A Retrospective Cubic Regularization for Unconstrained Optimization Using Nonmonotone Techniques

Yajing Wang, Qinghua Zhou*
College of Mathematics and Information Science, Hebei University, Baoding, 071002, China

Abstract: In recent years, cubic regularization algorithm for unconstrained optimization has been defined as alternative to trust-region and line search schemes. These regularization techniques are based on the strategy of computing an approximate global minimizer of a cubic overestimator of the objective function. It can effectively handle with some worst cases and improve the iteration complexity. In this work, we investigate a new approach which combines retrospective adaptive cubic regularization algorithm and some nonmonotone linesearch strategy. Under some standard assumptions, the global convergence properties are given.

Keywords: Unconstrained optimization; Cubic regularization; Retrospective trust region method; Global convergence.

1. Introduction

We consider the unconstrained optimization problem

$$\min_{x \in R^n} \quad f(x) \tag{1}$$

where $f : R^n \to R$ is a continuously differentiable function.

It is well known that many methods have been proposed which are based on a regularization technique using a cubic overestimator of the objective function (see, e.g. [1-6]). In 2011, an adaptive regularization framework using cubic (ARC) has been proposed in [1]. Specially, at each iteration x_k, starting from the current point x_k, the cubic model is

$$m_k(p) = f(x_k) + p^T g(x_k) + \frac{1}{2} p^T B_k p + \frac{1}{3}\sigma_k \|p\|^3 \tag{2}$$

Where $g(x_k)$ is the gradient of f computed at x_k, B_k is symmetric approximation of the Hessian matrix, $\sigma_k \in R^+$ is an adaptive parameter. For the sake of simplicity, we set $g(x_k) = g_k$, $f(x_k) = f_k$. The trial step p_k is computed as an approximate minimizer of $m_k(p)$ in equation (2) and must yield a decrement at least as good as that provided by a Cauchy point.

Under certain conditions, first-order global convergence and second-order global convergence of the adaptive cubic regularization algorithm have been proved in [1], and find the solution to satisfy the second-order necessary condition. Some iteration complexity properties and an iteration complexity bound based on the Cauchy condition of the ARC is given in [7].

As we all know, trust region method is an effective method to solve the problem of unconstrained optimization and has been systematically introduced. In 2010, Bastin et al. proposed a retrospective trust-region method for unconstrained optimization [8]. On the basis of the basic trust region method, the backtracking technique is added to update the trust region radius, and the current iteration point is applied instead of the previous iteration point. The backtracking technology of retrospective trust region is an improvement in the idea of adaptive technology which is applied in the optimization process, especially in the noise objective function.

By comparing the ARC algorithm with the trust region, it is found that the number of iterations and the number of function value iterations of ARC are both lower than that of the basic trust region methods in [1]. Based on the similarity of the adaptive regularization parameter updating and the radius updating of the trust-region method, Sun proposed a new method [9] that is combined the adaptive cubic regularization method with retrospective trust-region method for unconstrained optimization which provides some more promising computational results. Considering the well known merits of nonmonotone strategies, we introduce a hybrid method combing them with the current retrospective trust region methods.

The rest outline of the paper is as follows: Section 2 proposes the hybrid method and gives some explanations, and global convergence results are presented in Section 3. Finally, in section 4, we give some concluding remarks.

2. The Nonmonotone Retrospective Adaptive Cubic Regularization Algorithm

Tommaso Bianconcini and Marco Sciandrone proposed the nonmonotone ARC algorithm (NMARC) [10]. They use nonmonotone line search techniques on the basis of the cubic regularization algorithm and assume p_k is a good descent direction which satisfies the following conditions.

Condition 2.1 There exist some constants c_1, $c_2 > 0$ such that

$$g_k^T p_k \le -c_1 \|g_k\|^2 \tag{3}$$

$$\|p_k\| \le c_2 \|g_k\| \tag{4}$$

In the case of $f(x_k + p_k) < f(x_k)$, they perform a monotone extrapolation phase (MEP) along p_k in order to attain a further reduction of the objective function. In the case of $f(x_k + p_k) \ge f(x_k)$, a suitable nonmonotone accepting criterion is used.

Algorithm MEP

Given γ_a, $\theta \in (0,1)$, integer L, the current point x_k, the direction p_k satisfying condition (2.1) and such that

$$f(x_k + p_k) \le f(x_k) + \gamma_a g(x_k)^T p_k \tag{5}$$

Set λ=1, $j = 0$.

If

$$f\left(x_k + \frac{\lambda}{\theta} p_k\right) \le f(x_k) + \gamma_a \frac{\lambda}{\theta} g(x_k)^T p_k \tag{6}$$

Then set $\lambda = \frac{\lambda}{\theta}$, $j = j+1$, otherwise set $\lambda_k = \lambda$ and exit.

If $j < L$ then go to step 2, otherwise set $\lambda_k = \lambda$ and exit.

Set

$$f\left(x_{m(k)}\right) = \max_{j \in S_{k,M}} f(x_j) \tag{7}$$

where $M \in N$ and $m(k) \in S_{k,M}$ respectively, and $S_{k,M}$ is the set of indexes of the last M successful iterations before the kth.

After specifically investigate the program in [9], we found there is logically wrong. That is they computed "p_k" for twice and the second one is useless in the program. Noticing that, we modify the above error and introduce some non-monotone strategies in the new algorithm.

Here are the details.

Algorithm 1:

Step 1: Given x_0, $\sigma_0 > 0$, $k = 0$, $0 < \eta_1 < 1$, $0 < \tilde{\eta}_1 < \tilde{\eta}_2 < 1$, $1 < \gamma_1 < \gamma_2 < \gamma_3$, $\gamma_a > 0$, $\gamma_b > 0$, $\gamma_c > 0$, $L, M \in N$, $\varepsilon > 0$.

Step 2: Compute g_k, if $\|g_k\| < \varepsilon$, then exit; otherwise, compute p_k, satisfying

$$m_k(p_k) \le m_k(p_k^c) \tag{8}$$

where $p_k^c = -\alpha_k g_k$,

$$\alpha_k = \arg\min m_k(-\alpha_k g_k) \tag{9}$$

Step 3:

a) If p_k satisfies the condition 2.1 then

(i) If $f(x_k + p_k) \le f(x_k) + \gamma_a g(x_k)^T p_k$

Then apply MEP to compute λ_k and go to step 4.

(ii) If $f(x_k + p_k) \le f\left(x_{m(k)}\right) + \gamma_a g(x_k)^T p_k$

Then set $\lambda_k = 1$ and go to step 4.

(1.b) If $f(x_k) - f(x_k + p_k) \ge \gamma_b \|p_k\|^2$

Then set $\lambda_k = 1$ and go to step 4.

c) If $f(x_k) - f(x_k + p_k) \ge \gamma_c \|p_k^c\|^2$

Then set $p_k = p_k^c$, $\lambda_k = 1$; otherwise $\lambda_k = 0$ and go to step 4.

Step 4: compute

$$Ared_k = \begin{cases} f\left(x_{m(k)}\right) - f(x_k + \lambda_k p_k) & if\ 1.a(ii)\ is\ performed \\ f(x_k) - f(x_k + \lambda_k p_k) & otherwise \end{cases}$$

$$Pred_k = f(x_k) - m_k(p_k)$$

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

Set $x_{k+1} = \begin{cases} x_k + \lambda_k p_k & if\ r_k \ge \eta_1 \\ x_k & otherwise \end{cases}$, k=k+1.

Step 5: if $x_k = x_{k-1}$, then choose $\sigma_k \in [\sigma_{k-1}, \gamma_3 \sigma_{k-1}]$, otherwise compute

$$A\tilde{red}_k = \begin{cases} f\left(x_{m(k-1)}\right) - f(x_k) & if\ 1.a(ii)\ is\ performed \\ f(x_{k-1}) - f(x_k) & otherwise \end{cases}$$

$$P\tilde{red}_k = m_k(-p_{k-1}) - m_k(0)$$

$$\tilde{r}_k = \frac{A\tilde{red}_k}{P\tilde{red}_k}$$

Where $f\left(x_{m(k-1)}\right)$ is defined by equation (7).

Set

$$\sigma_k = \begin{cases} (0, \sigma_{k-1}] & if\ \tilde{r}_k > \tilde{\eta}_2 & very\ successful\ iteration \\ (\sigma_{k-1}, \gamma_1 \sigma_{k-1}] & if\ \tilde{r}_k \in [\tilde{\eta}_1, \tilde{\eta}_2) & successful\ iteration \\ (\gamma_1 \sigma_{k-1}, \gamma_2 \sigma_{k-1}] & if\ \tilde{r}_k < \tilde{\eta}_1 & unsuccessful\ iteration \end{cases}$$

And go to step 2.

Above is the whole program of the new retrospective adaptive cubic regularization algorithm using nonmonotone line search technique.

3. Convergence Theory

We now investigate the convergence properties of the new algorithm, under some certain assumptions, the convergence of the proposed algorithm is proved.

A.1 $\|B_k\| \le \tau_B$ for all k and some $\tau_B \ge 0$.

A.2 the gradient g_k is uniformly continuous on the sequence of iterates $\{x_k\}$.

A.3 f is bounded below and is uniformly continuous on the sequence $\{x_k\}$.

Lemma 1 Suppose that the step p_k satisfies equation (8), then for all $k \ge 0$, we have that

$$f(x_k) - m_k(p_k) \ge f(x_k) - m_k(p_k^c) \ge \frac{\|g_k\|^2}{6\sqrt{2}\, max\left[1+\|B_k\|, 2\sqrt{\sigma_k \|g_k\|}\right]} = \frac{\|g_k\|}{6\sqrt{2}} min\left[\frac{\|g_k\|}{1+\|B_k\|}, \frac{1}{2}\sqrt{\frac{\|g_k\|}{\sigma_k}}\right] \quad (10)$$

Proof. See the proof of [1, lemma 2.1].

Lemma 2 Suppose that A.1 holds, and the step p_k satisfies (8), then we have that

$$\|p_k\| \le \frac{3}{\sigma_k} max\left(\tau_B, \sqrt{\sigma_k \|g_k\|}\right) \quad (11)$$

Proof. See the proof of [1, lemma 2.2].

Lemma3 Assume $f \in C^1(R^n)$ and that A.1 holds, suppose that I is an infinite index set such that $\lambda_k \ne 0$ and $\|g(x_k)\| \ge \varepsilon$ for all $k \in I$ and some $\varepsilon > 0$, and

$$\sqrt{\frac{\|g(x_k)\|}{\sigma_k}} \to 0 \text{ for } k \to \infty,\ k \in I \quad (12)$$

then $\|p_k\| \le 3\sqrt{\frac{\|g(x_k)\|}{\sigma_k}}$ for all $k \in I$ sufficiently large. (13)

Additionally, if $x_k \to x^*$, as $k \in I$, $k \to \infty$ for some $x^* \in R^n$, Then each iteration $k \in I$ that is sufficiently large is very successful, and

$$\sigma_{k+1} \le \sigma_k \text{ for all } k \in I \text{ sufficiently large.} \quad (14)$$

Proof. We can get $\sqrt{\sigma_k \|g_k\|} = \|g_k\| \sqrt{\frac{\sigma_k}{\|g_k\|}} \ge \varepsilon \sqrt{\frac{\sigma_k}{\|g_k\|}} \to \infty$ from $\sqrt{\frac{\|g_k\|}{\sigma_k}} \to 0$ for $k \in I$, $k \to \infty$, from the equation (11), we can obtain (13), and thus we proved the first part of the lemma. Next we prove the remainder of the lemma.

Firstly, due to the definition of the (7) and $\tilde{r}_k$, we observe that

$$\tilde{r}_k \ge \frac{f(x_{k-1}) - f(x_k)}{m_k(-p_{k-1}) - m_k(0)}$$

then, if $\tilde{r}_k \ge \tilde{\eta}_2$ holds, we can obtain

$$\frac{f(x_{k-1}) - f(x_k)}{m_k(-p_{k-1}) - m_k(0)} \ge \tilde{\eta}_2$$

from lemma 1, we can know

$$m_k(-p_{k-1}) - m_k(0) \ge \frac{\|g_k\|}{6\sqrt{2}} min\left\{\frac{\|g_k\|}{1+\|B_k\|}, \frac{1}{2}\sqrt{\frac{\|g_k\|}{\sigma_k}}\right\} \quad (15)$$

So we can get

$$\frac{m_k(-p_{k-1}) - f(x_{k-1})}{m_k(-p_{k-1}) - m_k(0)} \le 1 - \tilde{\eta}_2$$

$$\Rightarrow m_k(-p_{k-1}) - f(x_{k-1}) + (1-\tilde{\eta}_2)\left(m_k(0) - m_k(-p_{k-1})\right) \le 0$$

then we set

$$\rho_k = m_k(-p_{k-1}) - f(x_{k-1}) + (1-\tilde{\eta}_2)\left(m_k(0) - m_k(-p_{k-1})\right) \quad (16)$$

using equation (12), we obtain

$$m_k(-p_{k-1}) - m_k(0) \ge \frac{\|g_k\|}{12\sqrt{2}}\sqrt{\frac{\|g_k\|}{\sigma_k}}$$
$$\Rightarrow m_k(0) - m_k(-p_{k-1}) \le -\frac{\|g_k\|}{12\sqrt{2}}\sqrt{\frac{\|g_k\|}{\sigma_k}} \quad (17)$$

a sample Taylor expansion of $f(x_{k-1})$ around x_k gives that for each k,

$$f\left(x_{m(k-1)}\right) \ge f(x_{k-1}) = f\left(x_k + (-p_{k-1})\right) = f(x_k) - g_k^T(\xi_k)\, p_{k-1}$$

for some $\xi_k \in (x_{k-1}, x_k)$, then

$$m(-p_{k-1}) - f\left(x_{m(k-1)}\right) \le m_k(-p_{k-1}) - f(x_{k-1}) = \left(g(\xi_k) - g(x_k)\right)^T p_{k-1} + \frac{1}{2} p_{k-1}^T B_k p_{k-1} - \frac{1}{3}\sigma_k \|p_{k-1}\|^3 \le \left(g(\xi_k) - g(x_k)\right)^T p_{k-1} + \frac{1}{2} p_{k-1}^T B_k p_{k-1}$$

Using A.1 and equation (13), we obtain

$$m_k(-p_{k-1}) - f\left(x_{m(k-1)}\right) \le m_k(-p_{k-1}) - f(x_{k-1}) \le 3\sqrt{\frac{\|g_k\|}{\sigma_k}}\left\{\|g(\xi_k) - g(x_k)\| + \frac{3\tau_B}{2}\sqrt{\frac{\|g_k\|}{\sigma_k}}\right\} \quad (18)$$

for all $k \in I$ sufficiently large. From equations (16), (17), (18), it follows

$$\rho_k \le 3\sqrt{\frac{\|g_k\|}{\sigma_k}} \left\{ \begin{array}{l} \|g(\xi_k) - g(x_k)\| + \\ \frac{3\tau_B}{2}\sqrt{\frac{\|g_k\|}{\sigma_k}} - \varepsilon \frac{1-\tilde{\eta}_2}{36\sqrt{2}} \end{array} \right\} \tag{19}$$

due to the continuity of the gradient, we conclude that

$$\|g(\xi_k) - g(x_k)\| \to 0 \quad \text{for } k \in I,\ k \to \infty \tag{20}$$

from equations (19), (20), (12), we obtain $\rho_k < 0$ for all $k \in I$, $k \to \infty$. The inequality (14) now follows the instructions at step 5 of the new algorithm.

Lemma 4 Assume that $f \in C^1(R^n)$ and that A.1 holds, suppose furthermore that there are only finitely many successful iterations. Then $x_k = x_*$ for all sufficiently large k and $g(x_*) = 0$.

Proof. See the proof of [10, lemma 4.4].

Lemma 5 Assume that A.1 and A.3 hold, then the sequence $\{f(x_k)\}$ converges.

Proof. See the proof of [10, proposition 4.5].

Lemma 6 Assume that A.1 and A.3 hold, then

$$\liminf_{k\to\infty} \|g_k\| = 0$$

Proof. See the proof of [10, theorem 4.6].

Theorem 7 Assume that A.1-A.3 hold. Then

$$\lim_{k\to\infty} \|g_k\| = 0$$

Proof. See the proof of [10, theorem 4.7].

4. Conclusions

In this paper, we present a new method. The proposed algorithm is based on the idea of modifying the length of the step obtained using the cubic overstimating model and on the employment of a nonmonotone accepting criterion. The aim was that of improving the computational efficiency of the retrospective cubic regularization method preserving the global convergence. Recently, the adaptive cubic regularization algorithm is generalized to solve linear least-squares problems and nonlinear constrained optimization problems [11], and the convergence analyses are given and good numerical results are obtained. To solve the constrained optimization problem with adaptive cubic regularizations still to be further studied.

References

[1] C. Cartis, N. I. M. Gould and Ph. L. Toint. Adaptive cubic regularization methods for unconstrained optimization. Part I: motivation, convergence and numerical results. Mathematical Programming, 2011; 127 (2): 245-295.

[2] H.Y. Benson and D.F. Shanno. Interior-point methods for nonconvex nonlinear Programming: Cubic regularization, Computational Optimization & Applications, 2014; 58 (2): 323-346.

[3] A. Griewank, J. Fisher, and T. Bosse. Cubic overestimation and secant updating for unconstrained Optimization of C2,1 functions. Optimization Method and Soft- ware, 2014; 29 (5): 1075-1089.

[4] J.M.Martinez and M. Raydan. Separable cubic modeling and a trust-region strategy for unconstrained minimization with impact in global optimization. Journal of Global Optimization, 2015; 63 (2): 319-342.

[5] Y. Nesterov and B. T. Polyak. Cubic regularization of Newton method and its glo-bal performance. Mathematical Programming, 2006; 108 (1): 177-205.

[6] Ph. L. Toint, Nonlinear stepsize control, trust regions and regularizations for unconstrained optimization. Optimization Methods & Software, 2013; 28 (1): 82-95.

[7] C. Cartis, N. I. M. Gould and Ph. L. Toint. Adaptive cubic regularization methods for unconstrained optimization. Part II: worst-casefunction-and derivative evaluation complexity, Mathematical Programming, 2011; 130 (2): 295-319.

[8] Bastin, Fabian, Malmedy, et al. A Retrospective Trust-Region Method for Unconstrained Optimization. Mathematical Programming, 2010; 123 (2): 395-418.

[9] Sun Wenzhu. A retrospective adaptive cubic regularization method for unconstrained optimization. Hebei University of Technology, 2015: 10-16. In Chinese.

[10] T. Bianconcini, M. Sciandrone. A cubic regularization algorithm for unconstrained optimization using line search and nonmonotone techniques. Optimization Method & Software, 2016; 31 (5).

[11] C. Cartis, N. I. M. Gould and Ph.L. Toint. Evaluation complexity of adaptive cubic regularization methods for convex unconstrained Optimization. Optimization Methods & Software, 2012; 27 (2): 197-219.