International Journal of Recent Engineering Science (IJRES),
ISSN: 2349-7157, *Volume 2 Issue 3 May to June 2015*

A New Non-monotone Self-Adaptive Trust Region Method based on simple conic model for Unconstrained Optimization

Weili Zheng[1], Qinghua Zhou[2], Liran Yang[3]
College of Mathematics and Information Science, Hebei University,
Baoding, 071002, Hebei Province, China

Abstract

In this paper, we propose and analyze a new non-monotone self-adaptive trust region method based on simple conic model for unconstrained optimization. Unlike the traditional non-monotone trust region method, the sub-problem in our method is a simple conic model, and the Hessian of the objective function is approximated by a scalar matrix. The trust region radius is adjusted with a new self-adaptive adjustment strategy, which makes use of the information of the previous iteration and current iteration.

Keywords: *large scale optimization, non-monotone technique, self-adaptive trust region method, conic model, global convergence*

I. INTRODUCTION

Consider the following unconstrained optimization problem:

$$\min\ f(x),\quad x \in R^n, \qquad (1)$$

Where $f: R^n \to R$ is a twice continuously differentiable function. Throughout this paper, we use the following notation: $\|.\|$ is the Euclidean norm. $g(x) = \nabla f(x) \in R^n$ 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.

The conic model was first proposed by Davidon [3] and Sorensen [4] in the following form:

$$\min_d \varphi(x_k + d) = f(x_k) + \frac{g_k^T d}{1 + b_k^T d} + \frac{1}{2}\frac{d^T B_k d}{(1 + b_k^T d)^2}. \qquad (2)$$

Due to the robust properties and the global convergence of the trust region methods, Di and Sun [1] proposed a conic trust region sub-problem for solving problem (1) as follow:

$$\min_d \varphi(x_k + d) = f(x_k) + \frac{g_k^T d}{1 + b_k^T d} + \frac{1}{2}\frac{d^T B_k d}{(1 + b_k^T d)^2}, \qquad (3)$$

$$s.t. \|d\| \le \Delta_k$$

where $b_k \in R^n$ is a horizontal vector. They put forward the necessary and effectual optimization conditions for the trust region sub-problems. Sun and Xu [2] proposed a filter trust region method based on conic model for unconstrained optimization.

Some researchers showed that utilizing non-monotone techniques improve both the possibility of finding the global optimum and the rate of convergence [5, 6]. The general non-monotone form follows:

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

Where $m(0) = 0,\ 0 \le m(k) \le \min\{M, m(k-1)+1\}$, and $M \ge 0$ is an integer constant.

However, although the non-monotone technique has many advantages, Zhang et al. [6] found that it still has some drawbacks and they proposed a new non-monotone form C_k. Gu et al. [7] introduced another non-monotone form in 2008 and the new form was computed easier than C_k. They define

$$D_k = \begin{cases} f(x_k) & k = 1; \\ \eta_k D_{k-1} + (1-\eta_k) f(x_k) & k \ge 2. \end{cases} \tag{4}$$

Where $\eta_k \in (0,1)$.

Based on an interpolation of the secant equation and on the Wolfe's line search conditions, Andrei [10] used the scalar matrix $\gamma_k I$ to approximate the Hessian matrix, and derived a new scaled conjugate gradient algorithm. Zhou and Zhang [8] proposed a non-monotone adaptive trust region method based on simple quadratic model for unconstrained optimization.

The rest of this paper is organized as follow. In section 2, we propose our non-monotone adaptive trust region method based on simple conic model for unconstrained optimization. The global convergences of the algorithm are established in section 3. Finally, we give some conclusions in section 4.

II. NON-MONOTONE ADAPTIVE TRUST REGION ALGORITHM BASED ON SIMPLE CONIC MODEL

A. Algorithm model

In this subsection, we discuss how to construct the simple conic model at each iteration. Like Andrei [10], at the k-th iteration, we consider using $\gamma_k I$ as an approximation of B_k, then, sub-problem (3) becomes

$$\begin{aligned} &\min_d c(x_k + d) = f(x_k) + \frac{g_k^T d}{1 + b_k^T d} + \frac{\gamma(x_k) d^T d}{2(1 + b_k^T d)^2}. \\ &s.t. \ \ \|d\| \le \Delta_k \end{aligned} \tag{5}$$

The conic model $c(x_k + d)$ should satisfy the following interpolation conditions [9]:

$$c(x_{k-1}) = f(x_{k-1}), \nabla c(x_{k-1}) = g_{k-1}, \tag{6}$$

$$\gamma(x_k)=\frac{2}{d_{k-1}^T d_{k-1}}\left[\mu^2\left(f\left(x_{k-1}\right)-f\left(x_k\right)\right)+\mu(1+\eta_k)g_k^T d_{k-1}\right].$$

Where $\mu=1-b_k^T d_{k-1}$, $\eta_k=\dfrac{\delta_1-\mu^2\left[f(x_{k-1})-f(x_k)\right]-\mu g_k^T d_{k-1}}{\mu g_k^T d_{k-1}}$, $\delta_1>0$.

Updating the $\gamma(x_k)$ is to keep $\gamma(x_k)I$ positive definite (see [11]).

B. Solution of sub-problem (5)

In this subsection, we discuss the solution of sub-problem (5). The strict minimize of the conic model function $c(x_k+d)$ is

$$d_k^N=-\frac{g_k}{\gamma(x_k)+b_k^T g_k},$$

and $d_k^C=-\tau_k g_k$ (see [11]). As we know, Newton method has a local quadratic convergence, and we can expect that the numerical performance behaves better by using d_k^N as much as possible. We compute the sub-problem (5) as follows,

if $\left\|d_k^N\right\|\le\Delta_k$, then set $d_k=d_k^N$, otherwise, set $d_k=d_k^C$.

C. Algorithm

Now, we state the non-monotone adaptive trust region algorithm based on simple conic model for unconstrained optimization.

Algorithm 1

Step 0 $x_0\in R^n,\Delta_0>0,0<\eta_1<\eta_2<\eta_3<1,\ 0<\gamma_1<\gamma_2<1<\gamma_3,0<\varepsilon<1,\varepsilon>0,\theta>0.$

$b_0\in R^n,\delta_1>0,\alpha\in[0,1],\eta_k\in(0,1)$, Set $k=0,\gamma(x_0)=1,D_0=0.$

Step 1 If $\|g_k\|\le\varepsilon$, stop, and x_k is an approximate solution. Otherwise, go to step 2.

Step 2 Solve the conic trust region sub-problem(5) for d_k by (2.2).

Step 3 Compute

$$\begin{aligned}Ared(d_k)&=D_k-f(x_k+d_k),\\ \Pr ed(d_k)&=c(x_k)-c(x_k+d_k),\end{aligned}\quad \rho_k=\frac{Ared(d_k)}{\Pr ed(d_k)}.$$

Step 4 If $\rho_k\le\eta_1$, set $\Delta_k=\gamma_1\Delta_k$, go to step 2. Otherwise, go to step 5.

Step 5 Set $x_{k+1}=x_k+d_k$.

Step 6 Compute $\gamma(x_{k+1})$, if $\gamma(x_{k+1})\le\varepsilon$ or $\gamma(x_{k+1})\ge 1/\varepsilon$, set $\gamma(x_{k+1})=\theta.$

Step 7 Compute $\Delta_k = \|g_{k+1}\| / \gamma(x_{k+1})$,

updating the trust region radius Δ_{k+1} as follows.

$$\Delta_{k+1} = \begin{cases} \max\{\Delta_k, \gamma_3 \|d_k\|\}, \rho_k \geq \eta_3. \\ \Delta_k, \eta_2 \leq \rho_k < \eta_3, \\ \gamma_2 \Delta_k, \rho_k < \eta_2. \end{cases}$$

Step 8 Updating b_{k+1}, D_{k+1}, set $k = k+1$, go to step 1.

III. CONVERGENCE ANALYSIS

(A1) $f : R^n \to R$ is twice continuously differentiated and bounded below on the level set $L(x_1) = \{x | f(x) \leq f(x_1)\}$.

(A2) The sequence $\{x_k\}$ generated by algorithm 1 is contained in a bounded closed set Ω containing $L(x_1)$.

(A3) Suppose that there exist two positive constants $\Delta_{\max}$ and M_b such that

$$\Delta_k \leq \Delta_{\max}, \|b_k\| \leq M_b, \forall k. \quad (7)$$

Assumptions (A1) and (A2) mean there exist two positive constants M_g and M_H such that

$$\|g(x)\| \leq M_g, \|\nabla^2 f(x)\| \leq M_H . \forall x \in L(x_0). \quad (8)$$

Lemma 1 (See Lemma 1 in [11]) Suppose that assumptions (A1)-(A3) hold, and that d_k is the solution of (5). Then we have

$$\Pr ed(d_k) = c(x_k) - c(x_k + d_k) \geq \frac{1}{2} \delta_2 \|g_k\| \min\left\{\Delta_k, \frac{\|g_k\|}{\gamma(x_k)}\right\}, \quad (9)$$

where $\delta_2 = \dfrac{1}{1 + \Delta_{\max} M_b}$ is a constant.

Lemma 2 Suppose that the sequence $\{x_k\}$ is generated by algorithm 1. Then the inequality

$$f(x_{k+1}) \leq D_{k+1} \leq D_k, \forall k . \quad (10)$$

Holds, where D_k is defined in (4) Also, the sequence $\{x_k\}$ remains in $L(x_1)$.

Proof. From the definition of D_k, we have

$$D_{k+1} - f_{k+1} = h(D_k - f_{k+1}) \quad \text{and} \quad D_{k+1} - D_k = (1-h)(f_{k+1} - D_k). \tag{11}$$

We consider two cases:

Case1. $k \in I$.

From algorithm 2 and (9), we have

$$D_k - f_{k+1} \geq h_1[c_k(x_k) - c_k(x_k + d_k)] \geq \frac{1}{2} h_1 d_2 \| g_k \| \min\left\{ D_k, \frac{\|g_k\|}{|g(x_k)|} \right\} \geq 0.$$

Therefore,

$$D_{k+1} - f_{k+1} = h(D_k - f_{k+1}) \geq 0,$$

$$D_{k+1} - D_k = (1-h)(f_{k+1} - D_k) \leq 0. \tag{12}$$

Case2. $k \in J$.

If $k-1 \in I$, then from (9) and (12), we have $f_{k+1} \leq D_{k+1} \leq D_k$.

If $k-1 \in J$, let $M = \{i \,|\, 1 < i \leq k, k-i \in I\}$. If $M = \emptyset$, then from (4) (6) and Lemma 1, we have $f_{k+1} \leq f_k \leq \cdots \leq f_1 = D_1$. Now we will use mathematical induction to prove $D_{k+1} \leq D_k$.

For $k = 1$, $D_2 = hD_1 + (1-h) f_2 \leq h f_1 + (1-h) f_1 = f_1 = D_1$. For $k = n$, we suppose that we have $D_{n+1} \leq D_n$.

For $k = n+1$ $D_{n+2} = hD_{n+1} + (1-h) f_{n+2} \leq hD_n + (1-h) f_{n+1} = D_{n+1}$.

So we get $D_{k+1} \leq D_k$. From (11) and $0 < h < 1$, we know $f_{k+1} \leq D_k$. Thus,

$$D_{k+1} = hD_k + (1-h) f_{k+1} \geq h f_{k+1} + (1-h) f_{k+1} = f_{k+1}. \tag{13}$$

On the other hand, if $M \neq \emptyset$, let $m = \min\{i \,|\, i \in M\}$. Then from (6) and Lemma 1, we have $f_{k+1} \leq f_k \leq \cdots \leq f_{k-m+1}$. Obviously, $k-m \in I$, then we get $f_{k-m+1} \leq D_{k-m+1} \leq D_{k-m}$ from Case 1. Thus,

$D_{k-m+2} = hD_{k-m+1} + (1-h) f_{k-m+2} \leq hD_{k-m} + (1-h) f_{k-m+1} = D_{k-m+1}$, by the induction principle, we have $D_{k+1} \leq D_k$. Finally we can get (13).

Both Case 1 and Case 2 imply that $f_{k+1} \leq D_{k+1} \leq D_k$. So we complete the proof.

Lemma 3 (See Lemma 4 in [11]) The step 2-step 4-step 2 in algorithm 1 are well defined.

Lemma 4 Suppose that assumptions (A1)-(A1) hold, then we have

$$\left\|\left[f(x_k)-f(x_k+d_k)\right]-\left[c(x_k)-c(x_k+d_k)\right]\right\| \le M\Delta_k^2,$$

where $$M=\frac{M_g M_b}{1-\gamma}+\frac{M_H}{2}+\frac{1}{2(1-\gamma)^2}\max\left\{\frac{1}{\varepsilon},\theta\right\}.$$

Lemma 5 (See Lemma 5 in [11]) Suppose that assumptions (A1)-(A2) hold, and that there exist a constant $\varepsilon>0$ such that $\|g_k\|>\varepsilon$, then there exists a constant $\Delta_{lbd}>0$ such that $\Delta_k \ge \Delta_{lbd}$.

Lemma 6 Suppose that (A1) holds and the sequence $\{x_k\}$ is generated by Algorithm 1. Then, the sequence $\{x_k\}$ is convergent.

Proof. Lemma 2 together with (A1) imply that

$$\$l \quad s.t.\,"n\,Î\,¥ : l\ £\ f_{k+1}\ £\ D_{k+1}\ £\ D_k\ £\ ×××£\ D_1\ £\ f_1.$$

This shows that the sequence $\{x_k\}$ is convergent.

Theorem 7 Suppose that assumptions (A1)-(A3) hold. Then the sequence $\{x_k\}$ generated by algorithm 2 satisfies $\lim_{k\to+\infty}\inf\|g_k\|=0.$

Proof. If there are finitely many successful iterations, then the conclusion holds obviously from algorithm 1. First we can prove when $\|g_k\|>\varepsilon>0$, there must be $\lim\Delta_k=0.$

According to the step 6 of algorithm 1, we know that the sequence $\{\gamma(x_k)\}$ is uniformly bounded, i.e.,

$$0<\min\{\varepsilon,\theta\}\le\gamma(x_k)\le\max\left\{\frac{1}{\varepsilon},\theta\right\}=m,\forall k,$$

where $\gamma>0$ is a constant. So, we have

$$D_k-f(x_k+d_k)\ge\eta_1\,\mathrm{Pr}ed_k\ge\frac{1}{2}\eta_1\delta_2\|g_k\|\min\left\{\Delta_k,\frac{\|g_k\|}{\gamma(x_k)}\right\}\ge\frac{1}{2}\eta_1\delta_2\varepsilon\min\left\{\Delta_k,\frac{\varepsilon}{m}\right\}.$$

Because f_k is bounded, so D_k is also bounded. Noting the $Y=\{k\,|\,\rho_k\ge\eta_1\}$,

so we have

$$+\infty>\sum_{k=1}^{\infty}\left[D_k-f(x_k+d_k)\right]\ge\sum_{k\in Y}\left[D_k-f(x_k+d_k)\right]\ge\sum_{k\in Y}\eta_1\,\mathrm{Pr}ed_k\ge\sum_{k\in Y}\frac{1}{2}\eta_1\delta_2\varepsilon\min\left\{\Delta_k,\frac{\varepsilon}{m}\right\}.$$ So we have $\sum_{k\in Y}\min\left\{\Delta_k,\frac{\varepsilon}{m}\right\}<+\infty$, moreover, $|Y|\to+\infty$, thus $\lim_{k\to+\infty}\Delta_k=0.$

Next, we will prove $\lim_{k\to+\infty} \inf \|g_k\| = 0$. Actually, we prove this result by a contradiction. Suppose that when k is very big, and $\|g_k\| > \varepsilon > 0$,

$$\left|\frac{f(x_k)-f(x_{k+1})}{\Pr ed_k}-1\right| = \left|\frac{f(x_k)-f(x_{k+1})-(c(x_k)-c(x_k+d_k))}{\Pr ed_k}\right|$$

$$= \frac{M\Delta_k^2}{\Pr ed_k} \le \frac{2M\Delta_k^2}{\eta_1\delta_2 \min\left\{\Delta_k, \frac{\varepsilon}{m}\right\}}.$$

Then we have $\rho_k = \dfrac{D_k - f(x_k+d_k)}{\Pr ed_k} \ge \dfrac{k(x_k)-f(x_k+d_k)}{\Pr ed_k}$,

and $k\to\infty$, then $\lim_{k\to\infty} \dfrac{f(x_k)-f(x_k+d_k)}{\Pr ed_k} = 1.$

So $\rho_k \ge \eta_1$, from algorithm 1 and Lemma 5, when $k\to\infty$, exists a constant $\Delta_{lbd} > 0$, such that $\Delta_k \ge \Delta_{lbd} \ge 0$, this contradicts $\lim_{k\to+\infty} \Delta_k = 0$.

IV. CONCLUSIONS

In this paper, we propose a non-monotone adaptive trust region method based on simple conic model for unconstrained optimization. The global convergences of the proposed algorithm are established. Our method is efficient for solving large scale optimization problems. The sub-problem incorporates more information which is useful to the algorithm. The Hessian of the objective function or its approximation is approximated by a scalar matrix, which needs less memory and computational efforts.

ACKNOWLEDGMENTS

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

REFERENCES

[1] Di S, Sun W. A trust region method for conic model to solve unconstrained optimizations. Optim Methods Softw, 1996, 6: 237-263

[2] Sun W, Xu D. A filter-trust-region method based on conic model for unconstrained optimization. Sci China Math, 2012, 42: 527-543(in china)

[3] Davidon W C. Conic approximations and collinear scalings for optimizers. SIAM J Numer Anal, 1980, 17: 268-281

[4] Sorensen D C. The Q-superlinear convergence of a collinear scaling algorithm for unconstrained optimization. SIAM J Numer Anal, 1980, 17: 84-114

[5] 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.

[6] 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.

[7] N.Z. Gu, J. T. Mo, Incorporating nonmonotone strategies into the trust region for unconstrained optimization, Computers and

Mathmatics with Applications 55(2008) 2158-2172

[8] Zhou Q Y, Zhang C. A new nonmonotone adaptive trust region method based on simple quadratic models. J Appl Math Comput, 2012, 40: 111-123

[9] Sun W, Yuan Y. Optimization Theory and Methods: Nonlinear Programming. New York: Spinger, 2006

[10] Andrei N. Scaled conjugate gradient algorithms for unconstrained optimization. Comput Optim Appl, 2007, 38: 401-416

[11] Lijian ZHAO, Wenyu Sun. Nonmonotone adaptive trust region method based on simple conic model for unconstrained optimization Math. China 2014,9(5): 1211-1238