IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 3 Issue 4, April 2016.
www.ijiset.com
ISSN 2348 – 7968

A Non-monotone Conic Trust Region Method with Line Search for Unconstrained Optimization

Liran Yang, Qinghua Zhou*
Key Lab. of Machine Learning and Computational Intelligence,
College of Mathematics and Information Science, Hebei University,
Baoding, 071002, Hebei Province, China
yang_liran@sina.com, qinghua.zhou@gmail.com(*Corresponding author)

Abstract: In this paper, we propose and analyze a non-monotone conic trust region method with line search for unconstrained optimization problem. Different from the usual trust region methods, we use a new non-monotone Wolfe-type line search to get the new point if the trial step is not accepted. The algorithm can be regarded as a combination of non-monotone, line search and conic trust region method. Theoretical analysis indicates that the new method has a global convergence under reasonable assumptions.

Keywords: non-monotone; conic trust region method; unconstrained optimization; Wolfe-type line search; global convergence.

1. Introduction

We consider the unconstrained optimization problem:

$$\min f(x), \quad x \in R^n, \tag{1.1}$$

where $f(x): R^n \to R$ is continuously differentiable.

Traditional iterative methods for solving (1.1) are either line search method or trust region method. Trust region method has strong convergence and robustness, 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 [1-4].

However, when the trial step is not successful, one rejects it, reduces the trust region radius and resolves the sub-problem, which can be costly. Nocedal and Yuan [5] proposed a new type of trust region method which combine line search and trust region method. Recently, non-monotone techniques have been studied by many authors since Grippo et al. [6]. Many authors have generalized the non-monotone strategy to trust region methods and presented non-monotone trust region methods [7-9].

The traditional quadratic model methods often produce a poor prediction of the minimizer of the function, when the objective function has strong non-quadratic. In order to overcome the problem, Qu et al. [10] proposed a new trust region sub-problem based on the conic model for unconstrained optimization:

$$\begin{aligned} \min \quad & c_k(s) = f_k + \frac{g_k^T s}{1-h_k^T s} + \frac{1}{2}\frac{s^T B_k s}{(1-h_k^T s)^2}, \\ s.t. \quad & 1-h_k^T s > 0, \\ & \|s\| \le \Delta_k, \end{aligned} \tag{1.2}$$

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, $\|\cdot\|$ denotes to the Euclidean norm, $c_k(s)$ is called conic model, h_k is usually called horizontal vector which is the associated vector for conic model and Δ_k is conic trust region radius. If $h_k = 0$, the conic model reduces to a quadratic model. Therefore, the conic model methods are the generalization of the quadratic model methods.

In this paper, we combine the sub-problem (1.2) with non-monotone technique proposed in [8] and Wolfe-type line search [11] to propose our new algorithm. This paper is organized as follows. We describe our new non-monotone trust region method with line search based on conic model in Section 2. The properties of this new algorithm and the global convergence property are given in Section 3. Finally, some conclusions are addressed in Section 4.

2. Algorithm

In this section, we describe our new non-monotone conic trust region method with Wolfe-type line search algorithm. We obtain the trial step by solving the conic model sub-problem (1.2). Let s_k be the solution to (1.2). Then either x_{k+1} is accepted or the trust region radius is reduced according to the ratio r_k between the actual reduction of the objective function

$$Ared_k = D_k - f(x_k + s_k) \tag{2.1}$$

and the predicted reduction

$$Pred_k = -\frac{g_k^T s_k}{1-h_k^T s_k} - \frac{1}{2}\frac{s_k^T B_k s_k}{(1-h_k^T s_k)^2} \tag{2.2}$$

i.e.,

$$r_k = \frac{Ared_k(s_k)}{Pred_k(s_k)} \tag{2.3}$$

where

$$D_k = \begin{cases} f_k, & k = 0 \\ \eta_{k-1}D_{k-1} + (1-\eta_{k-1}) f_k, & k \geq 1 \end{cases} \tag{2.4}$$

and $\eta_{k-1} \in [\eta_{\min}, \eta_{\max}]$, $\eta_{\min} \in [0,1)$ and $\eta_{\max} \in [\eta_{\min}, 1)$. If $r_k \geq c_0$, where $c_0 > 0$, we accept s_k as a successful step and let $x_{k+1} = x_k + s_k$. Otherwise we compute i_k, the minimum positive integer i satisfying

$$\begin{cases} g(x_k + \lambda^i s_k)^T s_k \geq \varsigma g_k^T s_k \\ f(x_k + \lambda^i s_k) \leq D_k + \delta\lambda^i (g_k^T s_k + \gamma \|g_k\| \min\{\Delta_k, \frac{\|g_k\|}{\|B_k\|}\}) \end{cases} \tag{2.5}$$

where $\varsigma \in (0,1)$, $\delta \in (0,\frac{1}{2})$, $\gamma > 0$ and $\lambda \in (0,1)$ are constants, then we use the Wolfe-type

ISSN 2348 – 7968

line search to generate the next point by using $x_{k+1} = x_k + \lambda^{i_k} s_k$.

Algorithm 2.1

Step 1. Given $x_0 \in R^n$, $\Delta_0 > 0$, $h_0 = 0$, $0 < c_0 < 1$, $0 < c_1 < c_2 < 1 < c_3$, $0 < \lambda < 1$, $0 < \delta < \frac{1}{2}$, $0 < \varsigma < 1$, $\varepsilon > 0$, $B_0 \in R^{n \times n}$ is a symmetric matrix. Set $k = 0$ and choose parameters $\eta_{k-1} \in [\eta_{\min}, \eta_{\max}]$, $\eta_{\min} \in [0,1)$ and $\eta_{\max} \in [\eta_{\min}, 1)$.

Step 2. Compute g_k. If $\|g_k\| < \varepsilon$, stop. Otherwise, go to Step 3.

Step 3. Solve the sub-problem (1.2) for s_k. Compute D_k, $Ared_k$, $Pred_k$ and r_k.

Step 4. If $r_k \geq c_0$, set $x_{k+1} = x_k + s_k$ and go to the Step 6; otherwise, go to Step 5.

Step 5. Compute i_k, the minimum positive integer i satisfying (2.5). Set $a_k = \lambda^{i_k}$, $x_{k+1} = x_k + a_k s_k$.

Step 6. Compute Δ_{k+1} as

$$\Delta_{k+1} \begin{cases} \in [c_1 \|s_k\|, c_2 \Delta_k], & if\ r_k < c_0 \\ \Delta_k, & if\ r_k \geq c_0, and \|s_k\| < \Delta_k \\ \in [\Delta_k, c_3 \Delta_k], & if\ r_k \geq c_0, and \|s_k\| = \Delta_k \end{cases}$$

Step 7. Update h_k and the symmetric matrix B_{k+1}. Set $k = k+1$, go to Step 2.

We define

$$I = \{k \mid r_k \geq c_0\} \text{ and } J = \{k \mid r_k < c_0\}.$$

3. Convergence

In this section, we will prove the global convergence property of Algorithm 2.1. The following assumptions are necessary to analyze the convergence property.

(H1) The level set $L(x_0) = \{x \in R^n \mid f(x) \leq f(x_0)\}$ is bounded for any given $x_0 \in R^n$.

(H2) There exists a constant $K_1 > 0$, such that $K_1 \|s\|^2 \leq s^T B_k s$ for all k.

(H3) There exists a constant $\sigma \in (0,1)$, such that, for all k, $\|h_k\| \Delta_k \leq \sigma$.

Lemma 1. If s_k is the solution of sub-problem (1.2) and suppose that (H1-H3) hold. Then there exist a positive scalar v such that

$$Pred_k \geq v\|g_k\|\min\{\Delta_k, \tfrac{\|g_k\|}{\|B_k\|}\} \geq v\|g_k\|\min\{\|s_k\|, \tfrac{\|g_k\|}{\|B_k\|}\} \quad (3.1)$$

$$g_k^T s_k \leq -(1-\sigma)v\|g_k\|\min\{\Delta_k, \tfrac{\|g_k\|}{\|B_k\|}\} \leq 0 \quad (3.2)$$

hold for all k.

Proof. The proof is analogous to Theorem 3.1 in [10].

Lemma 2. Let $\{x_k\}$ be the sequence generated by Algorithm 2.1. Then, for all k and $0 < \gamma < v$, we have

$$f_{k+1} \leq D_{k+1} \leq D_k \quad (3.3)$$

Proof. If $k \in I$, i.e., $r_k \geq c_0$. By the definition of r_k and $r_k \geq c_0$, we have

$$D_k - f_{k+1} \geq c_0 Pred_k \geq c_0 v\|g_k\|\min\{\Delta_k, \tfrac{\|g_k\|}{\|B_k\|}\} \geq 0 \quad (3.4)$$

Therefore, we have $D_k \geq f_{k+1}$.

If $k \in J$, i.e., $r_k < c_0$. From (2.5) and (3.2), we have

$$f_{k+1} - D_k \leq \delta a_k (g_k^T s_k + \gamma\|g_k\|\min\{\Delta_k, \tfrac{\|g_k\|}{\|B_k\|}\}) \leq 0 \quad (3.5)$$

So, we have $D_k \geq f_{k+1}$.

From (2.4), (3.4) and (3.5), we have

$$D_{k+1} - D_k = \eta_k (D_k - f_{k+1}) + f_{k+1} - D_k = (\eta_k - 1)(D_k - f_{k+1}) \leq 0 \quad (3.6)$$

$$D_{k+1} - f_{k+1} = f_{k+1} + \eta_k (D_k - f_{k+1}) - f_{k+1} = \eta_k (D_k - f_{k+1}) \geq 0 \quad (3.7)$$

These inequalities yield $f_{k+1} \leq D_{k+1} \leq D_k$.

Remark: Lemma 2 indicates that $\{D_k\}$ is not increasing monotonically and is convergent.

Lemma 3. (See Lemma 4 in [11]) Suppose that the sequence $\{x_k\}$ is generated by Algorithm 2.1. For any $k \in J$, the non-monotone line search terminates in a finite number of steps.

Lemma 4. Suppose that (H1-H3) hold, and the sequence $\{x_k\}$ is generated by Algorithm 2.1. Then there is a constant $0 < \omega < 1$ such that $a_k > \omega$ holds for all $k \in J$.

Proof. There exists a constant $M > K_1$ such that $\|\nabla^2 f(x)\| \leq M$. We consider two cases:

Case 1. If $a_k = 1$, the conclusion holds obviously.

Case 2. If $a_k < 1$, from the Algorithm 2.1, we know that a_k is the largest step-size which assures the descent of the objective function at the current point x_k. Notice that $0 < \lambda < 1$, we can obtain $\lambda^{-1} a_k > a_k$. So we have

$$f(x_k + \lambda^{-1} a_k s_k) > D_k + \delta \lambda^{-1} a_k (g_k^T s_k + \gamma \|g_k\| \min\{\Delta_k, \tfrac{\|g_k\|}{\|B_k\|}\}) \tag{3.8}$$

By Taylor's expansion, we obtain

$$f(x_k + \lambda^{-1} a_k s_k) = f_k + \lambda^{-1} a_k g_k^T s_k + \tfrac{1}{2} \lambda^{-2} a_k^2 s_k^T \nabla^2 f(\xi_k) s_k \tag{3.9}$$

where $\xi_k \in (x_k, x_k + \lambda^{-1} a_k s_k)$. Using (3.8) and (3.9), we get

$$\lambda^{-1} a_k g_k^T s_k + \tfrac{1}{2} \lambda^{-2} a_k^2 s_k^T \nabla^2 f(\xi_k) s_k > \delta \lambda^{-1} a_k (g_k^T s_k + \gamma \|g_k\| \min\{\Delta_k, \tfrac{\|g_k\|}{\|B_k\|}\}) \tag{3.10}$$

$$\begin{aligned} &-(1-\delta) g_k^T s_k \\ &< \tfrac{1}{2} \lambda^{-1} a_k s_k^T \nabla^2 f(\xi_k) s_k - \delta\gamma \|g_k\| \min\{\Delta_k, \tfrac{\|g_k\|}{\|B_k\|}\} \\ &< \tfrac{1}{2} \lambda^{-1} a_k \|s_k\|^2 M - \delta\gamma \|g_k\| \min\{\Delta_k, \tfrac{\|g_k\|}{\|B_k\|}\} \end{aligned} \tag{3.11}$$

Using the formula (3.11) and above assumptions, we have

$$\frac{(1-\delta) K_1 \|s_k\|^2}{2(1+\sigma)} + \delta\gamma \|g_k\| \min\{\Delta_k, \tfrac{\|g_k\|}{\|B_k\|}\} < \tfrac{1}{2} \lambda^{-1} a_k M \|s_k\|^2 \tag{3.12}$$

We obtain $a_k > \omega$ for all $k \in J$, where $\omega = \frac{2\lambda}{M\|s_k\|^2} (\frac{(1-\delta) K_1 \|s_k\|^2}{2(1+\sigma)} + \delta\gamma \|g_k\| \min\{\Delta_k, \frac{\|g_k\|}{\|B_k\|}\})$.

Lemma 5. (See Lemma 4 in [11]) Suppose that (H1-H3) hold and ∇f is Lipschitz continuous, i.e., there exists a constant $L > 0$ such that $\|\nabla f(x) - \nabla f(y)\| \le L \|x - y\|$. And the sequence $\{x_k\}$ is generated by Algorithm 2.1. Then

$$\pi (\|g_k\| \min\{\Delta_k, \tfrac{\|g_k\|}{\|B_k\|}\})^2 \le D_k - D_{k+1} \tag{3.13}$$

where $\pi = \frac{(1-\eta_k)\delta(1-\sigma)v(v-\gamma)}{L\Delta^2}$.

Lemma 6. (See Lemma 3.7 in [12]) Suppose that (H1-H3) hold, and the sequence $\{x_k\}$ is generated by Algorithm 2.1. Suppose that there is a positive number $\varepsilon > 0$ such that $\|g_k\| \ge \varepsilon$ for all k. Then

$$\lim_{k \to \infty} \min\{\Delta_k, \tfrac{\varepsilon}{M_k}\} = 0 \tag{3.14}$$

where $M_k = 1 + \max_{0 \le i \le k} \|B_i\|$.

Theorem 7. Suppose that (H1-H3) hold, and the sequence $\{x_k\}$ generated by Algorithm 2.1, and there exists a constant $\Delta > 0$ such that $\Delta_k < \Delta$. Then we have

$$\lim_{k\to\infty} \|g_k\| = 0. \tag{3.15}$$

Proof. From Lemma 5, we have

$$0 \le \pi(\|g_k\| \min\{\Delta_k, \tfrac{\|g_k\|}{\|B_k\|}\})^2 \le D_k - D_{k+1} \tag{3.16}$$

And from Lemma 5, we also know

$$\lim_{k\to\infty} (D_k - D_{k+1}) = 0 \tag{3.17}$$

Therefore, we can obtain

$$\lim_{k\to\infty} \|g_k\| \min\{\Delta_k, \tfrac{\|g_k\|}{\|B_k\|}\} = 0 \tag{3.18}$$

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

Hence, the proof is completed.

4. Conclusions

In this paper, we propose a new non-monotone trust region method with line search based on the conic model. It is useful to take advantage of non-monotone Wolfe-type line search which can find an iterative point easily if the trial step is not accepted. The global convergence result of the new proposed method is proved under some mild conditions.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (61473111) and the Natural Science Foundation of Hebei Province (Grant No. A2014201003, F2014201100).

References

[1]. R. Fletcher, Practical Method of Optimization, Unconstrained Optimization,vol.1, John Wiley, New York, 1980.

[2]. M. J. D. Powell, On the global convergence of trust region algorithms for unconstrained optimization, Mathematical Programming, 29(3), 1984, 297-303.

[3]. Y. Yuan, W. Sun, Optimization Theory and Methods, Science Press of China, 1997.

[4]. X. Zhang, J. Zhang, L. Liao, An adaptive trust region method and its convergence, Science in China A. 45(5), 2002, 620–631.

[5]. J. Nocedal, Y. Yuan, Combining trust region and linear search techniques, in: Y. Yuan (Ed.), Advances in Nonlinear Programming, Kluwer Academic Publishers, Dordrecht, 4, 1998, 153-175.

[6]. L. Grippo, F. Lamparillo, S. Lucidi, A nonmonotone line search technique for Newton's method, SIAM J. Numer. Anal. 23(4), 1986, 707-716.

[7]. N. Deng, Y. Xiao, F. Zhou, Nonmontonic trust region algorithm, Journal of Optimization Theory and Application, 76(2), 1993, 259-285.

[8]. N. Z. Gu, J. T. Mo, Incorporating nonmonotone strategies into the trust region method for unconstrained optimization, Journal of Computers& Mathematics with Applications, 55(9), 2008, 2158-2172.

[9]. J. Mo, C. Liu, S. Yan, A nonmonotone trust region method based on nonincreasing technique of weighted average of the successive function values, Journal of Computation and Applied Mathematics, 209(1), 2007, 97-108.

[10]. S. J. Qu, K. C. Zhang, J. Zhang, A nonmonotone trust region method of conic model for unconstrained optimization, Journal of Computational and Applied Mathematics, 220(1-2), 2008, 119-128.

[11]. X. T. Zeng, A new family of trust region algorithms with a non-monotone line search technique, Journal of Binzhou University, 28(6), 2012, 77-83.

[12]. Z. C. Cui, B. Y. Wu, S. J. Qu, Combining nonmonotone conic trust region and line search techniques for unconstrained optimization, Journal of Computational and Applied Mathematics, 235(8), 2011, 2432-2441.