A New Non-monotone Trust Region Technique for Equality Constrained Optimization

Xiao WU, Qinghua ZHOU*
College of Mathematics and Information Science, Hebei University, Baoding, 071002, CHINA
wuxiao616@163.com, qinghua.zhou@gmail.com
*Corresponding author

Abstract: In this paper, we propose a new non-monotone trust region algorithm for equality constrained optimization problems. We incorporate a new non-monotone strategy into trust region algorithm to construct a more relaxed trust region procedure. The global convergence is subsequently proved under some mild conditions.

Keywords: Non-monotone; Trust Region Methods; Constrained Optimization; Global Convergence

1. Introduction

In this paper, we consider the following equality constrained optimization problem

$$\begin{aligned} &\min_{x\in R^n} \ f(x) \\ &s.t. \ \ h_i(x)=0,\ i=1,2,\cdots,m \end{aligned} \tag{1}$$

where $f(x):R^n \to R$ and $h_i(x):R^n \to R\ (i=1,2,\cdots,m)(m\le n)$ are assumed to be continuously differentiable functions.

Many authors have given a lot of trust region algorithms to solve above problem (see [4,8,9]). These methods have the same feature: to enforce strict monotonicity for merit function at every iteration. Paper [1] shows that strict monotonic trust region methods are not always effective for some problems.

In 1986, the non-monotone line search technique was first proposed by Grippo et al. [7]. It shows that the non-monotone technique is helpful to avoid Maratos effect which is a common occurrence in difficult nonlinear problems. During the last few years, the non-monotone technique have been incorporate into trust region method to deal with unconstrained and constrained optimization problems [2,3,10,11,12,13]. These papers show that non-monotone technique can improve convergence rate in the case that a monotone technique is forced to creep along the bottom of a narrow curved valley; also they can improve the possibly of finding the global optimum.

In this paper we extend the non-monotone technique to trust region method for equality constrained optimization problems.

The rest of this paper is organized as follows: in Section 2, we describe a new non-monotone trust region algorithm. In Section 3, we prove that the proposed algorithm is globally convergent. Finally, some conclusions are expressed in Section 4.

2. Algorithm

Before describing the new algorithm, we introduce some notations: $g(x)=\nabla f(x), A(x)=\nabla h(x)=(\nabla h_1(x),\nabla h_2(x),\cdots,\nabla h_m(x))\in R^{n\times m}$. We define the matrix

$$P(x)=I-A(x)(A(x)^TA(x))^{-1}A(x)^T \tag{2}$$

where $A(x)$ has full column rank.

We know that a point x is called a stationary point of problem (1) if it satisfies the Kuhn-Tucker condition

$$\|h(x)\|+\|P(x)g(x)\|=0 \tag{3}$$

Now we discuss our new non-monotone trust region algorithm for solving problem (1). At k th iteration, if x_k does not satisfy the Kuhn-Tucker condition, we compute a trial step d_k by solving the following quadratic programming sub-problem

$$\begin{aligned} &\min_{d\in R^n} \ g_k^Td+\frac{1}{2}d^TB_kd \\ &s.t. \ \ h_k+A_k^Td=0 \\ &\qquad \|d\|\le\Delta_k \end{aligned} \tag{4}$$

where B_k is an $n\times n$ symmetric matrix which is the Hessian of the Lagrangian function at (x_k,λ_k) or an approximation to it, $\Delta_k>0$ is a trust region radius.

For testing whether the point x_k+d_k is accepted as the next iteration, we use the augmented Lagrangian merit function

$$\Phi(x,\lambda,\sigma)=f(x)+\lambda(x)^Th(x)+\sigma\|h(x)\|^2 \tag{5}$$

where $\lambda(x)$ satisfies

$$\min_{\lambda \in R^m} \| g(x) - A(x)\lambda \|^2 \quad (6)$$

and $\sigma > 0$ is the penalty parameter.
Now, we define

$$\bar{R}_k = \eta_k \Phi_{l(k)} + (1-\eta_k)\Phi_k \quad (7)$$

where

$\Phi_{l(k)} = \max_{0 \le j \le m(k)} \{\Phi(x_{k-j}, \lambda_{k-j}, \sigma_{k-j})\}$

and $0 \le m(k) \le \min\{m(k-1)+1, N\}$, $m(0)=0, N>0, 0 \le \eta_{\min} \le \eta_{\max} \le 1$ and $\eta_k \in [\eta_{\min}, \eta_{\max}]$.

The actual reduction is

$$Ared_k = \bar{R}_k - \Phi(x_k + d_k, \lambda(x_k+d_k), \sigma_k)$$

$$Pred_k = -(g_k + A_k\lambda_k)^T d_k - \frac{1}{2} d_k^T B_k d_k - (\lambda(x_k+d_k) - \lambda_k)^T (h_k + A_k^T d_k) + \sigma_k (\| h_k \|^2 - \| h_k + A_k^T d_k \|^2) \quad (8)$$

Therefore, the ratio is calculated

$$\rho_k = \frac{\bar{R}_k - \Phi(x_k + d_k, \lambda(x_k+d_k), \sigma_k)}{Pred_k}$$

Now, we can outline our new non-monotone trust region algorithm.

Algorithm 1

Step 1 Given $x_0 \in R^n, B_0 \in R^{n\times n}, \Delta_0 > 0, \varepsilon \ge 0$, $0 < \mu_1 \le \mu_2 < 1, 0 < \gamma_1 \le \gamma_2 < 1$, $0 \le \eta_{\min} \le 1$ $\eta_{\max} < 1, N > 0, \sigma_0 > 0, \eta > 0$. Set $k=0, m(0)=0$.

Step 2 If $\| h_k \| + \| P_k g_k \| \le \varepsilon$, stop.

Step 3 Solve the sub-problem (4) to determine d_k. If $d_k = 0$, then stop; otherwise, calculate $Pred_k$. If

$$Pred_k \ge \frac{\sigma_k}{2}(\| h_k \|^2 - \| h_k + A_k^T d_k \|^2) \quad (9)$$

does not hold, set

$$\sigma_k = \eta + 2\frac{(g_k + A_k\lambda_k)^T d_k + \frac{1}{2} d_k^T B_k d_k + (\lambda(x_k+d_k) - \lambda_k)^T (h_k + A_k^T d_k)}{\| h_k \|^2 - \| h_k + A_k^T d_k \|^2} \quad (10)$$

Step 4 Compute $Ared_k, Pred_k$ and ρ_k. If $\rho_k \ge \mu_1$, then set $x_{k+1} = x_k + d_k$.

Step 5 Set

$$\Delta_{k+1} \in \begin{cases} [\Delta_k, \infty), & if\ \rho_k \ge \mu_2; \\ [\gamma_2\Delta_k, \Delta_k), & if\ \mu_1 \le \rho_k \le \mu_2; \\ [\gamma_1\Delta_k, \gamma_2\Delta_k), & if\ \rho_k < \mu. \end{cases} \quad (11)$$

Step 6 Update the matrix B_k to generate B_{k+1}. Set $\sigma_{k+1} = \sigma_k, k = k+1$ and return to Step 2.

3. Convergence Analysis

To prove the global convergence of the new algorithm, the following assumptions are proved throughout this paper:

Assumptions

(H1) There exists a convex set $\Omega \in R^n$ such that $x_k, x_k + d_k \in \Omega$ for all k.

(H2) f and $h_i \in C^2(\Omega), i = 1,2,\cdots,m$.

(H3) The matrix $A(x) = \nabla h(x)$ has full column rank for all $x \in \Omega$.

(H4) $f(x), h(x), A(x), \nabla f(x), \nabla^2 f(x), (A(x)^T A(x))^{-1}$, and each $\nabla^2 h_i(x), i=1,2,\cdots,m$ are all uniformly bounded in norm in Ω.

(H5) The matrices $\{B_k, k=1,2,\cdots\}$ have a uniform upper bound, i.e. there exist $b_1 > 0$ such that $\| B_k \| \le b_1$ for all $k \in N$.

In what follows, we introduce some basic Lemmas which play important role in the analysis of our new algorithm.

Lemma 1. Under the assumptions, there exists a positive constant b_2 such that

$$\| h_k \|^2 - \| h_k + A_k^T d_k \|^2 \ge b_2 \| h_k \| \min\{\| h_k \|, \Delta_k\} \quad (12)$$

$$Pred_k \ge \frac{1}{2}\sigma_k b_2 \| h_k \| \min\{\| h_k \|, \Delta_k\}. \quad (13)$$

Proof. The proof can be found from Lemma 7.2 in [4].

Lemma 2. Let d_k be a step generated by Algorithm 1, and let $\hat{d}_k$ be its normal components, under the assumptions, there exists a positive constant b_3 such that

$$\| \hat{d}_k \| \le b_3 \| h_k \|. \quad (14)$$

Proof. We have

$$\begin{aligned} \| \hat{d}_k \| &= \| A_k (A_k^T A_k)^{-1} A_k^T \hat{d}_k \| \\ &= \| A_k (A_k^T A_k)^{-1} (h_k + A_k^T \hat{d}_k - h_k) \| \\ &\le \| A_k (A_k^T A_k)^{-1} \| [\| h_k + A_k^T \hat{d}_k \| + \| h_k \|] \end{aligned}$$

Now, since

$\| h_k + A_k^T \hat{d}_k \| \le \| h_k \|$

Hence

$\| \hat{d}_k \| \le b_3 \| h_k \|$

where $b_3 = 2\| A_k (A_k^T A_k)^{-1} \|$.

Lemma 3. Under the assumptions, there exists a positive constant b_4 such that

$$\| \Phi_k - \Phi_{k+1} - Pred_k \| \le b_4 \sigma_k \| d_k \|^2. \quad (15)$$

Proof. See Lemma 7.4 and 7.5 in [4].

Lemma 4. Let d_k be a step generated by Algorithm 1, and let $\hat{d}_k$ be its normal components, then there exists positive constants b_1 and b_5 such that

$$Pred_k \geq \frac{1}{4}\| P_k(g_k + B_k\hat{d}_k)\| \min\left(\bar{\Delta}_k, \frac{\| P_k(g_k+B_k\hat{d}_k)\|}{2b_1}\right)$$
$$-b_5\|d_k\|\|h_k\| - |(g_k+B_kd_k)^T\hat{h}_k|$$
$$+\sigma_k(\|h_k\|^2 - \|h_k + A_k^Td_k\|^2)$$

where $\bar{\Delta}_k = \sqrt{\Delta_k^2 - \|\hat{d}_k\|^2}$ and $\hat{h}_k = A_k(A_k^TA_k)^{-1}h_k$.

Proof. By the same way as in the proof of Lemma 6.7 in [5], we have the conclusion.

Lemma 5. Let $\hat{h}_k$ be as in Lemma 4. Then there exist constant b_6 and b_7 such that

$$|(g_k+B_kd_k)^T\hat{h}_k| \leq [b_6\|d_k\| + b_7\|d_{k-t_k}\|]\|h_k\|$$

where d_{k-t_k} is the last acceptable step.

Proof. The proof is similar to that of Lemma 6.8 in [5].

Lemma 6.(See Lemma 5 in [6]) Under the assumptions, if $\|P_kg_k\| + \|h_k\| \neq 0$, then there exists a integer k_0 and a positive constant $\bar{\sigma}$ such that for all $k \geq k_0, \sigma_k = \bar{\sigma}$.

Lemma 7. Under the assumptions and there exists an infinite set N, we have

$$\lim_{\substack{k\in N\\ k\to\infty}} \Phi_{l(k)} = \lim_{\substack{k\in N\\ k\to\infty}} \Phi_{k+1} = \lim_{\substack{k\in N\\ k\to\infty}} \bar{R}_k .$$

Proof. Using definition of $\bar{R}_k$ and $\Phi_{l(k)}$, we observe that

$$\bar{R}_k = \eta_k\Phi_{l(k)} + (1-\eta_k)\Phi_k \leq \eta_k\Phi_{l(k)} + (1-\eta_k)\Phi_{l(k)} = \Phi_{l(k)}$$

And from the definition of $\Phi_{l(k)}$, we have $\Phi_k \leq \Phi_{l(k)}$, for any $k \in N$. Hence,

$$\Phi_k = \eta_k\Phi_k + (1-\eta_k)\Phi_k$$
$$\leq \eta_k\Phi_{l(k)} + (1-\eta_k)\Phi_k = \bar{R}_k$$

Then we have

$$\Phi_k \leq \bar{R}_k \leq \Phi_{l(k)}$$

This fact, along with Lemma 4.7 in [3], leads us to have the conclusion.

Theorem 1. Under the assumptions. If Algorithm 1 fails to satisfy the termination condition, then

$$\lim_{k\to\infty}\|h_k\| = 0 \qquad (16)$$

Proof. We consider two cases.

Case 1. $\liminf_{k\to\infty}\Delta_k = \Delta > 0$. Suppose that $\limsup_{k\to\infty}\|h_k\| \geq \varepsilon > 0$. Then there exists an infinite sequence of indices $\{k_j\}$ such that $\|h_k\| \geq \frac{\varepsilon}{2}$ for all $k \in \{k_j\}$.

For any such k, from Lemma 1 we have

$$Pred_k \geq \frac{1}{2}\sigma_kb_2\|h_k\|\min\{\|h_k\|, \Delta_k\}$$
$$\geq \frac{1}{4}\sigma_kb_2\varepsilon\min\{\frac{\varepsilon}{2}, \Delta_k\}$$

Then

$$\bar{R}_k - \Phi_{k+1} \geq \mu_1 Pred_k \geq \frac{\mu_1}{4}\sigma_kb_2\varepsilon\min\{\frac{\varepsilon}{2},\Delta_k\}$$

Using Lemma 7, we get

$\liminf_{k\to\infty}\Delta_k = 0$.

This is a contradiction, the contradiction shows that (16) hold.

Case 2. $\liminf_{k\to\infty}\Delta_k = 0$, assume that $\lim_{k_j\to\infty}\Delta_{k_j} = 0$, which means $\rho_{k_j} < \eta_1$ for all k_j. Assume that (16) does not hold, similar to case 1, we have

$$Pred_k \geq \frac{1}{4}\sigma_kb_2\varepsilon\min\{\frac{\varepsilon}{2},\Delta_{k_j}\} = \frac{\varepsilon}{4}\sigma_kb_2\Delta_{k_j} .$$

From Lemma 3, we can obtain

$$\left|\frac{\Phi_{k_j} - \Phi_{k_j+1}}{Pred_{k_j}} - 1\right| = \left|\frac{\Phi_{k_j} - \Phi_{k_j+1} - Pred_{k_j}}{Pred_{k_j}}\right| \leq \frac{b_4\sigma_k\Delta_{k_j}^2}{\frac{\varepsilon}{4}\sigma_kb_2\Delta_{k_j}} \to 0$$

This implies that

$$\frac{\Phi_{k_j} - \Phi_{k_j+1}}{Pred_{k_j}} \geq \eta_1 .$$

We know that

$\bar{R}_k \geq \Phi_k$.

Then

$$\rho_{k_j} = \frac{\bar{R}_{k_j} - \Phi_{k_j+1}}{Pred_{k_j}} \geq \frac{\Phi_{k_j} - \Phi_{k_j+1}}{Pred_{k_j}} \geq \eta_1 .$$

This is a contradiction, the contradiction shows that (16) hold.

Theorem 2. Under the assumptions. If Algorithm 1 fails to satisfy the termination condition, then

$$\liminf_{k\to\infty}\|P_kg_k\| = 0 . \qquad (17)$$

Proof. We consider two cases.

Case 1. $\liminf_{k\to\infty}\Delta_k = \Delta > 0$. Suppose that there exist an $\varepsilon > 0$ and an integer k_0 such that $\|P_kg_k\| \geq \varepsilon$ for all $k \geq k_0$. By (H5) and Lemma 2, we can get

$$\|P_k(g_k + B_k\hat{d}_k)\| \geq \|P_kg_k\| - \|P_kB_k\hat{d}_k\|$$
$$\geq \|P_kg_k\| - b_1b_3\|h_k\|$$
$$= \|P_kg_k\| - b_5\|h_k\|$$

where $b_5 = b_1b_3$.

From Theorem 1 there exist k_1 sufficiently large such that for all $k \geq k_1$, we have

$\| h_k \| < \frac{1}{2b_5}\varepsilon$.

Thus for $k \geq \max[k_0, k_1]$

$$\| P_k(g_k + B_k\hat{d}_k) \| \geq \frac{1}{2}\varepsilon . \quad (18)$$

From Lemma 4 and Lemma 5

$$Pred_k \geq \frac{1}{4}\| P_k(g_k + B_k\hat{d}_k) \| \min\left(\bar{\Delta}_k, \frac{\| P_k(g_k + B_k\hat{d}_k) \|}{2b_1}\right)$$
$$- b_5 \| d_k \| \| h_k \| - [b_6 \| d_k \| + b_7 \| d_{k-t_k} \|] \| h_k \|$$

Using (16) and (18), we have

$$Pred_k \geq \frac{1}{8}\| P_k(g_k + B_k\hat{d}_k) \| \min\left(\frac{1}{2}\Delta_k, \frac{\| P_k(g_k + B_k\hat{d}_k) \|}{2b_1}\right)$$
$$\geq \frac{\varepsilon}{32}\min\left(\Delta_k, \frac{\varepsilon}{2b_1}\right)$$

Then

$$\bar{R}_k - \Phi_{k+1} \geq \mu_1 Pred_k \geq \frac{\mu_1}{32}\varepsilon \min\left(\Delta_k, \frac{\varepsilon}{2b_1}\right)$$

Using Lemma 7, we have

$\liminf_{k\to\infty} \Delta_k = 0$

This is a contradiction, the contradiction shows that (17) hold.

Case 2. $\liminf_{k\to\infty} \Delta_k = 0$, assume that $\lim_{k_j\to\infty} \Delta_{k_j} = 0$, which means $\rho_{k_j} < \eta_1$ for all k_j. Assume that (17) does not hold, similar to case 1, we have

$$Pred_k \geq \frac{\varepsilon}{32}\min\left(\Delta_{k_j}, \frac{\varepsilon}{2b_1}\right) = \frac{\varepsilon}{32}\Delta_{k_j}$$

From Lemma 3, we can obtain

$$\left| \frac{\Phi_{k_j} - \Phi_{k_j+1}}{Pred_{k_j}} - 1 \right| = \left| \frac{\Phi_{k_j} - \Phi_{k_j+1} - Pred_{k_j}}{Pred_{k_j}} \right| \leq \frac{b_4 \sigma_k \Delta_{k_j}^2}{\frac{\varepsilon}{32}\Delta_{k_j}} \to 0 .$$

Similar Case 2 in Theorem 1, we can obtain

$\rho_{k_j} \geq \eta_1$.

This is a contradiction, then (17) hold.

Theorem 3. Under the assumptions, Algorithm 1 produces iterates $\{x_k\}$, which satisfy

$\liminf_{k\to\infty} (\| h_k \| + \| P_k g_k \|) = 0$.

Proof. By Theorem 1 and Theorem 2 we can get the conclusion.

4. Conclusions

In this paper, we propose a new non-monotone trust region algorithm for solving equality constrained optimization problems. After we analyzed the properties of the new algorithm, the global convergence theory is proved. We believe that there is considerable scope for modifying and adapting the basic ideas introduced in this paper. In the near future, we would like to combine the new algorithm with line search algorithm in order to sufficiently use the information which the algorithm has already derived.

Acknowledgements

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

References

[1] N.Y. Deng, Y. Xiao, F.J. Zhou. Nonmonotonic trust region algorithm, J. Optim. Theory Appl. 76(1993)259-284.

[2] Xiaowu Ke, Jiye Han. A nonmonotone trust region algorithm for equality constrained optimization, Science in China. 38(1995)683-695.

[3] X.J. Tong, S.Z. Zhou, Global convergence of nonmonotone trust region algorithm for nonlinear optimization, Appl. Math,: J. Chinese Univ. 15(2000)201-210.

[4] J.E. Dennis, M.M. El-Alem, M.C. Maciel, A global convergence theory for general trust region-based algorithms for equality constrained optimization, SIAM J. Optim. 7(1997)177-207.

[5] El-Alem, M. M., A global convergence theory for the Celis-Dennis-Tapia trust region algorithm for constrained optimization, SIAM J. Numer. Anal. 28(1991)266-290.

[6] Z.S. Yu, C.X. He, Y. Tian. Global and local convergence of a nonmonotone trust region algorithm for equality constrained optimization, Applied Mathematical Modelling. 34(2010)1194-1202.

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

[8] El-Alem, M., A robust trust-region algorithm with a nonmonotonic penalty parameter scheme for constrained optimization, SIAM J. Optim. 5(1995)348-378.

[9] J.E. Dennis, L.N. Vicente. On the convergence theory of trust-region-based algorithms for equality constrained optimization, SIAM J. Optim. 7(1997)927-950.

[10] Z.W. Chen, X.S. Zhang, A nonmonotone trust region algorithm with nonmonotone parameter for constrained optimization, J. Comput. Appl. Math. 172(2000)7-39.

[11] Ph.L. Toint, A nonmonotone trust region algorithm for nonlinear programming subject to convex constraints, Math. Prog. 77(1997)69-94.

[12] D.T. Zhu, A nonmonotonic trust region technique for nonlinear constrained optimization, J. Comput. Math. 13(1995)20-31.

[13] M. Ahookhosh, K. Amini., (2012). An efficient nonmonotone trust-region method for unconstrained optimization. Numer Algor. 59, 523-540.

Subscriptions and Individual Articles:

User	**Hard copy:**
Institutional:	800 (HKD/year)
Individual:	500 (HKD/year)

Individual Article: 20 (HKD)