

A non-monotone trust region method for constrained optimizationQinghua Zhou¹, Hao Zhao²¹College of Mathematics and Information Science, Hebei University, Baoding, 071002, Hebei Province, China.²The School of Electronic and Information Engineering, Xi'an Jiaotong University, 710049, Shanxi Province, China.***Corresponding Author:**

Qinghua Zhou

Email: 48458272@qq.com

Abstract: In this paper, we propose a non-monotone trust region method for solving equality constrained optimization problem. Unlike the traditional trust region methods, our algorithm uses a new non-monotone strategy. Theoretical analysis indicates that the new method preserves the global convergence under some mild conditions.

Keywords: non-monotone technique; trust region method; equality constrained optimization; global convergence.

INTRODUCTION

Trust region method is a well-known method to solve the equality constrained optimization problem:

$$\begin{aligned} \min \quad & f(x), \\ \text{s.t.} \quad & c(x) = 0, \end{aligned} \quad (1.1)$$

where $f(x): R^n \rightarrow R$, $c(x) = (c_1(x), c_2(x), \dots, c_m(x))^T$, $c_i(x): R^n \rightarrow R^m$ ($m \leq n$), $i = (1, 2, \dots, m)$. $f(x)$ and $c_i(x)$ are twice continuously differentiable.

Trust region methods have strong convergence, and have been proved to be efficient for both unconstrained and constrained optimization problems in [1-5].

In 1986, the non-monotone technique was originally proposed by Grippo et al. [6] for unconstrained optimization problems based on Newton's method, in which the step-size α_k satisfies the following condition:

$$f(x_k + \alpha_k d_k) \leq f(x_{l(k)}) + \beta \alpha_k \nabla f(x_k)^T d_k, \quad (1.2)$$

where $\beta \in (0, \frac{1}{2})$, $f(x_{l(k)}) = \max_{0 \leq j \leq m_k} f(x_{k-j})$, $m_0 = 0$, $0 \leq m_k \leq \min\{m_{k-1} + 1, M\}$ ($k \geq 1$), and $M \geq 0$ is an integer. It pointed out the non-monotone techniques are helpful to overcome the case that the sequence follows the bottom of curved narrow valleys, a common occurrence in difficult nonlinear problems. Since then, many non-monotone algorithms are proposed to solve the unconstrained and constrained optimization problems [6–13]. Numerical tests show that the performance of the non-monotone technique is superior to those of the monotone cases. In 1993, Deng et al. in [8] made some changes and applied it to the trust region method, and proposed a non-monotone trust region method for unconstrained optimization. Theoretical analysis and numerical results show that algorithms with non-monotone strategy are more effective than algorithms without it. From then on a variety of the non-monotone trust region methods have been presented [9, 13].

Recently, Gu and Mo [13] introduced another non-monotone strategy. They replaced the maximum function value in (1.2) with D_k . In detail, their method finds a step-size α_k satisfying the following condition:

$$f(x_k + \alpha_k d_k) \leq D_k + \beta \alpha_k \nabla f(x_k)^T d_k, \quad (1.3)$$

where

$$D_k = \begin{cases} f_k, & k = 0, \\ \eta_k D_{k-1} + (1 - \eta_k) f_k, & k \geq 1. \end{cases} \quad (1.4)$$

and $\eta_k \in [\eta_{\min}, \eta_{\max}]$, $\eta_{\min} \in [0, 1)$ and $\eta_{\max} \in [\eta_{\min}, 1)$ are two chosen parameters. Numerical results showed that this non-monotone technique was superior to (1.2).

In this paper, we further extend the non-monotone technique to equality constrained optimization. In order to introduced our algorithm, we will do some notations as follows: $g(x) = \nabla f(x)$, $A(x) = (\nabla c_1(x), \nabla c_2(x), \dots, \nabla c_m(x)) \in R^{n \times m}$. Assuming that $A(x)$ has full column rank, we define the projective matrix

$$Z(x) = I - A(x)(A(x)^T A(x))^{-1} A(x)^T \in R^{n \times n} \quad (1.5)$$

and the Lagrange function

$$L(x, \lambda) = f(x) + \lambda^T c(x) \quad (1.6)$$

where $\lambda(x) = (A(x)^T A(x))^{-1} A(x)^T g(x)$.

At each iteration, we calculate the trust region trial step as follows (see [14]):

First, we calculate

$$v_k = -\alpha_k A(x)(A(x_k)^T A(x_k))^{-1} c(x_k) \quad (1.7)$$

$$\text{where } \alpha_k = \begin{cases} 1, & c_k = 0, \\ \min\{1, \frac{\Delta_k}{\|A_k (A(x_k)^T A(x_k))^{-1} c_k\|}\}, & \text{otherwise.} \end{cases}$$

Then, we solve the trust region sub-problem:

$$\begin{aligned} \min \quad & (Z_k g_k)^T \omega + \frac{1}{2} \omega^T (Z_k B_k Z_k) \omega \\ \text{s.t.} \quad & \|\omega\| \leq \Delta_k, \end{aligned} \quad (1.8)$$

where $B_k \in R^{n \times n}$ is a symmetric matrix which is the Hessian matrix or its approximation of $L(x, \lambda)$ at the current point x_k , $\Delta_k > 0$ is called the trust radius. Let ω_k be the solution of (1.8) and

$$h_k = Z_k \omega_k \quad (1.9)$$

The trust region trail step is taken as

$$d_k = h_k + v_k \quad (1.10)$$

To test whether the point $x_k + d_k$ can be accepted as the next iteration, we use the Fletcher's exact penalty function as the merit function as follows:

$$\psi(x, \lambda, \sigma) = f(x) + \lambda^T c(x) + \sigma \|c(x)\|^2 \quad (1.11)$$

where $\sigma > 0$ is the penalty parameter.

We define

$$F_k = \begin{cases} \psi(x_k, \lambda_k, \sigma_k), & \text{if } k = 0, \\ \eta_k F_{k-1} + (1 - \eta_k) \psi(x_k, \lambda_k, \sigma_k), & \text{if } k \geq 1, \end{cases} \quad (1.12)$$

where $\eta_k \in [\eta_{\min}, \eta_{\max}]$, $\eta_{\min} \in [0, 1)$ and $\eta_{\max} \in [\eta_{\min}, 1)$ are two chosen parameters.

The paper is organized as follows. In Section 2, we describe our new non-monotone trust region method for constrained optimization. The properties of this new algorithm and the global convergence property are given in Section 3. Finally, some conclusions are given in Section 4.

New non-monotone trust region algorithm

A point x is called a stationary point of problem (1.1) if it satisfies

$$\|Z(x)^T g(x)\| + \|c(x)\| = 0, \quad (2.1)$$

The traditional actual reduction from x_k to $x_k + d_k$ is defined by

$$TAred_k = \psi(x_k, \lambda_k, \sigma_k) - \psi(x_k + d_k, \lambda_{k+1}, \sigma_k). \quad (2.2)$$

While the non-monotone actual reduction and the predicted one are defined by the two following equalities respectively,

$$Ared_k = F_k - \psi(x_k + d_k, \lambda_{k+1}, \sigma_k) \quad (2.3)$$

$$\begin{aligned} Pred_k = & -g_k^T d_k - \frac{1}{2} d_k^T B_k d_k - \nabla \lambda_k^T (c_k + A_k^T d_k) - \lambda_k^T A_k^T d_k \\ & + \sigma_k (\|c_k\|^2 - \|c_k + A_k^T d_k\|^2) \end{aligned} \quad (2.4)$$

After obtaining d_k , then the ratio r_k is computed by

$$r_k = \frac{Ared_k}{Pred_k}, \quad (2.5)$$

Algorithm 2

Step 1. Given $x_0 \in R^n$, $\Delta_0 > 0$, $\sigma_0 > 0$, $0 < \mu < 1$, $0 < c_1 < c_2 < 1$, $c_3 > 0$, $B_0 \in R^{n \times n}$,

Choose parameters $\eta_{\min} \in [0, 1)$ and $\eta_{\max} \in [\eta_{\min}, 1)$, Set $k = 0$.

Step 2. If $\|Z(x)^T g(x)\| + \|c(x)\| = 0$, stop. Otherwise, go to Step 3.

Step 3. Solve the trust region trial step d_k .

Step 4. Set $\sigma_k = \sigma_k$, if $Pred_k \geq \frac{1}{2} \sigma_k (\|c_k\|^2 - \|c_k + A_k^T d_k\|^2)$, and then set

$$\sigma_k = \max\{\sigma_k, 2 \frac{g_k^T d_k + \frac{1}{2} d_k^T B_k d_k + \nabla \lambda_k^T (c_k + A_k^T d_k) + \lambda_k^T A_k^T d_k}{\|c_k\|^2 - \|c_k + A_k^T d_k\|^2}\} \quad (2.6)$$

Step 5. Compute F_k and r_k .

Step 6. Set $x_{k+1} = x_k + d_k$, if $r_k \geq \mu$. Otherwise, set $x_{k+1} = x_k$.

Step 7. Compute Δ_{k+1} as follows:

$$\Delta_{k+1} = \begin{cases} [c_1 \|d_k\|, c_2 \Delta_k], & \text{if } r_k < \mu, \\ \Delta_k, & \text{if } r_k \geq \mu, \|d_k\| < \Delta_k, \\ [\Delta_k, c_3 \Delta_k], & \text{if } r_k \geq \mu, \|d_k\| = \Delta_k. \end{cases} \quad (2.7)$$

go to Step 2

Step 8. Update B_k and choose $\eta_k \in [\eta_{\min}, \eta_{\max}]$. Set $k = k + 1$, go to Step 2.

CONVERGENCE

In this section, we will prove the global convergence property of Algorithm 2. The following assumptions are necessary to analyze the convergence property.

(H1) The sequence $\{x_k\}$ and $\{x_k + d_k\}$ are contained in a compact set Ω .

(H2) The matrix B_k is uniformly bounded, i.e., there exists a constant $M_1 > 0$, such that, for all k , $\|B_k\| \leq M_1$

(H3) $A(x)$ is of column full rank for all $x \in \Omega$.

Lemma 1. Assume (H1-H3) hold, then there exists a positive constant K_1 such that

$$\begin{aligned} \|c_k\|^2 - \|c_k + A_k^T d_k\|^2 & \geq K_1 \|c_k\|^2 \min\{\|c_k\|^2, \Delta_k\} \\ Pred_k & \geq \frac{1}{2} \sigma_k K_1 \|c_k\| \min\{\|c_k\|, \Delta_k\} \end{aligned} \quad (3.1)$$

Lemma 2. Let $\zeta_k(d) = g_k^T d - \frac{1}{2} d^T B_k d$, assume that (H1-H3) hold. Then, there exists a positive constant K_2 such that

$$\zeta_k(d_k) \leq \zeta_k(v_k) - K_2 \|Z_k(g_k + B_k v_k)\| \min\{\frac{\|Z_k(g_k + B_k v_k)\|}{M+1}, \Delta_k\} \quad (3.2)$$

Lemma 3. Assume (H1-H3) hold, then there exists a positive constant K_3 such that

$$|Ared_k - Pred_k| \leq K_3 \sigma_k \|d_k\|^2. \quad (3.3)$$

Lemma 4. Assume (H1-H3) hold, then there exists a positive constant K_4 such that

$$\begin{aligned} Pred_k &\geq K_2 \|Z_k(g_k + B_k v_k)\| \min\left\{\frac{\|Z_k(g_k + B_k v_k)\|}{M+1}, \Delta_k\right\} \\ &\quad - K_4 \|c_k\| + \sigma_k (\|c_k\|^2 - \|c_k + A_k^T d_k\|^2). \end{aligned} \quad (3.4)$$

The above lemmas (Lemma 1-4) are useful to analyze the convergence of the Algorithm 2, and the proofs are similar to [3].

Lemma 5. Suppose the sequence $\{x_k\}$ generated by Algorithm 2, we have

$$\psi_{k+1} \leq F_{k+1} \text{ for all } k \quad (3.5)$$

Proof. We obtain $F_{k+1} - \psi(x_{k+1}, \lambda_{k+1}, \sigma_{k+1}) = \eta_{k+1} (F_k - \psi(x_{k+1}, \lambda_{k+1}, \sigma_{k+1}))$ from the definition of F_k . By $r_k \geq \mu$ and Lemma 1, we have

$$\begin{aligned} F_k - \psi(x_{k+1}, \lambda_{k+1}, \sigma_{k+1}) \\ \geq \mu Pred_k \geq \frac{\mu}{2} \sigma_k K_1 \|c_k\| \min\{\|c_k\|, \Delta_k\} \geq 0 \end{aligned} \quad (3.6)$$

So

$$\begin{aligned} F_{k+1} - \psi(x_{k+1}, \lambda_{k+1}, \sigma_{k+1}) \\ = \eta_{k+1} (F_k - \psi(x_{k+1}, \lambda_{k+1}, \sigma_{k+1})) \\ \geq \mu \eta_{k+1} Pred_k \geq \frac{\mu}{2} \sigma_k K_1 \|c_k\| \min\{\|c_k\|, \Delta_k\} \geq 0 \end{aligned} \quad (3.7)$$

Therefore, we have $\psi_{k+1} \leq F_{k+1}$.

Lemma 6. Suppose that the (H1-H3) hold and the sequence $\{x_k\}$ generated by Algorithm 2. Then the algorithm is well defined.

Proof. Similar to Theorem 2 in [12].

Lemma 7. Under (H1), if $\|Z(x)^T g(x)\| + \|c(x)\| \neq 0$, then there exist a integer k_0 and a positive constant σ^* such that for all $k \geq k_0$, $\sigma_k = \sigma^*$.

Proof. Similar to Lemma 7.11 in [3].

Theorem 8. Suppose that Assumptions (H1-H3) hold. Let the sequence $\{x_k\}$ generated by Algorithm 2, then we have

$$\liminf_{k \rightarrow \infty} \|c_k\| = 0. \quad (3.8)$$

Proof. We assume that Formula (3.15) is not true, that is, there exists a constant $\varepsilon > 0$, such that

$$\|c_k\| \geq \varepsilon \text{ for all } k. \quad (3.9)$$

From (1.12), (3.6) and Lemma 7, we obtain that

$$\begin{aligned} F_{k+1} &= \eta_{k+1} F_k + (1 - \eta_{k+1}) \psi(x_{k+1}, \lambda_{k+1}, \sigma_{k+1}) \\ &\leq \eta_{k+1} F_k + (1 - \eta_{k+1}) (F_k - \frac{1}{2} \mu \sigma_k K_1 \|c_k\| \min\{\|c_k\|, \Delta_k\}) \\ &= F_k - (1 - \eta_{k+1}) [\frac{1}{2} \mu \sigma_k K_1 \|c_k\| \min\{\|c_k\|, \Delta_k\}], \end{aligned} \quad (3.10)$$

By assumptions, we know that $\{F_k\}$ is bounded, let $k \rightarrow \infty$, we have

$$\begin{aligned}
 +\infty &> \sum_{k=1}^{\infty} F_k - F_{k+1} = F_1 - \lim_{k \rightarrow \infty} F_{k+1} \\
 &\geq \sum_{k \in I} F_k - F_{k+1} \\
 &\geq \sum_{k \in I} \frac{1}{2} (1 - \eta_{k+1}) \mu \sigma_k K_1 \|c_k\| \min\{\|c_k\|, \Delta_k\} \\
 &\geq \sum_{k \in I} \frac{1}{2} (1 - \eta_{k+1}) \mu \sigma^* K_1 \|c_k\| \min\{\|c_k\|, \Delta_k\} \\
 &\geq \sum_{k \in I} \frac{1}{2} (1 - \eta_{\max}) \mu \sigma^* K_1 \|c_k\| \min\{\|c_k\|, \Delta_k\}
 \end{aligned} \tag{3.11}$$

From (3.9) and (3.11), we have

$$\lim_{k \rightarrow \infty} \Delta_k = 0. \tag{3.12}$$

But we get $r_k > \mu$, and therefore we have $\Delta_{k+1} > \Delta_k$, which is contradiction to (3.12). Theorem 8 has been proved.

Theorem 9. If (H1) holds, we have

$$\liminf_{k \rightarrow \infty} \|Z_k^T g_k\| = 0. \tag{3.13}$$

Proof. Similar to the proof of Theorem 4 in [12].

Theorem 10. If (H1-H3) hold, we have

$$\liminf_{k \rightarrow \infty} \|Z_k^T g_k\| + \|c_k\| = 0. \tag{3.14}$$

Proof. Based on Theorem 8 and Theorem 9, we get the Theorem 10 is true.

CONCLUSIONS

In this paper, we present a new non-monotone trust region method based on average of the successive penalty function values for nonlinear optimization. The form of the new method is very simple. Under some mild conditions, we proved the global convergence result of the proposed method.

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, A2014201100).

REFERENCES

1. El-Alem, M. (1999). A Global convergence theory for Dennis. *El-Alem and Maciel's class of trust region*.
2. Byrd, R. H., Schnabel, R. B., & Shultz, G. A. (1987). A trust region algorithm for nonlinearly constrained optimization. *SIAM Journal on Numerical Analysis*, 24(5), 1152-1170.
3. Dennis, J. E., & Vicente, L. N. (1997). On the convergence theory of trust-region-based algorithms for equality-constrained optimization. *SIAM Journal on Optimization*, 7(4), 927-950.
4. Powell, M. J. D., & Yuan, Y. (1990). A trust region algorithm for equality constrained optimization. *Mathematical Programming*, 49(1), 189-211.
5. Yuan, Y.X. (1994). Convergence of trust region methods, *Chinese Journal of Numerical Mathematics and Applications*, 16(4), 92-106.
6. Grippo, L., Lampariello, F., & Lucidi, S. (1986). A nonmonotone line search technique for Newton's method. *SIAM Journal on Numerical Analysis*, 23(4), 707-716.
7. Chen, Z., & Zhang, X. (2004). A nonmonotone trust-region algorithm with nonmonotone penalty parameters for constrained optimization. *Journal of Computational and Applied Mathematics*, 172(1), 7-39.
8. Deng, N. Y., Xiao, Y., & Zhou, F. J. (1993). Nonmonotonic trust region algorithm. *Journal of optimization theory and applications*, 76(2), 259-285.
9. Li, Z. F., & Deng, N. Y. (1999). A class of new nonmonotone trust region algorithm and its convergence. *Acta Mathematicae Applicatae Sinica*, 22(3), 139-145.
10. Ke, X., & Han, J. (1995). A Nonmonotone Trust Region Algorithm for Equality Constrained Optimization. In *Chinese Science Abstracts Series A* (Vol. 6, No. 14, pp. 3-4).
11. Mo, J., Liu, C., & Yan, S. (2007). A nonmonotone trust region method based on nonincreasing technique of weighted average of the successive function values. *Journal of Computational and Applied Mathematics*, 209(1), 97-108.

12. Yu, Z., He, C., & Tian, Y. (2010). Global and local convergence of a nonmonotone trust region algorithm for equality constrained optimization. *Applied Mathematical Modelling*, 34(5), 1194-1202.
13. Gu, N. Z., & Mo, J. T. (2008). Incorporating nonmonotone strategies into the trust region method for unconstrained optimization. *Computers & Mathematics with Applications*, 55(9), 2158-2172.
14. Zhang, J., & Zhu, D. (1994). A projective quasi-Newton method for nonlinear optimization. *Journal of computational and applied mathematics*, 53(3), 291-307.