Non-monotone Trust Region Technique for Equality Constrained Optimization

Xiao WU, Qinghua ZHOU*
College of Mathematics and Information Science, Hebei University, Baoding, CHINA

Abstract: In this article, we propose and analyze a new trust region algorithm for solving equality constrained optimization problems. We incorporate a non-monotone strategy into trust region algorithm to construct a more relaxed trust region procedure and employ a differentiable exact penalty function. Under some reasonable conditions, the global convergence is established.

Keywords: Nonmonotone; Trust region methods; Constrained optimization; Exact penalty function; Global convergence

1. Introduction

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

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

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

Many authors have studied problem (1) (see[2,3,6,10-12]). These methods are monotonic algorithm. In 1982, Chamberlain in [1] proposed the watchdog technique for constrained optimization to overcome the Maratos effect. Inspired by this idea, Grippo, Lamparillo and Lucidi introduced a nonmonotone line search technique for Newton's method in [4].Their conclusions were overall approachable for the nonmonotone method, especially when applied to highly nonlinear problemsand in presence of narrow curved valley.

The nonmonotone methods are distinguished by the fact that they do not enforce strict monotonicity to the objective function values at successive iterations. Some researchers showed that utilizing non-monotone technique may improve both the possibility of finding the global optimum and the rate of convergence (see [13]). Due to the high efficiency of nonmonotone techniques, many authors are interested in working on employing nonmonotone strategies in various branches of optimization procedures (see [7,14]).

Although the nonmonotone technique has many advantages, it suffers from some drawbacks. Ahookhosh et al. introduced a modifiednonmonotone strategy and employs it in a trust region framework in [8]. Their analysis of the new algorithm showed that it inherited both stability of trust region methods and effectiveness of the nonmonotone strategy.

In this paper we extend the nonmonotone technique [8] 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 nonmonotone 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: $g(x)=\nabla f(x), A(x)=\nabla h(x)=(\nabla h_1(x), \nabla h_2(x),\ldots,\nabla h_m(x)) \in R^{n\times m}$. We define the matrix

$$\begin{aligned} P(x) &= I - A(x)(A(x)^T A(x))^{-1} A(x)^T \\ &= I - A(x)A(x)^+ \end{aligned} \tag{2}$$

where $A(x)$ has full column rank and $A(x)^+ = (A(x)^T A(x))^{-1} A(x)^T$.

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 nonmonotone 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 subproblem [2]

$$\begin{aligned} &\min_{d\in R^n} \ g_k^T d + \frac{1}{2} d^T B_k d \\ &s.t. \quad h_k + A_k^T d \le \theta_k \\ &\qquad \|d\| \le \Delta_k \end{aligned} \tag{4}$$

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, q_k is any number which satisfies

$$\min_{\|d\|\le\beta_1\Delta_k}\|h_k + A_k^T d\| \le q_k \le \min_{\|d\|\le\beta_2\Delta_k}\|h_k + A_k^T d\| \qquad (5)$$

and where β_1 and β_2 are two given constants that satisfy $0 \le \beta_2 \le \beta_1 < 1$.

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

$$\Phi(x,\lambda,\sigma) = f(x) + \lambda(x)^T h(x) + \sigma\|h(x)\|^2 \qquad (6)$$

where $\lambda(x)$ satisfies

$$\min_{\lambda\in R^m}\|g(x) - A(x)\lambda\|^2 \qquad (7)$$

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

Now, we define

$$\bar{R}_k = \eta_k\Phi_{l(k)} + (1-\eta_k)\Phi_k \qquad (8)$$

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

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

$$\begin{aligned} Pred_k = & -(g_k + A_k\lambda_k)^T d_k - \frac{1}{2}d_k^T B_k d_k \\ & - (\lambda(x_k + d_k) - \lambda_k)^T(h_k + A_k^T d_k) \qquad (9) \\ & + \sigma_k(\|h_k\|^2 - \|h_k + A_k^T d_k\|^2) \end{aligned}$$

Therefore, the ratio is calculated

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

Now, we can outline our new nonmonotone trust region algorithm.

Algorithm 1

Step 1 Given $x_0 \in R^n, B_0 \in R^{n\times n}, \Delta_0 > 0, \varepsilon \ge 0, 0 < \mu_1 \le \mu_2 < 1, 0 < \gamma_1 \le \gamma_2 < 1, 0 \le \eta_{\min} \le \eta_{\max} < 1, N > 0, \sigma_0 > 0, \eta > 0, 0 < \beta_2 \le \beta_1 < 1$. Set $k = 0$, $m(0) = 0$.

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

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

$$Pred_k \ge \frac{\sigma_k}{2}(\|h_k\|^2 - \|h_k + A_k^T d_k\|^2) \qquad (10)$$

does not hold, set

$$\begin{aligned} \sigma_k = & 2[(g_k + A_k\lambda_k)^T d_k + \frac{1}{2}d_k^T B_k d_k + \\ & (\lambda(x_k + d_k) - \lambda_k)^T(h_k + A_k^T d_k)]/ \qquad (11) \\ & (\|h_k\|^2 - \|h_k + A_k^T d_k\|^2) + \eta \end{aligned}$$

which ensures that the new value of expression (11) satisfies condition (10).

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

Step 5 Set

$$\Delta_{k+1} \in \begin{cases} [\Delta_k, \bar{\Delta}], & \text{if } r_k \ge \mu_2; \\ [\gamma_2\Delta_k, \Delta_k], & \text{if } \mu_1 \le r_k < \mu_2; \\ [\gamma_1\Delta_k, \gamma_2\Delta_k], & \text{if } r_k < \mu_1. \end{cases} \qquad (12)$$

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

In this paper, we consider the following assumptions that will be used to analyze the convergence properties of the new algorithm.

Assumptions

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

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

(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 $\nabla^2 h_i(x), i = 1,2,\ldots,m$ are all uniformly bounded in norm in Ω.

(H5) The matrices $\{B_k, k = 1,2,\ldots\}$ 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

$$\|h_k\|^2 - \|h_k + A_k^T d_k\|^2 \ge \min\left\{\|h_k\|, \frac{b_2\Delta_k}{\|A_k^+\|}\right\} \qquad (13)$$

$$Pred_k \ge \frac{1}{2}\sigma_k\|h_k\|\min\left\{\|h_k\|, \frac{b_2\Delta_k}{\|A_k^+\|}\right\}. \qquad (14)$$

Proof. The inequality (13) can be found from Lemma 3.3 in [2]. The second result of the lemma is similar to Lemma 1 in [5].

Lemma 2. If Algorithm 1 does not terminate, then

$$(g_k + A_k\lambda_k)^T d_k + \frac{1}{2}d_k^T B_k d_k$$

$$\le -\frac{1}{4}\|P_k\bar{g}_k\|^2\min\left\{\frac{1}{\|B_k\|}, \frac{\Delta_k}{\|P_k\bar{g}_k\|}\right\} + \|B_k\|\|d_k\|\|\bar{d}_k\|$$

holds for all k, where

$$P_k = P(x_k) = I - A(x_k)(A(x_k)^T A(x_k))^{-1} A(x_k)^T \quad (15)$$

$$\bar{d}_k = (I - P_k) d_k \quad (16)$$

$$\bar{g}_k = g_k + B_k \bar{d}_k \quad (17)$$

$$\bar{\Delta}_k = (\Delta_k^2 - \|\bar{d}_k\|^2)^{\frac{1}{2}} \quad (18)$$

Proof. The proof is similar to Lemma 3 in [5], we omit it for convenience.

Lemma 3. Under Assumption, if Algorithm 1 does not terminate, there exists a positive constant m_1 such that the inequality

$$Pred_k - \frac{\sigma_k}{2}(\|h_k\|^2 - \|h_k + A_k^T d_k\|^2) \geq \frac{1}{4}\|P_k \bar{g}_k\|^2 \min\left\{\frac{1}{\|B_k\|}, \frac{\bar{\Delta}_k}{\|P_k \bar{g}_k\|}\right\} - m_1 \|d_k\| \|h_k\| + \frac{\sigma_k}{2}\|h_k\| \min\left\{\|h_k\|, \frac{b_2 \Delta_k}{\|A_k^+\|}\right\}$$

hold for all k.

Proof. Using Lemma 1, 2 and (9), we obtain

$$\begin{aligned} Pred_k - \frac{\sigma_k}{2}(\|h_k\|^2 - \|h_k + A_k^T d_k\|^2) &\geq \frac{1}{4}\|P_k \bar{g}_k\|^2 \min\left\{\frac{1}{\|B_k\|}, \frac{\bar{\Delta}_k}{\|P_k \bar{g}_k\|}\right\} - \|B_k\|\|d_k\|\|\bar{d}_k\| - (\lambda(x_k + d_k) - \lambda_k)^T \\ &\quad (h_k + A_k^T d_k) + \frac{\sigma_k}{2}\|h_k\| \min\left\{\|h_k\|, \frac{b_2 \Delta_k}{\|A_k^+\|}\right\} \end{aligned} \quad (19)$$

According to Lemma 1, (15) and (16)

$$\begin{aligned} \|\bar{d}_k\| &\leq \|A_k A_k^+ d_k\| \\ &= \|(A_k^+)^T (h_k + A_k^T d_k - h_k)\| \\ &\leq 2\|A_k^+\|\|h_k\| \end{aligned} \quad (20)$$

By Assumptions and (7), there exists a positive constant m_2 such that the inequality

$$\|\lambda(x_k + d_k) - \lambda_k\| \leq m_2 \|d_k\| \quad (21)$$

holds for all k. Hence, from Lemma1, (19), (20) and (21), we have

$$\begin{aligned} Pred_k - \frac{\sigma_k}{2}(\|h_k\|^2 - \|h_k + A_k^T d_k\|^2) &\geq \frac{1}{4}\|P_k \bar{g}_k\|^2 \min\left\{\frac{1}{\|B_k\|}, \frac{\bar{\Delta}_k}{\|P_k \bar{g}_k\|}\right\} \\ &\quad - 2\|A_k^+\|\|B_k\|\|d_k\|\|h_k\| - m_2\|d_k\|\|h_k\| \\ &\quad + \frac{\sigma_k}{2}\|h_k\| \min\left\{\|h_k\|, \frac{b_2 \Delta_k}{\|A_k^+\|}\right\} \end{aligned}$$

Therefore, under assumptions, there exists $m_1 > 0$ such that the result of Lemma 3 is true.

Lemma 4. Under Assumptions, there exist positive constants m_3 and m_4, such that, on the iterations that satisfy the condition

$$\|h_k\| \geq m_3 \Delta_k \quad (22)$$

we have

$$Pred_k - \frac{\sigma_k}{2}(\|h_k\|^2 - \|h_k + A_k^T d_k\|^2) \geq m_4 \Delta_k \quad (23)$$

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

Lemma 5. Under the assumptions, there exists a positive constant m_5 such that

$$|\Phi_k - \Phi_{k+1} - Pred_k| \leq m_5 \|d_k\|^2. \quad (24)$$

Proof. Using (6), (9) and Taylor expansion, we have

$$\begin{aligned} &|\Phi_k - \Phi_{k+1} - Pred_k| \\ &= |f_k + g_k^T d_k + \frac{1}{2} d_k^T B_k d_k - f_{k+1} + \lambda_{k+1}^T (h_k + A_k^T d_k) \\ &\quad - \lambda_{k+1}^T h_{k+1} + \sigma\|h_k + A_k^T d_k\|^2 - \sigma\|h_{k+1}\|^2| \\ &\leq \frac{1}{2}|d_k^T (B_k - \nabla^2 f(x_k + \xi_1 d_k)) d_k| \\ &\quad + \frac{1}{2}|d_k^T (\nabla^2 h(x_k + \xi_2 d_k) \lambda_{k+1}) d_k| \\ &\quad + \sigma\|d_k\|^2 (\|\nabla^2 h(x_k + \xi_2 d_k)\|\|h_k\| \\ &\quad + \|\nabla^2 h(x_k + \xi_2 d_k)\|\|A_k\|\|d_k\| \\ &\quad + \frac{1}{4}\|\nabla^2 h(x_k + \xi_2 d_k)\|^2 \|d_k\|^2) \end{aligned}$$

where $\xi_1, \xi_2 \in (0,1)$.

By Assumptions, there exists a positive constant m_5 such that

$$|\Phi_k - \Phi_{k+1} - Pred_k| \leq m_5 \|d_k\|^2.$$

Lemma 6. (See Lemma 5 in [9]) Under the assumptions, if $\|P_k g_k\| + \|h_k\| \neq 0$, then there exists a integer k_0 and a positive constant $\bar{\sigma}$ such that for all $k \geq k_0, \sigma_k \leq \bar{\sigma}$.

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

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

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

$$\begin{aligned} \bar{R}_k &= \eta_k \Phi_{l(k)} + (1 - \eta_k)\Phi_k \\ &\leq \eta_k \Phi_{l(k)} + (1 - \eta_k)\Phi_{l(k)} = \Phi_{l(k)}. \end{aligned}$$

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

$$\begin{aligned} \Phi_k &= \eta_k \Phi_k + (1 - \eta_k)\Phi_k \\ &\leq \eta_k \Phi_{l(k)} + (1 - \eta_k)\Phi_k = \bar{R}_k. \end{aligned}$$

Then we have

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

This fact, along with Lemma 4.7 in [7], leads us to have the conclusion.

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

$$\lim_{k\to\infty} \|h_k\| = 0 \quad (25)$$

Proof. We prove it by contradiction. Assume that (25) is not true, that is, there exists a constant $\bar{h} > 0$ such that

$$\|h_k\| \ge \bar{h}, \forall k. \quad (26)$$

It follows from Lemma 1 and Step 4 that

$$\bar{R}_k - \Phi_{k+1} \ge \frac{\eta_1 \sigma_k}{2}\|h_k\| \min\left\{\|h_k\|, \frac{b_2 \Delta_k}{\|A_k^+\|}\right\} \quad (27)$$

If $\|h_k\| \ge m_3 \Delta_k$, by Lemma 4 and (27) , we have

$$\bar{R}_k - \Phi_{k+1} \ge \eta_1 m_4 \Delta_k. \quad (28)$$

Otherwise, (27) implies that

$$\bar{R}_k - \Phi_{k+1} \ge \frac{\eta_1 \sigma_k}{2} m_3^2 \Delta_k^2 \quad (29)$$

Hence (28) and (29) imply

$$\bar{R}_k - \Phi_{k+1} \ge \eta_1 \min\{m_4 \Delta_k, \frac{\eta_1 \sigma_k}{2} m_3^2 \Delta_k^2\}, \forall k \quad (30)$$

Using Lemma 7 and (30), we obtain

$$\lim_{k\to\infty} \Delta_k = 0, \quad \lim_{k\to\infty} \|d_k\| = 0 \quad (31)$$

without loss of generality, it is assumed by

$$\Delta_k \le \frac{\bar{A}}{b_2}\bar{h}, \quad \forall k \quad (32)$$

By (26), (31) ,(32), Lemma 3 and Lemma 5, we can deduce $\Delta_{k+1} \ge \Delta_k$ for all k sufficiently large, which is in contradiction with (31). Therefore, (25) holds for all k.

Theorem 2.(See Theorem 3.9 in [2])Under theassumptions, Algorithm 1 produces iterates $\{x_k\}$, which satisfy $\lim_{k\to\infty} \inf(\|h_k\| + \|P_k g_k\|) = 0$.

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.

5. Acknowledgement

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

References

[1] R.M. Chamberlain and M.J.D. Powell,"The watchdog technique for forcing convergence inalgorithm for constrained optimization," Mathematical Programming Study, 1982,pp. 1-17.

[2] M.J.D. Powell and Y. Yuan, "A trust region algorithm for equality constrained optimization," Math. Prog., vol. 1, 1991, pp. 189-211.

[3] M. El-Alem, "A global convergence theory for the Celis-Dennis-Tapia trust region algorithm for constrained optimization," SIAM J. Numer. Anal,vol. 1, 1991, pp. 266-290.

[4] L. Grippo, F. Lampariello, and S. Lucidi. "A nonmonotone line search technique for Newton's method," SIAM J.Numer.Anal.,vol. 4, 1986, pp. 707-716.

[5] X. W.Ke and J. Y. Han, "A nonmonotone trust region algorithm for equality constrained optimization," Science in China,vol. 6, 1995, pp. 683-695.

[6] J. E. Dennis, M. El-Alem, and M. Maciel, "A global convergence theory for general trust-region-based algorithms for equality constrained optimization," SIAM Journal on Optimization, vol. 1, 1997, pp. 177-207.

[7] X.J. Tong and S.Z. Zhou, "Global convergence of nonmonotonic trust region algorithm for nonlinear optimization,"vol. 2, 2000, pp. 201-210.

[8] M. Ahookhosh and K. Amini, "An efficient nonmonotone trust-region method for unconstrained optimization," NumerAlgor, vol. 4, 2012, pp. 523-540.

[9] Z.S. Yu, C.X. He,and Y. Tian. "Global and local convergence of a nonmonotone trust region algorithm for equality constrained optimization,"Applied Mathematical Modelling, vol. 5, 2010, pp. 1194-1202.

[10] R. H. Byrd,R. Schnabel, and G. A. Shultz, "A trust region algorithm for nonlinearly constrained optimization," SIAM J. Numer. Anal,vol. 5, 1987, pp. 1152-1169.

[11] J. Z. Zhang and D. T. Zhu,"Projected quasi-Newton algorithm with trust region for constrained optimization," J. Optim. Theory Anal, 1990.

[12] A. Vardi,"A trust region algorithm for equality constrained minimization: convergence properties and implementation," SIAM J. Numer. Anal., vol. 3, 1985, pp. 575-591.

[13] Ph.L.Toint, "A nonmonotone trust region algorithm for nonlinear programming subject to convex constraints," Math. Prog., vol. 3, 1997, pp. 69-94.

[14] N.Y. Deng, Y. Xiao, and F.J. Zhou. "Nonmonotonic trust region algorithm," J. Optim. Theory Appl., vol. 2, 1993, pp. 259-284.