Scholars Bulletin ***(Mathematics)***
An Official Publication of "Scholars Middle East Publishers",
Dubai, United Arab Emirates
Website: http://scholarsbulletin.com/
ISSN 2412-9771 (Print)
ISSN 2412-897X (Online)

Non-Monotone Conic Trust Region Method Combined with Line Search Strategy

Yunfeng Zhang[1], Qinghua Zhou[1,2]
[1]College of Mathematics and Information Science, Hebei University, Baoding, China
[2]School of Applied Mathematics, Beijing Normal University, Zhuhai, China

***Corresponding author**
Qinghua Zhou

Article History
Received: 29.09.2017
Accepted: 10.10.2017
Published: 30.10.2017

DOI:
10.21276/sb.2017.3.10.1

Abstract: In this paper, we propose a non-monotone adaptive trust region algorithm based on conic model for solving unconstrained optimization problems. Unlike the traditional non-monotone trust-region method, our proposed algorithm avoids resolving the sub-problem whenever a trial step is rejected. Instead, it performs a non-monotone Armijo-type line search in direction of the rejected trial step to construct a new point. The algorithm can be regarded as a combination of non-monotone, line search and conic trust region method. Theoretical analysis indicates that the new approach preserves the global convergence to the first-order critical points under classical assumptions.
Keywords: Unconstrained optimization; Conic trust-region method; Armijo-type line search; Non-monotone technique; global convergence

INTRODUCTION

Consider the unconstrained minimization problem

$$\min f(x), \text{ subject to } x \in R^n, \qquad (1)$$

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

One of most often used methods for solving problem (1) is trust region method. It is known by having strong convergence and robustness and it can be applied to ill-conditioned problems. Another advantage of trust region is that there is no need to require the approximate Hessian matrix of the trust region sub-problem to be positive definite. So trust region methods have been studied by many researchers [1, 2]. However, when the trial step is not successful, the next processes of trust region method including reduces the trust region radius and resolves the sub-problem, and so on, can be costly.

The traditional quadratic model methods often produce a poor prediction of the minimizer of the function, when the objective function has strong non-quadratic quality. In 1980, Davidon [3] proposed the conic model methods for unconstrained optimization problems. A typical conic model sub-problem is as follows:

$$\min \quad m_k(d) = f_k + \frac{g_k^T d}{1 + h_k^T d} + \frac{1}{2}\frac{d^T B_k d}{\left(1 + h_k^T d\right)^2}, \qquad (2)$$

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

where $f_k = f(x_k)$, $g_k = \nabla f(x_k)$, $B_k \in R^{n\times n}$ is a symmetric matrix which is the Hessian matrix or its approximation of $f(x)$ at the current point x_k, Δ_k is conic trust region radius, $\|\cdot\|$ denotes the Euclidean norm and h_k is usually called horizontal vector which is the associated vector for conic model. If $h_k = 0$, the conic model reduces to a quadratic model.

In [4], Zhou & Zhang proposed a simple quadratic trust region sub-problem using a scalar approximation of the minimizing function's Hessian. Based on the Taylor's theorem, $\gamma(x_k)I$ is considered as an approximation of B_k in problem (2), where $\gamma(x_k)I$ is a positive scalar. As a result, the new sub-problem could be also resolved easily. Furthermore, they use the same idea into the conic model (see [5] for details), and construct the following new sub-problem

$$\min \quad m_k(d) = f_k + \frac{g_k^T d}{1 + h_k^T d} + \frac{1}{2}\frac{\gamma_k d^T d}{\left(1 + h_k^T d\right)^2},$$

$$s.t. \quad 1 + h_k^T d > 0, \qquad (3)$$

$$\|d\| \le \Delta_k,$$

which is called as the simple conic trust region sub-problem. Where

$$\gamma_k = \begin{cases} \hat{\gamma}_{k+1}, & if\, \hat{\gamma}_{k+1} > 0, \\ \dfrac{2\beta^2\delta}{d_k^T d_k}, & otherwise, \end{cases} \quad h_{k+1} = \frac{1-\beta}{g_k^T d_k} g_k \qquad (4)$$

and $\hat{\gamma}_{k+1} = \dfrac{2}{d_k^T d_k}\left[\beta^2\left(f_k - f_{k+1}\right) + \beta g_{k+1}^T d_k\right]$, $\delta > 0$ is a small constant, and

$$\beta = \begin{cases} \dfrac{f_k - f_{k+1} + \sqrt{p}}{-g_{k+1}^T d_k}, & if\ p \ge 0, \\ 1, & otherwise, \end{cases} \quad \text{where } p = \left(f_k - f_{k+1}\right)^2 - \left(g_{k+1}^T s_k\right)\left(g_k^T s_k\right).$$

M. Ahookhosh and S. Ghaderi in [6] proposed a novel non-monotone strategy based on a weighted average of former successive iterates. In detail, the non-monotone item T_k is defined as follows

$$T_k = \begin{cases} f_{l(k)} & if \quad k < N, \\ \max\left\{\overline{T}_k, f_k\right\} & if \quad k \ge N, \end{cases} \quad \overline{T}_k = \left(1 - \eta_{k-1}\right) f_k + \eta_{k-1}\overline{T}_{k-1} + \xi_k\left(f_{k-N} - f_{k-N-1}\right), \quad (5)$$

where $\xi_k = \eta_{k-1}\eta_{k-2}\ldots\eta_{k-N-1} = \dfrac{\eta_{k-1}}{\eta_{k-N-2}}\eta_{k-2}\ldots\eta_{k-N-1}\eta_{k-N-2} = \dfrac{\eta_{k-1}}{\eta_{k-N-2}}\xi_{k-1}$; $f_{l(k)} = \max\limits_{0 \le j \le m(k)}\left\{f_{k-j}\right\}$, $k = 0, 1, 2, \ldots$, where $m(0) = 0, m(k) \le \min\left\{m(k-1) + 1, M\right\}$ for positive integer M . It is clear that the new term uses a stronger term $f_{l(k)}$ for first $k < N$ iterations and then employs the relaxed convex term proposed above. In this work, we combine the above ideas into conic models.

This paper organized as follows. In Section 2, we describe the novel adaptive conic trust region line search algorithm. In Section 3, we first give its properties, prove that the new algorithm is well defined, and then the global convergence is investigated. Finally, some conclusions are given in Section 4.

NOVEL ADAPTIVE CONIC TRUST-REGION LINE SEARCH ALGORITHM

In this section, we describe a new non-monotone adaptive conic trust region method with line search techniques.

In our algorithm, at each iterative point x_k, the trial step is obtained by solving the conic model sub-problem. Let d_k be the solution of (3). If $\gamma_k + h_k^T g_k \ne 0$, then the unique minimizer point of the conic function $m_k(d)$ in (3) is

$$d_k^N = -\frac{g_k}{\gamma_k + h_k^T g_k}.$$

The Cauchy point of the function is

$$d_k^C = -\tau_k g_k,$$

where

$$\tau_k = \begin{cases} \dfrac{\Delta_k}{\|g_k\|}, & \gamma_k + h_k^T g_k \le 0, \\ \min\left\{\tau^*, \dfrac{\Delta_k}{\|g_k\|}\right\}, & \gamma_k + h_k^T g_k > 0, \end{cases} \qquad \tau^* = \frac{1}{\gamma_k + h_k^T g_k}.$$

We define the actual reduction as $ared(d) = T_k - f(x_k + d)$, the predicted reduction as $pred(d) = m_k(0) - m_k(d)$, and the ratio of the actual reduction to the predicted reduction as

$$r_k = \frac{ared(d)}{pred(d)} = \frac{T_k - f(x_k + d)}{m_k(0) - m_k(d)}. \qquad (6)$$

If $r_k \ge \mu$, then we accept the trial step and set $x_{k+1} = x_k + d_k$. Otherwise, we determine the step-length $\alpha_k \in \{s, \rho s, \rho^2 s \ldots\}$ by subsequent Armijo-type line search

$$f(x_k + \alpha_k d_k) \le T_k + \sigma \alpha_k g_k^T d_k, \qquad (7)$$

where s is a positive constant, $\rho \in (0,1)$ and $\sigma \in (0,1/2)$. In this case, we set $x_{k+1} = x_k + \alpha_k d_k$. Now, we can outline our new non-monotone trust-region line search algorithm as follows:

Algorithm1 New Adaptive Conic Non-monotone Trust-Region Line Search Algorithm

step1. Given $x_0 \in R^n$, $\Delta_0 > 0$, $\Delta_{max} > 0$, $0 < \mu_1 < \mu < \mu_2 < 1$, $0 < \rho < 1$, $0 < \sigma < 1/2$, $0 < c_1 < 1 < c_2$, $\delta > 0$, $\theta > 0$, $\omega > 0$, $\varepsilon > 0$. Set $k = 0$, $\gamma_0 = 1$, $h_0 = 0$.

step2. Compute $g(x_k)$. If $\|g(x_k)\| \le \omega$, stop. Otherwise, go to Step 3.

step3. Solve the sub-problem (3) to determine a trial step d_k.

step4. Compute T_k and r_k. If $r_k \ge \mu$, set $x_{k+1} = x_k + d_k$ and go to Step 7. Otherwise, go to Step5.

step5. Find the step-length α_k satisfying in (7), and set $x_{k+1} = x_k + \alpha_k d_k$.

step6. If $\gamma_{k+1} \le \varepsilon$ or $\gamma_{k+1} \ge \dfrac{1}{\varepsilon}$, set $\gamma_{k+1} = \theta$.

Step7. Compute $\Delta = \lambda_{k+1} \max\left\{\dfrac{1}{\gamma_{k+1}}, \dfrac{1}{\gamma_{k+1} + h_{k+1}^T g_{k+1}}\right\}\|g_{k+1}\|$, where

$$\lambda_{k+1} = \begin{cases} c_1 \lambda_k, & if\ r_k < \mu_1; \\ \lambda_k, & if\ \mu_1 \le r_k \le \mu_2; \\ c_2 \lambda_k, & otherwise. \end{cases}$$

Set $\Delta_{k+1} = \min\{\Delta, \Delta_{max}\}$.

Step8. Update h_{k+1} and γ_{k+1}. Set $k = k+1$ go to Step1.

Remark 2.1

(1) The object of Step7 avoids uphill direction and keeps the sequence $\{\gamma_k\}$ uniformly bounded. In fact, for all k,

$$0 < \min(\varepsilon, \theta) \le \gamma_k \le \max\left(\frac{1}{\varepsilon}, \theta\right) \qquad (8)$$

(2) In order to guarantee the global convergence, we choose a sufficiently small constant $0 < \ell < 1$ such that $\Delta_k \|h_k\| \le \ell$, $\forall k$.

Algorithm2

step1. If $\gamma_k + h_k^T g_k \neq 0$, compute d_k^N. Then if $\left\| d_k^N \right\| \leq \Delta_k$, set $d_k = d_k^N$ and return; otherwise, go to Step2.

step2. Set $d_k = d_k^C$ and return.

CONVERGENCE ANALYSIS

In this section, we discuss some convergence properties of the new algorithm, and prove the global convergence.

For convenience, we define two index sets as below,

$I = \{k : r_k \geq \mu\}$ and $J = \{k : r_k < \mu\}$.

The following assumptions are used to analyze the convergence properties of Algorithm:

(H1) The objective function f is twice continuously differentiable and bound below on level set

$$L(x_0) = \{x \in R^n \mid f(x) \leq f(x_0), x_0 \in R^n\}.$$

(H2) Suppose that there exist two positive constant $\Delta_{\max}$ and M_h such that

$\Delta_k \leq \Delta_{\max}, \|h_k\| \leq M_h, \forall k$.

(H3) Suppose that there exist two positive constant M_G and M_g such that

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

Lemma 3.1 If d_k is the solution of sub-problem (3) and assumption (H1)-(H3) hold. Then there exist a positive scalar ν such that, for all k,

$$m_k(0) - m_k(d) \geq \frac{1}{2}\nu \|g_k\| \min\left\{\Delta_k, \frac{\|g_k\|}{\gamma_k}\right\},$$

and

$$g_k^T d_k \leq -\frac{1}{2}\nu(1+\ell)\|g_k\| \min\left\{\Delta_k, \frac{\|g_k\|}{\gamma_k}\right\},$$

where $\nu = \dfrac{1}{1+\Delta_{\max} M_h}$.

Proof. A proof of this lemma can be observed in [5].

Lemma 3.2 Suppose that all conditions of Lemma 3.1 hold. Then we have

$\left| [f_k - f_{k+1}] - [m_k(0) - m_k(d)] \right| \leq O\left(\|d_k\|^2\right)$.

Proof. We consider two cases:

Case 1. $k \in I$. When $\|d_k\|$ is sufficiently close to zero, since $\{\|h_k\|\}$ is bounded, we have $1/(1 + h_k^T d_k) = 1 + O(\|d_k\|)$. By the boundedness of $\|g_k\|$ and γ_k, we have

$$\frac{g_k^T d_k}{1 + h_k^T d_k} = g_k^T d_k + O\left(\|d_k\|^2\right), \frac{\gamma_k d_k^T d_k}{\left(1 + h_k^T d_k\right)^2} = \gamma_k d_k^T d_k + O\left(\|d_k\|^2\right).$$

By Taylor's expansion, Remark 2.1, and (H3), we have

$$\begin{aligned} &\left| [f_k - f(x_k + d_k)] - [m_k(0) - m_k(d_k)] \right| \\ &= \left| -g_k^T d_k - \frac{1}{2} d_k^T \nabla^2 f(x_k + \theta_k d_k) d_k + g_k^T d_k + \frac{1}{2}\gamma(x_k) d_k^T d_k + O\left(\|d_k\|^2\right) \right| \end{aligned}$$

$$\le \frac{1}{2}\left[M_G + \max\left\{ \frac{1}{\varepsilon}, \theta \right\} \right] \|d_k\|^2 + O\left(\|d_k\|^2\right) = O\left(\|d_k\|^2\right)$$

where $\theta_k \in (0,1)$ is a constant.

Case 2. $k \in J$. By Taylor's expansion, Remark 2.1, and (H2)-(H3), we have

$$\left| \left[f_k - f(x_k + \alpha_k d_k) \right] - \left[m_k(0) - m_k(\alpha_k d_k) \right] \right|$$

$$= \left| -\alpha_k g_k^T d_k - \frac{1}{2}\alpha_k^2 d_k^T \nabla^2 f(x_k + \theta_k \alpha_k d_k) d_k + \frac{\alpha_k g_k^T d_k}{1 + \alpha_k h_k^T d_k} + \frac{\alpha_k^2 \gamma(x_k) d_k^T d_k}{2\left(1 + \alpha_k h_k^T d_k\right)^2} \right|$$

$$\le \left| \frac{1}{2}\alpha_k^2 d_k^T \nabla^2 f(x_k + \theta_k \alpha_k d_k) d_k + \frac{\alpha_k^2 g_k^T d_k h_k^T d_k}{1 + \alpha_k h_k^T d_k} - \frac{\alpha_k^2 \gamma(x_k) d_k^T d_k}{2\left(1 + \alpha_k h_k^T d_k\right)^2} \right|$$

$$\le \alpha_k^2 \left[\frac{M_g M_h}{1 - \ell} + \frac{M_G}{2} + \frac{1}{2(1-\ell)^2} \max\left\{ \frac{1}{\varepsilon}, \theta \right\} \right] \|d_k\|^2 = O\left(\|d_k\|^2\right),$$

where $\theta_k \in (0,1)$ is a constant.

Lemma 3.3 ([6]) Suppose that the sequence $\{x_k\}$ is generated by Algorithm1, then we get

$$f_k \le T_k \le f_{l(k)}, \qquad (9)$$

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

Corollary 3.1 Suppose (H1)-(H3) hold and the sequence $\{x_k\}$ is generated by Algorithm1, then

$$\lim_{k\to\infty} T_k = \lim_{k\to\infty} f_k .$$

Proof. A proof of this Corollary can be observed in [7].

Corollary 3.2([6]) Suppose (H1)-(H3) hold and the sequence $\{x_k\}$ is generated by Algorithm1, then

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

Lemma 3.4 ([6]) Suppose that (H1)-(H3) holds, the sequence $\{x_k\}$ is generated by Algorithm1 is contained in the level set $L(x_0)$, and the sequence $\{f_{l(k)}\}$ is not increasing monotonically and convergent.

Lemma 3.5 Suppose that the sequence $\{x_k\}$ is generated by Algorithm1. Then, the Algorithm1 is well-defined.

Proof. We consider two cases:

Case 1. $k \in I$.

First we prove that when p is sufficiently large, $r_k \ge \mu$ holds. Let d_k be the solution of sub-problem (3). Using Lemma 3.1 and Lemma 3.2, we have

$$\left| \frac{f_k - f(x_k + d_k)}{m_k(0) - m_k(d_k)} - 1 \right| = \left| \frac{f_k - f(x_k + d_k) - (m_k(0) - m_k(d_k))}{m_k(0) - m_k(d_k)} \right|$$

$$\le \frac{O\left(\|d_k\|^2\right)}{\frac{1}{2}\nu \|g_k\| \min\left\{ \Delta_k, \frac{\|g_k\|}{\gamma_k} \right\}}$$

Now, as $k \to \infty$, then $\|d_k\| \to 0$ and consequently, the right hand side of the preceding inequality tends to zero. Now, using (9), we have

$r_k = \frac{T_k - f(x_k + d)}{m_k(0) - m_k(d)} \geq \frac{f_k - f(x_k + d)}{m_k(0) - m_k(d)} \geq \mu$.

Therefore, when k is sufficiently large, $r_k \geq \mu$.

Case 2. $k \in J$.

We prove that the line search terminates in the finite number of steps. For establishing a contradiction, assume that there exists $k \in J$ such that

$$f(x_k + \rho^i s_k d_k) > T_k + \sigma \rho^i s_k g_k^T d_k, \ \forall i \in \mathrm{N} \cup \{0\}. \qquad (10)$$

From Lemma 3.3, we have $f_k \leq T_k$. This fact, along with (10), implies that

$$\frac{f(x_k + \rho^i s_k d_k) - f_k}{\rho^i s_k} > \sigma g_k^T d_k, \ \forall i \in \mathrm{N} \cup \{0\}.$$

Since f is a differentiable function, by taking a limit, as $i \to \infty$, we obtain

$g_k^T d_k \geq \sigma g_k^T d_k$.

Using the fact that $\sigma \in (0, 1/2)$, this inequality leads us to $g_k^T d_k \geq 0$ which contradicts Lemma 3.1. Therefore, Algorithm 1 is well-defined.

Theorem 3.1 Suppose that (H1)-(H3) hold and the sequence $\{x_k\}$ is generated by Algorithm, then $\liminf_{k \to \infty} \|g_k\| = 0$.

Proof. If there are finitely many successful iterations, then the conclusion holds obviously from Algorithm 1.

Now, we consider the case in which there are infinitely successful iterations. We suppose that the conclusion does not hold, i.e., there exists a constant $0 < \omega < 1$ such that for all k sufficiently large, we have $\|g(x_k)\| > \omega$. Then, from Algorithm 1 and Lemma3.1, we have

$$f(x_{k+1}) \leq T_k - \mu\, pred(d_k) \leq T_k - \frac{1}{2}\mu \nu \|g_k\| \min\left\{\Delta_k, \frac{\|g_k\|}{\gamma_k}\right\}. \qquad (11)$$

From the definition of T_k , we consider:

Case1. $k < N$, $T_k = f_{l(k)}$.

Then $f_{l(k)} - f(x_{k+1}) \geq \mu\, pred(d_k) = \frac{1}{2}\nu \|g_k\| \min\left\{\Delta_k, \frac{\|g_k\|}{\gamma_k}\right\}$. For all $M < k < N$, we can write

$$f_{l(l(k)-1)} - f(x_{l(k)}) \geq \mu\, pred\left(d_{l(k)-1}\right). \qquad (12)$$

By Lemma 3.4, we take limit on both sides of (12) and get $\lim_{k \to \infty} pred\left(d_{l(k)-1}\right) = 0$.

From Lemma 3.1,

$$pred\left(d_{l(k)-1}\right) \geq \frac{1}{2}\nu \left\|g_{l(k)-1}\right\| \min\left\{\Delta_{l(k)-1}, \frac{\left\|g_{l(k)-1}\right\|}{\gamma_{l(k)-1}}\right\} \geq \frac{1}{2}\nu\omega \min\left\{\Delta_l, \frac{\omega}{\max(1/\varepsilon, \theta)}\right\}.$$

Then we conclude $pred\left(d_{l(k)-1}\right) \to 0$ as $k \to \infty$. On the other hand, from the algorithm we know that $\tilde{d}_{l(k)-1}$ is unacceptable. i.e., $ared\left(d_{l(k)-1}\right) \leq \mu_1 pred\left(d_{l(k)-1}\right)$.

From Lemma 3.1,

$$pred\left(\tilde{d}_{l(k)-1}\right) \geq \frac{1}{2}\nu \left\|g_{l(k)-1}\right\| \min\left\{\Delta_{l(k)-1}, \frac{\left\|g_{l(k)-1}\right\|}{\gamma_{l(k)-1}}\right\} \geq \frac{1}{2}\nu\omega \min\left\{\Delta_l, \frac{\omega}{\max\left(1/\varepsilon, \theta\right)}\right\}. \quad (13)$$

From Lemma 3.2, we have

$$\left|f_{l(l(k)-1)} - f(x_{l(k)} + \tilde{d}_{l(k)-1}) - pred\left(\tilde{d}_{l(k)-1}\right)\right| \leq O\left(\left\|\tilde{d}_{l(k)-1}\right\|^2\right). \quad (14)$$

Combing (13) and (14), we conclude

$$\frac{f_{l(l(k)-1)} - f(x_{l(k)} + \tilde{d}_{l(k)-1})}{pred\left(\tilde{d}_{l(k)-1}\right)} \to 1.$$

It follows that

$$\frac{ared\left(\tilde{d}_{l(k)-1}\right)}{pred\left(\tilde{d}_{l(k)-1}\right)} \geq \frac{f_{l(l(k)-1)} - f(x_{l(k)} + \tilde{d}_{l(k)-1})}{pred\left(\tilde{d}_{l(k)-1}\right)} \geq \mu_1,$$

this is a contradiction.

Case2. $k \geq N$, $T_k = \overline{T}_k$.

By using (5), (11), we have

$$\begin{aligned}\overline{T}_{k+1} &= (1-\eta_k) f_{k+1} + \eta_k \overline{T}_k + \xi_{k+1}(f_{k+1-N} - f_{k-N}) \\ &\leq (1-\eta_k)\left(\overline{T}_k - \mu\, pred(d_k)\right) + \eta_k \overline{T}_k + \xi_{k+1}(f_{k+1-N} - f_{k-N}) \\ &\leq \overline{T}_k - (1-\eta_k)\mu\, pred(d_k).\end{aligned}$$

Then $\overline{T}_k - \overline{T}_{k+1} \geq (1-\eta_k)\mu\, pred(d_k) = \frac{1}{2}\mu(1-\eta_k)\nu\omega \min\left\{\Delta_k, \frac{\omega}{\max(1/\varepsilon, \theta)}\right\}$.

From Corollary 3.1-3.2, Lemma 3.4, then

$$\lim_{k\to\infty} \min\left\{\Delta_k, \frac{\omega}{\max(1/\varepsilon, \theta)}\right\} = 0,$$

which implies that $\lim_{k\to\infty} \Delta_k = 0$. It follows from the proof of Lemma 3.5 that $r_k \geq \mu_2$ for k large enough. By the description of the Algorithm 1, it implies that there exists a positive constant λ^* such that $\lambda_k \geq \lambda^*$ for all sufficiently large k. On the other hand,

$$\lim_{k\to\infty} \lambda_{k+1} \max\left\{\frac{1}{\gamma_{k+1}}, \frac{1}{\gamma_{k+1} + h_{k+1}^T g_{k+1}}\right\}\left\|g_{k+1}\right\| = 0.$$

Then

$$\max\left\{\frac{1}{\gamma_{k+1}}, \frac{1}{\gamma_{k+1} + h_{k+1}^T g_{k+1}}\right\}\left\|g_{k+1}\right\| \geq \frac{\left\|g_{k+1}\right\|}{\gamma_{k+1}} \geq \frac{\omega}{\max(1/\varepsilon, \theta)},$$

$\lim_{k\to\infty} \lambda_{k+1} = 0$, for all sufficiently large k, which is a contradiction.

CONCLUSION

In this paper, a variant non-monotone adaptive conic trust region algorithm for solving unconstrained optimization problem is proposed. Unlike traditional conic trust region method, the proposed algorithm does not reject a failed trial step, but performs a non-monotone line search in direction of the rejected trial step in order to avoid resolving the trust region sub-problem instead. We analyzed the properties of the algorithm and proved the global convergence theory under some mild conditions.

ACKNOWLEDGEMENT

The author would be very grateful to the anonymous referees for their valuable comments and suggestions.

REFERENCES

1. Fletcher, R. (1980). *Practical Method of Optimization, Unconstrained Optimization*,vol.1, John Wiley, New York.
2. Powell, M. J. D. (1984). On the global convergence of trust region algorithms for unconstrained optimization. *Mathematical Programming, 29*(3), 297-303.
3. Davidono, W. C. (1980). Conic approximation and collinear scaling for optimizers. *SIAM Journal on Numerical Analysis, 17*(2), 268-281.
4. Zhou, Q., & Zhang, C. (2012). A new nonmonotone adaptive trust region method based on simple quadratic models. *J. Appl. Math. Comput, 40,* 111–123.
5. Zhou, Q., & Zhang, C. (2015). An adaptive trust region method based on simple conic models. *J. Math. Model. Algor, 14*, 453-467.
6. Ahookhosh, M., & Ghaderi, S. (2016). Two globally convergent non-monotone trust-region methods for unconstrained optimization. *Journal of Computational and Applied Mathematics, 50*(1), 529-555.
7. Cui, Z., Wu, B., & Qu, S. (2011). Combing non-monotone conic trust region and line search technique for unconstrained optimization. *Journal of Computational and Applied Mathematics, 235*, 2432-2441.
8. Zhao, L., Sun, W., & de Sampaio, R. J. B. (2014). Non-monotone adaptive trust region method based on simple conic model for optimization. *Front. Math. China, 9*(5), 1211-1238.