AMO - Advanced Modeling and Optimization, Volume 17, Number 1, 2015

An Adaptive Conic Cubic Overestimation Method for Unconstrained Optimization

Lijuan Zhao

Abstract

In this paper, we propose an Adaptive Conic Cubic Overestimation (ACCO) method for unconstrained optimization. ACCO model not only is an extension of conic model which was first proposed by Davidon and Sorensen, but also is a generalization of Adaptive Cubic Overestimation (ACO) model which was initialed by Cartis, Gould and Toint. Global convergence to first order critical point and local linear convergence are proved under some mild conditions. Furthermore, the worst case global iteration complexity of the ACCO method is analyzed in this paper.

Key words. adaptive conic cubic overestimation model, adaptive cubic overestimation model, iteration complexity, trust region method, conic model, unconstrained optimization

1. Introduction

In this paper, we consider the following unconstrained optimization problem

$$\min_{x\in R^n} \quad f(x), \tag{1.1}$$

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

After the two kinds of globalization method: trust region method (see [1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 17, 18]) and line search method (see [8, 12, 13, 14, 15, 16]) for problem (1.1). An Adaptive Cubic Overestimation (short for ACO) method has been proposed in [19] as an alternative globalization method for problem (1.1). Assume the Hessian $H(x_k) \overset{def}{=} \nabla^2 f(x_k)$ of the objective function is globally Lipschitz continuous, the Lipschitz constant is L, then

$$f(x_k+s) \le f(x_k) + g(x_k)^T s + \frac{1}{2}s^T H(x_k)s + \frac{L}{6}\|s\|^3 \overset{def}{=} m_k^c(s),\ \forall s \in R^n. \tag{1.2}$$

In [20], Cartis et al don't insist on the Lipschitz continuous condition of $H(x)$, they replaced the Hessian $H(x_k)$ by a symmetric approximation matrix B_k on each iteration, instead of (1.2), they proposed a more general ACO model as follows

$$m_k(s) \overset{def}{=} f(x_k) + g(x_k)^T s + \frac{1}{2}s^T B_k s + \frac{\sigma_k}{3}\|s\|^3, \tag{1.3}$$

This work was supported by the project of Nanjing Vocational Institute of Railway Technology (grant No. Y14031).

Department of Social Science Teaching, Nanjing Vocational Institute of Railway Technology, Nanjing, 210031, China. Email: zzlljj210@163.com

Advanced Modeling and Optimization. ISSN: 1841-4311

where σ_k is a dynamic positive parameter adjusted by trust region method. It can be regarded as the reciprocal of the trust region radius. It also relaxes the need of computing a global minimizer of cubic model and the availability of Lipschitz constant. The cubic overestimation model is first introduced by Nesterov and Polyak in [21] for Newton's method with a provable better global complexity bound. ACO method retains the excellent convergence properties and the worst-case iteration complexity bound. Recently, many authors have done some work on this field, see [29, 30, 31, 32, 33] etc.

Note that the first three items of (1.3) forms a quadratic model. It is well-known that conic model which was first proposed by Davidon and Sorensen (see [22, 23]) is an extension of quadratic model, and has some advantages compared with quadratic model. Firstly, if the objective function has strong non-quadratic behavior or it's curvature changes severely, the quadratic model often produces a poor prediction of the minimizer of the function. In this case, the conic model approximates the objective function value better than the quadratic model, because it has more freedom in the model. Secondly, the quadratic model does not take into account the information concerning the function values and gradient values in the previous iterations which is useful for algorithms. However, for unconstrained optimization, the conic model possesses richer interpolation information and satisfies four interpolation conditions of the function values and gradient values at the current and the previous point. Using the conic model may improve the performance of the algorithms. Thirdly, the initial and limited numerical results show that the conic model method gives an improvement over the quadratic model method. Finally, the conic model method has the similar global and local convergence properties as the quadratic model. Since 1980s, a lot of scholars have been taking great interest in conic model and extended it to constrained optimization, see [5, 24, 25, 26, 27, 28] etc.

In this paper, we propose an Adaptive Conic Cubic Overestimation (short for ACCO) model for unconstrained problem (1.1) as follows

$$c_k(s) \overset{def}{=} f(x_k) + \frac{g(x_k)^T s}{1 + h_k^T s} + \frac{s^T B_k s}{2(1 + h_k^T s)^2} + \frac{\sigma_k}{3}\|s\|^3, \tag{1.4}$$

where $h_k \in R^n$ is a horizontal vector. When $h_k = 0$, an ACCO model is an ACO model; When $h_k = 0$ and $\sigma_k = 0$, an ACCO model is a quadratic model. So we can regard an ACCO model not only as an extension of an ACO model, but also is a generalization of a quadratic model. Because an ACCO model incorporates more information which is useful for algorithms, we believe an ACCO model outperforms an ACO model and a quadratic model.

The paper is organized as follows. In Section 2, we give our ACCO method and algorithm for unconstrained optimization. Section 3 discusses the optimality conditions of the ACCO model. Section 4 shows the global convergence of ACCO algorithm to first order critical point and local linear convergence under some mild conditions. The global iteration complexity is given in Section 5. Section 6 concludes.

Notation: $\|\cdot\|$ is Euclidean norm.

2. Adaptive Conic Cubic Overestimation Method and Algorithm

We consider the following adaptive conic cubic overestimation model of $f(x)$

around x_{k+1}

$$c(x_{k+1}+s) = f(x_{k+1}) + \frac{g_{k+1}^T s}{1+h_{k+1}^T s} + \frac{s^T B_{k+1} s}{2(1+h_{k+1}^T s)^2} + \frac{\sigma_{k+1}}{3}\|s\|^3, \tag{2.1}$$

where $g_{k+1} \stackrel{def}{=} \nabla f(x_{k+1})$ is the gradient of the objective function at x_{k+1}, B_{k+1} is a $n \times n$ symmetric matrix, which is the Hessian of the objective function or it's approximation. The adaptive conic cubic overestimation model $c(x_{k+1}+s)$ should satisfy the following interpolation conditions

$$c(x_{k+1}) = f(x_{k+1}), \quad \nabla c(x_{k+1}) = g_{k+1}, \tag{2.2}$$

$$c(x_k) = f(x_k), \qquad \nabla c(x_k) = g_k. \tag{2.3}$$

The gradient of $c(x_{k+1}+s)$ about s is

$$\nabla c(x_{k+1}+s) = \frac{1}{1+h_{k+1}^T s}\Big(I - \frac{h_{k+1}s^T}{1+h_{k+1}^T s}\Big)\Big(g_{k+1} + \frac{B_{k+1}s}{1+h_{k+1}^T s}\Big) + \sigma_{k+1}\|s\|s. \tag{2.4}$$

Set $s_k = x_{k+1} - x_k$. The first interpolation condition in (2.3) leads to

$$f(x_k) = f(x_{k+1}) - \frac{g_{k+1}^T s_k}{1-h_{k+1}^T s_k} + \frac{s_k^T B_{k+1} s_k}{2(1-h_{k+1}^T s_k)^2} + \frac{\sigma_{k+1}}{3}\|s_k\|^3. \tag{2.5}$$

Set $\gamma_k = 1 - h_{k+1}^T s_k$, then we get

$$2[3(f(x_{k+1}) - f(x_k)) + \sigma_{k+1}\|s_k\|^3]\gamma_k^2 - 6(g_{k+1}^T s_k)\gamma_k + 3 s_k^T B_{k+1} s_k = 0. \tag{2.6}$$

The second interpolation condition in (2.3) leads to

$$\frac{1}{1-h_{k+1}^T s_k}\Big(I + \frac{h_{k+1}s_k^T}{1-h_{k+1}^T s_k}\Big)\Big(g_{k+1} - \frac{B_{k+1}s_k}{1-h_{k+1}^T s_k}\Big) - \sigma_{k+1}\|s_k\|s_k = g_k, \tag{2.7}$$

thus we get that

$$\begin{aligned} B_{k+1}s_k &= \gamma_k g_{k+1} - \gamma_k^2\Big(I + \frac{h_{k+1}s_k^T}{\gamma_k}\Big)^{-1}(g_k + \sigma_{k+1}\|s_k\|s_k) \\ &= \gamma_k g_{k+1} - \gamma_k^2(I - h_{k+1}s_k^T)(g_k + \sigma_{k+1}\|s_k\|s_k) \\ &= \gamma_k g_{k+1} - \gamma_k^2(g_k + \sigma_{k+1}\|s_k\|s_k - s_k^T g_k h_{k+1} - \sigma_{k+1}\|s_k\|^3 h_{k+1}) \\ &\stackrel{def}{=} y_k \end{aligned} \tag{2.8}$$

where, usually, B_{k+1} is usually updated by BFGS formula

$$B_{k+1} = B_k + \frac{y_k y_k^T}{s_k^T y_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} \tag{2.9}$$

and $y_k = \gamma_k g_{k+1} - \gamma_k^2(g_k + \sigma_{k+1}\|s_k\|s_k - s_k^T g_k h_{k+1} - \sigma_{k+1}\|s_k\|^3 h_{k+1})$. If B_0 is positive definite, the BFGS updating formula keeps the positive definiteness of B_k. From (2.8), we obtain

$$s_k^T B_{k+1} s_k = \gamma_k s_k^T g_{k+1} - \gamma_k^2(s_k^T g_k + \sigma_{k+1}\|s_k\|^3 - s_k^T g_k s_k^T h_{k+1} - \sigma_{k+1}\|s_k\|^3 s_k^T h_{k+1}). \tag{2.10}$$

Combining (2.6) and (2.10) yields that

$$3(\sigma_{k+1}\|s_k\|^3+s_k^Tg_k)\gamma_k^2-2[3(f(x_{k+1})-f(x_k))+\sigma_{k+1}\|s_k\|^3]\gamma_k+3g_{k+1}^Ts_k=0, \quad (2.11)$$

which is a quadratic equation about γ_k. The equality (2.11) has real roots if and only if

$$\rho=[3(f(x_{k+1})-f(x_k))+\sigma_{k+1}\|s_k\|^3]^2-9(\sigma_{k+1}\|s_k\|^3+s_k^Tg_k)g_{k+1}^Ts_k\geq 0. \quad (2.12)$$

Taking $\rho:=\sqrt{\rho}$ gives

$$\gamma_k = \frac{\sigma_{k+1}\|s_k\|^3+3(f(x_{k+1})-f(x_k))-\rho}{3(\sigma_{k+1}\|s_k\|^3+s_k^Tg_k)} \quad (2.13)$$

$$= \frac{9(\sigma_{k+1}\|s_k\|^3+s_k^Tg_k)g_{k+1}^Ts_k}{\sigma_{k+1}\|s_k\|^3+3(f(x_{k+1})-f(x_k))+\rho}. \quad (2.14)$$

Some special cases satisfying $\gamma_k=1-h_{k+1}^Ts_k$ are

$$h_{k+1}=\frac{1-\gamma_k}{g_k^Ts_k}g_k, \quad (2.15)$$

$$h_{k+1}=\frac{1-\gamma_k}{s_k^Ts_k}s_k, \quad (2.16)$$

$$h_{k+1}=(1-\gamma_k)\Big[\frac{\beta g_k}{g_k^Ts_k}+\frac{(1-\beta)s_k}{s_k^Ts_k}\Big],\ \forall\beta\in R. \quad (2.17)$$

Next, we formally state our Adaptive Conic Cubic Overestimation method (abbreviated as ACCO) as follows.

Algorithm 2.1 $\{ACCO\}$

Step 0. *Given* x_0, $\gamma_2\geq\gamma_1>1$, $1>\eta_2\geq\eta_1>0$, $\sigma_0>0$, $\epsilon>0$, $B_0>0$. *Set* $k:=0$, *compute* $g_0=\nabla f(x_0)$, $f(x_0)$.

Step 1. *If* $\|g_k\|<\epsilon$, *stop with approximate solution* x_k.

Step 2. *Solve* $c_k(s)$ *inaccurately so that the step* s_k *satisfies*

$$c_k(s_k)\leq c_k(s_k^C), \quad (2.18)$$

where the Cauchy point

$$s_k^C=-\alpha_k^Cg_k, \quad \text{and} \quad \alpha_k^C=\arg\min_{\alpha\in R^+}c_k(-\alpha g_k). \quad (2.19)$$

Step 3. *Calculate* $\rho_k=\frac{f(x_k)-f(x_k+s_k)}{c_k(0)-c_k(s_k)}$. *If* $\rho_k\geq\eta_1$, *then* $x_{k+1}=x_k+s_k$. *Otherwise, set* $x_{k+1}=x_k$.

Step 4. *Set*

$$\sigma_{k+1}\in\begin{cases}(0,\sigma_k] & \text{if } \rho_k>\eta_2; \quad \text{[very successful iteration]}\\ [\sigma_k,\gamma_1\sigma_k] & \text{if } \eta_1\leq\rho_k\leq\eta_2; \quad \text{[successful iteration]}\\ [\gamma_1\sigma_k,\gamma_2\sigma_k] & \text{otherwise.} \quad \text{[unsuccessful iteration]}\end{cases}$$

Step 5. Compute $f(x_{k+1})$ and g_{k+1}. Set

$$\begin{aligned}\rho_k &= [3(f(x_{k+1})-f(x_k))+\sigma_{k+1}\|s_k\|^3]^2-9(\sigma_{k+1}\|s_k\|^3 \\ &\quad +s_k^T g_k)g_{k+1}^T s_k. \qquad (2.20)\end{aligned}$$

$$\begin{aligned}\gamma_k &= \frac{\sigma_{k+1}\|s_k\|^3+3(f(x_{k+1})-f(x_k))-\sqrt{\rho_k}}{3(\sigma_{k+1}\|s_k\|^3+s_k^T g_k)} \qquad (2.21)\\ &= \frac{9(\sigma_{k+1}\|s_k\|^3+s_k^T g_k)g_{k+1}^T s_k}{\sigma_{k+1}\|s_k\|^3+3(f(x_{k+1})-f(x_k))+\sqrt{\rho_k}}. \qquad (2.22)\end{aligned}$$

$$y_k=\gamma_k g_{k+1}-\gamma_k^2(g_k+\sigma_{k+1}\|s_k\|s_k-s_k^T g_k h_{k+1}-\sigma_{k+1}\|s_k\|^3 h_{k+1}). \qquad (2.23)$$

$$h_{k+1}=(1-\gamma_k)\Big[\frac{\beta g_k}{g_k^T s_k}+\frac{(1-\beta)s_k}{s_k^T s_k}\Big],\ \forall\beta\in R. \qquad (2.24)$$

Update B_k to obtain B_{k+1} by BFGS formula, or symmetric-Broyden formula, or symmetric rank-one formula. Set $k:=k+1$, go to Step 1.

Remark:

(1). In Algorithm 2.1, σ_k is a dynamic positive parameter that may be regarded as the reciprocal of the trust region radius. It relaxes the need of computing a global minimizer of cubic model and the availability of Lipschitz constant.

(2). For the global convergence and iteration complexity analysis, we assume that there exist two positive constants $0<c_1<\frac{1}{2}$ and $c_2>1$ such that

$$c_1\le 1+h_k^T s\le c_2,\ \forall s\in R^n. \qquad (2.25)$$

The constants c_1 has two purposes: the first one is to prevent the denominator $1+h_k^T s$ from being too small such that the model function might be unbounded. The second one is the need of the global convergence analysis and iteration complexity analysis. Perhaps there are some constants c_2 satisfying $0<c_2<1$, for the convenience of analysis, we choose a larger c_2 such that $c_2>1$.

(3). For the complexity analysis, we require that there exists a constant $h_{max}>0$ such that

$$\|h_k\|\le h_{max},\ \forall k. \qquad (2.26)$$

3. Optimality conditions for the Adaptive Conic Cubic Overestimation Method

Define the adaptive conic cubic overestimation model as follows

$$c(s)=f+\frac{g^T s}{1+h^T s}+\frac{s^T B s}{2(1+h^T s)^2}+\frac{\sigma}{3}\|s\|^3. \qquad (3.1)$$

The following lemma gives the optimality conditions of adaptive conic cubic overestimation model, which is an extension of Lemma 1 in [24] and theorem 3.1 in [19]. For simplicity, we drop the subscript.

Lemma 3.1 *The gradient of the function expressed by (3.1) is*

$$\nabla c(s) = \frac{1}{\nu^2}(I + hs^T)^{-1}(\nu g + Bs) + \lambda s \tag{3.2}$$

$$= \frac{1}{\nu}\Big(I - \frac{hs^T}{\nu}\Big)\Big(g + \frac{Bs}{\nu}\Big) + \lambda s \tag{3.3}$$

$$= \frac{1}{\nu^3}(\nu I - hs^T)(\nu g + Bs) + \lambda s, \tag{3.4}$$

where $\nu = 1 + h^T s$, $\lambda = \sigma\|s\|$. *This gradient vanishes at* s_*, *i.e.,* s_* *is a critical point of* $c(s)$, *if and only if*

$$(B + \nu_*^2\lambda_* I + gh^T + \nu_*^2\lambda_* hs_*^T)s_* = -g. \tag{3.5}$$

The value of $c(s)$ *is the same at all critical points* s_* *and equals*

$$c(s_*) = f + \frac{g^T s_* - \nu_*^2\lambda_*\|s_*\|^2}{2\nu_*} + \frac{\sigma}{3}\|s_*\|^3 \tag{3.6}$$

$$= f - \frac{s_*^T B s_* + 2\nu_*^3\lambda_*\|s_*\|^2}{2\nu_*^2} + \frac{\sigma}{3}\|s_*\|^3, \tag{3.7}$$

where $\nu_* = 1 + h^T s_*$, $\lambda_* = \sigma\|s_*\|$. *The Hessian of* $c(s)$ *at* s_* *is*

$$\begin{aligned}
&\frac{1}{\nu_*^2}(B + hg^T + gh^T + 2(f - c(s_*))hh^T) + \lambda_* I + \lambda_*\Big(\frac{s_*}{\|s_*\|}\Big)\Big(\frac{s_*}{\|s_*\|}\Big)^T \\
= \; &\frac{1}{\nu_*^2}(I + hs_*^T)^{-1}B(I + s_*h^T)^{-1} + \lambda_* I + \lambda_*\Big(\frac{s_*}{\|s_*\|}\Big)\Big(\frac{s_*}{\|s_*\|}\Big)^T \\
= \; &\frac{1}{\nu_*^4}(\nu_* I - hs_*^T)B(\nu_* I - s_*h^T) + \lambda_* I + \lambda_*\Big(\frac{s_*}{\|s_*\|}\Big)\Big(\frac{s_*}{\|s_*\|}\Big)^T.
\end{aligned}$$

A critical point is a global minimizer if and only if $B \geq 0$ *(positive semidefinite), it is unique if and only if* $B > 0$ *(positive definite).*

4. Convergence analysis

In this section, we will discuss the convergence properties of Algorithm 2.1. The following assumptions are required.

Assumption 4.1

A1. $f(x) : R^n \to R$ is twice continuously differentiable and bounded below by f_{low}.

A2. There exists a positive constant k_B such that $\|B_k\| \leq k_B$.

A3. $g(x)$ is Lipschitz continuous, i.e.,

$$\|g(x) - g(y)\| \leq k_H\|x - y\|, \; \forall\, x, y \in R^n, \text{ and some } k_H \geq 1.$$

Assumption A1 implies that there exists a constant $c_3 > 0$ such that

$$\|g(x)\| \leq c_3. \tag{4.1}$$

The next lemma shows that the adaptive conic cubic overestimation model $c_k(s)$ descends sufficiently, and gives a lower bound of the decrease of the model function $c_k(s)$.

Lemma 4.1 *Suppose that s_k satisfies the conditions (2.18)–(2.19). Then for all $k \geq 0$, we have that*

$$c_k(0) - c_k(s_k) \geq c_k(0) - c_k(s_k^C) \geq \frac{\|g_k\|}{3(\sqrt{2}+1)c_2^2} \min\Big\{\frac{c_1^2\|g_k\|}{\|B_k\|}, \frac{1}{2}\sqrt{\frac{c_2\|g_k\|}{\sigma_k}}\Big\}. \tag{4.2}$$

Proof. From (2.25), we know

$$c_1 \leq 1 - \alpha h_k^T g_k \leq c_2, \forall \alpha > 0. \tag{4.3}$$

Then

$$\begin{aligned} c_k(s_k^C) - c_k(0) &\leq c_k(-\alpha g_k) - c_k(0) \\ &= \frac{-\alpha\|g_k\|^2}{1-\alpha h_k^T g_k} + \frac{\alpha^2 g_k^T B_k g_k}{2(1-\alpha h_k^T g_k)^2} + \frac{\alpha^3\sigma_k}{3}\|g_k\|^3 \\ &\leq \alpha\|g_k\|^2\Big(-\frac{1}{1-\alpha h_k^T g_k} + \frac{\alpha\|B_k\|}{2(1-\alpha h_k^T g_k)^2} + \frac{\alpha^2\sigma_k}{3}\|g_k\|\Big) \\ &\leq \alpha\|g_k\|^2\Big(-\frac{1}{c_2} + \frac{\alpha\|B_k\|}{2c_1^2} + \frac{\alpha^2\sigma_k}{3}\|g_k\|\Big). \end{aligned} \tag{4.4}$$

So $c_k(s_k^C) \leq c_k(0)$ if and only if

$$-\frac{1}{c_2} + \frac{\alpha\|B_k\|}{2c_1^2} + \frac{\alpha^2\sigma_k}{3}\|g_k\| \leq 0 \quad \text{and} \quad \alpha > 0. \tag{4.5}$$

So $\alpha \in [0, \overline{\alpha_k}]$, where $\overline{\alpha_k} = \frac{3}{2\sigma_k\|g_k\|}\Big(-\frac{\|B_k\|}{2c_1^2} + \sqrt{\frac{\|B_k\|^2}{4c_1^4} + \frac{4\sigma_k\|g_k\|}{3c_2}}\Big)$. By rationalizing the numerators of $\overline{\alpha_k}$, we have

$$\overline{\alpha_k} = \frac{2}{c_2}\Big(\frac{\|B_k\|}{2c_1^2} + \sqrt{\frac{\|B_k\|^2}{4c_1^4} + \frac{4\sigma_k\|g_k\|}{3c_2}}\Big)^{-1}.$$

Define $\theta_k = \Big[\frac{c_2}{\sqrt{2}-1}\max\Big\{\frac{\|B_k\|}{c_1^2}, \frac{2\sqrt{\sigma_k\|g_k\|}}{\sqrt{c_2}}\Big\}\Big]^{-1}$, it follows that

$$\begin{aligned} \sqrt{\frac{\|B_k\|^2}{4c_1^4} + \frac{4\sigma_k\|g_k\|}{3c_2}} &\leq \frac{\|B_k\|}{2c_1^2} + \frac{2\sqrt{\sigma_k\|g_k\|}}{\sqrt{3c_2}} \\ &\leq 2\max\Big\{\frac{\|B_k\|}{2c_1^2}, \frac{2\sqrt{\sigma_k\|g_k\|}}{\sqrt{3c_2}}\Big\} \\ &\leq \sqrt{2}\max\Big\{\frac{\|B_k\|}{c_1^2}, \frac{2\sqrt{\sigma_k\|g_k\|}}{\sqrt{c_2}}\Big\} \end{aligned} \tag{4.6}$$

by using of $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}\,(a, b > 0)$. Noticing

$$\frac{\|B_k\|}{2c_1^2} \leq \max\Big\{\frac{\|B_k\|}{c_1^2}, \frac{2\sqrt{\sigma_k\|g_k\|}}{\sqrt{c_2}}\Big\}. \tag{4.7}$$

Combining (4.6) and (4.7) yields that

$$\Big(\frac{\|B_k\|}{2c_1^2} + \sqrt{\frac{\|B_k\|^2}{4c_1^4} + \frac{4\sigma_k\|g_k\|}{3c_2}}\Big)^{-1} \geq \frac{1}{\sqrt{2}+1}\Big(\max\Big\{\frac{\|B_k\|}{c_1^2}, \frac{2\sqrt{\sigma_k\|g_k\|}}{\sqrt{c_2}}\Big\}\Big)^{-1},$$

so we can easily get that $\overline{\alpha_k} \geq \theta_k \geq 0$. From (4.4), we obtain

$$c_k(s_k^C) - c_k(0) \leq \theta_k\|g_k\|^2\Big(-\frac{1}{c_2} + \frac{\theta_k\|B_k\|}{2c_1^2} + \frac{\theta_k^2\sigma_k\|g_k\|}{3}\Big). \tag{4.8}$$

In view of

$$\theta_k\|B_k\| = \frac{(\sqrt{2}-1)\|B_k\|}{c_2\max\Big\{\frac{\|B_k\|}{c_1^2}, \frac{2\sqrt{\sigma_k\|g_k\|}}{\sqrt{c_2}}\Big\}} \leq \frac{c_1^2}{c_2}, \tag{4.9}$$

$$\begin{aligned}\theta_k^2\sigma_k\|g_k\| &= \frac{\sigma_k\|g_k\|}{(\frac{c_2}{\sqrt{2}-1})^2\Big[\max\Big\{\frac{\|B_k\|}{c_1^2}, \frac{2\sqrt{\sigma_k\|g_k\|}}{\sqrt{c_2}}\Big\}\Big]^2} \\ &\leq \frac{\sigma_k\|g_k\|}{\frac{c_2^2}{2}\max\Big\{\frac{\|B_k\|^2}{c_1^4}, \frac{4\sigma_k\|g_k\|}{c_2}\Big\}} \\ &\leq \frac{1}{2c_2}.\end{aligned} \tag{4.10}$$

So $c_k(s_k^C) - c_k(0) \leq -\frac{1}{3c_2}\theta_k\|g_k\|^2$, thus

$$c_k(0) - c_k(s_k^C) \geq \frac{\|g_k\|}{3(\sqrt{2}+1)c_2^2}\min\Big\{\frac{c_1^2\|g_k\|}{\|B_k\|}, \frac{1}{2}\sqrt{\frac{c_2\|g_k\|}{\sigma_k}}\Big\}.$$

From (2.18), we know that

$$c_k(0) - c_k(s_k) \geq c_k(0) - c_k(s_k^C) \geq \frac{\|g_k\|}{3(\sqrt{2}+1)c_2^2}\min\Big\{\frac{c_1^2\|g_k\|}{\|B_k\|}, \frac{1}{2}\sqrt{\frac{c_2\|g_k\|}{\sigma_k}}\Big\}. \tag{4.11}$$

□

The following lemma gives an upper bound of the norm of step s_k, which is an extension of Lemma 2.2 in [19].

Lemma 4.2 *Suppose that s_k satisfies the conditions (2.18)–(2.19). Then*

$$\|s_k\| \leq \frac{3}{\sigma_k}\max\Big\{\frac{\|B_k\|}{c_2^2}, \sqrt{\frac{\sigma_k\|g_k\|}{c_2}}\Big\}. \tag{4.12}$$

Proof. From (2.25) and Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}c_k(s_k)-c_k(0) &= \frac{g_k^Ts_k}{1+h_k^Ts_k}+\frac{s_k^TB_ks_k}{2(1+h_k^Ts_k)^2}+\frac{\sigma_k}{3}\|s_k\|^3\\ &\geq -\frac{\|s_k\|\|g_k\|}{c_2}-\frac{\|B_k\|\|s_k\|^2}{2c_2^2}+\frac{\sigma_k}{3}\|s_k\|^3\\ &\geq \Big(\frac{2}{9}\sigma_k\|s_k\|^3-\frac{\|B_k\|\|s_k\|^2}{2c_2^2}\Big)+\Big(\frac{1}{9}\sigma_k\|s_k\|^3-\frac{\|s_k\|\|g_k\|}{c_2}\Big)\end{aligned}$$

When $\|s_k\|\geq \frac{3}{\sigma_k}\max\Big\{\frac{\|B_k\|}{c_2^2},\sqrt{\frac{\sigma_k\|g_k\|}{c_2}}\Big\}$, we have

$$\frac{2}{9}\sigma_k\|s_k\|^3-\frac{\|B_k\|\|s_k\|^2}{2c_2^2}\geq 0,\quad \frac{1}{9}\sigma_k\|s_k\|^3-\frac{\|s_k\|\|g_k\|}{c_2}\geq 0.$$

So $c_k(s_k)\geq c_k(0)$. However, from Lemma 4.1, we obtain $c_k(s_k)\leq c_k(0)$, so

$$\|s_k\|\leq\frac{3}{\sigma_k}\max\Big\{\frac{\|B_k\|}{c_2^2},\sqrt{\frac{\sigma_k\|g_k\|}{c_2}}\Big\}.$$

We complete the proof. □

The next lemma shows that under some conditions, the iterate becomes very successful, which is an extension of Lemma 2.3 in [19].

Lemma 4.3 *Suppose that Assumptions A2 and A3 hold, and that there exists a constant* $\epsilon>0$ *such that* $\|g_k\|\geq\epsilon$, *and* $\lim_{k\to\infty}\sqrt{\frac{\|g_k\|}{\sigma_k}}=0$. *Then for sufficiently large* k, *the* k*-th iteration is very successful, i.e.,* $\rho_k>\eta_2$, $\sigma_{k+1}\leq\sigma_k$.

Proof. $\sqrt{\frac{\sigma_k\|g_k\|}{c_2}}=\|g_k\|\sqrt{\frac{\sigma_k}{c_2\|g_k\|}}\geq\frac{\epsilon}{\sqrt{c_2}}\sqrt{\frac{\sigma_k}{\|g_k\|}}\to\infty$, so Lemma 4.2 yields that $\|s_k\|\leq 3\sqrt{\frac{\|g_k\|}{c_2\sigma_k}}$, thus $\|s_k\|\to 0\,(k\to\infty)$. In order to show that the k-th iteration is very successful, i.e., $\rho_k\geq\eta_2$, we need to prove $\frac{f(x_k)-f(x_k+s_k)}{c_k(0)-c_k(s_k)}>\eta_2$. From Lemma 4.1, we have $c_k(0)>c_k(s_k)$. That is to say, we need to show $f(x_k+s_k)-c_k(s_k)+(1-\eta_2)(c_k(s_k)-c_k(0))<0$. Define $r_k=f(x_k+s_k)-c_k(s_k)+(1-\eta_2)(c_k(s_k)-c_k(0))$. Next, we prove that $r_k<0$. From Lemma 4.2, we have

$$\begin{aligned}f(x_k+s_k)-c_k(s_k) &= g(\xi_k)^Ts_k-\frac{g_k^Ts_k}{1+h_k^Ts_k}-\frac{s_k^TB_ks_k}{2(1+h_k^Ts_k)^2}-\frac{\sigma_k}{3}\|s_k\|^3\\ &\leq \frac{1}{c_1}\Big(\|g(\xi_k)-g_k\|\|s_k\|+\|g(\xi_k)\|\|s_k\||h_k^Ts_k|+\frac{k_B\|s_k\|^2}{2c_1}\Big)\\ &\leq \frac{3}{c_1}\Big[\|g(\xi_k)-g_k\|+c_3\max\{|c_1-1|,|c_2-1|\}\\ &\quad+\frac{3k_B}{2c_1}\sqrt{\frac{\|g_k\|}{c_2\sigma_k}}\Big]\sqrt{\frac{\|g_k\|}{c_2\sigma_k}},\end{aligned}\tag{4.13}$$

where ξ_k lies on the segment $[x_k,x_k+s_k]$. Lemma 4.1 and $\lim_{k\to\infty}\sqrt{\frac{\|g_k\|}{\sigma_k}}=0$ yields that

$$c_k(0)-c_k(s_k)\ \geq\ \frac{\epsilon}{3(\sqrt{2}+1)c_2^2}\min\Big\{\frac{c_1^2\epsilon}{k_B},\frac{\sqrt{c_2}}{2}\sqrt{\frac{\|g_k\|}{\sigma_k}}\Big\}$$

$$\geq \quad \frac{\epsilon}{6(\sqrt{2}+1)c_2^{\frac{3}{2}}}\sqrt{\frac{\|g_k\|}{\sigma_k}}. \tag{4.14}$$

Combining (4.13) and (4.14) results in

$$r_k \leq \frac{3}{c_1}\Big[\|g(\xi_k)-g_k\|+c_3\max\{|c_1-1|,|c_2-1|\}+\frac{3k_B}{2c_1}\sqrt{\frac{\|g_k\|}{c_2\sigma_k}}-\frac{(1-\eta_2)\epsilon}{6(\sqrt{2}+1)c_2}\Big]\sqrt{\frac{\|g_k\|}{c_2\sigma_k}}.$$

Because $g(x)$ is Lipschitz continuous and $\|s_k\| \to 0\,(k \to \infty)$, so $\|g(\xi_k) - g_k\| \to 0\,(k \to \infty)$, thus

$$r_k < 0, \text{ when } k \text{ is sufficiently large.}$$

So $\rho_k > \eta_2$, i.e., the k-th iteration is successful, so $\sigma_{k+1} \leq \sigma_k$. □

The following lemma shows that Algorithm 2.1 is well-definied, i.e., the inner iteration can't cycle infinitely.

Lemma 4.4 *Suppose that Assumptions A2 and A3 hold. Then Algorithm 2.1 is well-definied, i.e., the iteration between Step 1 and Step 5 of Algorithm 2.1 can not cycle infinitely.*

Proof. Suppose Algorithm 2.1 doesn't stop at x_k, i.e., there exists a constant $\epsilon > 0$ such that $\|g_k\| \geq \epsilon$. Suppose the iterations produced by Algorithm 2.1 cycle infinitely between Step 1 and Step 5. i.e.,

$$x_{k+i} = x_k,\ \|g_{k+i}\| = \|g_k\| \geq \epsilon,\ \sigma_{k+i} \in [\gamma_1\sigma_{k+i-1}, \gamma_2\sigma_{k+i-1}],\ i = 0, 1, \cdots$$

thus

$$\rho_{k+i} < \eta_1,\ \sigma_{k+i} \geq \gamma_1^i\sigma_k\,(\gamma_1 > 1),\ i = 0, 1, \cdots$$

So $\sqrt{\frac{\|g_{k+i}\|}{\sigma_{k+i}}} \leq \sqrt{\frac{\|g_k\|}{\gamma_1^i\sigma_k}} \to 0,\ i \to \infty$, where $\gamma_1 > 1$. From Lemma 4.3, we know that the $(k+i)$-th iteration is successful, so $\rho_{k+i} > \eta_2$, which is a contradiction. We complete the proof. □

Next, we show that provided that there are only finitely many successful iterations, the subsequent iterations are first order critical points. The detailed proof see Lemma 2.4 in [19].

Theorem 4.5 *Suppose that Assumptions A1, A2, and A3 hold, and that there are only finitely many successful iterations. Then for sufficiently large k, $x_k = x^*$ and $g(x^*) = 0$.*

The following lemma shows that if the objective function is bounded below, then at least an accumulate point is a first order point. The detailed proof see theorem 2.5 in [19].

Theorem 4.6 *Suppose that Assumptions A1, A2, and A3 hold. Then*

$$\liminf_{k\to\infty} \|g_k\| = 0. \tag{4.15}$$

The following lemma shows that under some strengthen conditions, all the accumulate points are first order points. The detailed proof see theorem 2.6 in [19].

Theorem 4.7 *Suppose that Assumptions A1, A2, and A3 hold. Then*

$$\lim_{k\to\infty} \|g_k\| = 0. \tag{4.16}$$

To conclude this section, we discuss the local convergence rate of Algorithm 2.1.

Theorem 4.8 *Suppose that the iterate sequence $\{x_k\}$ generated by Algorithm 2.1 converges to x^*, $g(x^*) = 0$, and $H(x^*)$ is Lipschitz continuous in a neighborhood of x^*. If*

$$\lim_{k\to\infty} \frac{\|(B_k - H_k)(x_k - x^*)\|}{\|x_k - x^*\|} = 0, \tag{4.17}$$

then the iterate sequence $\{x_k\}$ converges to x^* linearly.

Proof. Due to $x_k \to x^*\,(k \to \infty)$, so $\|s_k\| \to 0\,(k \to \infty)$. From $\|g_k\| \le c_3$, $\|h_k\| \le h_{max}$, $\|B_k\| \le k_B$, $\sigma_k \le \sigma_{max}$ (which will be proved in Lemma 5.2 later), we know that when $k \to \infty$ there exist two constants c_4, $c_5 > 0$ such that

$$\|(B_k + \nu_k^2\lambda_k I + g_k h_k^T + \nu_k^2\lambda_k h_k s_k^T)^{-1}\| \le c_4.$$

$$\|\nu_k^2\lambda_k I + g_k h_k^T + \nu_k^2\lambda_k h_k s_k^T\| \le c_5.$$

From (5.11), we have

$$\begin{aligned}
\|x_k + s_k - x^*\| &= \|x_k - x^* - (B_k + \nu_k^2\lambda_k I + g_k h_k^T + \nu_k^2\lambda_k h_k s_k^T)^{-1} g_k\| \\
&\le c_4\|g_k - g(x^*) - B_k(x_k - x^*) \\
&\quad -(\nu_k^2\lambda_k I + g_k h_k^T + \nu_k^2\lambda_k h_k s_k^T)(x_k - x^*)\| \\
&\le c_4[\|g_k - g(x^*) - B_k(x_k - x^*)\| \\
&\quad +\|(\nu_k^2\lambda_k I + g_k h_k^T + \nu_k^2\lambda_k h_k s_k^T)(x_k - x^*)\|] \\
&\le c_4\Big[\int_0^1 (H(x_k + \theta(x_k - x^*)) - H_k)(x_k - x^*)d\theta \\
&\quad +(H_k - B_k)(x_k - x^*) + c_5\|x_k - x^*\|\Big] \\
&\le c_4[L_1\theta\|x_k - x^*\|^2 + o(\|x_k - x^*\|) + c_5\|x_k - x^*\|] \\
&= c_6\|x_k - x^*\|.
\end{aligned}$$

So the iterate sequence $\{x_k\}$ converges to x^* linearly. $\square$

5. Complexity analysis

The most advantage of ACCO method is that the corresponding algorithm can enjoy a better complexity bound. As mentioned before, ACCO model is an extension of ACO model. Whether the ACCO model still keeps the property of complexity bound? In this section, we will discuss the complexity of ACCO model for unconstrained optimization.

Denote

$$S_j \overset{def}{=} \{k \le j : \text{iteration}\, k\, \text{is successful}\},$$

$$U_j \overset{def}{=} \{i \le j : \text{iteration}\, i\, \text{is unsuccessful}\}.$$

Let $|S_j|$ and $|U_j|$ be the respective cardinalities.

Define

$$S_g^\epsilon = \{k \in S : \|g_k\| \geq \epsilon\}.$$

$$U_g^\epsilon = \{k \in U : \|g_k\| \geq \epsilon\}.$$

Next, we will give an upper bound of $|S_g^\epsilon|$, i.e., the estimation of the times of the successful iterations to satisfy $\|g_k\| \leq \epsilon$, and $|S_g^\epsilon| + |U_g^\epsilon|$, i.e., the estimation of the times of the successful and unsuccessful iterations to satisfy $\|g_k\| \leq \epsilon$.

First, we give some technical lemmas.

Lemma 5.1 *Suppose that Assumptions A2 and A3 hold, $\|g_k\| \neq 0$ and*

$$\sqrt{\sigma_k \|g_k\|} \geq \frac{27(\sqrt{2}+1)\sqrt{c_2}(k_H + k_B + c_3 h_{max})}{(1-\eta_2)c_1^2} \stackrel{def}{=} k_{HB}. \tag{5.1}$$

Then the k-th iteration is successful, i.e., $\rho_k > \eta_2$, $\sigma_{k+1} \leq \sigma_k$.

Proof. Due to $\|g_k\| \neq 0$, from Lemma 4.1, we know $c_k(0) > c_k(s_k)$. So

$$\rho_k > \eta_2 \Longleftrightarrow r_k = f(x_k + s_k) - c_k(s_k) + (1-\eta_2)(c_k(s_k) - c_k(0)) < 0.$$

Next, we prove $r_k < 0$. Combining (2.25) and (5.1) yields that

$$\begin{aligned} \sqrt{\frac{\sigma_k\|g_k\|}{c_2}} &\geq \frac{27(\sqrt{2}+1)(k_H + k_B + c_3 h_{max})}{(1-\eta_2)c_1^2} \\ &\geq \frac{27(\sqrt{2}+1)(k_H + k_B + c_3 h_{max})}{(1-\eta_2)c_2^2} \\ &\geq \frac{\|B_k\|}{c_2^2}. \end{aligned}$$

From Lemma 4.2, we have

$$\|s_k\| \leq 3\sqrt{\frac{\|g_k\|}{c_2\sigma_k}}. \tag{5.2}$$

From Taylor's expansion, (2.25) and (5.2), we obtain

$$\begin{aligned} f(x_k + s_k) - c_k(s_k) &= g(\xi_k)^T s_k - \frac{g_k^T s_k}{1 + h_k^T s_k} - \frac{s_k^T B_k s_k}{2(1+h_k^T s_k)^2} - \frac{\sigma_k \|s_k\|^3}{3} \\ &\leq \frac{\|g(\xi_k) - g(x_k)\|\|s_k\| + \|g(\xi_k)\|\|h_k\|\|s_k\|^2}{c_1} + \frac{k_B\|s_k\|^2}{2c_1^2} \\ &\leq \frac{k_H + k_B + c_3 h_{max}}{2c_1^2}\|s_k\|^2 \\ &\leq \frac{9(k_H + k_B + c_3 h_{max})\|g_k\|}{2c_1^2 c_2 \sigma_k}. \end{aligned}$$

Combining (5.1), $k_H \geq 1$ and $\eta_1 \in (0,1)$ yields that

$$\begin{aligned} 2\sqrt{\frac{\sigma_k\|g_k\|}{c_2}} &\geq \frac{54(\sqrt{2}+1)(k_H+k_B+c_3h_{max})}{(1-\eta_2)c_1^2} \\ &\geq \frac{k_B}{c_1^2} \\ &\geq \frac{\|B_k\|}{c_1^2}, \end{aligned}$$

so

$$\frac{c_1^2\|g_k\|}{\|B_k\|} \geq \frac{1}{2}\sqrt{\frac{c_2\|g_k\|}{\sigma_k}}. \tag{5.3}$$

Lemma 4.1 and (5.3) results in

$$c_k(s_k) - c_k(0) \leq -\frac{\|g_k\|^{\frac{3}{2}}}{6(\sqrt{2}+1)c_2^{\frac{3}{2}}\sqrt{\sigma_k}}. \tag{5.4}$$

Then

$$\begin{aligned} r_k &\leq \frac{9(k_H+k_B+c_3h_{max})\|g_k\|}{2c_1^2c_2\sigma_k} - \frac{(1-\eta_2)\|g_k\|^{\frac{3}{2}}}{6(\sqrt{2}+1)c_2^{\frac{3}{2}}\sqrt{\sigma_k}} \\ &\leq \frac{\|g_k\|}{\sigma_k}\Big[\frac{9(k_H+k_B+c_3h_{max})}{2c_1^2c_2} - \frac{(1-\eta_2)\sqrt{\sigma_k\|g_k\|}}{6(\sqrt{2}+1)c_2^{\frac{3}{2}}}\Big]. \end{aligned}$$

From (5.1), we know $r_k \leq 0$, so $\rho_k \geq \eta_2$, $\sigma_{k+1} \geq \sigma_k$. We complete the proof. □

The above lemma is an extension of Lemma 3.2 in [20]. However, our result is weaker than Cartis et al's, because our lower bound on $\sqrt{\sigma_k\|g_k\|}$ is smaller than theirs.

Lemma 5.2 *Suppose that Assumptions A1, A2, and A3 hold, and there exists a constant $\epsilon > 0$ such that $\|g_k\| \geq \epsilon$. Then*

$$\sigma_k \leq \max\Big(\sigma_0, \frac{\gamma_2}{\epsilon}k_{HB}^2\Big) \overset{def}{=} \sigma_{max}. \tag{5.5}$$

Proof. From Lemma 5.1, we know

$$\sigma_k > \frac{k_{HB}^2}{\epsilon} \Longrightarrow \rho_k > \eta_2,\ \sigma_{k+1} \leq \sigma_k, \tag{5.6}$$

so there must exist k' such that $\sigma_{k'} \leq \frac{k_{HB}^2}{\epsilon}$. From Step 4 of Algorithm 2.1, we get $\sigma_{k+1} \leq \gamma_2\sigma_k$ ($\gamma_2 > 1$) whenever the k-th iteration is successful or not. So

$$\sigma_{k'+i} \leq \gamma_2\frac{k_{HB}^2}{\epsilon},\ i = 0, 1, \cdots \tag{5.7}$$

Thus, if $\sigma_0 \leq \frac{k_{HB}^2}{\epsilon}$, from (5.7), we know $\sigma_k \leq \gamma_2\frac{k_{HB}^2}{\epsilon}$. On the contrary, if $\sigma_0 \geq \frac{k_{HB}^2}{\epsilon}$, σ_k ($k = 1, 2, \cdots$) will keep on decreasing until there exists k' such that $\sigma_{k'} \leq \frac{k_{HB}^2}{\epsilon}$. We choose $k > k'$ then $\sigma_k \leq \gamma_2\frac{k_{HB}^2}{\epsilon}$. We complete the proof. □

Lemma 5.3 *Suppose that for each very successful iteration, there exists a constant $\gamma_3 \in (0,1)$ such that $\gamma_3\sigma_k < \sigma_{k+1} \le \sigma_k$. Let $\overline{\sigma} > 0$ such that $\sigma_k \le \overline{\sigma}$ for all $k \le j$. Then*

$$|U_j| \le \Big\lceil -\frac{\log\gamma_3}{\log\gamma_1}|S_j| + \frac{1}{\log\gamma_1}\log(\frac{\overline{\sigma}}{\sigma_0})\Big\rceil. \tag{5.8}$$

Proof. The proof see Lemma 2.1 in [20]. □

Theorem 5.4 *Suppose that Assumptions A1, A2, and A3 hold. Then for all $k \in S_g^\epsilon$,*

$$|S_g^\epsilon| \le \lceil c_7\epsilon^{-2}\rceil, \tag{5.9}$$

where $c_7 = \frac{(f(x_0)-f_{low})6(\sqrt{2}+1)c_2^{\frac{3}{2}}}{\eta_1}\max\{k_{HB}\sqrt{\gamma_2}, \sqrt{\sigma_0}\}$.

If σ_0 in Algorithm 2.1 is chosen such that $\sigma_0\epsilon \ge \gamma_2 k_{HB}^2$. Then

$$|S_g^\epsilon| \le \lceil c_8\epsilon^{-\frac{3}{2}}\rceil, \tag{5.10}$$

where $c_8 = \frac{(f(x_0)-f_{low})6(\sqrt{2}+1)c_2^{\frac{3}{2}}\sqrt{\sigma_0}}{\eta_1}$.

Additionally, assume that on each very successful iteration k, $\sigma_{k+1} \ge \gamma_3\sigma_k$, $\gamma_3 \in (0,1]$. Then

$$|S_g^\epsilon| + |U_g^\epsilon| \le \lceil c_9\epsilon^{-2}\rceil, \tag{5.11}$$

where $c_9 = (1 - \frac{\log\gamma_3}{\log\gamma_1})c_7 + \frac{\max\{1,\gamma_2 k_{HB}^2\}}{\log\gamma_1}$.

Proof. From the definition of k_{HB}, Lemma 4.1 and Lemma 5.2, we have

$$\begin{aligned}
c_k(0) - c_k(s_k) &\ge \frac{\|g_k\|}{3(\sqrt{2}+1)c_2^2}\min\Big\{\frac{c_1^2\|g_k\|}{\|B_k\|}, \frac{1}{2}\sqrt{\frac{c_2\|g_k\|}{\sigma_k}}\Big\} \\
&\ge \min\Big\{\frac{c_1^2\epsilon^2}{3(\sqrt{2}+1)c_2^2 k_B}, \frac{\sqrt{c_2}\epsilon^{\frac{3}{2}}}{6(\sqrt{2}+1)c_2^2\sqrt{\sigma_k}}\Big\} \\
&\ge \min\Big\{\frac{c_1^2\epsilon^2}{3(\sqrt{2}+1)c_2^2 k_B}, \frac{\epsilon^2}{6(\sqrt{2}+1)c_2^{\frac{3}{2}}\max\{\sqrt{\sigma_0\epsilon}, k_{HB}\sqrt{\gamma_2}\}}\Big\} \\
&= \min\Big\{\frac{\epsilon^2}{6(\sqrt{2}+1)c_2^{\frac{3}{2}}\sqrt{\sigma_0\epsilon}}, \frac{\epsilon^2}{6(\sqrt{2}+1)c_2^{\frac{3}{2}}k_{HB}\sqrt{\gamma_2}}\Big\}
\end{aligned} \tag{5.12}$$

case a: If $\sqrt{\sigma_0\epsilon} < k_{HB}\sqrt{\gamma_2}$, from (5.12), we have

$$c_k(0) - c_k(s_k) \ge \frac{\epsilon^2}{6(\sqrt{2}+1)c_2^{\frac{3}{2}}k_{HB}\sqrt{\gamma_2}},$$

so for all $k \in S_g^\epsilon$,

$$f(x_k) - f(x_k+s_k) \ge \eta_1(c_k(0) - c_k(s_k)) \ge \frac{\eta_1\epsilon^2}{6(\sqrt{2}+1)c_2^{\frac{3}{2}}k_{HB}\sqrt{\gamma_2}}.$$

Summing up all the very successful iterations satisfying $\|g_k\| \geq \epsilon$ yields that

$$\begin{aligned}\frac{|S_g^\epsilon|\eta_1\epsilon^2}{6(\sqrt{2}+1)c_2^{\frac{3}{2}}k_{HB}\sqrt{\gamma_2}} &\leq \sum_{k\in S_g^\epsilon}[f(x_k)-f(x_{k+1})] \\ &\leq \sum_{k=0}^{j}[f(x_k)-f(x_{k+1})] \\ &\leq f(x_0)-f_{low}.\end{aligned}$$

So

$$|S_g^\epsilon| \leq \frac{(f(x_0)-f_{low})6(\sqrt{2}+1)c_2^{\frac{3}{2}}k_{HB}\sqrt{\gamma_2}}{\eta_1}\epsilon^{-2}. \tag{5.13}$$

case b: If $\sqrt{\sigma_0\epsilon} \geq k_{HB}\sqrt{\gamma_2}$, from (5.12), we have

$$c_k(0)-c_k(s_k) \geq \frac{\epsilon^{\frac{3}{2}}}{6(\sqrt{2}+1)c_2^{\frac{3}{2}}\sqrt{\sigma_0}}$$

so for all $k \in S_g^\epsilon$,

$$f(x_k)-f(x_k+s_k) \geq \eta_1(c_k(0)-c_k(s_k)) \geq \frac{\eta_1\epsilon^{\frac{3}{2}}}{6(\sqrt{2}+1)c_2^{\frac{3}{2}}\sqrt{\sigma_0}}.$$

Summing up all the very successful iterations satisfying $\|g_k\| \geq \epsilon$ yields that

$$\begin{aligned}\frac{|S_g^\epsilon|\eta_1\epsilon^{\frac{3}{2}}}{6(\sqrt{2}+1)c_2^{\frac{3}{2}}\sqrt{\sigma_0}} &\leq \sum_{k\in S_g^\epsilon}[f(x_k)-f(x_{k+1})] \\ &\leq \sum_{k=0}^{j}[f(x_k)-f(x_{k+1})] \\ &\leq f(x_0)-f_{low}.\end{aligned}$$

So

$$|S_g^\epsilon| \leq \frac{(f(x_0)-f_{low})6(\sqrt{2}+1)c_2^{\frac{3}{2}}\sqrt{\sigma_0}}{\eta_1}\epsilon^{-\frac{3}{2}}. \tag{5.14}$$

From Lemma 5.2, we know

$$\sigma_k \leq \max\{\sigma_0, \frac{\gamma_2}{\epsilon}k_{HB}^2\} \leq \frac{\max\{\sigma_0,\gamma_2 k_{HB}^2\}}{\epsilon} \overset{def}{=} \overline{\sigma}.$$

From Lemma 5.3, we get

$$|S_g^\epsilon| + |U_g^\epsilon| \leq \Big[(1-\frac{\log\gamma_3}{\log\gamma_1})c_7 + \frac{\max\{1,\gamma_2 k_{HB}^2\}}{\log\gamma_1}\Big]\epsilon^{-2} \tag{5.15}$$

□

6. Conclusions

In this paper, we propose an adaptive conic cubic overestimation model method for unconstrained optimization problem. It incorporates an adaptive cubic overestimation model and a quadratic model as special cases. Global convergence to first order critical point and local linear convergence are guaranteed under some mild conditions. The algorithm and theory presented in this paper can be extended to constrained nonlinear optimization problem, which will be our next work.

References

[1] M.J.D. Powell, On the global convergence of trust region algorithms for unconstrained minimization, Mathematical Programming, 29, 297-303, 1984.

[2] J.P. Bulteau and J.P. Vial, Curvilinear path and trust region in unconstrained optimization: a convergence analysis, Mathematical Programming Study, 30, 82-101, 1987.

[3] G.A. Shultz, R.B. Schnabel and R.H. Byrid, A family of trust region based algorithms for unconstrained minimization with strong global convergence properties, SIAM Journal on Numerical Analysis, 22, 47-67, 1985.

[4] H.P. Wu and Q. Ni , A new trust region algorithm with conic model, Numerical Mathematics A Journal of Chinese Universities, 30, 57-63, 2008.

[5] Q. Ni, Optimality conditions for trust region subproblems involving a conic model, SIAM Journal on Optimization, 15, 828-837, 2005.

[6] J.H. Fu and W.Y. Sun, Nonmonotone adaptive trust-region method for unconstrained optimization problems, Applied Mathematics and Computation, 163, 489-504, 2005.

[7] Q.M. Han, W.Y. Sun and J.Y. Han, An adaptive conic trust-region method for unconstrained optimization, Optimization Methods and Software, 20, 665-677, 2005.

[8] W.Y. Sun, J.Y. Han and J. Sun, Global convergence of nonmonotone descent methods for unconstrained optimization problems, Journal of Computational and Applied Mathematics, 146, 89-98, 2002.

[9] Y. Ji, Y.J. Li, K.C. Zhang and X.L. Zhang, A new nonmonotone trust-region algorithm with conic model for unconstrained optimization, Journal of Computational and Applied Mathematics, 233, 1746-1754, 2010.

[10] S.J. Qu, S.D. Jiang and Y. Zhu, A conic trust region method and its convergence properties, Computers and Mathematics with Applications, 57, 513-528, 2009.

[11] J.P. Bulteau and J.P. Vial, A restricted trust region algorithm for unconstrained optimization, Journal of Optimization Theory and Applications, 47, 413-435, 1985.

[12] L. Grippo, F. Lampariello and S. Lucidi, A truncated Newton method with nonmonotone line search for unconstrained optimization, Journal of Optimization Theory and Applications, 60, 401-419, 1989.

[13] H.C. Zhang and W.W. Hager, A nonmonotone line search technique and its application to unconstrained optimization, SIAM Journal on Optimization, 14, 1043-1056, 2004.

[14] Ph.L. Toint, An assessment of nonmonotone line search techniques for unconstrained optimization, SIAM Journal on Scientific Computing, 17, 725-739, 1996.

[15] M.C. Ferris, S. Lucid and M. Roma, Nonmonotone curvilinear line search methods for unconstrained optimization, Computational Optimization and Applications, 6, 117-136, 1996.

[16] Y.X. Yuan and J. Nocedal, Combining trust region and line search techniques, Advances in Nonlinear Programming, Kluwer, 153–175, 1998.

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

[18] W.Y. Sun and Y.X. Yuan, Optimization Theory and Methods: Nonlinear Programming, Springer, New York, 2006.

[19] C. Cartis, N.I.M. Gould and Ph.L. Toint, Adaptive cubic overestimation methods for unconstrained mimimization. Part I: motivation, convergence and numerical results, Mathematical Programming, 127, 245-295, 2011.

[20] C. Cartis, N.I.M. Gould and Ph.L. Toint, Adaptive cubic overestimation methods for unconstrained mimimization. Part II: worst case function and derivative evaluation complexity, Mathematical Programming, 2010(online).

[21] Y. Nesterov and B.T. Polyak, Cubic regularization of Newton method and its global performance, Mathematical Programming, 108, 177-205, 2006.

[22] W. Davidon, Conic Approximations and Collinear Scalings for Optimizers, SIAM Journal on Numerical Analysis, 17, 268-281, 1980.

[23] D.C. Sorensen, The Q-superlinear convergence of a collinear scaling algorithm for unconstrained optimization, SIAM Journal on Numerical Analysis, 17, 84-114, 1980.

[24] S. Di and W.Y. Sun, A trust region method for conic model to solve unconstraind optimizaions, Optimization Methods and Software, 6, 237-263, 1996.

[25] J.H. Fu, W.Y. Sun and R.De Sampaio, An adaptive approach of conic trust region method for unconstrained optimization problems, Journal of Applied Mathematics and Computing, 19, 165-177, 2005.

[26] L.J. Zhao and W.Y. Sun, A conic affine scaling dogleg method for nonlinear optimization with bound constraints, submitted to Asia and Pacific Journal of Operational Research.

[27] W.Y. Sun, J.Y. Sun and Y.X. Yuan, Conic trust region method for linearly constrained optimization, Journal of Computational Mathematics, 21, 295-304, 2003.

[28] W.Y. Sun and Y.X. Yuan, A conic trust region method for nonlinearly constrained optimization, Annals of Operations Research, 103, 175-191, 2001.

[29] K. Ueda and N. Yamashita, Convergence properties of the regularized Newton method for the unconstrained nonconvex optimization, Applied Mathematics and Optimization, 62, 27-46, 2010.

[30] S. Lu, Z.X. Wei and L. Li, A trust region algorithm with adaptive cubic regularization methods for nonsmooth convex minimization, Computational Optimization and Applications, 51, 551-573, 2012.

[31] N.I.M. Gould, D.P. Robinson and H.S. Thorne, On solving trust region and other regularised subproblem in optimization, Mathematical Programming, 2, 21-57, 2010.

[32] C. Cartis, N.I.M. Gould and Ph.L. Toint, On the complexity of steepest descent, Newton's and regularized Newton's methods for nonconvex unconstrained optimization, SIAM Journal on Optimization, 20, 2833-2852, 2010.

[33] C. Cartis, N.I.M. Gould and Ph.L. Toint, Trust region and other regularisation of linear least squares problems, BIT, 49, 21-53, 2009.