A New Non-monotone Adaptive Trust Region Method for Unconstrained Optimization Problems

Changyuan Li[1], Qinghua Zhou[1,*], Xiao Wu[1], Xiaodian Sun[2]

[1]*College of Mathematics and Information Science, Hebei University, Baoding, Hebei Province, China, 071002.*

[2]*Laboratory for Statistical Genomics and Systems Biology, Department of Environmental Health, University of Cincinnati College of Medicine, 3223 Eden Ave. ML56, Cincinnati OH 45267-0056, USA*

*(mbydlcy@126.com, qinghua.zhou@gmail.com(*corresponding author), wuxiao616@163.com, sunxd@uc.edu)*

---***---

Abstract - ***In this paper, we propose and analyze a new non-monotone adaptive trust region method for unconstrained optimization problems. Actually, we combined a new strategy of non-monotone line search with the ratio of actual reduction and the predicted reduction within a traditional trust region method. Then some properties of the new algorithm are analyzed. Theoretical analysis shows that the new proposed method has a global convergence under some mild conditions.***

Keywords: *adaptive trust region method, unconstrained optimization, non-monotone approach, global convergence.*

1.INTRODUCTION

Consider the following unconstrained optimization problem:

$$\min f(x), \quad x \in R^n, \tag{1}$$

where $f: R^n \to R$ is a twice continuously differentiable function. Through out this paper, we use the following notation:

- $\|\cdot\|$ is the Euclidean norm.
- $g(x) \in R^n$ and $H(x) \in R^{n \times n}$ are the gradient and Hessian matrix of f evaluated at x, respectively.
- $f_k = f(x_k), g_k = g(x_k), H_k = \nabla^2 f(x_k)$ and B_k is a symmetric matrix which is either H_k or an approximation of H_k.

Traditional iterative methods for solving (1) are either line search method or trust region method. It is well known that trust region method is a kind of important and efficient methods for nonlinear optimization. This method is based on the following idea: at each iterate x_k, a trial step d_k is usually computed by solving the quadratic sub-problem:

$$\begin{aligned} \min \ & m_k(d) = f_k + g_k^T d + \frac{1}{2} d^T B_k d, \\ s.t. \ & \|d\| \le \Delta_k. \end{aligned} \tag{2}$$

A crucial issue in solving sub-problems is a strategy for choosing the trust radius Δ_k. In the traditional trust-region method, the radius Δ_k is determined based on a comparison between the model and the objective function. This leads the traditional trust-region method to define the following ratio

$$r_k = \frac{f(x_k) - f(x_k + d_k)}{pred_k}, \tag{3}$$

where the numerator is called the actual reduction and the denominator is called the predicted reduction which is defined as

$$pred_k = m_k(0) - m_k(d_k). \tag{4}$$

It is clear that there is an appropriate agreement between the model and the objective function over the current region whenever r_k is close to 1, so it is safe to expand the trust region radius r_k in the next iterate. In addition, if r_k is a so small positive number or a negative number, the agreement is not appropriate and so the trust region r_k should be shrunk.

It is well known that the standard trust region method is very sensitive on initial radius, see [2, 10]. Furthermore, the radius Δ_k in (2) is independent from any information about g_k and B_k. These facts cause an increase in the number of sub-problems in some questions that need solving which decreases the efficiency of these methods. In order to reduce the number of sub-problems that need solving, Zhang et al. proposed a strategy to determine the trust region radius [11]. Inspired by Zhang's method, Shi and Guo proposed a trust region method which can automatically adjust the trust region radius in [2]. They also proposed a new adaptive radius for the trust region method in [1]. Actually, they choose parameters $m, r \in (0,1)$, and q_k to satisfy the following angel condition

$$\cos\langle -g_k, q_k\rangle = -\frac{g_k^T q_k}{\|g_k\| \cdot \|q_k\|} \geq t, \qquad (5)$$

where $t \in (0,1]$, and set

$$s_k = \frac{g_k^T q_k}{q_k^T \overline{B_k} q_k}, \qquad (6)$$

in which $\overline{B_k}$ is generated by the procedure: $q_k^T \overline{B_k} q_k = q_k^T B_k q_k + i\|q_k\|^2$, and i is the smallest nonnegative integer such that

$$q_k^T \overline{B_k} q_k = q_k^T B_k q_k + i\|q_k\|^2 > 0, \qquad (7)$$

so, they proposed a new trust region radius as follows

$$\Delta_k = a_k \|q_k\|, \qquad (8)$$

where $a_k = r^p s_k$, and p is the least positive integer number so that

$$r_k \geq m, \qquad (9)$$

they proved that the new adaptive trust region method has global, super-linear and quadratic convergence properties and is a numerically efficient method.

On the other hand, in 1982, Chamberlain et al. in [9] proposed the watchdog technique for constrained optimization to overcome the Maratos effect. Motivated by this idea, Grippo et al. introduced a non-monotone line search technique for Newton's method in [14, 15]. Due to the high efficiency of non-monotone techniques, many authors are interested in working on the combination of non-monotone techniques and trust region methods [5, 12, 13, 16, 18]. Let

$$f_{l(k)} = f(x_{l(k)}) = \max_{0 \leq j \leq m(k)} \{f_{k-j}\}, \quad k = 0, 1, 2, \ldots \qquad (10)$$

where $m(k) = \min\{M, k\}$ and $M \geq 0$ is an integer constant. Actually, the most common non-monotone ratio is defined as follows:

$$\hat{r}_k = \frac{f_{l(k)} - f(x_k + d_k)}{pred_k}.$$

It is obviously that this ratio is more relaxed in comparison with (3) and some researchers showed that utilizing non-monotone techniques may improve both the possibility of finding the global optimum and the rate of convergence [14, 19]. However, although the non-monotone technique has many advantages, it contains some drawbacks [7, 8, 19]. To overcome those disadvantages, Ahookhosh et al. in [7, 8] proposed a new non-monotone technique to replace (10). They define

$$R_k = h_k f_{l(k)} + (1 - h_k) f_k, \qquad (11)$$

where $h_{\min} \in [0,1)$, $h_{\max} \in [h_{\min}, 1]$ and $h_k \in [h_{\min}, h_{\max}]$. At the same time, they have the new non-monotone ratio:

$$\overline{r}_k = \frac{R_k - f(x_k + d_k)}{pred_k}. \tag{12}$$

The form (11) has the following advantages when compared with (10):

- Fully employing useful properties of the current objective function value f_k.
- By choosing an adaptive h_k, we can own the better convergence results. Because it has been proved that the best convergence results are obtained by stronger non-monotone strategy when iterations are far from the optimum, and by weaker non-monotone strategy when iterations are close to it [19].

In recently years, some researchers have combined the adaptive trust region method with the non-monotone technique (10) and good numerical results have been achieved [3, 4, 6, 10]. However, the adaptive trust region method with the non-monotone technique (11) has not been studied, which is the focus of this paper. In this paper, we incorporate a more efficient adaptive trust region method proposed by Shi and Guo in [1] with the non-monotone technique (11) in order to propose the new non-monotone adaptive trust region method. The global convergence of this method is analyzed under some suitable conditions.

The rest of this paper is organized as follows. In Section 2, we introduce the new adaptive trust region method. In Section 3, we analyze the new method is well-defined and prove the global convergence. Some conclusions are given in Section 4.

2.NEW ADAPTIVE TRUST REGION METHOD

In this paper, we consider the following assumptions (the same with [6]) that will be used to analyze the convergence properties of the below new algorithm:

(H1) The objective function $f(x)$ has lower bound on R^n and the gradient $g(x) = \nabla f(x)$ of $f(x)$ is uniformly continuous on open convex set Ω that contains the level set $L_0 = \{x \in R^n \mid f(x) \le f(x_0)\}$, where $x_0 \in R^n$ is given.

(H2) The matrix B_k is a uniformly bounded matrix, i.e. there exists a constant $M_0 > 0$ such that $\|B_k\| \le M_0$ for all $k \in \infty$.

The new non-monotone adaptive trust region method can be described as follows:

Algorithm 1 (A new non-monotone adaptive trust region method)

Step 1 An initial point $x_0 \in R^n$ and a symmetric matrix $B_0 \in R^{n \times n}$ are given. The constants $0 < m < 1$, $0 < r < 1$, $M \ge 0$ and $e > 0$ are also given. Compute $f(x_0)$ and set $k = 0$ and $p = 0$.

Step 2 Compute g_k. If $\| g_k \| \le e$ then stop, else go to Step 3.

Step 3 Choose q_k to satisfy (5).

Step 4 Solve (2) to determine d_k, and set $\overline{x_{k+1}} = x_k + d_k$.

Step 5 Compute $m(k)$, $f_{l(k)}$, R_k, $pred_k$ and $\overline{r}_k$. If $\overline{r}_k < m$, then set $p = p + 1$ and go to Step 3.

Step 6 Set $x_{k+1} = \overline{x_{k+1}}$, generate B_{k+1} by a quasi Newton updating formula, set $k = k + 1$ and go to Step 2.

Obviously, in the case of $M = 0$, the new algorithm reduces to the adaptive trust region algorithm proposed by Shi and Guo [1]; in the other case of $h_k = 1$, it reduces to the non-monotone adaptive trust region algorithm presented in [3] and [6].

We need the following lemmas in order to prove the convergence of the new algorithm.

Lemma 1 (see [2]) Assume that Algorithm 1 generates an infinite sequence $\{x_k\}$. Then

$$pred_k \ge -\frac{1}{2} a_k g_k^T q_k.$$

Lemma 2 (see [11]) Suppose that the sequence $\{x_k\}$ be generated by Algorithm 1, then we have

$$\left| f(x_k) - f(x_k + d_k) - pred_k \right| = O(\|d_k\|^2).$$

Lemma 3 Suppose that (H1) and (H2) hold. Then Steps 4 and 5 of the new non-monotone adaptive trust region algorithm are well-defined, i.e., in each iteration, these steps are terminated after finite iterates.

Proof. It is similar to Lemma 2.3 in [6]. We can prove that for p sufficiently large, $r_k \ge \mu$ holds. Now, by (10), we have

$$\overline{r}_k = \frac{R_k - f(x_k + d_k)}{pred_k} = \frac{\eta_k f_{l(k)} + (1-\eta_k) f_k - f(x_k + d_k)}{pred_k} \ge \frac{f_k - f(x_k + d_k)}{pred_k} = r_k \ge \mu$$

Thus, Steps 4 and 5 of the new non-monotone adaptive trust region algorithm are well-defined.

Lemma 4 Suppose that the sequence $\{x_k\}$ is generated by Algorithm 1. Then, for all $k \in \mathbb{N}$, we have $x_k \in L(x_0)$ and $\{f_{l(k)}\}$ is a decreasing sequence.

Proof. Using definition of R_k and $f_{l(k)}$, we observe that

$$\begin{aligned} R_k &= \eta_k f_{l(k)} + (1-\eta_k) f_k \\ &\le \eta_k f_{l(k)} + (1-\eta_k) f_{l(k)} \\ &= f_{l(k)}. \end{aligned} \tag{13}$$

By induction, we will show that $x_k \in L(x_0)$, for all $k \in \mathbb{N}$. The result evidently holds for $k = 0$. Assume that $x_k \in L(x_0)$, then we show that $x_{k+1} \in L(x_0)$. From definition of Algorithm 1, we have that $r_k \ge \mu > 0$, so by $pred_k > 0$, we have

$$R_k \ge \mu pred_k + f_{k+1} \ge f_{k+1}. \tag{14}$$

So by (13) and (14), we know that

$$f_{k+1} \le R_k \le f_{l(k)} \le f_0, \quad \forall k \in \mathbb{N}. \tag{15}$$

Obviously, $x_{k+1} \in L(x_0)$, thus, the sequence $\{x_k\}$ is contained in $L(x_0)$.

The rest of proof that the sequence $\{f_{l(k)}\}$ is decreased is similar to Lemma 2.1 in [8].

Lemma 5 Suppose that the sequence $\{x_k\}$ be generated by the Algorithm 1, then we have

$$f_{k+1} \le R_{k+1}, \quad \forall k \in \mathbb{N}. \tag{16}$$

Proof. From the definition of $f_{l(k+1)}$, we have $f_{k+1} \le f_{l(k+1)}$, for any $k \in \mathbb{N}$. Hence, (16) holds by the following inequality

$$\begin{aligned} f_{k+1} &= \eta_{k+1} f_{k+1} + (1-\eta_{k+1}) f_{k+1} \\ &\le \eta_{k+1} f_{l(k+1)} + (1-\eta_{k+1}) f_{k+1}. \\ &= R_{k+1}, \quad \forall k \in \mathbb{N} \end{aligned}$$

Lemma 6 Suppose that (H1) holds and the sequence $\{x_k\}$ is generated by Algorithm 1. Then the sequence $\{f_{l(k)}\}$ is convergent.

Proof. Lemma 4 together with (H1) imply that

$$\exists l \ \ s.t. \ \forall n \in \mathbb{N} : l \le f_{k+n} \le f_{l(k+n)} \le \cdots \le f_{l(k+1)} \le f_{l(k)}$$

This shows that the sequence $\{f_{l(k)}\}$ is convergent.

3.GLOBAL CONVERGENCE

It is well-known that trust-region methods have strong global convergence [17, 19]. In this section, we discuss some convergence properties of the new trust region algorithm, and show that our proposed method has global convergence.

In order to attain the global convergence, we need some additional conclusions as follows:

Lemma 7 (See Lemma 3.1 in [6]) Suppose that $\{x_k\}$ is generated by Algorithm 1 and $\|d_k\| \le \Delta_k$, then there exists a constant $c > 0$ such that the trial step d_k satisfies

$$\|d_k\| \le c\|g_k\|.$$

Lemma 8 (See Lemma 7 in [7]) Suppose that the sequence $\{x_k\}$ be generated by Algorithm 1, then we have

$$\lim_{k\to\infty} f(x_{l(k)}) = \lim_{k\to\infty} f(x_k). \tag{17}$$

Corollary 9 Suppose that the sequence $\{x_k\}$ be generated by Algorithm 1, then we have

$$\lim_{k\to\infty} R_k = \lim_{k\to\infty} f(x_k). \tag{18}$$

Proof. From (13) and (16), we observe

$$f_k \le R_k \le f_{l(k)}.$$

This completes the proof by using Lemma 8.

Theorem 10 Suppose that (H1) and (H2) hold, then Algorithm 1 either stops at stationary point of $f(x)$ or generates an infinite sequence $\{x_k\}$ such that

$$\lim_{k\to\infty} \frac{-g_k^T q_k}{\|q_k\|} = 0. \quad (19)$$

Proof. If Algorithm 1 doesn't stop at a stationary point, we use the reduction to absurdity to prove (19) holds. Suppose that Algorithm 1 generates an infinite sequence $\{x_k\}$ and (19) doesn't holds. Which implies that there exist a $\varepsilon > 0$ and an infinite subset $K \subseteq \{0,1,2,\ldots\}$, such that

$$\frac{-g_k^T q_k}{\|q_k\|} \ge \varepsilon, \quad \forall k \in K. \quad (20)$$

From Remark 2.2 in [6], we have

$$\exists M_1 > 0 \quad s.t. \|\overline{B_k}\| \le M_1, \quad \forall k,$$

thus,

$$q_k^T \overline{B_k} q_k \le M_1 \|q_k\|^2, \quad \forall k. \quad (21)$$

Let $K_1 = \{k \in K | \alpha_k = s_k\}$ and $K_2 = \{k \in K | \alpha_k < s_k\}$. Obviously, $K = K_1 \cup K_2$ is an infinite subset of the set $\{0,1,2,\ldots\}$. Now, we prove that neither K_1 nor K_2 can be an infinite set which contradicts (20). To do so, we consider the following two cases:

Case 1: $k \in K_1$. Let K_1 be an infinite subset of K, by using of Lemma 1 and (21), we get

$$\begin{aligned} R_k - f(x_k + d_k) &\ge \mu pred_k \ge -\frac{1}{2}\mu\alpha_k g_k^T q_k \\ &\ge -\frac{1}{2}\mu s_k g_k^T q_k = \frac{1}{2}\mu \frac{(g_k^T q_k)^2}{q_k^T \overline{B_k} q_k} \\ &\ge \frac{\mu}{2M_1}\left(\frac{g_k^T q_k}{\|q_k\|}\right)^2 \ge \frac{\mu}{2M_1}\varepsilon^2, \quad k \in K_1 \end{aligned}$$

As $k \to \infty$, the above inequality along with Corollary 9 implies

$$0 \ge \frac{\mu}{2M_1}\varepsilon^2.$$

Which is a contradiction and shows that K_1 cannot be an infinite subset of K.

Case 2: $k \in K_2$. Let K_2 be an infinite subset of K, Lemma 1 implies

$$\begin{aligned} R_k - f(x_k + d_k) &\ge \mu pred_k \ge -\frac{1}{2}\mu\alpha_k g_k^T q_k \\ &= -\frac{1}{2}\mu\Delta_k \frac{g_k^T q_k}{\|q_k\|} \ge \frac{1}{2}\mu\Delta_k \varepsilon. \end{aligned}$$

As $k \to \infty$, the above inequality along with Corollary 9 implies

$$\lim_{k\to\infty} \Delta_k = 0, \quad k \in K_2. \quad (22)$$

Now, suppose that $\overline{d_k}$ is an optimal solution of the following sub-problem

$$\begin{aligned} \min\ & m_k(\overline{d_k}) = f_k + g_k^T \overline{d_k} + \frac{1}{2}\overline{d_k}^T B_k \overline{d_k}, \\ s.t.\ & \|\overline{d_k}\| \le \overline{\Delta_k} \end{aligned}.$$

where $\overline{\Delta_k} = \Delta_k / \rho$, $\overline{\alpha_k} = \alpha_k / \rho$.

Then, following the steps of Algorithm 1, we have

$$\frac{R_k - f(x_k + \overline{d_k})}{pred_k} < \mu, \quad k \in K_2. \quad (23)$$

On the other hand, (22) implies that

$$\lim_{k\to\infty} \overline{\Delta_k} = 0, \quad k \in K_2. \quad (24)$$

Now, using Lemma 1, Lemma 2, (20) and (24), we have

$$\begin{aligned} |r_k - 1| &= \left|\frac{f(x_k) - f(x_k + d_k) - pred_k}{pred_k}\right| \\ &= \frac{O(\|d_k\|^2)}{pred_k} \le \frac{O(\overline{\Delta}_k^{\,2})}{-\frac{1}{2}\overline{\alpha}_k g_k^T q_k} \\ &= \frac{O(\overline{\Delta}_k^{\,2})}{-\frac{1}{2}\overline{\Delta}_k g_k^T q_k \big/ \|q_k\|} \le \frac{O(\overline{\Delta}_k^{\,2})}{\frac{1}{2}\overline{\Delta}_k \varepsilon}, \quad k \in K_2. \end{aligned}$$

As $k \to \infty$, this inequality tends to zero. Therefore, in this case the monotone ratio is well defined. From the above facts and (16), we observe

$$\frac{R_k - f(x_k + d_k)}{pred_k} \ge \frac{f(x_k) - f(x_k + d_k)}{pred_k} \ge \mu. \quad (25)$$

Thus, we can indicate that the new non-monotone ratio is also well defined. However, for a sufficiently large $k \in K_2$, (25) contradicts (23), it shows that K_2 cannot be an infinite subset of K.

Therefore, there is no infinite subset of K such that (20) holds, so the proof is completed.

Theorem 11 Suppose that conditions of Theorem 10 holds and q_k satisfies (5), then the Algorithm 1 either stops finitely or generates an infinite sequence $\{x_k\}$ such that

$$\lim_{k \to \infty} \|g_k\| = 0.$$

Proof. The proof is similar to Theorem 3.4 in [6], we omit it for convenience.

4.CONCLUSIONS AND FUTURE WORKS

In this paper, we proposed a new non-monotone adaptive trust region method for unconstrained optimization problems. We analyzed the properties of the new algorithm and proved the global convergence theory under some mild conditions. When we investigate the nature of different non-monotone strategies, we think there are still some drawbacks if only the information of the maximum and the current iteration point be considered. In the near future, we will examine the effectiveness of different non-monotone strategies and design more proper ones.

ACKNOWLEDGEMENTS

This work is supported by the National Natural Science Foundation of China (61473111) and the National Science Foundation of Hebei Province (Grant No. A2014201003, A2014201100).

REFERENCES

[1] Z.J. Shi, J.H. Guo, *A new trust region methods with adaptive radius*, Comput Optim Appl (2008) 41: 225-242.

[2] Z.J. Shi, J.H. Guo, *A new trust region methods for unconstrained optimization*, Computational and Applied Mathematics 213 (2008) 509-520.

[3] Z.J. Shi, J.H. Guo, *Nonmonotone adaptive trust region method*, European Journal of Operational Research 208 (2011) 28-36.

[4] J.H. Fu, W.Y. Sun, *Nonmonotone adaptive trust-region method for unconstrained optimization problems*, Applied Mathematics and Computation 163 (2005) 489-504.

[5] W.Y. Sun, *Nonmonotone trust region method for solving optimization problems*, Applied Mathematics and Computation 156 (2004) 159-174.

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

[7] M. Ahookhosh, K. Amini, *An efficient nonmonotone trust-region method for unconstrained optimization*, Numer Algor(2012) 59: 523-540.

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

[9] R.M. Chamberlain, M.J.D. Powell, *The watchdog technique for forcing convergence in algorithm for*

constrained optimization, Mathematical Programming Study 16 (1982) 1-17.

[10] J.L. Zhang, X.S. Zhang, L.Z. Liao, *A nonmonotone adaptive trust region method and its convergence*, Computers and Mathematics with Applications 45 (2003) 1469-1477.

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

[12] Ph.L. Toint, *An assessment of nonmonotone linesearch technique for unconstrained optimization*, Society for Industrial and Applied Mathematics, 17 (1996) 725-739.

[13] Ph. L. Toint, *Non-monotone trust-region algorithm for nonlinear optimization subject to convex constraints*, Mathmatical Programming 77 (1997) 69-94.

[14] L. Grippo, F. Lampariello, S. Lucidi, *A nonmonotone line search technique for Newton's method*, Society for Industrial and Applied Mathematics 23 (1986) 707-716.

[15] L. Grippo, F. Lampariello, S. Lucidi, *A truncated Newton method with nonmonotone linesearch for unconstrained optimization*, Journal of Optimization Theory and Application 69 (1989) 401-419.

[16] J.T. Mo, K.C. Zhang, Z.X. Wei, *A nonmonotone trust region method for unconstrained optimization*, Applied Mathematics and Computation 171(2005) 371-384.

[17] J. Nocedal, Y.X. Yuan, *Combining trust region and line search techniques*, in: Y. Yuan (Ed.), Advanced in nonlinear programming, Kluwer Academic Publishers, Dordrecht, 1998, pp. 153-175.

[18] N.Y. Deng, Y. Xiao and F.J. Zhou, *Nonmonotone trust region algorithm*, Journal of Optimization Theory and Application 76 (1993) 259-285.

[19] H.C. Zhang, W.W. Hager, *A nonmonotone line search technique and its application to unconstrained optimization*, SIAM J. Optim. 14 (4) (2004) 1043-1056.

BIOGRAPHIES

Changyuan Li: Postgraduate student at the College of Mathematics and Information Science, Hebei University, China. His research interest includes non-monotone trust region methods and derivative-free methods.

Qinghua Zhou: Professor in HeBei University. Major interest includes Optimization Theory and its Applications.

Xiao Wu: Postgraduate student at College of Mathematics and Information Science, Hebei University, China. Her research interest mainly concerns non-monotone line search and trust region methods .

Xiaodian Sun: Research Scientist at University of Cincinnati College of Medicine. Her research interest includes dynamic modeling and inference, time series, statistical learning