x_1^*, x_2^*, x_3^*) \approx (-1.10717473,\ 4.62615368,\ 0.50038327), \quad f(x^*) \approx 1.79449337 \cdot 10^{-3}.$$

Table 1 shows the results of the numerical experiment. In particular, the investigated methods are compared in terms of the number of iterations performed to find a solution with a given accuracy.

Table 1. Number of iterations for solving of the test problems.

Example	x_0	Kurchatov–type m-d (34)	Gauss–Newton–Kurchatov m-d (3)	Secant–type m-d (33)	Gauss–Newton–Secant m-d (5)
1	(1, 0.1)	6	5	7	5
	(3, 1)	12	9	12	10
	(0.5, 0.5)	12	10	15	10
2	(1, 0.1)	17	14	31	11
	(3, 1)	23	18	44	15
	(0.5, 0.5)	17	14	24	13
3	(-0.9, 4.3, 0.4)	6	6	13	7
	(-0.5, 4, 1.5)	23	10	39	14
	(-3, 5.8, 2.5)	17	10	19	11

5. Conclusions

It follows from the theoretical results, practical calculations, and comparison of the results obtained that the combined differential-difference (3) and (5) methods converge faster than the Kurchatov type method (34) and the Secant type method (33). As it has been proved, in the case of zero residual, the method (3) has a quadratic order of convergence and does not require the calculation of derivatives from a non-differentiable part of the operator. Then the method (3), as well as the method (5), are effective methods for solving nonlinear least squares problems with non-differentiable operator.

[1] Shakhno S. M. Gauss–Newton–Kurchatov method for solving nonlinear least squares problems. Mat. metody phys.-mech. polia. **60**, 52–62 (2017), (in Ukrainian).

[2] Argyros I. K., Hilout S. On an improved convergence analysis of Newton's method. Applied Mathematics and Computation. **225**, 372–386 (2013).

[3] Argyros I. K. Convergence and applications of Newton–type iterations. Springer–Verlag, New York (2008).

[4] Dennis J. E. (Jr.) Schnabel R. B. Numerical methods for unconstrained optimization and nonlinear equations. SIAM, Philadelphia (1996).

[5] Ortega J. M., Rheinboldt W. C. Iterative solution of nonlinear equations in several variables. Academic Press, New York etc. (1970).

[6] Shakhno S. Some numerical methods for nonlinear least squares problems. In: Alefeld G., Rohn J., Rump S., Yamamoto T. (eds.) Symbolic-algebraic Methods and Verification Methods, Springer, Vienna. Pp.235–243 (2001).

[7] Ren H., Argyros I. K. Local convergence of a secant type method for solving least squares problems. Appl. Math. Comp. **217** (8), 3816–3824 (2010).

[8] Ren H., Argyros I. K., Hilout S. A derivative free iterative method for solving least squares problems. Numer. Algor. **58**, 555–571 (2011).

[9] Shakhno S. M., Gnatyshyn O. P. Iterative-difference methods for solving nonlinear least-squares problem. In: Arkeryd L., Bergh J., Brenner P. et al. (eds.) Progress in Industrial Mathematics at ECMI 98, B. G. Teubner, Stuttgart. Pp. 287–294 (1999).

[10] Shakhno S. M., Gnatyshyn O. P. On an iterative algorithm of order 1.839 . . . for solving the nonlinear least squares problems. Appl. Math. Comp. **161** (1), 253–264 (2005).

[11] Cătinaş E. On some iterative methods for solving nonlinear equations. Rev. Anal. Numér. Théor. Approx. **23** (1), 47–53 (1994).

[12] Hernández-Verón M. A., Rubio M. J. On the local convergence of Newton–Kurchatov–type method for non-differentiable operators. Appl. Math. Comp. **304**, 1–9 (2017).

[13] Iakymchuk R., Shakhno S., Yarmola H. Combined Newton-Kurchatov method for solving nonlinear operator equations. Proc. Appl. Math. Mech. **16**, 719–720 (2016).

[14] 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. **201**, 32–43 (2014).

[15] Shakhno S. M., Yarmola H. P. Two-point method for solving nonlinear equation with nondifferentiable operator. Matematychni Studii. **36**, 213–220 (2011), (in Ukrainian).

[16] Shakhno S. M. Combined Newton–Kurchatov method under the generalized Lipschitz conditions for the derivatives and divided differences. J. Comp. Appl. Math. **2** (119), 78–89 (2015).

[17] Magrenan A. A., Argyros I. K. A contemporary study of iterative methods. Convergence, Dynamics and applications. Academic Press, London (2018).

[18] Deuflhard P. Newton methods for nonlinear problems. Affine invariance and adaptive algorithms. Springer–Verlag, Berlin (2004).

[19] Shakhno S. M. On the difference method with quadratic convergence for solving nonlinear operator equations. Matematychni Studii. **26**, 105–110 (2006), (in Ukrainian).

[20] Shakhno S. Secant method under the generalized Lipschitz conditions for the first-order divided differences. Mathematical bulletin of the Shevchenko scientific society. **4**, 293–303 (2007), (in Ukrainian).

[21] Ulm S. On generalized divided differences. I, II. Izv. Akad. Nauk Est. SSR. Fiz. mat. **16**, 13–26, 146–156 (1967), (in Russian).

Аналіз локальної збіжності методу Гаусса–Ньютона–Курчатова

Аргирос І. К.[1], Шахно С. М.[2], Ярмола Г. П.[2]

[1] *Університет Кемерона,*
Лоутон, OK 73505, США
[2] *Львівський національний університет імені Івана Франка,*
вул. Університетська, 1, 79000, Львів, Україна

У роботі представлено аналіз локальної збіжності методу Гаусса–Ньютона–Курчатова для розв'язання нелінійних задач про найменші квадрати з декомпозицією оператора. Метод використовує суму похідної від диференційовної частини оператора і поділену різницю від недиференційовної частини замість обчислення повного якобіана. Доведено теорему, яка встановлює умови, радіус та порядок збіжності методу, запропонованого у [1]. Однак радіус збіжності, в загальному випадку, невеликий, що обмежує вибір початкових точок. Використовуючи більш чіткі оцінки похибок при слабших гіпотезах [2], наведено аналіз методу Гаусса–Ньютона–Курчатова з такими перевагами перед відповідними результатами у [1]: ширша область збіжності, точніші оцінки похибок і, принаймні, точніша інформація про місце розташування точного розв'язку. Чисельні приклади підтерджують теоретичні результати.

Ключові слова: *метод Гаусса–Ньютона–Курчатова, локальна збіжність, похідна Фреше, умова Ліпшиця, центральна умова Ліпшиця, область збіжності.*