Scholars Journal of Engineering and Technology (SJET) **ISSN 2321-435X (Online)**
Sch. J. Eng. Tech., 2014; 2(4B):586-590 **ISSN 2347-9523 (Print)**
©Scholars Academic and Scientific Publisher
(An International Publisher for Academic and Scientific Resources)
www.saspublisher.com

Review Article

A Review on the Non-Monotone Trust Region Methods

Yuanyuan Wu, Qinghua Zhou*
College of Mathematics and Computer Sciences, Hebei University, Baoding, 071002, Hebei Province, China

***Corresponding author**
Qinghua Zhou
Email: qinghua.zhou@gmail.com

Abstract: Recently, non-monotone technique attracts the attention of many scholars and there are various articles about this subject. The non-monotone line search technique performs well especially in the case of the bottom of a curved narrow valley, which is a common occurrence in nonlinear problems. The purpose of this paper is to discuss the main developments of non-monotone techniques used in trust region algorithms for unconstrained optimization.
Keywords: trust region; non-monotone technique; line search

INTRODUCTION

Trust region method is a robust iterative method for solving the unconstrained problem as following:

Minimize $f(x)$

where $x \in R^n$ and f is a twice differentiable real-valued function from R to R^n. $\|\cdot\|$ is the Euclidean norm. At each given iteration point x_k, the algorithm calculates a trial step s_k by solving the sub problem:

$$\text{Minimize } \phi_k(s) = f(x_k) + g_k^T s + \frac{1}{2} s^T B_k s \qquad (1)$$

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

where $\phi_k(s)$ is a quadratic approximation of the objective function. $g_k = \nabla f(x_k)$ is the gradient of f at x_k. B_k is a symmetric matrix which is either an approximation for the Hessian matrix or the exact Hessian matrix of the objective function. s_k is called trial step. Δ_k is the trust region radius. After getting the trial step, the ratio r_k which plays an important role in deciding the acceptance of the trial step has to be computed:

$$r_k = \frac{f(x_k) - f(x_k + s_k)}{\phi_k(0) - \phi_k(s_k)} \qquad (2)$$

When r_k is greater than $u \in (0,1)$ which is a preconditioned constant, the trial step s_k is accepted. Therefore we get a new iteration point $x_{k+1} = x_k + s_k$. The trust region radius has to be updated to Δ_{k+1} according to the scheduled rules at the same time. Otherwise the trial is considered failed and we have to calculate s_k again in a shrunken trust region radius.

DEVELOPMENTS OF THE NON-MONOTONE TECHNIQUES

The ideal of non-monotone technique can be traced back to the watchdog technique proposed by Chamberlain which is aimed to overcome Marotos effect for constrained optimization by relaxing some standard linear search condition [9]. Inspired by this idea, Grippo *et al*. establish a generalization of Armijo's rule and demonstrate the global convergence to a stationary point in [10]. The non-monotone Armijo rule is described as follows:

$$f(x_k + \beta^p \alpha d_k) \le \max_{0 \le j \le m(k)} \left[f(x_{k-j}) \right] + \gamma \beta^p \alpha g_k^T d_k \qquad (3)$$

where the parameter $\alpha > 0, \lambda \in (0,1), \beta \in (0,1)$. $m(k)$ is preconditioned and satisfy the condition: $m(0) = 0, m(k) \le \min\left[m(k-1)+1, M\right], M$ is a constant stands for a nonnegative integer. p_k is the smallest nonnegative integer which satisfies (3). The new trail step is compared with the maximum point in the previous steps. Then the new iteration point can be obtained from the equation $x_{k+1} = x_k + \lambda_k d_k$ $\lambda_k = \beta^{p_k}\alpha$. The search direction d_k has to meets two demands instead of just one angle criterion:

$$g_k^T d_k \le -c_1 \|g_k\|^2, \|d_k\| \le c_2 \|g_k\|$$

where c_1,c_2 are positive constants. In paper [12], Grippo *et al*. combine the relaxed Armijo rule with truncated Newton method and analysis the advantages of the algorithm. In paper [13], Grippo and his co-writers illustrate an even further relaxation of the original linear search rule. Concisely speaking, some steps could be accepted automatically without satisfying (3), if the corresponding d_k is short enough. The value of objective function is checked once in a while to decide whether a proper decrease has been obtained when compared to $r_k = \max\limits_{j=0,1,\dots p} f_{l-j}$. l represents the last iteration at which the value of the objective function is evaluated and checked to be decreased sufficiently. The mechanism to get r_k is rather complicated, but it has a great inspiration for other studies.

In paper [11], Philippe L Toint concludes the two algorithm proposed in [10, 13] and their convergence. Later after that, several studies about the non-monotone technique referred in [10] appear [8, 15, 18, 19, 20, 24, 25]. Besides the relaxation of Armijo's rule, Wenyu Sun *et al*. explored the non-monotone Goldstein linear search rule and the non-monotone Wolfe linear search rule in [24]. They prove the convergence under mild assumptions such as Newton-type search directions, bounded level set and so on. They also combined forcing functions referred in [20] with the non-monotone line search technique and give a general line search rule for unconstrained minimization problems which is called the non-monotone F-rule.

DEVELOPMENTS OF TRUST REGION METHOD

Trust region method has been studied by lots of researchers due to its strong convergence to a stationary point which satisfies the second-order necessary conditions, see [1-7]. In order to guarantee the convergence, the sequence of the objective function value has been forced to be monotonous. This monotonicity property may be considered as natural, but not without drawbacks. Some studies [10,12,18] indicate that the monotonicity of the objective function may slow down the convergence rate of the algorithm, for example the minimization of Rosenbrock function. The reason is that the algorithm may be trapped in the neighborhood of narrow curved valleys. To overcome this deficiency, the non-monotone technique is introduced to the trust region method.

In [8, 16], Xiao, Zhou and their co-writers first applied non-monotone technique to trust region algorithm for unconstrained optimization. Zhou and Xiao [17], Xiao and Chu [21] develop the idea further. The main creation of [8] is changing the constant u which is the criterion to decide whether the iteration is successful or not to a variable $\overline{u_k}$ according to following rule:

$$u_k = \begin{cases} \dfrac{\left(f_{l(k)} - f(x_k)\right) - \gamma\Delta_k \|g_k\|}{\phi_k(s_k) - \phi_k(0)}, & M > 0 \\ u, & M = 0 \end{cases} \tag{4}$$

$$\overline{u_k} = \min\{u, u_k\} \tag{5}$$

where $f_{l(k)} = f\left(x_{l(k)}\right) = \max\limits_{0 \le j \le m(k)} \left\{f_{k-j}\right\}, m(0) = 0, m(k) = \min\left\{m(k-1)+1, M\right\}$. Thus $\overline{u_k}$ is not necessarily positive. Therefore, the objective function value $f(x_{k+1})$ at the new acceptable iteration point may not monotonically decrease compared to $f(x_k)$ while it does decrease when compared to $f_{l(k)}$. When M=0, the algorithm becomes traditional trust region method. Equation (2), (4), (5) and the acceptability criterion $r_k > \overline{u_k}$ together imply that $f_{l(k)} - f(x_{k+1}) \ge \gamma\Delta_k \|g_k\|$. In fact, this inequality is quite similar to the relaxed Armijo's rule in [10]. Because the use of $\overline{u_k}$, the acceptability criterion would not be too strict, for example, the trial step is acceptable when $r_k \in [u, u_k)$. They prove the convergence properties of this algorithm under certain assumptions including:

$$\|s_k\| \le c_2 \|g_k\| \qquad (6)$$

and propose a specialized algorithm for the subproblem which ensures the solution of the approximate model satisfies (6).

Toint in [22] once points out that condition (6) may prevent large steps in case of saddle points or the bottom of valleys. He insists that the validity of condition (6) still has to be discussed. Toint also advances a new non-monotone trust region algorithm in [22] .The key technique of it is a modification on the acceptance rule of the trial point. The non-monotone ratio in [22] is as follows:

$$r_{1,k} = \begin{cases} \dfrac{f_{l(k)} - f(x_k + s_k)}{\sum_{i=l(k)}^{k} d_i}, & if\ l > 0 \\ 0, & otherwise \end{cases}, r_{2,k} = \frac{f(x_k) - f(x_k + s_k)}{\phi_k(0) - \phi_k(s_k)}, r_k = \max\{r_{1,k}, r_{2,k}\}$$

where $d_k = \phi_k(0) - \phi_k(s_k)$. At variance with the scheme in [8], the proposed non-monotone method requires the storage of the last p terms of the sequence $\{d_k\}$. p stands for how far the current iterate to search in the past for its associated reference iteration. To overcome some unwanted occurrences like successive increasing function values of successful iterations, Toint proposes an adaptive mechanism for the associated function value used in the ration. Dai and Y.H in [15] have also suggested that a standard Armijo line search rule should be adopted when condition (3) is not satisfied.

As is known, trust region method is sensitive to the initial radius Δ_0 and the way to update Δ_k . Zhang *et al*. in [14] utilize adaptive trust region radius technique to the non-monotone algorithm and study its convergence properties. They try to find the solution of the subproblem in Δ_k which is constructed of the information of g_k and B_k as follows:

$$\Delta_k = c^p \|g_k\| \overline{M_k}, 0 < c < 1, \overline{M_k} = \|\widehat{B}_k^{-1}\|$$

p is a nonnegative integer, $\widehat{B}_k$ can be obtained through Schnabel and Eskow modified Cholesky factorization and is a positive definite matrix. They define the ratio in [14] is as follows:

$$r_k = \frac{Ared_k}{\Pr ed_k} = \frac{f_{l(k)} - f(x_k + s_k)}{\phi(0) - \phi_k(s_k)} \qquad (7)$$

It is the widely used in many papers. Wenyu Sun in [18] explore the ration (7) in depth and prove convergence properties of the algorithm based on it. In a subsequent paper [23], Jinhua Fu and Wenyu Sun propose another non-monotone adaptive trust region method. They make a modification on $\Pr ed_k$ of (7) as follows:

$$\Pr ed_k = f_{l(k)} - \phi_k(s_k)$$

They solve the subproblem through the truncated conjugate gradient method instead of solving it exactly as in [14].

In 2004, H.C. Zhang and W.W. Hager propose a new non-monotone algorithm based on weighted average of the function values [26]. They use the following C_k to replace $\max_{0 \le j \le m(k)} \left[f(x_{k-j}) \right]$ in the right hand of (3).

$$C_k = \begin{cases} f(x_k), & k = 1 \\ \dfrac{\left(\eta_{k-1} Q_{k-1} C_{k-1} + f(x_k)\right)}{Q_k}, & k \ge 2 \end{cases}$$

$$Q_k = \begin{cases} 1, & k = 1 \\ \eta_{k-1} Q_{k-1} + 1, & k \ge 2 \end{cases}$$

where $\eta_{k-1} \in [\eta_{\min}, \eta_{\max}]$, $\eta_{\min} \in [0, \eta_{\max})$, $\eta_{\max} \in [0,1]$. Jiangtao Mo *et al*. in [27] propose that Q_k should be abandoned. They constructs D_k to instead of $f_{l(k)}$ in (7). D_k is described as follows:

$$D_k = \begin{cases} f(x_k) & k = 1 \\ \eta D_{k-1} + (1-\eta) f(x_k) & k \geq 2 \end{cases}$$

In [30], Jinghui Liu and his co-writers propose another non-monotone line search rule and establish a ration r_k based on D_k in [30]. They use line search technique to resolve the subproblem when the iteration is unsuccessful. M. Ahookhosh*et al*. in [28,29] introduce a new formula instead of $f_{l(k)}$ in (7), say:

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

where $\eta_k \in [\eta_{\min}, \eta_{\max}]$, $\eta_{\min} \in [0,1)$, $\eta_{\max} \in [\eta_{\min}, 1]$. They advise that a weaker non-monotone strategy should be used when the iteration point is close to the minimum while a stronger strategy should be adopted in the opposite case. This idea to adjust the nonmonotonicity of the algorithm can be realized by selecting η_k .The idea is also true to the step length.

ANALYSIS

The non-monotone techniques have lots of advantages, however, it also suffers from some drawbacks. One is that s_k produced at each iteration is essentially abandoned according to the maximum function value $f_{l(k)}$ in the previous iterations. To overcome it, some researchers begin to change to the reference function value $f_{l(k)}$ in $\mathrm{Pr}\,ed_k$ of the ratio r_k, see [26-30]. In general, there are two ways about choosing reference function value: one is based on the previous maximum function value $f_{l(k)}$, the other is a linear combination of several function values. There are not sufficient studies on comparing calculation efficiency between them. Another drawback is that non-monotone line search method seems unsuitable for functions that do not have strong nonlinearity. For example, the Brown and Dennis function. Papers [15, 22] provide some advices to deal with it. Besides, we realize that the ratio r_k of actual descent and predicted descent is essential both in selecting the new iteration point and updating the trust region radius Δ_k. In fact, the key content of many papers is the modification on r_k. Nevertheless, there is no explicit research about how the mechanism of obtaining $m(k)$ and the construction of $\mathrm{Pr}\,ed_k$ affect the efficiency of the non-monotone trust region algorithm.

From the above analysis we can learn that although most of the researches about the non-monotone trust region algorithm are promising, there are still lots of problems to tackle with.

Acknowledgements

This work is supported by the National Nature Science Foundation of China (Grant No. 11101115) and the Natural Science Foundation of Hebei Province (Grant No. A2014201003, A2014201100).

REFERENCES

1. Fletcher R; Practical Method of Optimization. Unconstrained Optimization, Volume 1, 1980.
2. More JJ; Recent developments in algorithms and software for trust region methods. Mathematical Programming: The State of the Art,1983; 258-287.
3. Powell MJD; Convergence properties of class minimization algorithms. Nonlinear Programming, 1975; 2:l-27
4. Powell MJD; On the global convergence of trust region algorithms for unconstrained optimization. Math. Prog, 1984; 29:297-303.
5. Schultz GA, Schnabel RB, Byrd RH; A family of trust region-based algorithms for unconstrained minimization with strong global convergence. SIAM J. Namer. Anal, 1985; 22:47-67.
6. Yuan Y; On the convergence of trust region algorithms, Mathematics Numerica Sinica, 1996; 16:333-346.
7. Yuan Y, Sun W; Optimization: Theorem and Application, (in Chinese), Science Publish House of China, 1997.
8. Deng NY, Xiao Y, Zhou FJ; Nonmonotonic trust region algorithm. J. Optim. Theory Appl, 1993; 76:259–285.

9. Chamberlain RM, Powell MJD, Lemarechal C, Pedersen HC; The watchdog technique for forcing convergence in algorithms for constrained optimization. Math. Program. Study. 1982;16:1–17.
10. Grippo L, Lamparillo F, Lucidi S; A nonmonotone line search technique for Newton's method. SIAM Journal of Numerical Analysis,1986; 23:707–716.
11. Toint PhL; An assessment of nonmonotone linesearch technique for unconstrained optimization. SIAM Journal of Scientific Computing, 1996; 17: 725–739.
12. Grippo L, Lamparillo F, Lucidi S; A truncated Newton method with nonmonotone line search for unconstrained optimization. Journal of Optimization Theory and Applications, 1989; 60:401–419.
13. Grippo L, Lamparillo F, Lucidi S; A class of nonmonotone stabilization methods in unconstrained optimization. Numerical Mathematics, 1991; 59:779–805.
14. Zhang JL, Zhang XS; A nonmonotone adaptive trust region method and its convergence. Computers and Mathematics with Applications, 2003; 45:1469–1477.
15. Dai YH; On the nonmonotone line search, J. Optim. Theory Appl, 2002; 112:315–330.
16. Xiao Y, Zhou FJ; Nonmonotone trust region methods with curvilinear path in unconstrained optimization. Computing, 1992; 48:303–317.
17. Zhou F, Xiao Y; A class of nonmonotone stabilization trust region methods. Computing, 1994; 53(2):119–136.
18. Sun WY; Nonmonotone trust region method for solving optimization problems. Appl. Math. Comput, 2004; 156(1):59–174.
19. Grippo L, Sciandrone M; Nonmonotone globalization techniques for the Barzilai-Borwein gradient method. Comput. Optim, 2002; 23:143–169.
20. Ortega JM, Rheinboldt WC; Iterative Solution of Nonlinear Equations in Several Variables. Academic Press, 1970.
21. Xiao Y, Chu EKW; Nonmonotone trust region methods. MonashUniversity, Technical Report, 1995; 95/17.
22. Toint PhL; Non-monotone trust region algorithm for nonlinear optimization subject to convex constraints. Math. Program, 1997; 77:69–94.
23. Fu J, Sun W; Nonmotone adaptive trust region method for unconstrained optimization problems. Appl. Math. Comput, 2005; 163(1):489–504.
24. Sun WY, Han JY, Sun J; Global convergence of nonmonotone descent methods for unconstrained optimization problems. J. Comput. Appl. Math, 2002; 146: 89–98.
25. Liu G, Han J, Sun D; Global convergence of BFGS algorithm with nonmonotone linesearch. Optim, 1995; 34:147–159.
26. Zhang HC, Hager WW; A nonmonotone line search technique and its application to unconstrained optimization". SIAM J. Optim, 2004; 14 (4):1043–1056.
27. Mo J，Liu C，Yan S; A nonmonotone trust region method based on nonincreasing technique of weighted average of the successive function values. Journal of Computational and Applied Mathematics 2007;209:97-108.
28. Ahookhosh M, Amini K; A nonmonotone trust region method with adaptive radius for unconstrained optimization problems. Computers and Mathematics with Applications, 2010; 60: 411-422.
29. Ahookhosh M, Amini K, Peyghami MR; A nonmonotone trust region line search method for large-scale unconstrained optimization. Applied Mathematical Modelling, 2012; 36:478-487.
30. Liu J, Ma C; A nonmonotone trust region method with new inexact line search for unconstrained optimization. Numerical Algorithms, 2013; 64(1):1-20.