

МЕТОД l_3 -КУБИЧЕСКОЙ РЕГУЛЯРИЗАЦИИПан Тай¹⁾, Бо Ю²⁾¹⁾ Университет технологий Даляня, Китай, Далянь, pangtai@mail.dlut.edu.cn²⁾ Университет технологий Даляня, Китай, Далянь, yubo@dlut.edu.cn

В данной статье представлен новый метод кубической регуляризации для задачи минимизации без ограничений. В качестве регуляризирующего члена используется сепарабельная норма $\|s\|_3^3$ вместо стандартной $\|s\|_2^3$, что позволяет более эффективно решать подзадачи. Доказана сходимость метода и получена оценка его наихудшей итерационной сложности $\mathcal{O}(\varepsilon^{-2})$. Для эффективного решения подзадач кубической регуляризации ℓ_3 предлагается алгоритм с тремя порогами. Проведены численные эксперименты, демонстрирующие эффективность предложенного метода.

Ключевые слова: кубическая регуляризация; гладкая минимизация без ограничений; норма ℓ_1 .

THE l_3 -CUBIC REGULARIZATION METHODPang Tai¹⁾, Bo Yu²⁾¹⁾ Dalian university of technology, China, Dalian, pangtai@mail.dlut.edu.cn²⁾ Dalian university of technology, China, Dalian, yubo@dlut.edu.cn

In this paper, a new cubic regularization method for unconstrained minimization is introduced. It uses separable $\|s\|_3^3$ instead of $\|s\|_2^3$ as the regularization term, so that it is possible to solve the subproblems more efficiently. Its convergence is proved and the worst-case iteration complexity is shown to be $\mathcal{O}(\varepsilon^{-2})$. To solve the ℓ_3 -cubic regularization subproblems efficiently, a 3-thresholding algorithm is proposed. Numerical experiments are done to show the efficiency of the proposed method.

Keywords: Cubic regularization; Smooth unconstrained minimization; ℓ_3 norm.

Introduction

The cubic regularized Newton (CRN) method,

$$x^{k+1} = x^k + \arg \min_s \left\{ m_k(s) \triangleq f_k + (g^k)^\top s + \frac{1}{2} s^\top H_k s + \frac{\omega_k}{6} \|s\|_3^3 \right\}, \quad (1.4)$$

for solving the unconstrained optimization problem

$$\min_{x \in \mathbb{R}^n} f(x) \quad (1.5)$$

where $f: \mathbb{R}^n \rightarrow \mathbb{R}$ is a twice differentiable function, ω_k is chosen such that:
 $\forall x \in \mathbb{R}^n$,

$$f(x) \leq m_k(x - x^k) \quad (1.6)$$

and f_k, g^k, H_k represent the function value, gradient and the Hessian of f at x^k , respectively. The class of methods was systematically studied by Nesterov and Polyak [1], and the similar idea was independently used by Griewank et al. [2], in which

$$m_k(s) = f_k + (g^k)^\top s + \frac{1}{2} s^\top H_k s + \frac{\sigma_k}{6} \|s\|_{G_k}^3, \quad (1.7)$$

where regular parameters σ_k and positive definite weight matrices G_k are iteratively chosen to ensure the overestimation property while preserving affine invariance.

Furthermore, Weiser et al. [3] also pursued the same line of thought as Griewank et al. with $G_k = G$. If the objective Hessian is globally Lipschitz continuous, global convergence to second-order critical points and asymptotically quadratic rate of convergence are also proved. Besides, if the objective Hessian is globally Lipschitz continuous and positive definite, CRN method has a better global iteration complexity bound [4] [5], to line-search and trust region strategies. CRN method addresses the drawback of Newton's method that can't handle degenerate Hessians and provides an alternative globalization strategy.

The paper is organized as follows. The ℓ_3 -cubic regularization method, its convergence, complexity results are given in Section 2, a 3-thresholding algorithm for solving the subproblems is given in Section 3, experimental results are in Section 4 and the conclusions follow in Section 5.

The ℓ_3 -cubic regularization method.

In this section, we introduce an ℓ_3 -cubic regularization (ℓ_3 -CR) method, prove its convergence and give some complexity results. The ℓ_3 cubic subproblem of ℓ_3 -CR method is as follows:

$$\min_{s \in \mathbb{R}^n} \left\{ C_k(s) \triangleq f_k + (g^k)^\top s + \frac{1}{2} s^\top H_k s + \frac{\omega_k}{6} \|s\|_3^3 \right\}. \quad (1.8)$$

Using the separable $\|s\|_3^3$ instead of inseparable $\|s\|_2^3$ will ensure that we can solve the new cubic subproblems more efficiently, although the second term is inseparable. Then, we give an adaptive ℓ_3 -CR method in Figure 1. We are now ready to prove the first convergence result for the method. In particular, we show that there is a subsequence of $\{g^k\}$ converging to zero.

Algorithm 1 ℓ_3 -cubic regularization method

Input: initial value x^0 ; parameters $\gamma_2 \geq \gamma_1 > 1$, $1 > \eta_2 \geq \eta_1 > 0$, k_{max} and $M_0 > 0$.
For $k = 0, 1, \dots, k_{max}$, do

1. Compute a step s^k by solving the ℓ_3 -cubic regularization subproblem such that

$$C_k(s^k) \leq C_k(s_c^k),$$

where $s_c^k = -\alpha_k^c g^k$ is the Cauchy point with

$$\alpha_k^c = \arg \min_{\alpha \in \mathbb{R}^+} C_k(-\alpha g^k) = \frac{-(g^k)^\top H_k g^k + \sqrt{((g^k)^\top H_k g^k)^2 + 2M_k \|g^k\|_3^3 \|g^k\|_2^2}}{M_k \|g^k\|_3^3};$$

2. Compute

$$\rho_k = \frac{f(x^k) - f(x^k + s^k)}{f(x^k) - C_k(s^k)};$$

3. Set

$$x^{k+1} = \begin{cases} x^k + s^k, & \text{if } \rho_k \geq \eta_1, \\ x^k, & \text{otherwise;} \end{cases}$$

4. Set

$$M_{k+1} \in \begin{cases} [0, M_k], & \text{if } \rho_k > \eta_2, & \text{(very successful step)} \\ [M_k, \gamma_1 M_k], & \text{if } \eta_1 \leq \rho_k \leq \eta_2, & \text{(successful step)} \\ [\gamma_1 M_k, \gamma_2 M_k], & \text{otherwise;} & \text{(unsuccessful step)} \end{cases}$$

5. If termination criterions are satisfied, then return x^k and break;

6. If $k = k_{max}$ without convergence, return failure information.
-

Fig.1. ℓ_3 -cubic regularization method.

Theorem 1: Suppose $\|H_k\|_1 \leq \kappa_2$ hold for all $k \geq 0$ and $\kappa_2 \geq 0$. If $\{f(x^k)\}$ is bounded below, then

$$\liminf_{k \rightarrow \infty} \|g^k\| = 0. \quad (1.9)$$

Our next result shows the Q-quadratic asymptotic convergence of the ℓ_3 -cubic regularization method.

Theorem 2: If $\nabla_s C_k(s^k) = 0$ and $\|H(x) - H(y)\|_1 \leq L_H \|x - y\|_3$ hold for all x, y sufficiently close to x^* , $x^k \rightarrow x^*$, as $k \rightarrow \infty$ and $H(x^*)$ is positive definite, then g^k converges to zero, and x^k to x^* , Q-quadratically, as $k \rightarrow \infty$.

If the subproblem satisfies $\|\nabla_s C_k(s^k)\| \leq \theta_k \|g^k\|$, then our subsequent result establishes Q-quadratic asymptotic convergence of the ℓ_3 -cubic regularization method.

Theorem 3: If $\|\nabla_s C_k(s^k)\| \leq \theta_k \|g^k\|$, $\nabla_s C_k(s^k)^\top s^k \leq 0$ and $\|H(x) - H(y)\|_1 \leq L_H \|x - y\|_3$ hold for all x, y sufficiently close to x^* , $x^k \rightarrow x^*$, as $k \rightarrow \infty$ and $H(x^*)$ is positive definite, then g^k converges to zero, and x^k to x^* , Q-quadratically, as $k \rightarrow \infty$.

We are now ready to give an improved complexity bound for the ℓ_3 -cubic regularization method.

Theorem 4: Let $\|H_k\| \leq \kappa_2$ for all $k \geq 0$ hold, and $\{f(x^k)\}$ be bounded below by f_{low} . Assume that $\|g^0\| > \epsilon$, $\epsilon \in (0, 1]$, $j_1 \leq \infty$ be the first iteration such that $\|g^{j_1+1}\| \leq \epsilon$. Then, the ℓ_3 -cubic regularization method takes at most

$$K_1^s \triangleq \lceil \kappa_4 \epsilon^{-2} \rceil \quad (1.10)$$

successful iterations or equivalently, gradient evaluations, to generate $\|g^{j_1+1}\| \leq \epsilon$, where

$$\kappa_4 = (f(x^0) - f_{\text{low}}) / (\eta_1 \alpha_c) \text{ and } \alpha_c = \left[\frac{9}{2} \max(\kappa_2, \sqrt{2\omega_0}, \sqrt{2\gamma_2 \kappa_3}) \right]^{-1}. \quad (1.11)$$

Additionally, assume that on each very successful iteration $\{s^k\}$, ω_{k+1} is chosen such that $\omega_{k+1} \geq \gamma_3 \omega_k$, $\gamma_3 \in (0, 1]$ is satisfied. Then

$$j_1 \leq \lceil \kappa_c \epsilon^{-2} \rceil \triangleq K_1 \quad (1.12)$$

and so the ℓ_3 -cubic regularization method takes at most K_1 (successful and unsuccessful) iterations, and function evaluations, to generate $\|g^{j_1+1}\| \leq \epsilon$, where

$$\kappa_c = \left(1 - \frac{\log \gamma_3}{\log \gamma_1} \right) \kappa_4 + \kappa_c^u, \quad \kappa_c^u = \frac{1}{\log \gamma_1} \max \left(1, \frac{\gamma_2 \kappa_3^2}{\omega_0} \right). \quad (1.13)$$

The algorithm for solving the ℓ_3 -cubic regularization subproblems.

In this section, we consider how to efficiently solve the ℓ_3 -cubic regularization subproblems

$$\min_{s \in \mathbb{R}^n} \left\{ C_{\omega_k}(s) = (g^k)^\top s + \frac{1}{2} s^\top H_k s + \frac{\omega_k}{6} \|s\|_3^3 \right\}. \quad (1.14)$$

Employing the first-order optimality conditions, we can obtain

$$g^k + H_k s + \frac{\omega_k}{6} \nabla(\|s\|_3^3) = 0. \quad (1.15)$$

Multiplying by any positive parameter μ on both sides of the equation, we have

$$s + \frac{\omega_k \mu}{6} \nabla(\|s\|_3^3) = s - \mu(g^k + H_k s). \quad (1.16)$$

We denote the operator as:

$$R_{\omega_k, \mu, 3}(\cdot) = \left[I + \frac{\omega_k \mu}{6} \nabla(\|\cdot\|_3^3) \right]^{-1}, \quad (1.17)$$

which implies

$$s = \left[I + \frac{\omega_k \mu}{6} \nabla(\|\cdot\|_3^3) \right]^{-1} (s - \mu(g^k + H_k s)) = R_{\omega_k, \mu, 3}(s - \mu(g^k + H_k s)) \quad (1.18)$$

Then we prove the following results on $R_{\omega_k, \mu, 3}(\cdot)$.

Theorem 5: As a mapping from $\mathbb{R}^n$ to $\mathbb{R}$, the resolvent operator $R_{\omega_k, \mu, 3}(\cdot)$ is well defined. It is a diagonally nonlinear analytically expressive operator, and can be specified by

$$R_{\omega_k, \mu, 3}(s) = (f_{\omega_k, \mu, 3}(s_1), f_{\omega_k, \mu, 3}(s_2), \dots, f_{\omega_k, \mu, 3}(s_n))^\top. \quad (1.19)$$

Where

$$f_{\omega_k, \mu, 3}(s_i) = \frac{\text{sign}(s_i)(-1 + \sqrt{1 + 2\omega_k \mu |s_i|})}{\omega_k \mu}, \quad i = 1, \dots, n. \quad (1.20)$$

With the thresholding representation, the 3-thresholding algorithm to solve the subproblems can be shown in Figure 2.

Algorithm 2 3-thresholding algorithm

```

1: Input: parameters  $\mu_j$  and  $j_{\max}$ ; initial values  $g^k$ ,  $H_k$ , and  $\omega_k$ .
2: for  $j = 0, 1, \dots, j_{\max}$  do
3:   Compute  $z = s^{j-1} - \mu_j(g^k + H_k s^{j-1})$ .
4:   for  $i = 1, \dots, n$  do
5:     Compute  $s_i^j = \frac{\text{sign}(z_i)(-1 + \sqrt{1 + 2\omega_k \mu_j |z_i|})}{\omega_k \mu_j}$ .
6:   end for
7:   if termination criteria are satisfied then
8:     return  $s^j$ ; break
9:   end if
10:  if  $j = j_{\max}$  and termination criteria is not achieved then
11:    return failure information.
12:  end if
13: end for
14: return  $s = s^j$ 

```

Fig. 2. 3-threshoding algorithm for solving subproblems.

Experimental results

We now turn our attention to investigating how our methods perform in practice. We have implemented the l_3 -cubic regularization method described in Sction 2. In Figure 3, we present the iteration-count performance profile for these algorithms.

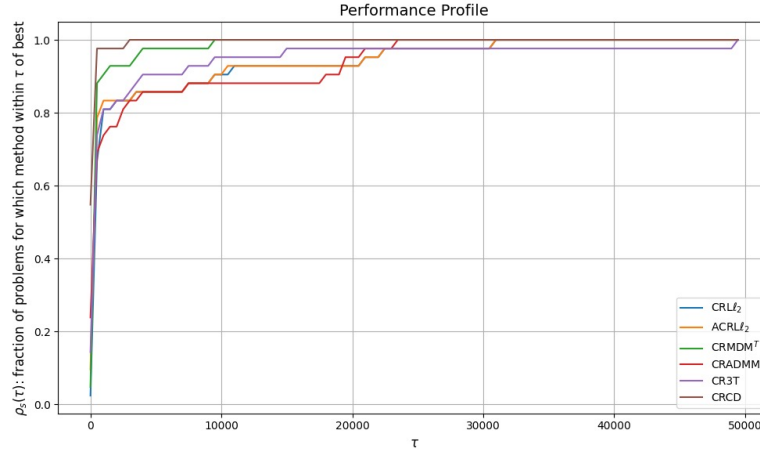


Fig. 3. Performance profiles (computing time) for $CRL\ell_2$, $ACL\ell_2$, $CR-MDM^\top$, $CRADMM$, $CR3T$ and $CRCD$ on the CUTEst problems.

Conclusion

In this paper, we have considered a new l_3 -cubic regularization method for unconstrained optimization and presented its convergence and complexity analysis. The method allows for the approximate solution of the key computational step, and are suitable for large-scale problems. We presented the l_3 -thresholding algorithm for solving the subproblems as well as their

convergence. Encouraging preliminary numerical experiments with small-scale problems were reported.

References

1. *Nesterov, Y., Polyak, B. T.* Cubic regularization of newton method and its global performance. *Mathematical Programming*. 2006. №108(1). C. 177-205.
2. *Griewank, Andreas.* The modification of Newton's method for unconstrained optimization by bounding cubic terms. Technical report NA/12. 1981.
3. *Weiser, M., Deufhard, P., & Erdmann, B.* Affine conjugate adaptive Newton methods for nonlinear elastomechanics. *Optimisation Methods and Software*. 2007. № 22(3). C. 413-431.
4. *Cartis, C., Gould, N. I., & Toint, P. L.* Adaptive cubic regularisation methods for unconstrained optimization. Part I: motivation, convergence and numerical results. *Mathematical Programming*. 2011. №127(2). C. 245-295.
5. *Nesterov, Y.* Accelerating the cubic regularization of Newton's method on convex problems. *Mathematical Programming*. 2008. №112(1). C. 159-181.