Adaptive Cubic Regularization Methods for Solving Non-convex Multi-objective Optimization Problem

Huanhuan Yang
College of Mathematics and Information Science, Hebei University, Baoding, 071002, China

Abstract: In this paper, we design an algorithm by adaptive cubic regularization methods for solving multi-objective optimization problem. The regularization technique is based on the strategy of computing an approximate global minimizer of a cubic overestimator of the objective function. The new method can effectively improve the iteration complexity. Theoretical analysis indicates the fact that the new method preserves the global convergence under some standard assumptions.

Keywords: Cubic regularization; Multi-objective optimization; Unconstrained optimization; Global convergence

1. Introduction

In this paper, we discuss the adaptive cubic regularization method about how to solve an unconstrained non-convex multi-objective problem, where a set of objective functions must be minimized simultaneously.
Actually, we consider the unconstrained optimization problem

$$\begin{array}{l} min \quad F(x) \\ x \in R^n \end{array} \tag{1}$$

Where $F(x)=(F_1(x),F_2(x),\cdots,F_m(x))^T$, $F_i(x): R^n \to R$, $i=1,2,\ldots,m$, and $F_i(x)$ is twice a continuously differentiable. At least one of $F_i(x)$ is a non-convex function.
F is twice continuously differentiable on R^n, for $x \in R^n$, denoted by $\nabla F(x) \in R^{m\times n}$ the Jacobian matrix of the vectorial function F at x, denoted by $\nabla F_i(x) \in R^n$ the gradient vector of the scalar function F_i at x, and by $\nabla^2 F_i(x) \in R^{n\times n}$ the Hessian matrix of F_i at x. Furthermore we can assume that there are positive definite.
Definition 1.1 A point $x^* \in X$ is called Pareto optimal if there is no $x \in X$ such that $F(x) \le F(x^*)$, $F(x) \ne F(x^*)$. If x^* is Pareto optimal, then $F(x^*)$ is called efficient.
Definition 1.2 A point $x^* \in X$ is called weakly Pareto optimal or weakly efficient, if there is no $x \in R^n$ such that $F(x) < F(x^*)$, where the vector strict inequality $F(x) < F(x^*)$ must be understood componentwise sense.
The inequality sign $\le$ between vectors is to be understood in a componentwise sense. And means that searching to minimize point in the partial order induced by the positive orthant R_+^m.
In [1], the Newton direction d_k defined as the optimal solution of

$$\begin{cases} \min\limits_{d\in R_n} \quad \max\limits_{i=1,\cdots m}\left[F_i(x)+\nabla F_i(x)^T d+\frac{1}{2}d^T\nabla^2 F_i(x)d\right] \\ d \in R^n \end{cases} \tag{2}$$

Although (2) is a non-smooth problem, it can be framed as a convex quadratic optimization problem. Actually, the problem (2) is equivalent to

$$\begin{cases} min \quad t \\ F_i(x)+\nabla F_i(x)^T d+\frac{1}{2}d^T\nabla^2 F_i(x)d - t \le 0, i=1,2,\cdots m, \\ (t,d) \in R\times R^n \end{cases} \tag{3}$$

It is well known that many cubic regularization methods have been proposed to solve unconstrained optimization problems ([2-5]). In 2011, Coralia et al. proposed an adaptive cubic regularization method for unconstrained optimization [6]. At each iteration of that method, an approximate global minimizer of a local cubic regularization of the objective function is determined. It can ensures a significant improvement of the objective function only if the Hessian of the objective function remains locally Lipschitz continuous. The proposed cubic model is as follows:

$$m_k(s)=f(x_k)+s^T g_k+\frac{1}{2}s^T B_k s+\frac{1}{3}\sigma_k\|s\|^3 \quad (4)$$

Where $g_k=\nabla f(x_k)$, $\|\cdot\|=\|\cdot\|_2$, σ_k is an adaptive parameter, B_k is symmetric approximation of the Hessian matrix.

The adaptive cubic regularization method generates an approximate global minimum point of a local cubic regularization objective function at each iteration, where the parameter σ_k is used to adjust the degree of approximation of the cubic model and the objective function. The algorithm has good convergence and numerical performance in certain conditions; also it has better algorithmic complexity than the steepest descent method.

Comparing the ARC algorithm with the trust region algorithm, it is found that the trust region algorithm using the quadratic model for solve unconstrained optimization problem, while the ARC algorithm solves unconstrained optimization problem by cubic model. Furthermore, the ARC method and the trust-region algorithm also have some similarities. The adaptive regularization method uses adaptive parameters to realize the role of the trust region radius in the trust region algorithm (see [7] for reference), that is, if the new iteration point does not make the objective function sufficient decrease, the parameter σ_k will be decrease, otherwise, the parameter σ_k is increase.

The outline of this paper is as follows. In Section 2, we describe the adaptive cubic regularization algorithm. In Section 3, we introduce the convergence analysis. Finally, some conclusions are given in Section 4.

2. Adaptive Cubic Regularization Algorithm

To solve the optimization problem (4), Gabriel et al. [8] proposed trust region algorithm and defined a quadratic model for each objective function. In Section 1, we introduce adaptive cubic regularization method for single-objective optimization problem. Here we generalize the adaptive cubic regularization method to the problem of multi-objective optimization. We present the new model as follows:

$$m_k^{\ i}(d)=F_i(x_k)+\nabla F_i(x_k)d+\frac{1}{2}d^T\nabla^2 F_i(x_k)d+\frac{1}{3}\sigma_k\|d\|^3 \quad (5)$$

Where σ_k is an adaptive parameter, $\|\cdot\|=\|\cdot\|_2$.

The actual reduction defined as $F_i(x_k)-F_i(x_k+d_k)$ and the predicted reduction defined as $F_i(x_k)-m_k^{\ i}(d_k)$.

The ratio $r_k^{\ i}$ defined by $r_k^{\ i}=\dfrac{F_i(x_k)-F_i(x_k+d_k)}{F_i(x_k)-m_k^{\ i}(d_k)}$

The new algorithm can be described as follows:

Algorithm 2.1

Given x_0, $1<\gamma_1\le\gamma_2$, $0<\eta_1\le\eta_2\le 1$, and $\sigma_0>0$, set $k=0$.

Step 1. Compute $\nabla F_i(x_k)$, If $\|\nabla F_i(x_k)\|<\varepsilon$, stop. Otherwise, go to Step 2.

Step 2. Compute d_k, satisfying

$$m_k^{\ i}(d_k)\le m_k^{\ i}(d_k^c) \quad (6)$$

Where the Pareto Cauchy point

$d_k^c=-\alpha_k^c\nabla F_i(x_k)$ $\quad \alpha_k^c=arg \quad min\, m_k^i(-\alpha\nabla F_i(x_k))$

Step 3. Compute $F_i(x_k+d_k)$ and

$$r_k^{\ i}=\frac{F_i(x_k)-F_i(x_k+d_k)}{F_i(x_k)-m_k^{\ i}(d_k)}$$

Step 4. Set $x_{k+1}=\begin{cases}x_k+d_k & if\ r_k^{\ i}\ge\eta_1,\forall i.\\ x_k & otherwise.\end{cases}$

Step 5. Set $\sigma_{k+1}=\begin{cases}(0,\sigma_k] & if\quad \eta_2\le r_k^i;\\ (\sigma_k,\gamma_1\sigma_k] & if\quad \eta_1\le r_k^i\le\eta_2;\ \forall i.\\ [\gamma_1\sigma_k,\gamma_2\sigma_k] & if\quad r_k^i<\eta_1;\end{cases}$

Step 6. Set $k:=k+1$, and go to Step 2.

3. Global Convergence Analysis

In this section, we introduce some convergence properties of the new algorithm, and prove the global convergence.

Here are some standard assumptions.

(A1) $F\in(R^n,R^m)$, and $F_i(x)$ is lower bound.

(A2) Suppose that there exist two positive constants κ_1 and κ_2 such that $\|\nabla F_i(x)\|\le\kappa_1$ $\quad \|\nabla^2 F_i(x)\|\le\kappa_2$

(A3) For all k, $m_k^{\ i}(0)=F_i(x_k)$, $\nabla m_k^{\ i}(0)=\nabla F_i(x_k)$

(A4) The matrix H_k is uniformly bounded, that is, there exists a constant $k_{uml}\ge 1$ such that, for all $x\in R^n$, and for all $k:\|H_k\|\le k_{uml}-1$

Lemma 3.1 Assumption the step d_k satisfies (6), then for $k\ge 0$, we have

$$\begin{aligned}F_i(x_k)-m_k^{\ i}(d_k)&\ge F_i(x_k)-m_k^{\ i}(d_k^c)\\ &\ge\frac{\|g_k\|^2}{6\sqrt{2}\,max\left[1+\|B_k\|,2\sqrt{\sigma_k\|g_k\|}\right]}\\ &=\frac{\|g_k\|}{6\sqrt{2}}min\left[\frac{\|g_k\|}{1+\|B_k\|},\frac{1}{2}\sqrt{\frac{\|g_k\|}{\sigma_k}}\right]\end{aligned} \quad (7)$$

Proof. From (6), we can see that the first inequality is true. Next, we prove the second inequality.

For any $\alpha\ge 0$, use the Cauchy-Schwartz inequality, we have

$$m_k^i(d_k^c)-F_i(x_k)\le m_k^i(-\alpha g_k)-F_i(x_k)=$$
$$-\alpha\|\nabla F_i(x_k)\|^2+\frac{1}{2}\alpha^2\nabla F_i(x_k)^T\nabla^2F_i(x_k)\nabla F_i(x_k)+\frac{1}{3}\alpha^3\sigma_k\|\nabla F_i(x_k)\|^3$$
$$\le\alpha\|\nabla F_i(x_k)\|^2\left(-1+\frac{1}{2}\alpha\|\nabla^2F_i(x_k)\|+\frac{1}{3}\alpha^2\sigma_k\|\nabla F_i(x_k)\|\right) \quad (8)$$

According to $m_k^i(d_k^c)\le F_i(x_k)=m_k^i(0)$ and $\alpha\ge 0$, we get $-1+\frac{1}{2}\alpha\nabla^2F_i(x_k)+\frac{1}{3}\alpha^2\sigma_k\|\nabla F_i(x_k)\|\le 0$

The inequality (8) is equivalent to $\alpha\in[0,\alpha_k]$,

Where

$$\alpha_k=\frac{3}{2\sigma_k\|\nabla F_i(x_k)\|}\left(-\frac{1}{2}\|\nabla^2F_i(x_k)\|+\sqrt{\frac{1}{4}\|\nabla^2F_i(x_k)\|^2\frac{4}{3}\sigma_k\|\nabla F_i(x_k)\|}\right)$$

We can define α_k as

$$\alpha_k=2\left(\frac{1}{2}\|\nabla^2F_i(x_k)\|+\sqrt{\frac{1}{4}\|\nabla^2F_i(x_k)\|^2\frac{4}{3}\sigma_k\|\nabla F_i(x_k)\|}\right)^{-1}$$

In addition, we also define β_k as

$$\beta_k=\sqrt{2}\,max\left(1+\|\nabla F_i(x_k)\|,2\sqrt{\sigma_k\|\nabla F_i(x_k)\|}\right)^{-1}$$

The inequalities

$$\sqrt{\frac{1}{4}\|\nabla^2F_i(x_k)\|^2\frac{4}{3}\sigma_k\|\nabla F_i(x_k)\|}\le\frac{1}{2}\|\nabla F_i(x_k)\|+\frac{2}{\sqrt{3}}\sqrt{\sigma_k\|\nabla F_i(x_k)\|}$$
$$\le 2\,max\left(\frac{1}{2}\|\nabla F_i(x_k)\|,\frac{2}{\sqrt{3}}\sqrt{\sigma_k\|\nabla F_i(x_k)\|}\right)$$
$$\le\sqrt{2}\,max\left(1+\|\nabla F_i(x_k)\|,2\sqrt{\sigma_k\|\nabla F_i(x_k)\|}\right)$$

And $\frac{1}{2}\|\nabla F_i(x_k)\|\le max\left(1+\|\nabla F_i(x_k)\|,2\sqrt{\sigma_k\|\nabla F_i(x_k)\|}\right)$

Hence, we obtain $0<\beta_k\le\alpha_k$. We can combine β_k with (8), we get

$$m_k^i(d_k^c)-F_i(x_k)\le\frac{\|\nabla F_i(x_k)\|^2}{\sqrt{2}\,max\left(1+\|\nabla F_i(x_k)\|,2\sqrt{\sigma_k\|\nabla F_i(x_k)\|}\right)}\cdot\left(-1+\frac{1}{2}\beta_k\|\nabla^2F_i(x_k)\|+\frac{1}{3}\beta_k^2\sigma_k\|\nabla F_i(x_k)\|\right) \quad (9)$$

According to definition of β_k, we have $\beta_k\|\nabla^2F_i(x_k)\|\le 1$ and $\beta_k^2\sigma_k\|\nabla F_i(x_k)\|\le 1$. Then,

$$m_k^i(d_k^c)-F_i(x_k)\le\frac{\|\nabla F_i(x_k)\|^2}{\sqrt{2}\,max\left(1+\|\nabla F_i(x_k)\|,2\sqrt{\sigma_k\|\nabla F_i(x_k)\|}\right)}$$
$$\cdot\left(-1+\frac{1}{2}+\frac{1}{3}\right)\le-\frac{\|\nabla F_i(x_k)\|^2}{6\sqrt{2}\,max\left(1+\|\nabla F_i(x_k)\|,2\sqrt{\sigma_k\|\nabla F_i(x_k)\|}\right)}$$
$$F_i(x_k)-m_k^i(d_k^c)\ge\frac{\|\nabla F_i(x_k)\|^2}{6\sqrt{2}\,max\left(1+\|\nabla F_i(x_k)\|,2\sqrt{\sigma_k\|\nabla F_i(x_k)\|}\right)}$$

The proof is complete.

Lemma 3.2 Suppose that (A2) hold, and the step d_k satisfies(6), for all $k\ge 0$, we have $\|d_k\|\le\frac{3}{\sigma_k}max\left(\kappa_1,\sqrt{\sigma_k\|g_k\|}\right)$.

Proof. From the definition of (5), we have

$$m_k^i(d)-F_i(x_k)=d^T\nabla F_i(x_k)+\frac{1}{2}d^T\nabla^2F_i(x_k)d+\frac{1}{3}\sigma_k\|d\|^3$$
$$\ge-\|d\|\|\nabla F_i(x_k)\|-\frac{1}{2}\|d\|\|\nabla^2F_i(x_k)\|+\frac{1}{3}\sigma_k\|d\|^3$$
$$=\left(\frac{1}{9}\sigma_k\|d\|^3-\|d\|\|\nabla F_i(x_k)\|\right)+\left(\frac{2}{9}\sigma_k\|d\|^3-\frac{1}{2}\|d\|\|\nabla^2F_i(x_k)\|\right)$$

If $\|d\|\ge 3\sqrt{\frac{\|\nabla F_i(x_k)\|}{\sigma_k}}$, then $\frac{1}{9}\sigma_k\|d\|^3-\|d\|\|\nabla F_i(x_k)\|\ge 0$, while if $\|d\|\ge\frac{9}{4}\frac{\|\nabla F_i(x_k)\|}{\sigma_k}$, then $\frac{2}{9}\sigma_k\|d\|^3-\frac{1}{2}\|d\|\|\nabla^2F_i(x_k)\|\ge 0$.

Hence, when

$$\|d_k\|>\frac{3}{\sigma_k}max\left(\|\nabla^2F_i(x_k)\|,\sqrt{\sigma_k\|g_k\|}\right)>\frac{3}{\sigma_k}max\left(\kappa_1,\sqrt{\sigma_k\|g_k\|}\right),$$

Then $m_k^i(d)>F_i(x_k)$.

By (7), we get $m_k^i(d_k)\le F_i(x_k)$ and $\|d_k\|\le\frac{3}{\sigma_k}max\left(\kappa_1,\sqrt{\sigma_k\|g_k\|}\right)$.

The proof is the complete.

Lemma 3.3 Let (A1) and (A2) hold, suppose that H is an infinite index set, such that $\|\nabla F_i(x_k)\|\ge\varepsilon$, for all $k\in H$ and some $\varepsilon>0$, and $\sqrt{\frac{\|\nabla F_i(x_k)\|}{\sigma_k}}\to 0$, as $k\to\infty$, $k\in H$, then $\|d_k\|\le 3\sqrt{\frac{\|\nabla F_i(x_k)\|}{\sigma_k}}$ for all $k\in H$ sufficiently large.

Additionally, if $x_k\to x^*$, as $k\in H$, $k\to\infty$ for some $x^*\in R^n$, Then each iteration $k\in H$ that is sufficiently large is very successful, and $\sigma_{k+1}\le\sigma_k$ for all $k\in I$ sufficiently large.

Proof. A proof of this lemma can be observed in [6].

Lemma 3.4 Assume that (A1) and (A2) holds, and the sequence $\{x_k\}$ is generated by Algorithm 2.1, then the sequence $\{F_i(x_k)\}$ is monotonically decreasing and converges.

Proof. See the proof of [10] for reference.

Theorem 3.1 Let (A1) and (A2) holds, suppose furthermore that there are only a finitely many successful iterations. Then $x_k = x^*$, for all k large enough and $\nabla F_i\left(x^*\right) = 0$.

Proof. A proof of this lemma can be observed in [9].

Theorem 3.2 Suppose that (A1) and (A2) hold. Then $\liminf_{k\to\infty} \left\|\nabla F_i\left(x_k\right)\right\| = 0$.

Proof. Suppose that exists $\delta > 0$, for all k, and $i = 1\ldots,m,$ such that $\left\|\nabla F_i\left(x_k\right)\right\| \geq \delta$

Let H is the index set of successful iterations, using

$$F_i\left(x_k\right) - F_i\left(x_{k+1}\right) = r_k^i\left(F_i\left(x_k\right) - m_k^i\left(d_k\right)\right)$$

lemma 3.1, then $\geq \eta_1\left(F_i\left(x_k\right) - m_k^i\left(d_k\right)\right)$

$$\geq \frac{\eta_1\delta}{6\sqrt{2}} min\left(\frac{\delta}{1+\kappa_2}, \frac{1}{2}\sqrt{\frac{\left\|\nabla F_i\left(x_k\right)\right\|}{\sigma_k}}\right)$$

Then, adding all successful interactions up to the k-th

$$F_i\left(x_0\right) - F_i\left(x_{k+1}\right) = \sum_{l=0,l\in H}^{k}\left(F_i\left(x_l\right) - F_i\left(x_{l+1}\right)\right)$$

index, $\geq \sum_{l=0,l\in H}^{k} \frac{\eta_1\delta}{6\sqrt{2}} min\left(\frac{\delta}{1+\kappa_2}, \frac{1}{2}\sqrt{\frac{\left\|\nabla F_i\left(x_k\right)\right\|}{\sigma_k}}\right)$

$$\geq \theta_k \frac{\eta_1\delta}{6\sqrt{2}} min\left(\frac{\delta}{1+\kappa_2}, \frac{1}{2}\sqrt{\frac{\left\|\nabla F_i\left(x_k\right)\right\|}{\sigma_k}}\right)$$

Using lemma 3.3, we obtain $\sqrt{\frac{\left\|\nabla F_i\left(x_k\right)\right\|}{\sigma_k}} \to 0$

$k \to \infty, k \in H$

θ_k is the number of successful iterations. Then if $\lim_{k\to\infty,k\in H} \theta_k < +\infty$

From the theorem 3.1, we know that the conclusion is true. If $\lim_{k\to\infty,k\in H} \theta_k = +\infty$

Since $F_i\left(x\right)$ is monotonically decreasing, and $F_i\left(x\right)$ is bounded below. Then, we have $F_i\left(x\right)$ is convergent. Hence, it is contract with the $F_i\left(x\right)$. The proof is completed.

4. Conclusions

In this paper, we propose a new adaptive cubic regularization algorithm for multi-objective optimization problem. The new algorithm uses an adaptive estimation of the local Lipschitz constant, and an approximation of the global model minimization, which is computationally feasible even for large-scale problems. The aim is that of improving the computational efficiency of the cubic regularization method preserving the global convergence.

References

[1] Fliege J., Drummond L. M., Graña B. F. Svaiter. Newton's Method for Multi-objective Optimization. SIAM J. Optim, 2009, 20, 602–626.

[2] Toint ph.L. Nonlinear Stepsize Control, Trust Regions and Regularizations for Unconstrained Optimization. Optimization Methods & Software, 2013, 28 (1), 82-95.

[3] Cartis C., Gould N. I. M., Toint Ph. L. Evaluation Complexity of Adaptive Cubic Regularization Methods for Convex Unconstrained Optimization. Optimization Methods &Software, 2012, 27(2),197-219.

[4] Griewank A., Fisher J., Bosse T. Cubic Overestimation and Secant Updating for Unconstrained Optimization of C2,1 functions. Optimization Method & Software, 2014; 29(5).

[5] Martinez J.M., Raydan M. Separable Cubic Modeling and a Trust-region Strategy for Unconstrained Minimization with Impact in Global Optimization. Kluwer Academic Publishers, 2015, 63 (2), 319-342.

[6] Cartis C., Gould N. I. M., Toint Ph.L. Adaptive Cubic Regularization Methods for Unconstrained Optimization. Part I: Motivation, Convergence and Numerical Results. Mathematical Programming, 2011, 127 (2), 245-295.

[7] Conn A. R., Gould N. I. M, Toint Ph.L. Trust-Region Methods. SIAM, Philadelphia, 2000.

[8] Gabriel A., Pablo A., Maria C. Trust Region Globalization Strategy for the Nonconvex Unconstrained Multi-objective Optimization Problem. Math. Program, Ser. 2016, 159, 339-369.

[9] Cartis C., Gould N. I. M., Toint Ph.L. Adaptive Cubic Regularization Methods for Unconstrained Optimization. Part II: Worst-case Function-and Derivative-evaluation Complexity, Mathematical Programming, 2011, 130 (2), 295-319.

[10] Bianconcini T., Sciandrone M. A Cubic Regularization Algorithm for Unconstrained Optimization Using Line Search and Nonmonotone Techniques. Taylor &Francis, Inc, 2016, 31(5).