Using a trust region method with nonmonotone technique to solve unrestricted optimization problem

Hasan H. Dwail and Mushtak A.K. Shiker

1Department of Mathematics, College of Education for Pure Sciences, University of Babylon, Hilla-Iraq.
E-mail: 1mathhasan74@gmail.com, 2mmttmhh@yahoo.com

2Corresponding Author

Abstract. In this work, we combine adaptive trust region method TRM with nonmonotone strategy to introduce a new algorithm to solve systems of unconstrained optimization problems depending on the simple quadratic model. The subproblems will be solving easier with the new approximation of the Hessian function. This technique does not need large capacity and is characterized by ease of application. We established the algorithm convergence with standard assumptions. The numerical results compared the performance of the new approach with three famous algorithms in terms of the time consumed to find the solution in addition to the number of iterations and the number of functions evaluations. From these results we can show that the new method is the best among the traditional algorithms.

Keywords: Global convergence, Unconstrained optimization and Nonmonotone trust region.

1. Introduction

Assume that

\[ \min_{x \in \mathbb{R}^n} f(x), \]

is a large scale unconstrained optimization problem, where \( f(x) : \mathbb{R}^n \rightarrow \mathbb{R} \) is continuously differentiable function.

The TR model is a type of iterative procedure that produces a direction step \( d_k \), in all iterations, it is done by solving the following quadratic subproblem:

\[ \min q_k(d) = f(x_k) + g_k^T d + \frac{1}{2} d^T g_k d, \quad \text{such that } \|d\| \leq \Delta_k. \]

(2)

We will use the following abbreviations during the search: \( g_k = \nabla f(x_k), B_k \in \mathbb{R}^{n \times n} \) is a symmetric matrix, \( \|\cdot\| \) denotes Euclidean norm, and \( \Delta_k > 0 \) is the trust region radius (TRR). We will change the Armijo rule to accept the step length \( \alpha_k \) by

\[ f(x_k + \alpha_k d_k) \leq \mu \alpha_k g_k^T d_k \|g_k\|, \]

where \( \mu \in (0,1) \) \[1\]. The \( \rho_k \) ratio represents the actual reduction to the predicate reduction

\[ \rho_k = \frac{f(x_k) - f(x_{k+1})}{q_k(0) - q_k(d_k)}, \]

(4)

\( \rho_k \) is used to find out if the step length is rejected or not, then adjust the radius of the trust region. The trust region extended if the ratio closed enough to 1. If \( \rho_k \) is equal to a positive constant \( \mu \), but not closed to 1 then the trial step \( d_k \) will be accepted, leading to \( x_{k+1} = x_k + d_k \) and the TRR kept the...
same without altering. Otherwise, the TRR must be shrink and the subproblem (2) will be resolved to
determine an acceptable experimental indicate the continuation of the procedure [2, 7, 8]. To develop
the work of TRM, many known techniques has been modified, the most important is the LSM
technique, which reduced the issue of resolving subproblems in the event of a trial step rejected. Its
work is to find and produce an acceptable value for step length in each iteration [3].

In the 1980s, Gribo et al. established a manner to find nonmonotone lines for Newton method [9].
Later, based on the nonmonotone procedure proposed by Gribo et al. Deng et al. presented a
nonmonotone trust region method for unconstrained optimization [4].

Numerical results indicated that nonmonotone TRM was better than TRM, so many authors took
advantage of nonmonotone techniques both in LSM and TRM [5, 11- 14] to solve large scale
unconstrained optimization problem. For TRM, the choice of the radius is very important for the
strength and efficiency algorithm. Likewise, in conventional TRM, the radius is independent of goal
function information and it is selected agreeing to $\rho_k$. So, the main difficult of TRM is how to select
the radius at the $k^{th}$ iteration. If $\Delta_k$ is greater than $\mu$, then the number of resolved problems will be
increases as well as the computational cost. If $\Delta_k$ is less than $\mu$, the number of steps will increase, and
therefore the efficiency decreases, this would reduce the efficiency of the algorithms. Some
researcher’s collective nonmonotone TRM depended on simple quadratic models. The authors
introduced many techniques to solve various optimization and reliability problems (see 15- 21), but in
this work we aim to introduce a new nonmonotone adaptive TRM depended on simple quadratic
models with the adaptive procedure [6]. We custom a new scale approximation of the minimizing
function’s Hessian in the trust region subproblem, and then collective this new TRM with the
nonmonotone procedure of the new adaptive method.

The respite of this work is prepared as follows. In the second section, we offered nonmonotone
adaptive TR algorithm depending on simple quadratic models. In the third section, we present the
global convergence of the suggested technique. The fourth section contains the numerical results.
Finally, in the fifth section, we offer some conclusion remarks.

2- New algorithm depending on the simple quadratic model.

Here, we will discuss how to build a simple quadratic subproblem in each iteration. If we have the
initial point $x_0$ where $f(x_0)$ and $g_0$ can be immediately calculate and set $B_0= I$, where I represent a
unit matrix. The subproblem (2) can be formulated as:

$$
\text{min}\, q_0(d) = f_0 + g_0^T d + \frac{1}{2} d^T d, \|d\| \leq \Delta_0.
$$

(5)

Now we will give a procedure to calculate the approximate Hessian scalar of the function $f$ at $x$, from
$x_k = x_{k-1} + d_{k-1}$, we get $x_{k-1} = x_k - d_{k-1}$ [7, 8].

$$
f(x_{k-1}) = f(x_k) - g_k^T d_{k-1} + \frac{1}{2} d_{k-1}^T \nabla^2 f(x) d_{k-1},
$$

(6)

where $z$ lies on the line connecting $x_{k-1}$ and $x_k$. Possess an opinion the local character of the
searching process and the distance between $x_{k-1}$ and $x_k$ is suitable small, we select $z = x_k$ and
consider $\gamma(x_k)$ as an approximation of $\nabla^2 f(x_k)$, where $\gamma(x_k) \in R$. Therefore, the approximation of
the Hessian at point $x_k$ is calculated using the local information from points $x_k$ and $x_{k-1}$. So

$$
\gamma(x_k) = \frac{2}{d_k^T d_{k-1}} \left[ f(x_{k-1}) - f(x_k) + g_k^T d_{k-1} \right],
$$

(7)

Now, we select $\gamma(x_k) < 0$ to make Hessian matrix is definite, we can use $\delta > 0$ and obtain,

$$
\theta_k = \frac{1}{d_k^T d_{k-1}} \left[ f(x_k) - f(x_{k-1}) - g_k^T d_{k-1} + \delta \right].
$$
and modified $\gamma(x_k)$ by

$$\theta_k = \frac{2}{d_{k-1}^Td_{k-1}}[f(x_{k-1}) - f(x_k) + (1 - \theta_k)g_k^Td_{k-1}],$$

(8)

The next value for $\gamma(x_k)$ is positive. At the iteration point $x_k$, Therefore, subproblem (2) be able to be formed by way of follows,

$$min q_k(d) = f_k + g_k^Td + \frac{1}{2}\gamma(x_k)d^Td, \|d\| \leq \Delta_k,$$

(9)

the subproblem (2) can be easily solved as well. After finding a trial step $d_k$, To find out if he will be rejected or not, and in what way to set the new TRR, we calculate the ratio $\rho_k$ between

$$Ared_k = f(x_k) - f(x_{k+1}),$$

and

$$Pred_k = q_k(0) - q_k(d_k),$$

that is mean,

$$\rho_k = \frac{Ared_k}{Pred_k}$$

(10)

To optimize the efficiency, we present one more adaptive technique to describe a TRR $\Delta_k$, that is improved by the idea in [1]. We take $\Delta_k = \frac{\|g_k\|}{\gamma(x_k)}\|f_k\|$, and then alter it appropriately agreeing to the value of $\rho_k$.

2.1 Algorithm: New Adaptive TRM:

Step 0. Given $x_0 \in R^n$, $\Delta_0 > 0$, $0 < \mu < v_1 < v_2 < 1$, $0 < c < 1$, $c_1 > 1$, $0 < c_2 < 1$, $0 < \varepsilon < 1$, $\varepsilon > 0$, $\theta > 0$, $M > 0$, set $k = 0$, $\gamma(x_0) = 1$.

Step 1. Calculate $g_k$. If $\|g_k\| \leq \varepsilon$, stop. Or, go to Step 2.

Step 2. Find the Solve of the subproblem (9) for $d_k$.

Step 3. Calculate $\rho_k$.

Step 4. If $\rho_k < \mu$, take $\Delta_k = c\Delta_k$, go to Step 2; or go to Step 5.

Step 5. Take $x_{k+1} = x_k + d_k$.

Step 6. Calculate $\gamma(x_k)$ if $\gamma(x_k) \leq \varepsilon$ or $\gamma(x_k) \geq \frac{1}{3},$ set $\gamma(x_k) = \mu$.

Step 7. Calculate $\Delta_k = \frac{\|g_k\|}{\gamma(x_k)}\|f_k\|$.

$$\Delta_{k+1} = \begin{cases} c*\Delta & \text{if } \gamma_k < \mu \\ \Delta & \text{if } \gamma_k \geq \mu \end{cases}$$

Step 8. Set $k = k + 1$, go to Step 1.

2.2 Observation:

The goal of step 6 is to ensure strict directions and preserve the sequence \{γ(x_{k+1})\} uniformly bounded.

$$0 < min(\varepsilon, \mu) \leq \gamma(x_k) \leq max\left(\frac{1}{\varepsilon}, \mu\right),$$

(11)

3- Convergence analysis:
In this section of the paper, we will try to find the global convergence of algorithm 2.1. So we'll give some assumptions, including this one.

3.1 Supposition:
The level set \(L(x_0) = \{x \in \mathbb{R}^n | f(x) \leq f(x_0)\}\) is bounded.

3.2 Lemma:
Suppose the solution of (10), that is \(d_k\), so

\[
\|q_k(0) - q_k(d_k)\| \geq \frac{1}{2} \|g_k\| \|f_k\| \min \left\{ \Delta_k, \frac{\|g_k\|}{\gamma(x_k)} \right\},
\]

(proof):
If \(\Delta_k = \frac{\|g_k\|}{\gamma(x_k)} \|f_k\|\) and \(d_k = -g_k\)

\[
q_k(0) - q_k(d_k) = q_k(0) - q_k[-g_k] \\
= -g_k^T [-g_k] - \frac{1}{2} [g_k] \gamma(x_k)[-g_k] \\
= \frac{\|g_k\|^2}{\gamma(x_k)} - \frac{1}{2} \|g_k\|^T \frac{\|g_k\|}{\gamma(x_k)} \\
= \frac{\|g_k\|^2}{2\gamma(x_k)} \geq \frac{1}{2} \|g_k\| \min \left\{ \Delta_k, \frac{\|g_k\|}{\gamma(x_k)} \right\}
\]

That is complete proof.

3.3 Lemma:
The sequence \(\{x_k\}\) residues in \(L(x_0)\), and the sequence \(\{f(x_i(k))\}\) is no increasing and convergent in Algorithm 2.1.

(proof):
The proof is similar to the proof in the worksheet [9].

3.1 Theorem:
Presume that Supposition 3.1 holds. Suppose \(\{x_k\}\) be the sequence produced by Algorithm 2.1, so we get

\[
\lim_{k \to \infty} \inf \|g_k\| = 0,
\]

(proof):
By contradiction, we will get \(\|g_k\|\) is unbounded from zero. Assume that \(\exists \sigma \in (0,1) \) s.t

\[
\|g_k\| \geq \sigma, \forall k,
\]

(14)
It follows (10), \(\rho_k \geq \mu\) and Lemma 3.2 that

\[
f_{k+1} \leq f_{x_i(k)} - \mu \text{pred}_{k} \leq f_{x_i(k)} - \frac{1}{2} \mu \|g_k\| \min \left\{ \Delta_k, \frac{\|g_k\|}{\gamma(x_k)} \right\} \|f_k\|,
\]

(15)
then \(\forall k > M, \ let \ k = i(k) - 1. \) From (15),(11).
\[ f_{x_i(k)} \leq f_{x_i((k)-1)} - \frac{1}{2} \mu \|g^i(k)\| \min[\Delta_i(k)-1, \frac{\|g^i(k)\|}{f(x_i(k))-1}], \]

or

\[ f_{x_i(k)} \leq f_{x_i((k)-1)} - \frac{1}{2} \mu \sigma \min[\Delta_i, \frac{\sigma}{\max(\frac{1}{\varepsilon}, \mu)}], \]  \quad (16)

That means

\[ f_{x_i((k)-1)} - f_{x_i(k)} \geq \frac{1}{2} \mu \sigma \min[\Delta_i, \frac{\sigma}{\max(\frac{1}{\varepsilon}, \mu)}], \]  \quad (17)

from (17), we have \( f(x_i((k)-1)) - f(x_i(k)) \) is bounded a way from zero, that mean contradiction lemma 3-3, so (13) is true.

4. Numerical Results.

In this section we report the numerical tests to compare the results of a new technique TTRfb with three other algorithms:

TTRfB11: This algorithm is created by Quanyan Zhou · Chun Zhang [1].

TTRfB2: This algorithm is created by Shiker M. A. K and Sahib Z. [14].

TTRfB33: This algorithm is proposed by D. Tarzanagh et. al. [10].

The performance of all comparing algorithms runs on a computer with CPU–time 1.70 GHZ and 8.00 GB Ram, each algorithm codes are inscribed in MATLAB R2014a.

We take \( \mu_1 = 0.6; \mu_2 = 0.9; c = 0.5; p = 0.6; \) epsilon =\( 10^{-5} \) and number of total of iteration exceeds 20000. The problems that we took are:

\begin{align*}
  p_1: f &= 100 * (x_2 - x_1^2 + (1 - x_1) \\
  p_2: f &= (x_2 - x_1^2)^2 + (1 - x_1)^2 \\
  p_3: f &= (x_2 - x_1^2)^3 + (1 - x_1)^2 \\
  p_4: f &= (x_2 - x_1^2)^2 + (1 - x_1) \\
  p_5: f &= \sin(x_2)^2 - \cos(x_1)^3 \\
  p_6: f &= \sin(x_2)^{1/3} - \cos(x_1)^{1/3} \\
  p_7: f &= \sin(x_2)^{2} + \cos(x_1)^{3}
\end{align*}

The numerical results of all algorithms are recorded in table 4.1, 4.2 and 4.3. Table 4.1 consist the number of iterations, table 4.2 consists the number of functions evaluation and table 4.3 consists the CPU times.

| P  | TTRfB | TTRfB2 | TTRfB11 | TTRfB33 |
|----|-------|--------|---------|---------|
| P1 | 1     | 2      | 4       | 2       |
| P2 | 1     | 2      | 4       | 2       |
| P3 | 1     | 2      | 4       | 2       |
| P4 | 1     | 2      | 4       | 2       |
| P5 | 1     | 2      | 4       | 2       |
| P6 | 1     | 2      | 4       | 2       |
Table of functions evaluation

|   | TTRfB | TTRfB2 | TTRfB11 | TTRfB33 |
|---|-------|--------|---------|---------|
| P1 | 2     | 3      | 5       | 3       |
| P2 | 2     | 3      | 5       | 3       |
| P3 | 2     | 3      | 5       | 3       |
| P4 | 2     | 3      | 5       | 3       |
| P5 | 2     | 3      | 5       | 3       |
| P6 | 2     | 3      | 5       | 3       |
| P7 | 2     | 3      | 5       | 3       |

Table of CPU-Times

|   | TTRfB | TTRfB2 | TTRfB11 | TTRfB33 |
|---|-------|--------|---------|---------|
| P1 | 4.752 | 3.968  | 7.667   | 3.956   |
| P2 | 2.703 | 5.508  | 7.019   | 5.588   |
| P3 | 2.792 | 3.629  | 6.187   | 6.180   |
| P4 | 2.736 | 4.395  | 8.120   | 3.757   |
| P5 | 2.139 | 4.176  | 6.055   | 4.759   |
| P6 | 2.467 | 3.969  | 5.345   | 5.545   |
| P7 | 2.951 | 4.409  | 5.473   | 3.917   |

All the tables show that the results of the new algorithm (TTRfB) are better than the others in all terms. These tables show the quality of the performance of the new method that is in table 4.1 the number of iterations of (TTRfB) is less than the number of iteration of the other methods. In table 4.2 the number of functions evaluation is less than the other methods. And in table (4.3) the CPU time spent to reach the solution is less that the CPU time of the other method. So, that indicated that the new algorithm is better the other algorithms.

Conclusion:

The importance of this paper lies in the presentation of TRM. A new approach with adaptive nonmonotone TR has been introduced recognized on simple quadratic model. We custom a new approximation of Hessian scale for the minimization function of the TR subproblem and combining it
with nonmonotone technique to introduce the new adaptive method. We compare the performance of the new algorithm with three famous methods, the numerical results indicated the efficiency and robustness of our new algorithm and it is so promising to solve methods of unrestricted optimization.

Reference:

[1] Zhou, Q. and Zhang, C. (2012). "A new nonmonotone adaptive trust region method based on simple quadratic models", *J Appl Math Comput, V 40*, p 111–123.

[2] Amini, K., Shiker, M. A. K. and Kimiae, M. (2016). "A line search trust-region algorithm with nonmonotone adaptive radius for a system of nonlinear equations", *4 OR- Journal of operation, V 14*, p 133–152.

[3] Sheng, Z., yuan, G. and Cui, Z. (2018). "A new adaptive trust region algorithm for optimization problems", *Acta Mathematica Scientia, V 38B (2)*, p 479–496.

[4] Deng, N., Xiao, Y., Zhou, F. (1993), "Nonmonotonic trust region algorithm", *J. Optim. Theory Appl, V 26*, p 259–285.

[5] Raydan, M. (1997). "The Barzilia and Borwen Gradient method for the large scale unconstrained diminution problem", *SIAM J. OPTIM. V7*(1), p 26–33.

[6] Shi, Z. J and Guo, J. (2008). "A new trust region method with adaptive radius", *Compute Optima Appl, V 41*, p 225 - 242.

[7] Dwail, H. H., Mahdi, M. M., Wasi, H. A., Hashim, K. H., Dreeb, N. K., Hussein, A. H. and Shiker, M. A. K. (2020)," A new modified TR algorithm with adaptive radius to solve a nonlinear systems of equations", “in press”, accepted paper for publication in the University of Babylon (IVCUB)- Iraq.

[8] Dwail, H. H. and Shiker, M. A. K. (2020), "Using a new trust region algorithm with nonmonotone adaptive radius for solving nonlinear systems of equations", *International Journal of Advanced Science and Technology, V 29*(4), p 2351- 2360.

[9] Grippo, L., Lamparillo, F., and Lucidi, S. (1986). "A nonmonotone line search technique for Newton’s method", *SIAM J. Numer. Anal.* V 23, p 707–716.

[10] Ataeef, D. T., Reza M.P. and Mesgarani H, (2014)." A new nonmonotone trust region method for unconstrained optimization equipped by an efficient adaptive radius", *Optimization Methods and Software, V 29*(4), p 819-836.

[11] Mahdi, M. M. and Shiker, M. A. K., (2020), A new projection technique for developing a Liu-Storey method to solve nonlinear systems of monotone equations, “in press”, accepted paper for publication in *IOP Science, 5th International Scientific Conference for Iraqi Al-Khwarizmi Society, Iraq.*

[12] Mahdi, M. M. and Shiker, M. A. K., (2020), Three-term of new conjugate gradient projection approach under Wolfe condition to solve unconstrained optimization problems, “in press”, accepted paper for publication in *Journal of Advanced Research in Dynamical and Control Systems.*

[13] Shiker, M. A. K. and Amini, K. (2018)." A new projection-based algorithm for solving a large-scale nonlinear system of monotone equations", *Croatian Operational Research Review, V 9*(1), p 63–73.

[14] Shiker, M. A. K. and Sahib, Z, (2018)." A modified trust-region method for solving unconstrained optimization", *Journal of Engineering and Applied Sciences, V13*(22), p 9667-9671.

[15] Wasi, H. A. and Shiker, M. A. K. (2020)." A new conjugate gradient method for solving large scale systems of monotone equations", *International Journal of Advanced Science and Technology, V 29*(4), p 2303- 2314.

[16] Mahdi, M. M. and Shiker, M. A. K. (2020). "Three terms of derivative free projection technique for solving nonlinear monotone equations", “in press”, accepted paper for publication in *IOP Science, 5th International Scientific Conference for Iraqi Al-Khwarizmi Society, Iraq.*
[17] Mahdi, M. M. and Shiker, M. A. K. (2020). "Solving systems of nonlinear monotone equations by using a new projection approach", “in press”, accepted paper for publication in *IOP Science, 1st International Virtual Conference of the University of Babylon (IVCUB)- Iraq.

[18] Hussein, H. A. and Shiker, M. A. K. (2020).” A modification to Vogel's approximation method to Solve transportation problems”, “in press”, accepted paper for publication in *IOP Science, 5th International Scientific Conference for Iraqi Al-Khwarizmi Society, Iraq.

[19] Hussein, H. A., Shiker, M. A. K. and Zabiba, M. S. M. (2020).” A new revised efficient of VAM to find the initial solution for the transportation problem”, “in press”, accepted paper for publication in *IOP Science, 5th International Scientific Conference for Iraqi Al-Khwarizmi Society, Iraq.

[20] Hassan, Z. A. H. H. and Shiker, M. A. K. (2018)." Using of generalized baye’s theorem to evaluate the reliability of aircraft systems”. *Journal of Engineering and Applied Sciences*, (Special Issue13), p 10797-10801

[21] Hussein, H. A. and Shiker, M. A. K. (2020). "Two New Effective Methods to Find the Optimal Solution for the Assignment Problems", “in press”, accepted paper for publication in *Journal of Advanced Research in Dynamical and Control Systems*.