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

Yajing Wang
College of Mathematics and Information Science, Hebei University, Baoding, 071002, China

Abstract: In this paper, we propose and analyze a new nonmonotone adaptive retrospective trust region method for unconstrained optimization problems. Actually, we incorporate a new proposed nonmonotone technology with the adaptive retrospective trust region method. 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: Nonmonotone technology; Unconstrained optimization; Retrospective trust region; Global convergence

1. Introduction

In this paper we consider the unconstrained minimization problem

$$\min_{x \in R^n} f(x) \tag{1}$$

where $f : R^n \to R$ is a twice continuously differentiable function.

For convenience, we use the following notation:

1 $\|.\|$ is the Euclidean norm.

2 $g(x) \in R^n$ and $H(x) \in R^{n\times n}$ are the gradient and Hessian of f at x respectively.

3 $f_k = f(x_k)$, $g_k = g(x_k)$, $H_k = H(x_k) = \nabla^2 f(x_k)$, and B_k be a symmetric matrix approximation of H_k.

There are various methods to solve the problem (1) most of which are iterative methods that are either line search methods or trust region methods. 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 trust step d_k is usually computed by solving the quadratic sub-problem:

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

its ratio defined by

$$r_k^B = \frac{Are\,d_k}{Pre\,d_k} \tag{3}$$

Where $Are\,d_k = f(x_k) - f(x_k + d_k)$ is called the actual reduction and $\Pr e\ d_k = m_k(0) - m_k(d_k)$ is called the predicted reduction.

Because of their strong convergence and robustness, trust region methods have been studied by many authors [1–3] and some convergence properties are given in the literature [4-7].

It notices that 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, Shi and Guo proposed a trust region method which can automatically adjust the trust region radius in [8]. They choose $u_1, \rho, \tau \in (0,1)$, and q_k to satisfy the following inequality

$$-\frac{g_k^T q^k}{\|g_k\|\|q_k\|} \ge \tau \tag{4}$$

and set

$$s_k = -\frac{g_k^T q^k}{q_k^T \hat{B}_k q_k} \tag{5}$$

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

$$q_k^T \hat{B}_k q_k = q_k^T B_k q_k + i\|q_k\|^2 > 0 \tag{6}$$

then, they proposed a new trust region radius as follows

$$\Delta_k = \alpha_k \|q_k\| \tag{7}$$

where $\alpha_k = \rho^p s_k$ and p is the smallest natural number such that

$$r_k^B \geq u_1 \quad (8)$$

They proved that the new adaptive trust region method has global, superlinear and quadratic convergence properties and is a numerically efficient method.

Recently, a retrospective trust region method has been proposed [9]. Its retrospective ratio is

$$r_{k+1}^R = \frac{f(x_k) - f(x_k + d_k)}{m_{k+1}(0) - m_{k+1}(d_k)} \quad (9)$$

Moreover, the classical ratio r_k^B and the retrospective ratio r_{k+1}^R are simultaneously used for updating the trust-region radius. More precisely, r_k^B is used to accept or reject the trial step, while r_{k+1}^R is employed for updating Δ_k after each successful iteration.

On the other hand, the monotone algorithm has a default that it may result to the slow iterative schemes for highly nonlinear or badly-scaled problems. To avoiding this limitation, the idea of nonmonotone strategies has been proposed to overcome the Martos effect for constrained optimization. More recently, Masoud Ahookhosh et al. present a novel nonmonotone term [10]

$$T_k = \begin{cases} f_{l(k)} & if\ k < N \\ \max\left\{\overline{T}_k, f_k\right\} & if\ k \geq N \end{cases} \quad (10)$$

Where

$$f_{l(k)} = \max_{0 \leq j \leq m(k)} \left\{f_{k-j}\right\} \quad (11)$$

$m(0) = 0$, $m(k) \leq \min\left\{m(k-1)+1, N\right\}$ for nonnegative integer N.

$$\overline{T}_k = \begin{cases} (1-\eta_{k-1})f_k + \eta_{k-1}\overline{T}_{k-1} & if\ k < N \\ (1-\eta_{k-1})f_k + \eta_{k-1}(1-\eta_{k-2})f_{k-1} + \cdots & \\ +\eta_{k-1}\cdots\eta_{k-N}f_{k-N} & if\ k \geq N \end{cases} \quad (12)$$

$\overline{T}_0 = f_0$, $\eta_i \in [0,1)$ for $i = 1,2,\cdots,k$. $\overline{T}_k$ is a convex combination of the collected function values F_k in [10].

This paper is organized as follows: In Section 2, we describe the new algorithm, in Section 3, we prove that the new algorithm is well defined, and then the global convergence is investigated. Some conclusions are delivered in Section 4.

2. The New Algorithm

In this section, we will propose our new method. Before that, we first introduce a new nonmonotone ratio [11].

$$r_{k+1}^{NC} = \lambda r_k^{NB} + (1-\lambda) r_{k+1}^{NR} \quad \lambda \in \left[\lambda_{\min}, \lambda_{\max}\right] \in (0,1) \quad (13)$$

Where

$$r_k^{NB} = \frac{(1+\phi_k)T_k - f(x_k + d_k)}{m_k(0) - m_k(d_k)} \quad (14)$$

$$r_{k+1}^{NR} = \frac{(1+\phi_k)T_k - f(x_k + d_k)}{m_{k+1}(0) - m_{k+1}(d_k)} \quad (15)$$

Where T_k is given by (10) and

$$\phi_k = \begin{cases} \eta_k & if\ T_k \geq 0 \\ 0 & if\ T_k < 0 \end{cases} \quad (16)$$

where $\{\eta_k\}$ is a positive sequence satisfying the following condition:

$$\sum_{k=1}^{\infty} \eta_k \leq \eta < \infty \quad (17)$$

Then we will introduce the updating of the radius Δ_k. For given $\mu_1 \in (0,1)$, if $r_k^{NB} \geq \mu_1$, then we set $x_{k+1} = x_k + d_k$, and update the radius by $\Delta_{k+1} = \min\left\{\tilde{\Delta}_{k+1}, \Delta_{\max}\right\}$, where

$$\tilde{\Delta}_{k+1} = \begin{cases} v_{k+1}s_{k+1}\|q_{k+1}\| & if\ r_{k+1}^{NC} \geq \mu_1 \\ \min\left\{\gamma_0\|d_k\|, v_{k+1}s_{k+1}\|q_{k+1}\|\right\} & if\ r_{k+1}^{NC} < \mu_1 \end{cases} \quad \gamma_0 \in (0,1) \quad (18)$$

and for $0 < \sigma_0 < 1 < \sigma_1$ and $\mu_2 \in (\mu_1, 1)$,

$$v_{k+1} = \begin{cases} \min\left\{\sigma_1 v_k, v_{\max}\right\} & if\ r_{k+1}^{NC} \geq \mu_2 \\ v_k & if\ \mu_1 \leq r_{k+1}^{NC} < \mu_2 \\ \sigma_0 v_k & if\ r_{k+1}^{NC} < \mu_1 \end{cases} \quad (19)$$

Otherwise, set $x_{k+1} = x_k$ and shrink the radius by $\Delta_{k+1} = min\left\{\gamma_0\|d_k\|, v_{k+1}s_{k+1}\|q_{k+1}\|\right\}$, where $v_{k+1} = \sigma_0 v_k$ and q_{k+1} and s_{k+1} satisfy (4) and (5), respectively. In both cases, B_{k+1} is updated by a quasi-Newton formula. This procedure is repeated until the stopping criteria hold.

Now we can propose the new algorithm.

Algorithm 1 New nonmonotone adaptive retrospective trust region method

Step 0: given $x_0 \in R^n$, $0 < \mu_1 < \mu_2 < 1$, $0 < \gamma_0 < 1$, $0 < \sigma_0 < 1 < \sigma_1$, $0 < \varepsilon_{\min} < \varepsilon_{\max} < 1$, $0 < \lambda_{\min} < \lambda_{\max} < 1$, $v_0 > 0$, $v_{\max} > 0$, $v_{\min} > 0$, $\varepsilon > 0$, a symmetric matrix $B_0 \in R^{n \times n}$, and a positive integer N, set $k = 0$.

Step 1: Choose q_k so that (4) holds and compute s_k from (5). If $k = 0$, then set $\Delta_k = \min\left\{v_k s_k \|q_k\|, \Delta_{\max}\right\}$ and go to Step 2. If $r_{k-1}^{NB} < \mu_1$, then set $v_k = \sigma_0 v_{k-1}$, and $\Delta_k = \min\left\{\gamma_0\|d_{k-1}\|, v_k s_k \|q_k\|\right\}$, and go to Step 2. Else, compute r_k^{NR} and r_k^{NC} by (15) and (13), respectively. Update v_k using (19), and set $\Delta_k = \min\left\{\tilde{\Delta}_k, \Delta_{\max}\right\}$, where $\tilde{\Delta}_k$ is given by (18).

Step 2: if $\|g_k\| < \varepsilon$, then Stop.

Step 3: find d_k by (approximately) solving (2) and compute r_k^{NB} by (14).

Step 4: if $r_k^{NB} \ge \mu_1$, then set $x_{k+1} = x_k + d_k$. Else, set $x_{k+1} = x_k$.

Step 5: Update B_{k+1} by a quasi-Newton formula. Set $k = k+1$ and go to Step 1.

3. Convergence Analysis

Throughout this paper, we use the following two index sets in our analysis

$$I = \{k : r_k^{NB} \ge \mu_1\}, \quad J = \{k : r_k^{NB} < \mu_1\}$$

We refer to the *kth* iteration as a successful iteration when $x_{k+1} = x_k + d_k$, i.e., $k \in I$.

In order to analyze the new algorithm, we consider the following assumptions:

H1: For given η as in (17), the level set $L = \{x \in R^n \,|\, f(x) \le e^{\eta} f(x_0)\}$ is closed and bounded and $f(x)$ is a twice continuously differentiable function over L .

H2: B_k is uniformly bounded, i.e., there exists a positive constant M , so that $\|B_k\| \le M$, for all k .

Remark 1: Using Algorithm 2.6 in [12], one can approximately solve the sub-problem (2), so that a sufficient reduction is achieved in the model function, i.e.,

$$\mathrm{Pr}\,e(d_k) \ge \theta \|g_k\| \min\left\{\Delta_k, \frac{\|g_k\|}{\|B_k\|}\right\}$$

Where $\theta \in (0,1)$ is a constant.

Remark 2: using algorithm 1 in [10], we can obtain

$$f_k \le T_k \le f_{l(k)} \quad \text{for all } k \in N \cup \{0\}$$

Theorem 3: suppose that $\{x_k\}$ is generated by algorithm 1, when $T_k < 0$, then the sequence $f_{l(k)}$ is decreasing.

Proof. we consider two cases: 1) $k \in I$, 2) $k \in J$. In case 1), if $k \in I$,by the definition of I , we have

$$r_k^{NB} = \frac{(1+\phi_k)T_K - f(x_k + d_k)}{m_k(0) - m_k(d_k)} \ge u_1$$

$$u_1 \le \frac{(1+\phi_k)T_K - f(x_k + d_k)}{m_k(0) - m_k(d_k)} \le \frac{(1+\phi_k) f_{l(k)} - f(x_k + d_k)}{m_k(0) - m_k(d_k)}$$

$$= \frac{(1+\phi_k) f_{l(k)} - f(x_{k+1})}{m_k(0) - m_k(d_k)}$$

By the definition of μ_1 , we can have

$$\begin{aligned}&(1+\phi_k) f_{l(k)} - f(x_k + d_k) \ge u_1(m_k(0) - m_k(d_k)) \ge 0 \\ &\Rightarrow (1+\phi_k) f_{l(k)} - f(x_k + d_k) \ge 0\end{aligned} \quad (20)$$

Because $T_k < 0$, by the equation (16), we have

$$f_{l(k)} \ge f(x_k + d_k) = f_{k+1} \quad (21)$$

By the definition of T_k , $f_{l(k)}$ and equation (21), if $k \ge N$, we have

$$f_{l(k+1)} = \max_{0 \le j \le m(k+1)} \{f_{k-j+1}\} \le \max_{0 \le j \le m(k)+1} \{f_{k-j+1}\}$$

$$= \max\{f_{l(k)}, f_{k+1}\} \le f_{l(k)}$$

If $k < N$, we have $m(k) = k$, since for any k , we have $f_k \le f_0$, it is clear that $f_0 = f_{l(k)}$, the result holds.

In case 2), if $k \in J$, by the definition of J , we can have $x_k = x_{k+1}$, $f_k = f_{k+1}$, $f_{l(k)} = f_{l(k+1)} \le f_{l(k)}$, so in both cases, the result holds.

Lemma 4: Suppose that the sequence $\{x_k\}$ be generated by algorithm 1, then we have

$$|f(x_k) - f(x_k + d_k) - \mathrm{Pr}\,e\, d_k| = O(\|d_k\|^2) \quad (22)$$

Proof. We can obtain the result by using Taylor's expansion and Assumption H2.

Lemma 5: Let Assumptions H1 and H2 hold and $\{x_k\}$ be the sequence generated by Algorithm 1. Moreover, assume that there exists a constant $\delta \in (0,1)$, so that $\|g_k\| > \delta$, for all k . Then, for every k , there exists a nonnegative integer p , so that x_{k+p+1} is a successful iteration point, i.e., $k + p \in I$.

Proof. It is similar to lemma 4.2 in [11].

Lemma 6: For the sequence $\{x_k\}$, generated by Algorithm 1, we have $\{x_k\} \subseteq L$

Proof. Assume that $x_i \in L$, for all $i = 1, 2, \cdots, k$, then we show $x_{k+1} \in L$. To do so, we consider two cases.

In case 1, we assume $k \in J$, then we have $f(x_k) = f(x_{k+1})$, using H1, the result obviously holds; in case 2, we assume $k \in I$, using remark 2, we can obtain $f_{k+1} \le T_{k+1} \le f_{l(k+1)} \le f_0 \le e^{\eta} f_0$

So by the definition of L and η , we can obtain the result.

Theorem 7: Let $\{x_k\}$ be the sequence generated by Algorithm1. Then, we have:

$$f_{k+1} \le |f_0| \prod_{i=0}^{k} (1+\phi_i) - \omega_k \qquad \forall k \in I \quad (23)$$

Where $\omega_k = \theta \mu_1 \|g_k\| \min\left\{\Delta_k, \frac{\|g_k\|}{\|B_k\|}\right\}$ and θ is the same constant as mentioned in Remark 1.

Proof. Let $k \in I$, by the definition of I , we can have that $x_{k+1} = x_k + d_k$, using remark 1, we can have

$$(1+\phi_k)T_k - f(x_{k+1}) \geq \mu_1 \Pr ed_k$$
$$\geq \theta\mu_1 \|g_k\| \min\left\{\Delta_k, \frac{\|g_k\|}{\|B_k\|}\right\} \geq 0 \qquad (24)$$

without loss of generality, we assume that the first iteration is a successful iteration. i.e., $k=0$. Thus, using the definition of T_k and equation (24), we have

$$f_1 \leq (1+\phi_0)T_0 - \omega_0 = (1+\phi_0)f_0 - \omega_0 \leq (1+\phi_0)|f_0| - \omega_0$$

By induction hypothesis, let (23) hold for $1 \leq k \in I$, i.e.,

$$f_{k+1} \leq |f_0| \prod_{i=0}^{k}(1+\phi_i) - \omega_k \leq |f_0| \prod_{i=0}^{k}(1+\phi_i)$$

Due to lemma 5, there exists a positive integer p, so that $p+k$ is a successful iteration. We show that (23) holds for $k+p \in I$, i.e.,

$$f_{k+p+1} \leq |f_0| \prod_{i=0}^{k+p}(1+\phi_i) - \omega_{k+p} \qquad (25)$$

For this purpose, because the iterations $k+1, k+2, \cdots, k+p-1$ are unsuccessful iterations, we have $T_{k+1} = T_{k+2} = \cdots = T_{k+p}$, using remark 2 and equation (24), and lemma 6 , we have

$$f_{k+p+1} \leq (1+\phi_{k+p})T_{k+p} - \omega_{k+p} = (1+\phi_{k+p})T_{k+1} - \omega_{k+p}$$
$$\leq (1+\phi_{k+p}) f_{l(k+1)} - \omega_{k+p} \qquad (26)$$
$$\leq (1+\phi_{k+p})|f_0| \prod_{i=0}^{l(k+1)-1}(1+\phi_i) - \omega_{k+p}$$

By the definition of $l(k)$ and equation (26), we can have

$$f_{k+p+1} \leq (1+\phi_{k+p})|f_0| \prod_{i=0}^{k+p-1}(1+\phi_i) - \omega_{k+p}$$
$$= |f_0| \prod_{i=0}^{k+p}(1+\phi_i) - \omega_{k+p}$$

The proof has been completed.

Theorem 8: Let q_k be chosen so that (4) holds and $\lim_{k\to\infty} -\frac{g_k^T q_k}{\|q_k\|} = 0$. Then, we have $\lim_{k\to\infty}\|g_k\| = 0$

Proof. See [11] for reference.

4. Conclusions

In this paper, a new retrospective trust region method for solving unconstrained optimization problems is proposed. Our approach employs an adaptive rule for updating the radius and a relaxed nonmonotone technique in its structure. As the nonmonotone strategies have shown their efficiency in the structure of optimization methods, we construct the nonmonotone versions of the classical and retrospective ratios according to the relaxed nonmonotone term. Besides, under some suitable and standard assumptions, we analyzed the properties of the algorithm and proved the global convergence theory.

References

[1] Fu J., Sun W. Nonmonotone Adaptive Trust-region Method for Unconstrained Optimization Problems. Journal of Changsha Communications University, 2008, 163 (1), 489-504.

[2] More J. J. Recent Developments in Algorithms and Software for Trust Region Methods. Springer Berlin Heidelberg, 1983, 258-287.

[3] Sun W. Nonmontone Trust Region Method for Solving Optimization Problems. Applied Mathematics & Computation, 2004,156 (1), 159-174.

[4] Deng N. Y., Xiao Y., Zhou F. J. Nonmonotonic Trust Region Algorithm. Journal of Optimization Theory & Applications, 1993, 76 (2), 259-285.

[5] Powell M. J. D. On the Global Convergence of Trust Region Algorithms for Unconstrained Minimization. Mathematical Programming, 1984, 29(3), 297-303.

[6] Zhang J. Z., Xu C. X. Trust Region Dogleg Path Algorithms for Unconstrained Minimization. Annals of Operations Research, 1999, 87(1), 407-418.

[7] Shultz G. A. Schnabel R. B., Byrd R. H. A Family of Trust-Region-based Algorithms for Unconstrained Minimization with Strong Global Convergence. Siam Journal on Numerical Analysis, 1985, 22 (1), 47-67.

[8] Shi Z. J., Guo J. H. A New Trust Region Method for Unconstrained Optimization. Elsevier Science Publishers B. V, 2008, 213(2), 509-520.

[9] Bastin F., Toint P. L., Malmedy V. et al. A Retrospective Trust-Region Method for Unconstrained Optimization. Mathematical Programming, 2010, 123(2), 395-418.

[10] Ahookhosh M., Ghaderi S. Two Globally Convergent Nonmonotone Trust-region Methods for Unconstrained Optimization. Journal of Applied Mathematics & Computing, 2016, 50 (1-2), 529-555.

[11] Tarzanagh D. A., Peyghami M. R., Bastin F. A New Nonmonotone Adaptive Retrospective Trust Region Method for Unconstrained Optimization Problems. Journal of Optimization Theory & Application, 2015, 167(2), 676-692.

[12] Nocedal J., Yuan Y. X. Combining Trust Region and Line Search Techniques. Springer US, 1998, 14, 153-175.