Iranian Journal of Numerical Analysis and Optimization
Vol. 6, No. 1, (2016), pp 101-119

A nonmonotone trust-region-approach with nonmonotone adaptive radius for solving nonlinear systems

K. Amini, H. Esmaeili and M. Kimiaei

Abstract

This paper presents a trust-region procedure for solving systems of nonlinear equations. The proposed approach takes advantages of an effective adaptive trust-region radius and a nonmonotone strategy by combining both of them appropriately. It is believed that selecting an appropriate adaptive radius based on a suitable nonmonotone strategy can improve the efficiency and robustness of the trust-region framework as well as can decrease the computational cost of the algorithm by decreasing the number of subproblems that must be solved. The global convergence to first order stationary points as well as the local q-quadratic convergence of the proposed approach are proved. Numerical experiments show that the new algorithm is promising and attractive for solving nonlinear systems.

Keywords: Nonlinear equations; Trust-region framework; Adaptive radius; Nonmonotone technique.

1 Introduction

In this paper, we consider the nonlinear system of equations

$$F(x) = 0, \quad x \in \mathbb{R}^n, \tag{1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is a continuously differentiable mapping in the form $F(x) := (F_1(x), F_2(x), \cdots, F_n(x))^T$. If $F(x)$ has a zero, then the nonlinear

*Corresponding author
Received 5 April 2015; revised 4 July 2014; accepted 21 October 2015
K. Amini
Department of Mathematics, Faculty of Science, Razi University, Kermanshah, Iran. e-mail: kamini@razi.ac.ir

H. Esmaeili
Department of Mathematics, Bu-Ali Sina University, Hamedan, Iran. e-mail: esmaeili@basu.ac.ir

M. Kimiaei
Department of Mathematics, Asadabad Branch, Islamic Azad University, Asadabad, Iran. e-mail: morteza.kimiaei@gmail.com

system (1) is equivalent to the following nonlinear unconstrained least-squares problem

$$\begin{array}{ll} \min & f(x) := \frac{1}{2}\|F(x)\|^2 \\ \text{s.t.} & x \in \mathbb{R}^n, \end{array} \tag{2}$$

where $\|.\|$ denotes the Euclidean norm. The trust-region frameworks for solving system of nonlinear equations (1) are a popular class of iterative procedures that in each iteration generate a trial step d_k by computing an exact or approximate solution of the following subproblem

$$\begin{array}{ll} \min m_k(x_k + d) := \frac{1}{2}\|F_k + J_k d\|^2 = f_k + d^T J_k^T F_k + \frac{1}{2} d^T J_k^T J_k d \\ \text{s.t} \quad d \in \mathbb{R}^n \text{ and } \|d\| \le \Delta_k. \end{array} \tag{3}$$

where $f_k := f(x_k)$, $F_k := F(x_k)$, $J_k := F'(x_k)$, Jacobian of $F(x)$, and $\Delta_k > 0$ is trust-region radius. The ratio

$$r_k = \frac{f_k - f(x_k + d_k)}{m_k(x_k) - m_k(x_k + d_k)}. \tag{4}$$

plays a main role in trust region frameworks. Obviously, the model matches the original problem at the current iteration x_k whenever r_k is sufficiently close to 1. Then the agreement is weak or there is no agreement whenever r_k is near zero and there is not agreement when r_k is negative. Generally, if r_k is greater than a positive constant μ, the trial step d_k will be accepted and leading to $x_{k+1} := x_k + d_k$. In this case, it is safe to increase trust region radius in the next iteration. Otherwise, the trust-region radius should be shrunk and the subproblem (3) will be solved again to possibly find an acceptable trial point in the sequel of the process.

It is known that the traditional trust-region framework has some drawbacks: a) the very small trust-region radius Δ_k, increases the total number of the iterates, b) the remarkably large trust-region radius Δ_k increases the total number of solving subproblems, c) ratio (4) does not suffice to create the agreement between the quadratic model and the objective function in trust-region methods leading to increasing computational cost. Using the adaptive radius is an appropriate idea to overcome drawbacks (a) and (b). As a result, many researchers have investigated on finding the best trust-region radius Δ_k, but no one has actually claimed a general rule for generating the trust-region radius. Therefore, in order to decrease the total number of solving subproblems for an arbitrary problem, some adaptive processes determining the radius have been proposed, see [3, 33, 45]. For example, Zhang and Wang [44] proposed an adaptive radius by

$$\Delta_k = c^{p_k} \|F_k\|^{\delta},$$

where $0 < c < 1$ and $0.5 < \delta < 1$ are constant, also p_k is a non-negative integer starting from zero. The major advantage of this method is that the radius does not stay very large and therefore it is possible to prevent resolving the trust region subproblem. This proposal has some disadvantages: Firstly, the sequence generated by this method is superlinearly convergent with the convergence order 2δ. Secondly, the efficiency of the numerical results is largely dependent on the choice of δ. Finally, this method cannot adequately prevent the very small trust-region radius. To overcome these drawbacks, Fan and Pan in [15] suggest that

$$\Delta_k = c^{p_k} M\|F_k\|,$$

where it is also another satisfactory radius with a constant M, an integer p_k and $c \in (0,1)$. This choice for the trust region radius plays an important role in proving the quadratic convergence and also prevents some deal from introducing the intensely small trust-region radius. In this method, if $\|F_k\|$ is very small, then the constant M must be chosen so large that the radius is not too small. But for some problems in which $\|F_k\|$ is large, $M\|F_k\|$ will be very large and the number of solving subproblems may be increased. Thus, the amount of computation and the cost of solving problem will be increased.

One of the convenient ways to overcome the drawback (c), is the nonmonotone techniques that can improve the efficiency and the robustness of trust region algorithms, especially when it is applied to highly nonlinear problems, in the presence of narrow curved valley, see for examples [1,2,4,18–20,43,46]. Therefore, a nonmonotone strategy can be employed to increase the efficiency of the proposed procedures.

Contribution. The primary goal in the design of the new method is decreasing computational cost by combining two nonmonotone techniques and adaptive radius trust region. We hope that combining these two techniques can improve numerical performance and efficiency of algorithm. We attain this designed goal by building a new adaptive radius based on nonmonotone technique. The global convergence to first-order critical points along with q-quadratic convergence are being established. The numerical experiments confirm the efficiency and the robustness of the proposed method for solving systems of nonlinear equations.

Organization. This paper is organized as follows: The structure of algorithm will be described after a new adaptive trust-region radius and a nonmonotone technique are proposed in Section 2. In Section 3, we will investigate the global convergence and the quadratic convergence rate of the new algorithm under some suitable assumptions. Numerical results are reported in Section 4. Finally, we end up the paper by some conclusive remarks given in Section 5.

2 Motivation and algorithmic structure

It is well-known that the traditional approaches in unconstrained optimization generally use a globalization technique, like line search or trust-region, to guarantee the global convergence of the algorithm. These globalization techniques mostly enforce a monotonicity to the produced sequence of the objective function values which usually causes a short step to be produced and so a slow numerical convergence encountering highly nonlinear problems in the presence of a narrow curved valley, see [1, 2, 4, 9, 18, 43]. For example, the traditional trust-region framework exploits the ratio (4) which leads to

$$f_k - f_{k+1} \geq m_k(x_k) - m_k(x_k + d_k) > 0.$$

This condition clearly implies that the sequence $\{f_k\}$ should be monotone. In order to avoid this drawback of the Armijo-type line search globalization techniques, a nonmonotone strategy was introduced by Grippo, Lampariello and Lucidi in [18] for unconstrained optimization problems while they modified the Armijo condition by the following condition

$$f(x_k + \alpha_k d_k) \leq f_{l(k)} + \delta \alpha_k g_k^T d_k, \tag{5}$$

where $\delta \in (0, 1)$, $g_k := \nabla f(x_k)$, the gradient of $f(x)$ in x_k, and

$$f_{l(k)} = \max_{0 \leq j \leq m(k)} \{f_{k-j}\}, \quad k \in \mathbb{N} \cup \{0\}, \tag{6}$$

in which $m(0) := 0$ and $0 \leq m(k) \leq \min\{m(k-1) + 1, N\}$ with $N \geq 0$. The theoretical and numerical results have shown that the proposed technique has some remarkable effects and improves both the possibility of finding the global optimum and the rate of convergence for algorithms. Motivated by their work, Deng et al. in [9] made some changes in the ratio (4) which assesses the agreement between the quadratic model and the objective function in trust-region methods. In addition thy introduced the first nonmonotone trust region algorithm. This idea was developed further by Zhou and Xiao [38, 46], Xiao and Chu [37] and Toint [35, 36]. The most common nonmonotone ratio is defined as follows:

$$\widehat{r}_k := \frac{f_{l(k)} - f(x_k + d_k)}{m_k(x_k) - m_k(x_k + d_k)}. \tag{7}$$

To overcome disadvantages (a) and (b), according to the proposed method by Esmaeili and Kimiaei [11]. We define the new adaptive radius by

$$\Delta_k := c^{p_k} N F_{l(k)}, \tag{8}$$

in which $0 < c < 1$, p_k is the smallest nonnegative integer p such that $\widehat{r}_k \geq \mu$ and

$$NF_{l(k)} := \max_{0 \le j \le m(k)} \{\|F_{k-j}\|\}, \quad k \in \mathbb{N} \cup \{0\}, \tag{9}$$

in which $m(0) := 0$ and $0 \le m(k) \le \min\{m(k-1)+1, N\}$ with $N \ge 0$. The proposed adaptive trust region radius has some benefits. First, since the sequence $\{NF_{l(k)}\}$ is reduced slowly and is greater than the sequence $\{\|F_k\|\}$ (see (11)), it prevents introducing the intensely small trust-region radius as possible and thus prevents increasing the total number of iterates. Second, Due to the decreasing sequence $\{NF_{l(k)}\}$, Δ_k will not stay too large and it prevents increasing the number of solving subproblems. Hence, using controlling the radius of trust-region, the new method can prevent the production of larger trial step near the optimizer and smaller trial step far from the optimizer.

Our assumptions are identical to those utilized for the proposed approach:

(H1) The level set $L(x_0) := \{x \in \mathbb{R}^n \mid f(x) \le f(x_0)\}$ is bounded for any given $x_0 \in \mathbb{R}^n$ and $F(x)$ is continuously differentiable on compact convex set Ω containing the level set $L(x_0)$.

(H2) The matrix $\{J_k\}$ is bounded and uniformly nonsingular on Ω, i.e. there exists constants $0 < M_0 \le 1 \le M_1$ such that

$$\|J_k\| \le M_1 \quad \text{and} \quad M_0\|F_k\| \le \|J_k^T F_k\| \quad \forall k \in \mathbb{N} \cup \{0\}, \tag{10}$$

.

(H3) The decrease on the model m_k is at least as much as a fraction of that obtained by the Cauchy point, i.e. there exists a constant $\beta \in (0,1)$ such that

$$m_k(x_k) - m_k(x_k + d_k) \ge \beta \, \|J_k^T F_k\| \, \min\left\{\Delta_k, \frac{\|J_k^T F_k\|}{\|J_k^T J_k\|}\right\}, \tag{11}$$

for all $k \in \mathbb{N} \cup \{0\}$.

(H4) $J(x)$ is Lipschitz continuous in $L(x_0)$, with Lipschitz constant γ_L.

We now incorporate both of the two nonmonotone and adaptive radius terms into trust-region and outline the subsequent algorithm:

Algorithm 1 NATR (Nonmonotone Adaptive Trust-Region Algorithm)

Input: An initial point $x_0 \in \mathbb{R}^n$, $c, \mu \in (0,1)$, $N > 0$, $\epsilon > 0$ and k_{max}.
Output: x_b, f_b;
1 **Begin**
2 $\Delta_0 := \|F_0\|$; $f_{l(0)} := 1/2\|F_{l(0)}\|^2$; $NF_{l(0)} := \|F_0\|$; $m(0) := 0$; $k := 0$;
3 **While** $\|F_k\| \geq \epsilon$ && $k \leq k_{max}$ **do**
4 $p := 0$; $\widehat{r}_k := 0$;
5 **While** $\widehat{r}_k < \mu$ **do**
6 specify the trial point d_k by solving the subproblem (3) ;
7 compute $F(x_k + d_k)$;
8 $f(x_k + d_k) := 1/2\ \|F(x_k + d_k)\|^2$;
9 determine $\widehat{r}_k$ using (7);
10 **If** $\widehat{r}_k < \mu$ **then**
11 $p \leftarrow p + 1$;
12 determine Δ_k using (8);
13 **end**
14 **end**
15 $x_{k+1} := x_k + d_k$; $F_{k+1} := F(x_{k+1})$; $f_{k+1} := f(x_{k+1})$; $J_{k+1} := J(x_{k+1})$;
16 compute J_{k+1} and let $m(k+1) := \min\{m(k) + 1, N\}$;
17 calculate $NF_{l(k+1)}$ by (9) and set $f_{l(k+1)} := 1/2NF^2_{l(k+1)}$;
18 select $\Delta_{k+1} := NF_{l(k+1)}$;
19 $k \leftarrow k + 1$;
20 **end**
21 $x_b := x_k$; $f_b := f_k$;
22 **end**

In Algorithm 1, the cycle starting from Line 3 to Line 25 is called the outer cycle, and the cycle starting from Line 5 to Line 14 is called the inner cycle.

Remark 1. The inequality (11) is called the sufficient reduction condition, see [32] and has been investigated by many authors when they extended some inexact methods for solving subproblem (3), for example see [11,13–15]. For global convergence purpose, it is enough to find a vector d_k such that it gives a sufficient reduction in the quadratic model m_k. Well-known convergence results [31] show that the trial step d_k is required to give a reduction in the model m_k that is at least some fixed multiple of the decrease attained by the Cauchy step at each iteration.

Lemma 1. *Suppose that (H4) holds, the sequence $\{x_k\}$ is generated by Algorithm 1 and d_k is a solution of the subproblem (3) such that $\|F(x_k) + J(x_k)d_k\| \leq \|F(x_k)\|$. Then, we have*

$$|f(x_k + d_k) - m_k(x_k + d_k))| \leq O(\|d_k\|^2). \tag{12}$$

Proof. See [9]. □

The following lemma indicates that the inner cycle of Algorithm 1 terminates in a finite number of inner iterates.

Lemma 2. *Suppose that (H2)-(H4) hold and the sequence $\{x_k\}$ is generated by Algorithm 1. Then, the inner cycle of Algorithm 1 is well-defined.*

Proof. Assume that the inner cycle of Algorithm 1 cycles infinitely in the iteration k, i.e., $\Delta_k^p = c^p NF_{l(k)} \to 0$ as $p \to \infty$, equivalently, for any $\eta > 0$, we have $\Delta_k^p < \eta$ for sufficiently large p. Using the fact that x_k is not the optimum of (2), we can conclude that there exists a constant $\epsilon > 0$ such that $\|F_k\| \geq \epsilon$. Without loss of generality, let $\eta := \frac{M_0\epsilon}{M_1^2}$. This fact along with (H2) and (11) imply

$$\begin{aligned} m_k(x_k) - m_k(x_k + d_k^p) &\geq \beta \|J_k^T F_k\| \ \min\left\{\Delta_k^p, \frac{\|J_k^T F_k\|}{\|J_k^T J_k\|}\right\} \\ &\geq \beta M_0 \|F_k\| \ \min\left\{\Delta_k^p, \frac{M_0\epsilon}{M_1^2}\right\} \\ &\geq \beta M_0 \epsilon \ \min\{\Delta_k^p, \eta\} \\ &= \beta M_0 \epsilon \Delta_k^p, \end{aligned} \tag{13}$$

where d_k^p is a solution of subproblem (3) corresponding to p in k-th iterate. Now, Lemma 1 and (13) leads to

$$\begin{aligned} \left|\frac{f_k - f(x_k + d_k^p)}{m_k(x_k) - m_k(x_k + d_k^p)} - 1\right| &= \left|\frac{f(x_k + d_k^p) - m_k(x_k + d_k^p)}{m_k(x_k) - m_k(x_k + d_k^p)}\right| \\ &\leq \frac{O(\|d_k^p\|^2)}{\beta M_0 \epsilon \Delta_k^p} \leq \frac{O((\Delta_k^p)^2)}{\beta M_0 \epsilon \Delta_k^p} \to 0, \quad \text{as } \ p \to \infty. \end{aligned}$$

Therefore, there exists a sufficiently large p_k such that

$$r_k = \frac{f_k - f(x_k + d_k^{p_k})}{m_k(x_k) - m_k(x_k + d_k^{p_k})} \geq \mu.$$

Besides, from the definition $f_{l(k)}$, it is clear that $f_{l(k)} \geq f_k$. This fact along with the previous inequality immediately implies $\widehat{r}_k \geq r_k \geq \mu$ which means that the inner cycle of Algorithm 1 stops and so Algorithm 1 is well-defined. □

Lemma 3. *Suppose that (H1) holds and the sequence $\{x_k\}$ is generated by Algorithm 1. Then, for all $k \in \mathbb{N} \cup \{0\}$, we have $x_k \in L(x_0)$ and the sequence $\{NF_{l(k)}\}$ is decreasing and convergent.*

Proof. Using the definition of $NF_{l(k)}$, we have

$$\|F_k\| \leq NF_{l(k)}.$$

By induction, the result evidently holds for $k = 0$ because $NF_{l(0)} = \|F_0\|$. Assuming $x_i \in L(x_0)$ for $i = 1, 2, \ldots, k$, we show that $x_{k+1} \in L(x_0)$, for all $k \in \mathbb{N}$. It can be seen

$$\frac{NF_{l(k)}^2}{2} - \frac{\|F_{k+1}\|^2}{2} = f_{l(k)} - f_{k+1} \geq \mu(m_k(x_k) - m_k(x_k + d_k)) > 0,$$

so

$$\|F_{k+1}\| \leq NF_{l(k)} \leq \|F_0\|.$$

Thus, the sequence $\{x_k\}$ is contained in $L(x_0)$. It will be proved that the sequence $\{NF_{l(k)}\}$ is a decreasing sequence. We consider two following cases:

i) $k \geq N$. In this case, we have $m(k) = N$. So, the definition of $NF_{l(k)}$ along with this fact that $\|F_{k+1}\| \leq NF_{l(k)}$ implies that

$$\begin{aligned} NF_{l(k+1)} &= \max_{0 \leq j \leq N} \{\|F_{k+1-j}\|\} \leq \max\{\max_{0 \leq j \leq N} \{\|F_{k-j}\|\}, \|F_{k+1}\|\} \\ &= \max\{NF_{l(k)}, \|F_{k+1}\|\} = NF_{l(k)}. \end{aligned}$$

ii) $k < N$. In this case, we have $m(k) = k$. For any k, $\|F_k\| \leq \|F_0\|$,

$$NF_{l(k)} = F_0, \quad \forall k.$$

These cases show that the sequence $\{NF_{l(k)}\}$ is a decreasing sequence. According to assumption H_1 and $x_k \in L(x_0)$, one can see that the sequence $\{NF_{l(k)}\}$ is convergent. □

By Lemma 3 and since $f(x_k) = \frac{1}{2}\|F(x_k)\|^2$, we can conclude that the sequence $\{f_{l(k)}\}$ is also decreasing and convergent.

3 Convergence theory

In this section, we provide the global convergence and q-quadratic rate of results of the proposed algorithm.

Lemma 4. *Suppose that $\{x_k\}$ is the sequence generated by Algorithm 1. Then, we have*

$$\lim_{k \to \infty} NF_{l(k)} = \lim_{k \to \infty} \|F(x_k)\|.$$

Proof. By Lemma 3.2 in [1] and $f(x_k) = \frac{1}{2}\|F(x_k)\|^2$, we have

$$\lim_{k \to \infty} f_{l(k)} = \lim_{k \to \infty} f(x_k).$$

This implies that

$$\lim_{k \to \infty} NF_{l(k)} = \lim_{k \to \infty} \|F(x_k)\|.$$

□

In order to establish the global convergence of Algorithm 1, one needs to establish the following results.

Lemma 5. *Suppose that assumptions (H2) and (H3) hold, the sequence $\{x_k\}$ is generated by Algorithm 1 and d_k is a solution of the subproblem (3). Then, we have*

$$m_k(x_k) - m_k(x_k + d_k) \geq L_k \ \|F_k\|^2, \tag{14}$$

where $L_k := \beta M_0 \min\left\{c^{p_k}, \frac{M_0}{M_1^2}\right\}$.

Proof. Using (H2) and (11), we have

$$\begin{aligned} m_k(x_k) - m_k(x_k + d_k) &\geq \beta \|J_k^T F_k\| \min\left\{\Delta_k, \frac{\|J_k^T F_k\|}{\|J_k^T J_k\|}\right\} \\ &= \beta \|J_k^T F_k\| \min\left\{c^{p_k} F_{l(k)}, \frac{\|J_k^T F_k\|}{\|J_k^T J_k\|}\right\} \\ &\geq \beta M_0 \|F_k\| \min\left\{c^{p_k} \|F_k\|, \frac{M_0 \|F_k\|}{M_1^2}\right\} \\ &\geq \beta M_0 \|F_k\|^2 \min\left\{c^{p_k}, \frac{M_0}{M_1^2}\right\} \\ &= \ L_k \|F_k\|^2, \end{aligned}$$

where $L_k = \beta M_0 \min\left\{c^{p_k}, \frac{M_0}{M_1^2}\right\}$. Therefore, the proof is complete. □

At this point, the global convergence of Algorithm 1 based on the mentioned assumptions can be investigated.

Theorem 6. *Suppose that Assumptions (H1)-(H4) hold. Then, Algorithm 1 either stops at a stationary point of $f(x)$ or generates an infinite sequence $\{x_k\}$ such that*

$$\lim_{k\to\infty} \|F_k\| = 0. \tag{15}$$

Proof. By contradiction, let there exists a constant $\epsilon > 0$ and an infinite subset $K \subseteq \mathbb{N}$ satisfying

$$\|F_k\| > \epsilon, \quad \text{for all} \ \ k \in K. \tag{16}$$

Using (16), $\widehat{r}_k > \mu$ and Lemma 5, we can conclude that

$$f_{l(k)} - f_{k+1} = f_{l(k)} - f(x_k + d_k) \geq \mu[m_k(x_k) - m_k(x_k + d_k)] \geq \mu \|F_k\|^2 L_k \geq \mu \epsilon^2 L_k.$$

The left-hand side of above inequality tends to become zero when k goes to infinity and therefore L_k tends to 0. This means that $p_k \to \infty$ that clearly is a contradiction with Lemma 2. Therefore, the hypothesis (16) is not true and the proof is complete. □

To establish the quadratic convergence rate of the sequence generated by Algorithm 1, an additional assumption is required as follows (see [11, 14, 15, 40, 44]).

(H5) There exist constants $c_1 \geq 1$ and $\rho_1 \in (0, 1)$ such that

$$c_1\|x - x_*\| \leq \|F(x)\| = \|F(x) - F(x_*)\|, \quad \forall x \in N(x_*, \rho_1).$$

where x_* is a solution of (1) and $N(x_*, \rho_1) := \{x| \ \|x - x_*\| \leq \rho_1\}$.

Remark 2. By (H1) and (H4), the objective function $F(x)$ is continuously differentiable and $J(x)$ is Lipschitz continuous. So, there exist two constants $\gamma_L > 0$ and $\rho_2 \in (0, 1)$ such that

$$\|F(x) - F(y) + J(x)(x - y)\| \leq \gamma_L\|x - y\|^2, \quad \text{for all} \ \ x, y \in N(x_*, \rho_2).$$

For the purpose of our q-quadratic convergence, we simply choose $\rho := \min[\rho_1, \rho_2]$.

Theorem 7. *Suppose that Assumptions (H1)-(H5) hold and the sequence $\{x_k\}$ generated by Algorithm 1 converges to x_*. Then, for k sufficiently large, we have*

$$x_{k+1} = x_k + d_k^0,$$

where d_k^0 is the solution of (3) corresponding to $p_k = 0$. Furthermore, the sequence $\{x_k\}$ converges to x_* q-quadratically.

Proof. Let d_k^0 be a solution corresponding to $p_k = 0$ of the subproblem (3) , so d_k^0 is a feasible solution for (3). This along with Lemma 3 and Theorem 1 imply

$$\|d_k^0\| \leq \Delta_k^0 = NF_{l(k)} \to 0, \quad as \ \ k \to \infty. \tag{17}$$

On the other hand, since $p_k = 0$ and $M_0 \leq 1 \leq M_1$, we obtain

$$L_k := \frac{\beta M_0^2}{M_1^2}. \tag{18}$$

Because of the fact that Algorithm 1 is not stopped, it is clear that we have $\|F_k\| \geq \epsilon$. This fact together with Lemma 2, (17) and (18) suggests that

$$\begin{aligned}\left|\frac{f_k - f(x_k + d_k^0)}{m_k(x_k) - m_k(x_k + d_k^0)} - 1\right| &= \left|\frac{m_k(x_k + d_k^0) - f(x_k + d_k^0)}{m_k(x_k) - m_k(x_k + d_k^0)}\right| \\ &\leq \frac{O(\|d_k^0\|^2)}{L_k\|F_k\|^2} \leq \frac{O((\Delta_k^0)^2)}{\frac{\beta M_0^2}{M_1^2}\epsilon^2} \to 0, \quad \text{as} \ \ k \to \infty.\end{aligned}$$

This along with the fact $f_{l(k)} \geq f_k$, for sufficiently large k, implies

$$\widehat{r}_k \geq \mu.$$

Thus, for all sufficiently large k, the trial point d_k^0 is accepted by Algorithm 1, i.e. $x_{k+1} = x_k + d_k^0$.

At this point, the quadratic convergence of the sequence $\{x_k\}$ generated by Algorithm 1 is investigated. Regarding (H2), there exists a constant $M_1 > 0$ such that

$$\|J_k\| \leq M_1, \quad \text{for all } x \in \Omega. \tag{19}$$

Using (19) along with the mean value theorem, for all $x_k \in N(x_*, \rho)$, we can easily see that

$$\|F_k\| = \|F_k - F(x_*)\| = \|J(\xi)\| \; \|x_k - x_*\| \leq M_1 \|x_k - x_*\|, \tag{20}$$

for some $\xi \in [x_k, x_*]$. As a result of this fact and Lemma 4, for any sufficiently large k, it can be concluded that

$$F_{l(k)} \leq M_1 \|x_k - x_*\|,$$

and so

$$\|d_k^0\| \leq N F_{l(k)} \leq M_1 \|x_k - x_*\|. \tag{21}$$

From (H5), it is clear that

$$\|x_k - x_*\| \leq \frac{1}{c_1} \|F_k\| \leq \frac{1}{c_1} N F_{l(k)} \leq N F_{l(k)} = \Delta_k^0.$$

This fact directly implies that $x_k - x_*$ is a feasible point for the subproblem (3). Now, it is straightforwardly followed from Remark 2 and (21) that

$$\begin{aligned} \frac{1}{2}\|F_k + J_k d_k^0\|^2 = m_k(x_k + d_k^0) &\leq m_k(x_k + (x_* - x_k)) \\ &= \frac{1}{2}\|F(x_k + J_k(x_k - x_*))\|^2 \\ &= \frac{1}{2}\|F_k - F_* + J_k(x_k - x_*)\|^2 \\ &\leq \frac{\gamma_L^2}{2}\|x_k - x_*\|^4. \end{aligned} \tag{22}$$

Also (H6), (21) and (22), give us

$$\begin{aligned} c_1\|x_{k+1} - x_*\| \leq \|F(x_{k+1})\| &= \|F(x_k + d_k^0)\| \\ &\leq \|F_k + J_k d_k^0\| + O(\|d_k^0\|^2) \\ &\leq \gamma_L \|x_k - x_*\|^2, \end{aligned}$$

for any sufficiently large k. So

$$\|x_{k+1} - x_*\| = O(\|x_k - x_*\|^2).$$

Hence, the sequence $\{x_k\}$ generated by the Algorithm 1 is q-quadratically convergent. Therefore, the proof is completed. □

4 Numerical experiments

In this section, we report some numerical experiments obtained by running Algorithm 1 (NATR) in comparison with the nonmonotone trust-region algorithm (NTR), the adaptive trust-region algorithm from Zhang et al. in [44] (ATRZ), the nonmonotone version of it (NATRZ), the adaptive trust-region algorithm of Fan and Pan in [15] (ATRF) and the nonmonotone version of it (NATRF) on a set of nonlinear systems of equations with the dimension from 100 to 504 that are selected from the wide range of literatures . The problems 1-36 are chosen from Cruz et al. in [25] and the problems 37-42 are chosen from Lukšan and Vlček in [28]. For all of these codes, the trust-region subproblems are solved by Steihaug-Toint procedure, see [8]. The Steihaug-Toint algorithm terminates at $x_k + d$ when

$$\|\nabla m_k(x_k + d)\| \leq \min\left\{\frac{1}{10}, \|\nabla m_k(x_k + d)\|^{\frac{1}{2}}\right\}\|\nabla m_k(x_k + d)\|.$$

The Jacobian matrix J_k can be either evaluated analytically by a user-supplied function or approximated using finite-differences formula provided by the code. Since the exact computation cannot be appropriate for large scale problems, similar to [5], we used the following finite-differences formula to approximate the Jacobian matrix J_k

$$[J_k]_{\cdot j} \sim \frac{1}{h_j}(F(x_k + h_j e_j) - F_k),$$

where $[J_k]_{\cdot j}$ denotes the j-th column of J_k, e_j is the j-th vector of the canonic basis and

$$h_j := \begin{cases} \sqrt{\epsilon_m} & \text{if } \; x_{k_j} = 0, \\ \sqrt{\epsilon_m}\, sign(x_{k_j}) \max\{|x_{k_j}|, \frac{\|x_k\|_1}{n}\} & \text{otherwise.} \end{cases}$$

All codes are written in MATLAB 9 programming environment with double precision format in the same subroutine. In our numerical experiments, the algorithms were stopped whenever

$$\|F_k\| \leq 10^{-5},$$

or when the total number of iterates exceeded 1000. During implementations, It is checked that the codes be converged to the same point and only provided data for problems that all algorithms converged to the identical point while less than of 1 percent of problem was ignored.

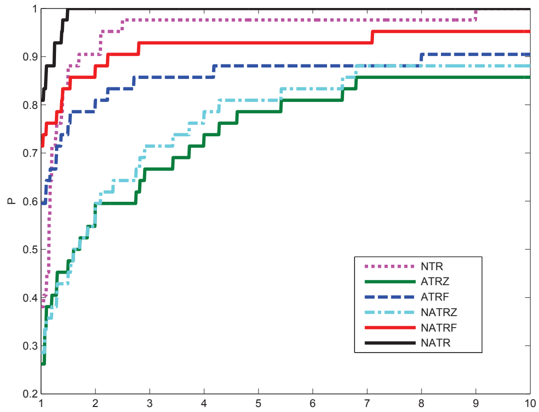

Figure 1: Iterates performance profile for the presented algorithms

While NATR algorithm takes advantages of the parameters $\mu = 10^{-6}$, $c = 0.5$. The NTR algorithm employs the parameters $\mu_1 = 0.1$, $\mu_2 = 0.9$ and updates trust-region radius like [8] by the following formula

$$\Delta_{k+1} := \begin{cases} c_1\|d_k\| & \text{if } r_k < \mu_1, \\ \Delta_k & \text{if } \mu_1 \leq r_k \leq \mu_2, \\ c_2\Delta_k & \text{if } r_k \geq \mu_2, \end{cases}$$

where $c_1 = 0.25$ and $c_2 = 0.3$. We also decide to follow the literature [34] in exploiting $\Delta_0 = 1$ as an initial trust-region radius for NTR. The parameters of ATRZ and ATRF have been chosen the same as in articles [44] and [15], respectively. Table 1 indicates the names and dimensions of the test problems considered. Figures 1 and 2 give the performance profiles for all of the algorithms with the choice of finite-differences Jacobian matrix for total number of iterations and total number of function evaluations, respectively. Performance profile gives, for every $\tau \geq 1$, the fraction of the number of problems for which the algorithm is within a factor of τ of the best [10].

Figure 1 clearly indicates that NATR outperforms NTR, ATRZ, ATRF, NATRZ and NATRF regarding the total number of iterates. In particular, NATR has the most wins in nearly 81% of the test problems with the greatest efficiency. Meanwhile, in the sense of the ability of completing a run successfully, it is the best among considered algorithms because it grows up faster than the others and reaches 1 more rapidly. However, as illustrated in Figure 2, NATR implements are remarkably better than the others where it has the most wins for almost 77% of performed tests concerning the total number of function evaluations. Furthermore, Figures 1 and 2 show similar patterns in

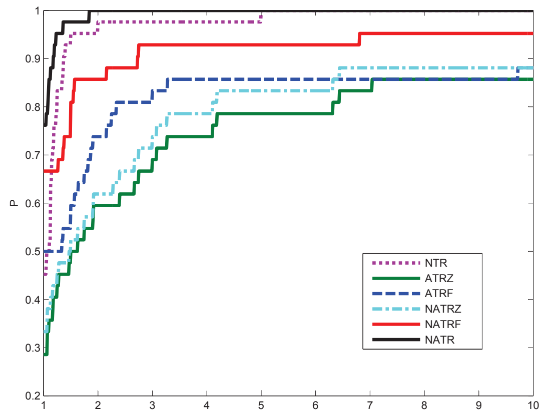

Figure 2: Function evaluations performance profile for the presented algorithms

the sense of the ability of completing a run successfully. As a result, this fact directly implies that the total number of solving the trust-region subproblems is the notably decreased thanks to using the NATR algorithm. These results imply that the proposed algorithm is an efficient and robust approach for solving systems of nonlinear equations.

5 Concluding remarks

It is well-known that trust-region methods for solving systems of nonlinear equations have a remarkable numerical reliability as well as strong theoretical convergence properties. Practical experiments of the trust-region framework indicate that applying nonmonotone adaptive techniques for determining trust-region radius declines the number of solving subproblems and employing nonmonotone strategies increases the efficiency and robustness of the algorithm. In this paper, by exploiting an effective adaptive trust region radius based on a reliable nonmonotone strategy, a new nonmonotone trust region algorithm is introduced for solving systems of nonlinear equations. Nevertheless, these modifications in the traditional trust-region procedure are favorably encouraging so that the global and q-quadratic convergence properties of the proposed algorithms are established. Numerical results on a set of nonlinear systems indicate that the number of iterates and the number of function evaluations are so close to each other that, by

Table 1: List of test functions

Problem name	Dim	Problem name	Dim
Exponential 1	500	Geometric	100
Exponential 2	500	Function 27	500
Extended Rosenbrock	500	Tridimensional valley	501
Chandrasekhar's H-equation	500	Complementary	500
Singular	500	Hanbook	500
Logarithmic	500	Tridiagnal system	500
Broyden tridiagonal	500	Five-diagonal system	500
Trigexp	500	Seven-diagonal system	504
Variable band 1	500	Extended cragg and levy	500
Variable band 2	500	Extended Wood	500
Function 15	500	Triadiagnal exponential	500
Strictly convex 1	500	Brent	500
Strictly convex 2	500	Thorech	500
Function 18	501	Broyden banded	500
Zero Jacobian	500	Discrete integral equation	500
Geometric programming	100	Countercurrent reactors 1	504
Function 21	501	Singular Broyden	500
Linear function-full rank 1	500	Structured Jacobian	500
Linear function-full rank 2	500	Extended Powell Singular	500
Brown almost linear	500	Generalized Broyden banded	500
Variable dimensioned	500	Extended powell badly scaled	500

applying the proposed algorithm, significant profits in computational costs can be obtained.

References

1. Ahookhosh, M. and Amini K. *A nonmonotone trust-region method with adaptive radius for unconstrained optimization*, Computers and Mathematics with applications 60 (2010) 411–422.

2. Ahookhosh, M., Amini K. and Peyghami, M.R. *A nonmonotone trust-region line search method for large-scale unconstrained optimization*, Applied Mathematical Modelling 36 (2012) 478–487.

3. Ahookhosh, M., Amini, K. and Kimiaei, M. *A globally convergent trust-region method for large-scale symmetric nonlinear systems*, Numerical Functional Analysis and Optimization 36 (2015) 830–855.

4. Ahookhosh, M., Esmaeili, H. and Kimiaei, M. *An effective trust-region-based approach for symmetric nonlinear systems*, International Journal of Computer Mathematics 90 (2013) 671–690.

5. Bellavia, S., Macconi, M. and Morini, B. *STRSCNE: A scaled trust-region solver for constrained nonlinear equations*, Computational Optimization and Applications 28, (2004) 31–50.

6. Bouaricha, A. and Schnabel, R.B. *Tensor methods for large sparse systems of nonlinear equations*, Mathematical Programming 82 (1998) 377–400.

7. Broyden, C.G. *The convergence of an algorithm for solving sparse nonlinear systems*, Mathematics of Computation 25 (114) (1971) 285–294.

8. Conn, A.R., Gould, N.I.M. and Toint, Ph.L. *Trust-Region Methods*, Society for Industrial and Applied Mathematics SIAM, Philadelphia, 2000.

9. Deng, N.Y., Xiao, Y. and Zhou, F.J. *Nonmonotonic trust region algorithm*, Journal of Optimization Theory and Applications 26 (1993) 259–285.

10. Dolan, E.D. and Moré, J.J. *Benchmarking optimization software with performance profiles*, Mathematical Programming 91 (2002) 201–213.

11. Esmaeili, H. and Kimiaei, M. *A new adaptive trust-region method for system of nonlinear equations*, Appl. Math. Model. 38 (11–12) (2014) 3003–3015.

12. Esmaeili, H. and Kimiaei, M. *An efficient adaptive trust-region method for systems of nonlinear equations,* International Journal of Computer Mathematics 92 (2015) 151–166.

13. Fan, J.Y. *Convergence rate of the trust region method for nonlinear equations under local error bound condition*, Computational Optimization and Applications. 34 (2005) 215–227.

14. Fan, J.Y. *An improved trust region algorithm for nonlinear equations*, Computational Optimization and Applications 48(1) (2011) 59–70.

15. Fan, J.Y. and Pan, J.Y *A modified trust region algorithm for nonlinear equations with new updating rule of trust region radius*, International Journal of Computer Mathematics 87 (14) (2010) 3186–3195.

16. Fasano, G., Lampariello, F. and Sciandrone, M. *A truncated nonmonotone Gauss-Newton method for large-scale nonlinear least-squares problems*, Computational Optimization and Applications 34 (3) (2006) 343–358.

17. Fischer, A., Shukla, P.K. and Wang, M. *On the inexactness level of robust Levenberg-Marquardt methods*, Optimization 59 (2) (2010) 273–287.

18. Grippo, L., Lampariello, F. and Lucidi, S. *A nonmonotone line search technique for Newton's method*, SIAM Journal on Numerical Analysis 23 (1986) 707–716.

19. Grippo, L., Lampariello, F. and Lucidi, S. *A truncated Newton method with nonmonotone linesearch for unconstrained optimization*, Journal of Optimization Theory and Applications 60 (3) (1989) 401–419.

20. Grippo, L., Lampariello, F. and Lucidi, S. *A class of nonmonotone stabilization method in unconstrained optimization*, Numer. Math. 59 (1991) 779–805.

21. Grippo, L. and Sciandrone, M. *Nonmonotone derivative-free methods for nonlinear equations*, Computational Optimization and Applications 37 (2007) 297–328.

22. Kanzow, C., Yamashita, N. and Fukushima, M. *Levenberg-Marquardt methods for constrained nonlinear equations with strong local convergence properties*, J Comput Appl Math. 172 (2004) 375–397.

23. Kanzow, C. and Kimiaei, M. *A new class of nonmonotone adaptive trust-region methods for nonlinear equations with box constraints*, Preprint 330, Institute of Mathematics, University of Wurzburg, Wurzburg, June 2015. http://www.mathematik.uni-wuerzburg.de/ kanzow/paper/NonmonotoneTR_FinalP.pdf

24. La Cruz, W. and Raydan, M. *Nonmonotone spectral methods for large-scale nonlinear systems*, Optimization Methods and Software 18 (5) (2003) 583–599.

25. La Cruz, W., Venezuela, C., Martínez, J.M. and Raydan, M. *Spectral residual method without gradient information for solving large-scale nonlinear systems of equations: Theory and experiments*, Technical Report RT–04–08, July 2004.

26. Li, D.H. and Fukushima, M. *A derivative-free line search and global convergence of Broyden-like method for nonlinear equations*, Optimization Methods and Software. 13 (2000) 181–201.

27. Lukšan, L. and Vlček, J. *Sparse and partially separable test problems for unconstrained and equality constrained optimization*, Technical Report, No. 767 (1999).

28. Lukšan, L. and Vlček, J. *Truncated trust region methods based on preconditioned iterative subalgorithms for large sparse systems of nonlinear equations*, Journal of Optimization Theory and Applications 95 (3) (1997) 637–658.

29. Macconi, M., Morini, B. and Porcelli, M. *A Gauss–Newton method for solving bound–constrained underdetermined nonlinear systems*, Optimization Methods and Software 24 (2) (2009) 219–235.

30. Nesterov, YU. *Modified Gauss–Newton scheme with worst case guarantees for global performance*, Optimization Methods and Software 22 (3) (2007) 469–483.

31. Nocedal, J. and Wright, S.J. *Numerical Optimization*, Springer, NewYork (2006).

32. Powell, M.J.D. *Convergence properties of a class of minimization algorithms, in Nonlinear Programming*, O.L. Mangasarian, R.R. Meyer, and S.M. Robinson, eds., Academic Press, NewYork, 1975, pp. 1–27.

33. Shi, Z.J. and Guo, J.H. *A new trust region method with adaptive radius*, Computational Optimization and Applications, 213 (2008) 509–520.

34. Toint, Ph.L.*Numerical solution of large sets of algebraic nonlinear equations*, Mathematics of Computation 46 (173) (1986) 175–189.

35. Toint, Ph.L. *Non-monotone trust region algorithm for nonlinear optimization subject to convex constraints*, Math. Program. 77 (1997) 69–94.

36. Toint, Ph.L. *An assessment of nonmonotone line search technique for unconstrained optimization*, SIAM J. Sci. Comput. 17 (1996) 725–739.

37. Xiao Y. and Chu, E.K.W. *Nonmonotone trust region methods*, Technical Report 95/17, Monash University, Clayton, Australia, 1995.

38. Xiao, Y. and Zhou, F.J. *Nonmonotone trust region methods with curvilinear path in unconstrained optimization*, Computing 48 (1992) 303–317.

39. Yamashita, N. and Fukushima, M. *On the rate of convergence of the Levenberg-Marquardt method*, Computing 15 (2001) 239–249.

40. Yuan, G., Lu, S. and Wei, Z. *A new trust-region method with line search for solving symmetric nonlinear equations*, International Journal of Computer Mathematics 88 (10) (2011) 2109–2123.

41. Yuan, G.L., Wei, Z.X. and Lu, X.W. *A BFGS trust-region method for nonlinear equations*, Computing 92 (4) (2011) 317–333.

42. Yuan, Y. *Trust region algorithm for nonlinear equations*, Information 1 (1998) 7–21.

43. Zhang, H.C. and Hager, W.W. *A nonmonotone line search technique for unconstrained optimization*, SIAM journal on Optimization 14 (4) (2004) 1043–1056.

44. Zhang, J. and Wang, Y. *A new trust region method for nonlinear equations*, Mathematical Methods of Operation Research 58 (2003) 283–298.

45. Zhang, X.S., Zhang, J.L. and Liao, L.Z. *An adaptive trust region method and its convergence*, Science in China 45 (2002) 620–631.

46. Zhou, F. and Xiao, Y. *A class of nonmonotone sta bilization trust region methods*, Computing 53 (2) (1994) 119–136.

یک الگوریتم ناحیه اطمینان غیریکنوا با شعاع تطبیقی غیریکنوا برای حل دستگاه معادلات غیر خطی

کیوان امینی[1] ، حمید اسمعیلی [2] و مرتضی کیمیایی[3]

[1] دانشگاه رازی، دانشکده علوم، گروه ریاضی
[2] دانشگاه بوعلی سینا، گروه ریاضی
[3] دانشگاه آزاد اسلامی، واحد اسدآباد، گروه ریاضی

دریافت مقاله ۱۶ فروردین ۱۳۹۴، دریافت مقاله اصلاح شده ۱۳ تیر ۱۳۹۴، پذیرش مقاله ۲۹ مهر ۱۳۹۴

چکیده : در این مقاله یک روش ناحیه اطمینان غیریکنوا برای حل دستگاه های معادلات غیر خطی معرفی می گردد که از یک شعاع تطبیقی مناسب استفاده می کند. استفاده همزمان از تکنیک های غیر یکنوا و یک شعاع اطمینان مناسب می تواند کارایی روش های ناحیه اطمینان را به طرز قابل ملاحظه ای افزایش دهد جایی که هزینه محاسباتی روش نیز به دلیل کاهش تعداد زیر مسائل حل شده کاهش می یابد. همگرایی سراسری و q– مجذوری روش تحت شرایط مناسب اثبات گردیده است. نتایج عددی ارائه شده نمایانگر کارایی و سرعت مناسب الگوریتم جدید در مقایسه با الگوریتم های مشابه می باشد.

کلمات کلیدی : دستگاه معادلات غیرخطی؛ الگوریتم ناحیه اطمینان؛ شعاع تطبیقی؛ تکنیک های غیریکنوا.