Журнал обчислювальної та прикладної математики | **2017** **№ 3 (126)** | Journal of Numerical & Applied Mathematics

UDC 519.6

CONVERGENCE ANALYSIS OF A TWO-STEP MODIFICATION OF THE GAUSS-NEWTON METHOD AND ITS APPLICATIONS

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

Резюме. У роботі досліджено збіжність двокрокової модифікації методу Гаусса-Ньютона за узагальнених умов Ліпшиця для похідних першого і другого порядків. Встановлено порядок і радіус збіжності методу, а також область єдиності розв'язку нелінійної задачі про найменші квадрати. Проведено чисельні експерименти на відомих тестових задачах.

Abstract. We investigate the convergence of a two-step modification of the Gauss-Newton method applying the generalized Lipschitz condition for the first- and second-order derivatives. 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 well-known test problems.

1. Introduction

Let us consider the nonlinear least squares problem [6]:

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

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 [4, 5]. In paper [11], 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.

For solving the problem (1), we consider a two-step modification of the Gauss-Newton method [1, 3]

$$\left\{ \begin{array}{l} x_{k+1} = x_k - [F^{'}(z_k)^T F^{'}(z_k)]^{-1} F^{'}(z_k)^T F(x_k), \\ y_{k+1} = x_{k+1} - [F^{'}(z_k)^T F^{'}(z_k)]^{-1} F^{'}(z_k)^T F(x_{k+1}),\ k = 0, 1, 2, ..., \end{array} \right. \tag{3}$$

Key words. Least squares problem, Gauss-Newton method, Lipschitz conditions with L average, radius of convergence, uniqueness ball.

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 [2] and Werner [17]. On each iteration, the method (3) computes the inversion of the matrix $[F^{'}(z_k)^T F^{'}(z_k)]^{-1}$ only once. Because of that, the computation cost of each iteration of the method (3) is roughly the same as of the Gauss-Newton method (2): for calculating y_{k+1}, it is only necessary to perform one backward substitution, which requires $O(n^2)$ floating-point operations (Flops), since the LL^T decomposition of the matrix $F^{'}(z_k)^T F^{'}(z_k)$, which costs $O(n^3)$ ($O(n^3/3)$ to be precise) Flops [6], is computed for x_{k+1}.

The main goal of this paper is to analyze the local convergence of the method (3). Bartish et al. [1] examined the local convergence of this method using the classical Lipschitz condition for derivatives of the second-order, but only for the problem (1) with zero residuals. Instead, we study the convergence of the above-mentioned method using the generalized Lipschitz conditions [15] for derivatives of the first- and second-orders; such conditions employ an integrable function $L(u)$ instead of the Lipschitz constant L. The Lipschitz condition with L average in the inscribe sphere makes us unify the convergence criteria containing the Kantorovich theorem and the Smale α-theory [5, 8, 12, 14, 15]. We prove the convergence of the method (3) for the problem (1) with zero as well as non-zero residuals. Furthermore, we find both the order and the radius of the convergence of the method (3) as well as the uniqueness ball of the solution of the problem (1). We have published some of the results without proofs as an extended abstract [7].

2. Preliminaries

For our study, we present different definitions of the Lipschitz conditions. Let us denote $B(x_*, r) = \{x \in D \subseteq \mathbb{R}^n : \|\mathrm{x} - \mathrm{x}_*\| \leq \mathrm{r}\}$ as an closed ball with the radius r ($r > 0$) at x_*.

Definition 1. The function $F : \mathbb{R}^n \to \mathbb{R}^m$ satisfies the classical Lipschitz condition on $B(x_*, r)$ if

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

where $x, y \in B(x_*, r)$ and L is the Lipschitz constant.

In Definition 1 L may not necessary be a constant, but it also can be an integrable function $L(u)$.

Definition 2 ([15])**.** The function $F : \mathbb{R}^n \to \mathbb{R}^m$ satisfies the Lipschitz condition with L average on $B(x_*, r)$ if

$$\|F(x) - F(y)\| \leq \int_0^{\|x-y\|} L(u)du, \quad \forall x \in B(x_*, r),$$

where $L(u)$ is a positive non-decreasing function.

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 [6] is defined as $A^{\dagger} = (A^T A)^{-1} A^T$.

Lemma 1 ([13,16]). *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 ([4]). *Let* $A, E \in \mathbb{R}^{m\times n}$. *Assume that* $C = A + E$, $\|EA^{\dagger}\| < 1$, *and* $rank(A) = n$, *then* $rank(C) = n$.

Lemma 3 ([15]). *Let* $h(t) = \dfrac{1}{t}\displaystyle\int_0^t L(u)du$, $0 \leq t \leq r$, *where* $L(u)$ *is a positive integrable function and monotonically non-decreasing on* $[0, r]$. *Then,* $h(t)$ *is monotonically non-decreasing with respect to* t.

Lemma 4 ([10]). *Let* $g(t) = \dfrac{1}{t^3}\displaystyle\int_0^t N(u)(t-u)^2du$, $0 \leq t \leq r$, *where* $N(u)$ *is a positive integrable function and monotonically non-decreasing on* $[0, r]$. *Then,* $g(t)$ *is monotonically non-decreasing with respect to* t.

3. Local Convergence Analysis of Method (3)

In this section, we investigate the convergence and the radius of the convergence ball of the method (3).

Theorem 1. *Let* $F : \mathbb{R}^n \to \mathbb{R}^m$, $m \geq n$, *be a twice Fréchet differentiable operator on a subset* $D \subseteq \mathbb{R}^n$. *Assume that the problem (1) has a solution* $x_* \in D$ *and a Fréchet derivative* $F'(x_*)$ *has full rank. Suppose that Fréchet derivatives* $F'(x)$ *and* $F''(x)$ *on* $B(x_*, R) = \{x \in D : \|x - x_*\| \leq R\}$ *satisfy the Lipschitz conditions with* L *and* N *average:*

$$\|F'(x) - F'(y)\| \leq \int_0^{\|x-y\|} L(u)du, \tag{4}$$

$$\|F''(x) - F''(y)\| \leq \int_0^{\|x-y\|} N(u)du, \tag{5}$$

where L *and* N *are positive non-decreasing functions on* $[0, 3R/2]$.

Furthermore, assume function

$$\begin{aligned} h_0(p) \quad = \quad & (\beta/8)\int_0^p N(u)(p-u)^2du + \beta p\Big(\int_0^{(3/2)p} L(u)du + \int_0^p L(u)du\Big) + \\ & +\sqrt{2}\alpha\beta^2\int_0^p L(u)du - p \end{aligned} \tag{6}$$

has a minimal zero r *on* $[0, R]$, *which also satisfies*

$$\beta\int_0^r L(u)du < 1. \tag{7}$$

Then, for all $x_0, y_0 \in B(x_*, r)$ *the sequences* $\{x_k\}$ *and* $\{y_k\}$*, which are generated by the method (3), are well defined, remain in* $B(x_*, r)$ *for all* $k \geq 0$*, and converge to* x_* *such that*

$$\rho(x_{k+1}) \leq \gamma\rho(x_k)^3 + \eta\rho(x_k)\rho(y_k) + \theta\rho(z_k), \tag{8}$$

$$\rho(y_{k+1}) \leq \gamma\rho(x_{k+1})^3 + (\eta/3)(\rho(x_k) + \rho(y_k) + \rho(x_{k+1}))\rho(x_{k+1}) + +\theta\rho(z_k), \tag{9}$$

$$r_{k+1} = \max\{\rho(x_{k+1}), \rho(y_{k+1})\} \leq q r_k \leq \cdots \leq q^{k+1} r_0, \tag{10}$$

where $\rho(x) = \|x - x_*\|$*,* $r_0 = \max\{\rho(x_0), \rho(y_0)\}$*,*

$$q = \gamma\rho(x_0)^2 + \theta + \eta, \tag{11}$$

$$\gamma = \frac{\beta\int_0^{\rho(x_0)} N(u)(\rho(x_0) - u)^2 du}{8\rho(x_0)^3\left(1 - \beta\int_0^{\rho(z_0)} L(u)du\right)}, \ \theta = \frac{\sqrt{2}\alpha\beta^2\int_0^{\rho(z_0)} L(u)du}{\rho(z_0)\left(1 - \beta\int_0^{\rho(z_0)} L(u)du\right)}, \tag{12}$$

$$\eta = \frac{\beta\int_0^{\rho(x_0)+\rho(y_0)/2} L(u)du}{(2\rho(x_0) + \rho(y_0))/3\left(1 - \beta\int_0^{\rho(z_0)} L(u)du\right)}, \tag{13}$$

$$\alpha = \|F(x_*)\|, \quad \beta = \|(F'(x_*)^T F'(x_*))^{-1} F'(x_*)^T\|. \tag{14}$$

Proof. Let choose arbitrary $x_0, y_0 \in B(x_*, r)$. For x_1, y_1 that are generated by (3), we have

$$x_1 - x_* = x_0 - x_* - \left[F'(z_0)^T F'(z_0)\right]^{-1} F'(z_0)^T F(x_0) =$$
$$= \left[F'(z_0)^T F'(z_0)\right]^{-1} F'(z_0)^T \left[F'(z_0)(x_0 - x_*) - F(x_0) + F(x_*)\right] +$$
$$+ \left[F'(x_*)^T F'(x_*)\right]^{-1} F'(x_*)^T F(x_*) - \left[F'(z_0)^T F'(z_0)\right]^{-1} F'(z_0)^T F(x_*) =$$
$$= \left[F'(z_0)^T F'(z_0)\right]^{-1} F'(z_0)^T \times$$
$$\times \left[\left(F'\left(\frac{x_0 + x_*}{2}\right)(x_0 - x_*) - F(x_0) + F(x_*)\right) + \right.$$
$$\left. + \left(F'(z_0) - F'\left(\frac{x_0 + x_*}{2}\right)\right)(x_0 - x_*)\right] +$$
$$+ \left[F'(x_*)^T F'(x_*)\right]^{-1} F'(x_*)^T F(x_*) - \left[F'(z_0)^T F'(z_0)\right]^{-1} F'(z_0)^T F(x_*);$$

$$y_1 - x_* = x_1 - x_* - \left[F'(z_0)^T F'(z_0)\right]^{-1} F'(z_0)^T F(x_1) =$$
$$= \left[F'(z_0)^T F'(z_0)\right]^{-1} F'(z_0)^T \left[F'(z_0)(x_1 - x_*) - F(x_1) + F(x_*)\right] +$$
$$+ \left[F'(x_*)^T F'(x_*)\right]^{-1} F'(x_*)^T F(x_*) - \left[F'(z_0)^T F'(z_0)\right]^{-1} F'(z_0)^T F(x_*) =$$
$$= \left[F'(z_0)^T F'(z_0)\right]^{-1} F'(z_0)^T \times$$
$$\times \left[\left(F'\left(\frac{x_1 + x_*}{2}\right)(x_1 - x_*) - F(x_1) + F(x_*)\right) + \right.$$
$$\left. + \left(F'(z_0) - F'\left(\frac{x_1 + x_*}{2}\right)\right)(x_1 - x_*)\right] +$$

$$+\left[F^{'}(x_*)^T F^{'}(x_*)\right]^{-1} F^{'}(x_*)^T F(x_*) - \left[F^{'}(z_0)^T F^{'}(z_0)\right]^{-1} F^{'}(z_0)^T F(x_*).$$

According to Lemma 1 from [17] with the value $\omega = 1/2$ we can write

$$\begin{aligned} & F(x) - F(y) - F^{'}\left(\frac{x+y}{2}\right)(x-y) = \\ & = \frac{1}{4}\int_0^1 (1-t)\left[F^{''}\left(\frac{x+y}{2} + \frac{t}{2}(x-y)\right) - \right. \\ & \left. - F^{''}\left(\frac{x+y}{2} + \frac{t}{2}(y-x)\right)\right](x-y)^2 dt. \end{aligned} \tag{15}$$

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

$$\begin{aligned} & \left\| F(x_*) - F(x_0) - F^{'}\left(\frac{x_0+x_*}{2}\right)(x_* - x_0)\right\| = \\ & = \frac{1}{4}\left\| \int_0^1 (1-t)\left[F^{''}\left(\frac{x_0+x_*}{2} + \frac{t}{2}(x_*-x_0)\right) - \right.\right. \\ & \left.\left. - F^{''}\left(\frac{x_0+x_*}{2} + \frac{t}{2}(x_0-x_*)\right)\right](x_*-x_0)^2 dt\right\| \le \\ & \le \frac{1}{4}\int_0^1 (1-t)\int_0^{t\|x_0-x_*\|} N(u)du \|x_0 - x_*\|^2 dt = \\ & = \frac{1}{8}\int_0^{\rho(x_0)} N(u)\left(1 - \frac{u}{\rho(x_0)}\right)^2 du \rho(x_0)^2 = \frac{1}{8}\int_0^{\rho(x_0)} N(u)(\rho(x_0)-u)^2 du, \end{aligned}$$

and also

$$\left\| F^{'}\left(\frac{x_0+y_0}{2}\right) - F^{'}\left(\frac{x_0+x_*}{2}\right)\right\| \le \int_0^{\rho(y_0)/2} L(u)du.$$

Using (4) and (14), we obtain that

$$\|(F^{'}(x_*)^T F^{'}(x_*))^{-1} F^{'}(x_*)^T\| \|F^{'}(x) - F^{'}(x_*)\| \le \beta \int_0^{\rho(x)} L(u)du.$$

According to Lemmas 1 and 2 and that $F^{'}(x)$ has full rank, for all $x \in B(x_*, r)$, the following inequalities hold

$$\|(F^{'}(x)^T F^{'}(x))^{-1} F^{'}(x)^T\| \le \frac{\beta}{1 - \beta\int_0^{\rho(x)} L(u)du}; \tag{16}$$

$$\begin{aligned} & \|(F^{'}(x)^T F^{'}(x))^{-1} F^{'}(x)^T - (F^{'}(x_*)^T F^{'}(x_*))^{-1} F^{'}(x_*)^T\| \le \\ & \le \frac{\sqrt{2}\beta^2 \int_0^{\rho(x)} L(u)du}{1 - \beta\int_0^{\rho(x)} L(u)du}. \end{aligned} \tag{17}$$

By the monotonicity of $L(u)$ and $N(u)$ with Lemmas 3 and 4, functions $\frac{1}{t}\int_0^t L(u)du$ and $\frac{1}{t^3}\int_0^t N(u)(t-u)^2 du$ are non-decreasing by t. Hence, from

(6) and (7) it follows that

$$q \le \frac{1}{r_0}\left[\frac{\beta \int\limits_0^{r_0} N(u)(r_0-u)^2 du}{8\left(1-\beta\int\limits_0^{r_0} L(u)du\right)} + \frac{\beta r_0 \int\limits_0^{(3/2)r_0} L(u)du + \sqrt{2}\alpha\beta^2 \int\limits_0^{r_0} L(u)du}{1-\beta\int\limits_0^{r_0} L(u)du}\right] <$$

$$< \frac{1}{r}\left[\frac{\beta \int\limits_0^{r} N(u)(r-u)^2 du}{8\left(1-\beta\int\limits_0^{r} L(u)du\right)} + \frac{\beta r \int\limits_0^{(3/2)r} L(u)du}{1-\beta\int\limits_0^{r} L(u)du} + \frac{\sqrt{2}\alpha\beta^2 \int\limits_0^{r} L(u)du}{1-\beta\int\limits_0^{r} L(u)du}\right] \le 1.$$

Thus, by Lemmas 1-4, conditions (4) and (5), and the afore-derived estimates, we obtain

$$\|x_1 - x_*\| \le \left\|\left[F'(z_0)^T F'(z_0)\right]^{-1} F'(z_0)^T\right\| \times$$
$$\times \left\|\left(F'\left(\frac{x_0+x_*}{2}\right)(x_0-x_*) - F(x_0) + F(x_*)\right) + \right.$$
$$\left. + \left(F'(z_0) - F'\left(\frac{x_0+x_*}{2}\right)\right)(x_0-x_*)\right\| +$$
$$+ \left\|\left[F'(x_*)^T F'(x_*)\right]^{-1} F'(x_*)^T F(x_*) - \left[F'(z_0)^T F'(z_0)\right]^{-1} F'(z_0)^T F(x_*)\right\| \le$$
$$\le \frac{\beta\rho(x_0)^3 \int_0^{\rho(x_0)} N(u)(\rho(x_0)-u)^2 du}{8\rho(x_0)^3\left(1-\beta\int_0^{\rho(z_0)} L(u)du\right)} +$$
$$+ \frac{\beta\rho(x_0)\rho(y_0)\int_0^{\rho(y_0)/2} L(u)du}{\rho(y_0)\left(1-\beta\int_0^{\rho(z_0)} L(u)du\right)} + \frac{\sqrt{2}\alpha\beta^2\rho(z_0)\int_0^{\rho(z_0)} L(u)du}{\rho(z_0)\left(1-\beta\int_0^{\rho(z_0)} L(u)du\right)} <$$
$$< \gamma\rho(x_0)^3 + \eta\rho(x_0)\rho(y_0) + \theta\rho(z_0) < qr_0 < r.$$

Similarly,

$$\|y_1 - x_*\| = \left\|\left[F'(z_0)^T F'(z_0)\right]^{-1} F'(z_0)^T\right\| \times$$
$$\times \left\|\left(F'\left(\frac{x_1+x_*}{2}\right)(x_1-x_*) - F(x_1) + F(x_*)\right) + \right.$$
$$\left. + \left(F'(z_0) - F'\left(\frac{x_1+x_*}{2}\right)\right)(x_1-x_*)\right\| +$$
$$+ \left\|\left[F'(x_*)^T F'(x_*)\right]^{-1} F'(x_*)^T F(x_*) - \left[F'(z_0)^T F'(z_0)\right]^{-1} F'(z_0)^T F(x_*)\right\| \le$$
$$\le \frac{\beta\rho(x_1)^3 \int_0^{\rho(x_1)} N(u)(\rho(x_1)-u)^2 du}{8\rho(x_1)^3\left(1-\beta\int_0^{\rho(z_0)} L(u)du\right)} +$$
$$+ \frac{\beta\rho(x_1)\rho(z_0')\int_0^{\rho(z_0')} L(u)du}{\rho(z_0')\left(1-\beta\int_0^{\rho(z_0)} L(u)du\right)} + \frac{\sqrt{2}\alpha\beta^2\rho(z_0)\int_0^{\rho(z_0)} L(u)du}{\rho(z_0)\left(1-\beta\int_0^{\rho(z_0)} L(u)du\right)} \le$$

$$\le \gamma\rho(x_1)^3 + (\eta/3)\rho(x_1)(\rho(x_0)+\rho(y_0)+\rho(x_1)) + \theta\rho(z_0) <$$
$$< \gamma\rho(x_0)^3 + (\eta/3)\rho(x_0)(2\rho(x_0)+\rho(y_0)) + \theta\rho(z_0) < qr_0 < r,$$

where $\rho(z_0^{'}) = (\rho(x_0)+\rho(y_0)+\rho(x_1))/2$. Therefore, $x_1, y_1 \in B(x_*, r)$ and both (8) and (9) follow for $k = 0$. Also, (10) is satisfied

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

Using mathematical induction, assume that $x_k, y_k \in B(x_*, r)$ and (8)–(10) hold for $k > 0$. Then, from (3) for $k+1$ we obtain that

$$\|x_{k+1} - x_*\| \le \frac{\beta\rho(x_k)^3 \int_0^{\rho(x_k)} N(u)(\rho(x_k)-u)^2 du}{8\rho(x_k)^3\left(1-\beta\int_0^{\rho(z_k)} L(u)du\right)} +$$
$$+\frac{\beta\rho(x_k)\rho(y_k)\int_0^{\rho(y_k)/2} L(u)du}{\rho(y_k)\left(1-\beta\int_0^{\rho(z_k)} L(u)du\right)} + \frac{\sqrt{2}\alpha\beta^2\rho(z_k)\int_0^{\rho(z_k)} L(u)du}{\rho(z_k)\left(1-\beta\int_0^{\rho(z_k)} L(u)du\right)} \le$$
$$\le \frac{\beta\rho(x_k)^3 \int_0^{\rho(x_0)} N(u)(\rho(x_0)-u)^2 du}{8\rho(x_0)^3\left(1-\beta\int_0^{\rho(z_0)} L(u)du\right)} + \tag{18}$$
$$+\frac{\beta\rho(x_k)\rho(y_k)\int_0^{\rho(y_0)/2} L(u)du}{\rho(y_0)\left(1-\beta\int_0^{\rho(z_0)} L(u)du\right)} + \frac{\sqrt{2}\alpha\beta^2\rho(z_k)\int_0^{\rho(z_0)} L(u)du}{\rho(z_0)\left(1-\beta\int_0^{\rho(z_0)} L(u)du\right)} \le$$
$$\le \gamma\rho(x_k)^3 + \eta\rho(x_k)\rho(y_k) + \theta\rho(z_k) \le qr_k < r.$$

and

$$\|y_{k+1} - x_*\| \le \frac{\beta\rho(x_{k+1})^3 \int_0^{\rho(x_{k+1})} N(u)(\rho(x_{k+1})-u)^2 du}{8\rho(x_{k+1})^3\left(1-\beta\int_0^{\rho(z_k)} L(u)du\right)} +$$
$$+\frac{\beta\rho(x_{k+1})\rho(z_k^{'})\int_0^{\rho(z_k^{'})} L(u)du}{\rho(z_k^{'})\left(1-\beta\int_0^{\rho(z_k)} L(u)du\right)} + \frac{\sqrt{2}\alpha\beta^2\rho(z_k)\int_0^{\rho(z_k)} L(u)du}{\rho(z_k)\left(1-\beta\int_0^{\rho(z_k)} L(u)du\right)} \le$$
$$\le \frac{\beta\rho(x_{k+1})^3 \int_0^{\rho(x_0)} N(u)(\rho(x_0)-u)^2 du}{8\rho(x_0)^3\left(1-\beta\int_0^{\rho(z_0)} L(u)du\right)} + \tag{19}$$
$$+\frac{\beta\rho(x_{k+1})\rho(z_k^{'})\int_0^{\rho(z_0^{'})} L(u)du}{\rho(z_0^{'})\left(1-\beta\int_0^{\rho(z_0)} L(u)du\right)} + \frac{\sqrt{2}\alpha\beta^2\rho(z_k)\int_0^{\rho(z_0)} L(u)du}{\rho(z_0)\left(1-\beta\int_0^{\rho(z_0)} L(u)du\right)} \le$$
$$\le \gamma\rho(x_{k+1})^3 + (\eta/3)(\rho(x_k)+\rho(y_k)+\rho(x_{k+1}))\rho(x_{k+1}) + \theta\rho(z_k) \le$$
$$\le qr_k < r.$$

where $\rho(z_k^{'}) = (\rho(x_k)+\rho(y_k)+\rho(x_{k+1}))/2$. According to (11) and both inequalities (8) and (9), 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, $x_{k+1}, y_{k+1} \in B(x_*, r)$ and (8)–(10) hold; and also $\lim_{k\to\infty} x_k = x_*$ and $\lim_{k\to\infty} y_k = x_*$. This completes the induction and the proof of Theorem 1. □

In case of zero residual ($\alpha = \|F(x_*)\| = 0$) the results of Theorem 1 are

Corollary 1. *Suppose that x_* satisfies (1), $F(x_*) = 0$, $F(x)$ is a twice Fréchet differentiable operator in $B(x_*, R)$, $F^{'}(x_*)$ has full rank, and both $F^{'}(x)$ and $F^{''}(x)$ satisfy the Lipschitz conditions with L and N average as in (4) and (5), respectively, where L and N are positive non-decreasing functions on $[0, 3R/2]$. Furthrmore, assume function H_0 has a minimal zero r on $[0, R]$, which also satisfies:*

$$\beta \int_0^r L(u)du < 1,$$

where

$$H_0(p) = (\beta/8) \int_0^p N(u)(p-u)^2 du + \beta p \Big(\int_0^{(3/2)p} L(u)du + \int_0^p L(u)du \Big) - p.$$

Then, the Gauss-Newton type method (3) is convergent for all x_0, $y_0 \in B(x_*, r)$ such that

$$\begin{aligned} \rho(x_{k+1}) &\leq \gamma\rho(x_k)^3 + \eta\rho(x_k)\rho(y_k), \\ \rho(y_{k+1}) &\leq \gamma\rho(x_{k+1})^3 + (\eta/3)(\rho(x_k) + \rho(y_k) + \rho(x_{k+1}))\rho(x_{k+1}), \\ r_{k+1} &= \max\{\rho(x_{k+1}), \rho(y_{k+1})\} \leq q r_k \leq \cdots \leq q^{k+1} r_0, \end{aligned}$$

where $\rho(x) = \|x - x_*\|$, $r_0 = \max\{\rho(x_0), \rho(y_0)\}$,

$$\begin{aligned} q = & \frac{\beta \int_0^{\rho(x_0)} N(u)(\rho(x_0) - u)^2 du}{8\rho(x_0)\left(1 - \beta \int_0^{\rho(z_0)} L(u)du\right)} + \\ & + \frac{\beta\rho(x_0) \int_0^{\rho(x_0)+\rho(y_0)/2} L(u)du}{(2\rho(x_0) + \rho(y_0))/3 \left(1 - \beta \int_0^{\rho(z_0)} L(u)du\right)} < 1, \end{aligned}$$

γ, η, β hold in (12)–(14).

Corollary 2. *Convergence order of the iterative method (3) in case of zero residual is equal to $1 + \sqrt{2}$.*

Proof. Assume that $a_k = \rho(x_k), b_k = \rho(y_k), k = 0, 1, 2, ...$ Since the residual is equal to zero, i.e. $\alpha = \|F(x_*)\| = 0$, so $\theta = 0$. From the inequalities (18) and (19), we have

$$a_{k+1} \leq a_k(\gamma a_k^2 + \eta b_k), \tag{20}$$

$$\begin{aligned} b_{k+1} &\leq a_{k+1}\left[\gamma a_{k+1}^2 + \eta/3(a_k + a_{k+1} + b_k)\right] \leq \\ &\leq a_{k+1}\left[(\gamma a_k + 2\eta/3)a_k + \eta b_k/3\right] \leq \\ &\leq a_{k+1}a_k\left[\gamma r + \eta\right] = a_{k+1}a_k\phi_1. \end{aligned} \tag{21}$$

From (20) and (21) 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 [17]

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

The positive root of the latter, which is $\rho_* = 1 + \sqrt{2}$, is the order of convergence of the iterative method (3). □

Theorem 2. ***(The uniqueness of solution)*** *Suppose x_* satisfies (1) and $F(x)$ has a continuous derivative $F^{'}(x)$ in the ball $B(x_*, r)$. Moreover, $F^{'}(x_*)$ has full rank and $F^{'}(x)$ satisfies the Lipschitz condition with L average (4). Let $r > 0$ satisfy*

$$\frac{\beta}{r}\int_0^r L(u)(r-u)du + \frac{\alpha\beta_0}{r}\int_0^r L(u)du \leq 1, \tag{22}$$

where α and β are defined in (14) and $\beta_0 = \|[F^{'}(x_*)^T F^{'}(x_*)]^{-1}\|$. Then, x_* is a unique solution of the problem (1) in $B(x_*, r)$.

The proof of this theorem is analogous to the one in [4].

4. Applications

In this section, we apply the obtained results to special cases, when, for instance, L is a Lipschitz constant. Then, we immediately receive results of the convergence analysis of the method (3).

Theorem 3. *Let $F : \mathbb{R}^n \to \mathbb{R}^m$, $m \geq n$, be a twice Fréchet differentiable operator in $D \subseteq \mathbb{R}^n$. Assume that (1) has a solution $x_* \in D$ and a Fréchet derivative $F^{'}(x_*)$ has full rank. Suppose that Fréchet derivatives $F^{'}(x)$ and $F^{''}(x)$ on $B(x_*, r) = \{x \in D : \|x - x_*\| \leq r\}$ satisfy the Lipschitz conditions:*

$$\|F^{'}(x) - F^{'}(y)\| \leq L\|x - y\|, \tag{23}$$

$$\|F^{''}(x) - F^{''}(y)\| \leq N\|x - y\|, \tag{24}$$

where $x, y \in B(x_*, r)$ and both L and N are positive numbers. Also, the radius $r > 0$ is a root of the equation

$$\beta N r^2 + 60\beta L r + 24\sqrt{2}\alpha\beta^2 L - 24 = 0. \tag{25}$$

Then, for all $x_0, y_0 \in B(x_*, r)$ the sequences $\{x_k\}$ and $\{y_k\}$, which are generated by the method (3), are well defined, remain in $B(x_*, r)$ for all $k \geq 0$, and converge to x_* such that

$$\rho(x_{k+1}) \leq \frac{(\beta/24)N\rho(x_k)^3 + \beta L\rho(x_k)\rho(y_k)/2 + \sqrt{2}\alpha\beta^2 L\rho(z_k)}{1 - \beta L\rho(z_k)}, \tag{26}$$

$$\rho(y_{k+1}) \leq \frac{(\beta/24)N\rho(x_{k+1})^3 + \beta L\rho(x_{k+1})(\rho(x_{k+1}) + \rho(x_k) + \rho(y_k))/2}{1 - \beta L\rho(z_k)} + \\ + \frac{\sqrt{2}\alpha\beta^2 L\rho(z_k)}{1 - \beta L\rho(z_k)}, \tag{27}$$

$$r_{k+1} = \max\{\rho(x_{k+1}), \rho(y_{k+1})\} \leq q r_k \leq \cdots \leq q^{k+1} r_0, \tag{28}$$

where $\rho(x) = \|x - x_*\|$, $r_0 = \max\{\rho(x_0), \rho(y_0)\}$,

$$0 < q = \frac{(\beta/24)N\rho(x_0)^2 + \beta L(\rho(x_0) + \rho(y_0)/2) + \sqrt{2}\alpha\beta^2 L}{1 - \beta L\rho(z_0)} < 1, \tag{29}$$

$z_k = (x_k + y_k)/2$ and both α and β are defined in (14).

Proof. Let choose arbitrary $x_0, y_0 \in B(x_*, r)$. According to Lemma 1 from [17] and the proof of Theorem 1, by setting $x = x_*$ and $y = x_0$ in (15), we receive

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

and also

$$\left\| F'\left(\frac{x_0 + y_0}{2}\right) - F'\left(\frac{x_0 + x_*}{2}\right)\right\| \leq L\rho(y_0)/2.$$

Using (23) and (14), we obtain that

$$\|(F'(x_*)^T F'(x_*))^{-1} F'(x_*)^T\| \|F'(x) - F'(x_*)\| \leq \beta L \rho(x).$$

According to that $F'(x)$ has full rank, for all $x \in B(x_*, r)$, the following inequalities hold

$$\|(F'(x)^T F'(x))^{-1} F'(x)^T\| \leq \frac{\beta}{1 - \beta L \rho(x)},$$

$$\|(F'(x)^T F'(x))^{-1} F'(x)^T - (F'(x_*)^T F'(x_*))^{-1} F'(x_*)^T\| \leq \frac{\sqrt{2}\beta^2 L \rho(x)}{1 - \beta L \rho(x)}.$$

Hence, from (25) it follows that

$$0 < q = \frac{(\beta/24)N\rho(x_0)^2 + 3\beta L(\rho(x_0) + \rho(y_0)/2) + \sqrt{2}\alpha\beta^2 L}{1 - \beta L \rho(z_0)} <$$
$$< \frac{(\beta/24)Nr^2 + 3\beta L r/2 + \sqrt{2}\alpha\beta^2 L}{1 - \beta L r} \leq 1.$$

Thus, by Lemmas 1-4, conditions (23) and (24), and the derived estimates in the proof of Theorem 1, we obtain

$$\|x_1 - x_*\| \leq \frac{(\beta/24)N\rho(x_0)^3 + \beta L\rho(x_0)\rho(y_0)/2 + \sqrt{2}\alpha\beta^2 L\rho(z_0)}{1 - \beta L\rho(z_0)} < q r_0 < r.$$

Similarly,

$$\|y_1 - x_*\| \leq \frac{(\beta/24)N\rho(x_1)^3}{1 - \beta L \rho(z_0)} +$$
$$+ \frac{\beta L \rho(x_1)(\rho(x_1) + \rho(x_0) + \rho(y_0))/2 + \sqrt{2}\alpha\beta^2 L\rho(z_0)}{1 - \beta L \rho(z_0)} \leq$$
$$\leq \frac{(\beta/24)N\rho(x_0)^3 + \beta L\rho(x_0)(2\rho(x_0) + \rho(y_0))/2 + \sqrt{2}\alpha\beta^2 L\rho(z_0)}{1 - \beta L\rho(z_0)} < q r_0 < r.$$

Therefore, $x_1, y_1 \in B(x_*, r)$ and both (26) and (27) follow for $k = 0$. Also, (28) is satisfied

$$r_1 = \max\{\|x_1 - x_*\|, \|y_1 - x_*\|\} \leq q r_0.$$

Using mathematical induction, assume that $x_k, y_k \in B(x_*, r)$ and (28) holds for $k > 0$. Then, for $k+1$ from (3) we obtain that

$$\|x_{k+1} - x_*\| \leq \frac{(\beta/24)N\rho(x_k)^3 + \beta L\rho(x_k)\rho(y_k)/2 + \sqrt{2}\alpha\beta^2 L\rho(z_k)}{1 - \beta L\rho(z_k)} \leq$$
$$\leq \frac{((\beta/24)N\rho(x_0)^2 + \beta L\rho(y_0)/2 + \sqrt{2}\alpha\beta^2 L)r_k}{1 - \beta L\rho(z_0)} = qr_k < r$$

and

$$\|y_{k+1} - x_*\| \leq \frac{(\beta/24)N\rho(x_{k+1})^3 + \beta L\rho(x_{k+1})(\rho(x_{k+1}) + \rho(x_k) + \rho(y_k))/2}{1 - \beta L\rho(z_k)} +$$
$$+ \frac{\sqrt{2}\alpha\beta^2 L\rho(z_k)}{1 - \beta L\rho(z_k)} < qr_k < r.$$

According to (29) and both inequalities (26) and (27), we receive

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

Thus, $x_{k+1}, y_{k+1} \in B(x_*, r)$ as well as (26), (27), and (28) hold. □

From (25) it follows that the convergence radius of the method (3) is

$$r = \frac{4(1 - \sqrt{2}\alpha\beta^2 L)}{5\beta L + \frac{1}{12}\sqrt{(60\beta L)^2 + 96\beta N(1 - \sqrt{2}\alpha\beta^2 L)}}.$$

For zero residual, Theorem 3 can be formulated as

Corollary 3. *Suppose that x_* satisfies (1), $F(x_*) = 0$, $F(x)$ is a twice Fréchet differentiable operator in $B(x_*, r)$, $F^{'}(x_*)$ has full rank, and both $F^{'}(x)$ and $F^{''}(x)$ satisfy the classic Lipschitz conditions as in (23) and (24), respectively. Moreover, the radius $r > 0$ is a unique positive root of the following equation*

$$\beta N r^2 + 60\beta L r - 24 = 0.$$

Then, the Gauss-Newton type method (3) is convergent for all x_0, $y_0 \in B(x_*, r)$ such that

$$\rho(x_{k+1}) \leq \frac{(\beta/24)N\rho(x_k)^3 + \beta L\rho(x_k)\rho(y_k)/2}{1 - \beta L\rho(z_k)},$$
$$\rho(y_{k+1}) \leq \frac{(\beta/24)N\rho(x_{k+1})^3 + \beta L\rho(x_{k+1})(\rho(x_{k+1}) + \rho(x_k) + \rho(y_k))/2}{1 - \beta L\rho(z_k)},$$
$$r_{k+1} = \max\{\rho(x_{k+1}), \rho(y_{k+1})\} \leq qr_k \leq \cdots \leq q^{k+1} r_0,$$

where $\rho(x) = \|x - x_*\|$, $r_0 = \max\{\rho(x_0), \rho(y_0)\}$,

$$0 < q = \frac{(\beta/24)N\rho(x_0)^2 + \beta L(\rho(x_0) + \rho(y_0)/2}{1 - \beta L\rho(z_0)} < 1,$$

$z_k = (x_k + y_k)/2$ and β is defined in (14).

From Corollary 3, the convergence radius is

$$r = \frac{4}{5\beta L + \frac{1}{12}\sqrt{(60\beta L)^2 + 96\beta N}} < \frac{2}{5\beta L}$$

that corresponds to the previously received results in [10] for nonlinear equations ($m = n$).

Under the classic Lipschitz condition Theorem 2 for the uniqueness of the solution can be written as follow

Theorem 4. *Suppose x_* satisfies (1) and $F(x)$ has a continuous derivative $F'(x)$ in $B(x_*, r)$. Moreover, $F'(x_*)$ has full rank and $F'(x)$ satisfies the classic Lipschitz condition as in (23). Let $r > 0$ satisfy*

$$\frac{\beta L r}{2} + \alpha\beta_0 L \le 1.$$

Then, x_* is a unique solution of the problem (1) in $B(x_*, r)$.

5. Numerical experiments

We carried out a set of experiments on widely used test problems and compared the number of iterations under which the Gauss-Newton method (2), the Secant method [11], and the method (3) converge to the solution. We used the same initial points for all methods and the following stopping criteria:

$$\|x_{k+1} - x_k\| \le \varepsilon \quad \text{and} \quad \|A_{k+1}^T F(x_{k+1})\| \le \varepsilon,$$

where

- $A_{k+1} = F'(x_{k+1})$ for the Gauss-Newton method (2);
- $A_{k+1} = F'(z_{k+1})$ for the method (3);
- $A_{k+1} = F(x_{k+1}, x_k)$ for the Secant method, $F(x_{k+1}, x_k)$ is the divided difference of the first order of F [11].

Tabl. 1. The number of iterations to the solution with the accuracy $\varepsilon = 10^{-12}$

Example	Gauss-Newton	Secant	M-d (3)
Rosenbrock func. ($n = m = 4$) $x_0 = (-1.2, 1, -1.2, 1)$	5	4	4
Box-3D func. ($n = 3$, $m = 10$) $x_0 = (0, 10, 20)$	7	9	6
Gnedenko-Veibull dist. ($n = 2$, $m = 8$) $x_0 = (1, 1)$	7	–	6
Freidenstein-Ross func. ($n = m = 2$) $x_0 = (0.5, -2)$	43	18	10
Wood func. ($n = 4$, $m = 6$) $x_0 = (-3, -1, -3, -1)$	52	75	50
Bard func. ($n = 3$, $m = 15$) $x_0 = (1, 1, 1)$	10	–	9

In Table 1 we present the amount of iterations spent by each methods to compute an approximation to the solution of the examples from [9, 11] with the accuracy $\varepsilon = 10^{-12}$. The additional initial point y_0 we calculated in the following way: $y_0 = x_0 + 0.01$. The symbol '–' indicates that the Secant method does not converge to the solution with the desired accuracy, however the method converges for the lower accuracy ($\varepsilon = 10^{-8}$).

6. Conclusions

We studied the local convergence of the Gauss-Newton type method (3) under the generalized and classic Lipschitz conditions for the first- and second-order derivatives. We determined the convergence order and the radius of the method (3) as well as proved the uniqueness ball of the solution of the nonlinear least squares problem (1). The method (3) is not only more efficient than the Gauss-Newton and Secant methods in terms of the convergence order, but also in terms of the amount of iterations to the solution on a variety of test problems. Furthermore, the method (3) has promising characteristics for parallelization, which we plan to utilize for constructing and developing new parallel methods for solving the problem (1).

Acknowledgements

The simulations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at PDC Centre for High Performance Computing (PDC-HPC).

We are in debt to the unknown referee, who made useful remarks which helped a lot to improve the paper.

Bibliography

1. 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).
2. 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).
3. 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).
4. 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.
5. 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.
6. Dennis J. M. Numerical Methods for Unconstrained Optimization and Nonlinear Equations / J. M. Dennis, R. B. Schnabel. – New-York: Prentice-Hall, 1983.
7. Iakymchuk R. On the Convergence Analysis of a Two-Step Modification of the Gauss-Newton Method / R. Iakymchuk, S. M. Shakhno // PAMM. – 2014. – Vol. 14. – P. 813-814.
8. Kantorovich L. V. Functional Analysis / L. V. Kantorovich, G. P. Akilov. – Oxford: Pergamon Press, 1982.
9. More J. J. Testing Unconstrained Optimization Software / J. J. More, B. S. Garbow, K. E. Hillstrom // ACM Transactions on Mathematical Software. – 1981. – Vol. 7, № 1. – P. 17-41.

10. Shakhno S. M. Local convergence of a two-step Newton-like method for solving nonlinear equations under the generalized Lipschitz conditions / S. M. Shakhno // Physical and Mathematical Modeling and Information Technology. – 2012. – Vol. 16. – P. 163-172. (in Ukrainian).
11. 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.
12. Smale S. Newtons method estimates from data at one point / S. Smale // The Merging of Disciplines: New Directions in Pure / editors: R. Ewing, K. Gross, and C. Martin. – New York: Springer-Verlag, 1986. – P. 185-196.
13. Steward G. W. On the continuity of the generalized inverse / G. W. Steward // SIAM J. Appl. Math. – 1960. – Vol. 17, № 1. – P. 33-45.
14. Wang X. H. Convergence of Newton's method and inverse function theorem in Banach space / X. H. Wang // Math. Comp. – 1999. – Vol. 68. – P. 169-186.
15. Wang X. H. Convergence of Newton's method and uniqueness of the solution of equations in Banach space / X. H. Wang // IMA Journal of Numerical Analysis. – 2000. – Vol. 20. – P. 123-134.
16. Wedin P.-Å. Perturbation theory for pseudo-inverses / P.-Å. Wedin // BIT Numerical Mathematics. – 1973. – Vol. 13, № 2. – P. 217-232.
17. Werner W. Über ein Verfahren der Ordnung $1+\sqrt{2}$ zur Nullstellenbestimmung / W. Werner // Numer. Math. – 1979. – Vol. 32. – P. 333-342.

Roman Iakymchuk,
Department of Computational Science and Technology,
School of Computer Science and Communication,
KTH Royal Institute of Technology,
Lindstedtsvägen 5, Stockholm, 100 44, Sweden;
Stepan Shakhno, Halyna Yarmola,
Faculty of Applied Mathematics and Informatics,
Ivan Franko National University of Lviv,
1, Universytets'ka Str., Lviv, 79000, Ukraine.

Received 05.08.2017; revised 07.10.2017