Noname manuscript No.
(will be inserted by the editor)

A sequential adaptive regularisation using cubics algorithm for solving nonlinear equality constrained optimization

Yonggang Pei · Shaofang Song · Detong Zhu

the date of receipt and acceptance should be inserted later

Abstract The adaptive regularisation algorithm using cubics (ARC) is initially proposed for unconstrained optimization. ARC has excellent convergence properties and complexity. In this paper, we extend ARC to solve nonlinear equality constrained optimization and propose a sequential adaptive regularisation using cubics algorithm inspired by sequential quadratic programming (SQP) methods. In each iteration of our method, the trial step is computed via composite-step approach, i.e., it is decomposed into the sum of normal step and tangential step. By means of reduced-Hessian approach, a new ARC subproblem for nonlinear equality constrained optimization is constructed to compute the tangential step, which can supply sufficient decrease required in the proposed algorithm. Once the trial step is obtained, the ratio of the penalty function reduction to the model function reduction is calculated to determine whether the trial point is accepted. The global convergence of the algorithm is investigated under some mild assumptions. Preliminary numerical experiments are reported to show the performance of the proposed algorithm.

Keywords nonlinear optimization · constrained optimization · adaptive regularization with cubics · global convergence

Mathematics Subject Classification (2020) 49M37 · 65K05 · 65K10 · 90C30

Y. Pei
Engineering Laboratory for Big Data Statistical Analysis and Optimal Control, College of Mathematics and Information Science, Henan Normal University
E-mail: peiyg@163.com

S. Song
Engineering Laboratory for Big Data Statistical Analysis and Optimal Control, College of Mathematics and Information Science, Henan Normal University
E-mail: 731713505@qq.com

D. Zhu
Mathematics and Science College, Shanghai Normal University
E-mail: dtzhu@shnu.edu.cn

1 Introduction

Line search methods and trust-region methods are two fundamental strategies for solving unconstrained optimization. Cartis, Gould and Toint [9] present a third alternative: the adaptive regularization algorithm using cubics (ARC). At each iteration of ARC, the objective function is locally replaced by a cubic approximation model, and an adaptive estimation of the local Lipschitz constant of the objective function's Hessian is employed. ARC has been shown to have excellent global and local convergence properties. The iteration complexity of ARC for unconstrained optimization is given in [10], and the numerical experiments performance is also encouraging when solving small-scale problems. There have been many methods ([2, 15, 28, 25, 11, 3, 16, 6, 24, 7, 26]) on using ARC (or its variants) for unstrained optimization. And some methods ([20, 21, 5, 19, 18]) are designed to solve or use ARC subproblems more efficiently.

For constrained optimization problems, some impressive algorithms are also proposed based on ARC framework. Cartis et al. [13] extended the ARC method to solve nonconvex optimization with convex constraints, and presented a class of adaptive regularization methods that use first- or higher-order local Taylor models of the objective regularized by a(ny) power of the step size and applied to convexly constrained optimization problems in [12]. They also presented a short-step ARC algorithm for solving general nonconvexly constrained optimization problems [14] and proposed an ARp algorithm for for nonconvex minimization problems with general inexpensive constraints even though they are conceptual regularization algorithms. Martínez [23] considered arbitrary (nonnecessarily Taylor-based) models of the objective function and nonnecessarily convex constraints. At each iteration a mild approximate KKT point is computed for the subproblem without requiring the fulfillment of any constraint qualification. Lars Lubkoll et al. [22] proposed an affine covariant composite step method, designed for equality constrained optimization with partial differential equations. A cubic hybrid approach was constructed for inequality constrained problem in [4]. Agarwal [1] researched a generalization of ARC to optimization on Riemannian manifolds. Birgin [8] introduced a practical method for smooth bound-constrained optimization that possesses worst-case evaluation complexity $O(\varepsilon^{-3/2})$ for finding an ε-approximate first-order stationary point when the Hessian of the objective function is Lipschitz continuous.

Motivated by the idea of SQP methods to tackle the equality constrained problem (1), we present a sequential adaptive regularization algorithm using cubics (SARC) for solving equality constrained optimization. It is different from the methods mentioned above. Obviously, the ARC subproblem (minimizing a cubic model of objective function) in the unconstrained optimization circumstance can not be applied directly because of the existence of constraints. We should consider a constrained cubic regularization subproblem in each iteration. In the algorithm to be proposed, a composite-step method is employed to handle linearized constraints, where the trial step is decomposed into the sum of the normal step and the tangential step in each iteration. The tangential step is obtained by solving a standard ARC subproblem, which is constructed via reduced-Hessian approach. The normal step is used to reduce the constraint violation, and the tangential step is used to provide sufficient reduction of the model. After the trial step is obtained, the exact penalty function is used as the merit function, and the ratio of the reduction of the penalty function to the reduction of the model is calculated to determine whether the trial point is accepted. Global convergence is proved under some suitable assumptions.

The remainder of this paper is organized as follows. In section 2, we describe the step computation and the ARC algorithm for equality-constrained. The global convergence to first-order critical point is presented in Section 3. Preliminary numerical results are reported in Section 4 and conclusion is offered in Section 5.

Throughout the paper, $\|\cdot\|$ denotes the Euclidean norm. The inner product of vectors $a, b \in \mathbb{R}^n$ is denoted by $a^T b$.

2 Cubic regularisation for nonlinear equality constrained

Consider the nonlinear equality constrained optimization problem as follows.

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x) \tag{1a}$$

$$\text{subject to} \quad c(x) = 0, \tag{1b}$$

where the objective function $f : \mathbb{R}^n \to \mathbb{R}$ and the equality constraints $c : \mathbb{R}^n \to \mathbb{R}^m$ are sufficiently smooth functions and $m \leq n$. The Lagrangian function for this problem is

$$\mathcal{L}(x, \lambda) = f(x) - \lambda^T c(x).$$

We use $A(x)$ to denote the Jacobian of $c(x)$, that is,

$$A(x)^T = [\nabla c_1(x), \nabla c_2(x), ..., \nabla c_m(x)],$$

where $c_i(x)$ is the ith component of $c(x)$.

As we know, the linearized constraints may be not consistent. So, we assume that A_k has full row rank for all k.

We consider the following constrained cubic regularization subproblem in each iteration,

$$\underset{p\in\mathbb{R}^n}{\text{minimize}} \quad f_k + g_k^T p + \frac{1}{2} p^T B_k p + \frac{1}{3}\sigma_k \|p\|^3 \tag{2a}$$

$$\text{subject to} \quad A_k p + c_k = 0, \tag{2b}$$

where $f_k = f(x_k)$, $g_k = \nabla f(x_k)$, $A_k = A(x_k)$, $c_k = c(x_k)$, B_k denotes the Hessian $\nabla_{xx}\mathcal{L}(x_k, \lambda_k)$ or its approximation, and $\sigma_k \in \mathbb{R}^+$ is an adaptive parameter.

However, we do not solve subproblem (2) directly. Like the composite-step approach, in iteration k we decompose the overall trial step p_k as follow,

$$p_k = n_k + t_k,$$

where n_k is called as a normal step to improve feasibility, and t_k is called as a tangential step whose purpose is to obtain sufficient reduction of the objective function's model.

Because t_k is used to produce the sufficient decrease in the objective function's model, the maximum size of n_k should be restricted. It is generated as follows.

$$n_k = \alpha_k n_k^c, \tag{3}$$

where n_k^c is a solution of

$$A_k n + c_k = 0,$$

and α_k satisfies

$$\alpha_k \in \left[\min\left\{1, \frac{\theta}{\sqrt{\sigma_k}\|n_k^c\|}\right\}, \min\left\{1, \frac{1}{\sqrt{\sigma_k}\|n_k^c\|}\right\}\right], \quad \text{for } \theta \in (0,1]. \tag{4}$$

Then, the tangential step t_k is computed by

$$t_k = N_k t_k^N, \tag{5}$$

where t_k^N is the solution(or its approximation) of the following standard ARC subproblem

$$\text{minimize} \quad \langle g_k^N, t^N\rangle + \frac{1}{2}\langle t^N, B_k^N t^N\rangle + \frac{1}{3}\sigma_k\|t^N\|^3, \tag{6}$$

where $B_k^N = N_k^T B_k N_k$, $g_k^N = N_k^T(g_k + B_k n_k)$ and the columns of N_k form an orthonormal basis of the null space of A_k (so that $A_k N_k = 0$).

Having obtained a trial step p_k, we use a merit function to decide whether it should be accepted. Here, we use l_2 penalty function

$$\phi(x,\mu) = f(x) + \mu\|c(x)\|$$

and its model

$$m(x,B,\mu,p) = f(x) + g^T(x)p + \frac{1}{2}p^T Bp + \frac{1}{3}\sigma\|p\|^3 + \mu\|c(x) + A(x)p\|. \tag{7}$$

For convenience, define

$$\begin{aligned} m^F(x,B,n) &= f(x) + g(x)^T n + \frac{1}{2}n^T Bn + \frac{1}{3}\sigma\|n\|^3, \\ m^N(x,n) &= \|c(x) + A(x)n\|, \\ m^H(x,B,t) &= (g(x) + Bn)^T t + \frac{1}{2}t^T Bt + \frac{1}{3}\sigma\|t\|^3. \end{aligned} \tag{8}$$

Define the overall model decrease from 0 to p_k as

$$\Delta m_k = m(x_k, B_k, \mu_k, 0) - m(x_k, B_k, \mu_k, p_k). \tag{9}$$

Then

$$\Delta m_k = \Delta m_k^H + \mu_k \Delta m_k^N + \Delta m_k^F + \frac{1}{3}\sigma_k(\|n_k\|^3 + \|t_k\|^3 - \|p_k\|^3),$$

where

$$\Delta m_k^N = m^N(x_k, 0) - m^N(x_k, n_k),$$

$$\Delta m_k^H = m^H(x_k, B_k, 0) - m^H(x_k, B_k, t_k),$$

and

$$\Delta m_k^F = m^F(x_k, B_k, 0) - m^F(x_k, B_k, n_k).$$

Note that Δm_k^N and Δm_k^H must be positive (see Lemma 1 and Lemma 3). To obtain significant decrease, we require that

$$\begin{aligned} \Delta m_k &= \Delta m_k^H + \mu_k \Delta m_k^N + \Delta m_k^F + \frac{1}{3}\sigma_k(\|n_k\|^3 + \|t_k\|^3 - \|p_k\|^3) \\ &\geq \nu \mu_k \Delta m_k^N \end{aligned} \tag{10}$$

for some $\nu \in (0, 1)$ and μ_k large enough. Here, we follow the rules similar to [27] updating μ_k as follows.

First, compute

$$\mu_k^c = -\frac{\Delta m_k^F + \Delta m_k^H + \frac{1}{3}\sigma_k(\|n_k\|^3 + \|t_k\|^3 - \|p_k\|^3)}{(1-\nu)\Delta m_k^N} \tag{11}$$

as the smallest value to satisfy the (10). Suppose that μ_k is the penalty parameter at the previous iteration.

Then μ_k is updated by

$$\mu_k = \begin{cases} \max\{\mu_k^c, \tau_1\mu_{k-1}, \mu_{k-1} + \tau_2\} & \text{if } \mu_{k-1} < \mu_k^c, \\ \mu_{k-1} & \text{otherwise,} \end{cases} \tag{12}$$

where $\tau_1 > 1, \tau_2 > 0$ are constants.

In order to determine whether the trial step is accepted, the descent ratio is defined as

$$\rho_k = \frac{\phi(x_k, \mu_k) - \phi(x_k + p_k, \mu_k)}{m(x_k, B_k, \mu_k, 0) - m(x_k, B_k, \mu_k, p_k)}. \tag{13}$$

Based on the above discussion, we gives the summary of the adaptive regularisation with cubic for equality constrained optimization algorithm framework as follows.

Algorithm 1: SARC for equality-constrained optimization.

Input: Input the initial point x_0, regularization parameter $\sigma_0 \geq 0$ and penalty parameter $\mu_{-1} > 0$. The constants η_1, η_2, γ_1, γ_2, γ_3, ν, τ_1 and τ_2 satisfy the conditions

$$0 < \eta_1 \leq \eta_2 < 1, \gamma_2 > \gamma_1 > 1, 0 < \nu < 1, \tau_1 > 1, \tau_2 > 0. \tag{14}$$

Set $k \leftarrow 0$;
while x_k *is not a* KKT *point* **do**
  Step 1. Compute a normal step n_k by (3) and a tangential step t_k by (5), and set $p_k = n_k + t_k$.
  Step 2. Compute μ_k^c by (11) and update the penalty parameter μ_k from (12).
  Step 3. Compute the ratio ρ_k by (13).
  if $\rho_k \geq \eta_1$ **then**
    $x_{k+1} = x_k + p_k$,
    update B_{k+1}.
  else
    $x_{k+1} = x_k$.
  end
  Step 4. Update the regularization parameter

$$\sigma_{k+1} \in \begin{cases} [0, \sigma_k] & \text{if } \rho_k \geq \eta_2, \\ [\sigma_k, \gamma_1 \sigma_k] & \text{if } \rho_k \in [\eta_1, \eta_2), \\ [\gamma_1 \sigma_k, \gamma_2 \sigma_k] & \text{if } \rho_k < \eta_1. \end{cases} \tag{15}$$

  Step 5. Set $k \leftarrow k + 1$;
end
return x_k.

3 Global convergence

In this section, we shall prove that the iterates generated by SARC is globally convergent to a first-order critical point for the problem (1). First of all, we state some assumptions necessary for the global convergence analysis. Some preliminary results are given in Subsection 3.1. The feasibility and the optimality are shown in the remain two subsections.

Let $\{x_k\}$ be the infinite sequence generated by Algorithm 1, where we assume that the algorithm does not stop at a KKT point.
(**G1**) The function $f(x)$ is uniformly bounded from below. The functions $g(x)$ and $c(x)$ are uniformly bounded and $c(x)$ is Lipschitz continuous at each x_k.
(**G2**) $g(x)$ and $\nabla c_i(x)$ exist and are Lipschitz continuous in some open set $\mathcal{C} \supseteq \{x_k\}$.
(**G3**) $\{B_k\}$ is uniformly bounded, namely, there exists some $\kappa_B \geq 0$ so that

for all k

$$\|B_k\| \leq \kappa_B.$$

(**G4**) There exists a constant $\kappa_{\rm n} > 0$ so that for all k

$$\|n_k^c\| \leq \kappa_{\rm n}\|c(x_k)\|. \tag{16}$$

(**G5**) $\{B_k\}$ is positive semi-definite on the null space of the Jacobian A_k for each k.

We note that,

$$n_k = 0 \quad \text{if} \quad c(x_k) = 0, \tag{17}$$

since $n_k^c = 0$ from (16) and (3).

Furthermore, the normal step n_k is within a fixed range of the maximum possible values from (4),

$$\frac{\theta}{\sqrt{\sigma_k}} \leq \|n_k\| \leq \frac{1}{\sqrt{\sigma_k}}. \tag{18}$$

3.1 Preliminary results

We have the following result about Δm_k^N after the normal step is calculated.

Lemma 1 *Suppose that n_k is given by* (3)*, where n_k^c and α_k satisfy assumption (G4) and* (4)*. Then*

$$\|c(x_k)\| \geq \Delta m_k^N \geq \min\left\{\|c(x_k)\|, \frac{\theta}{\kappa_{\rm n}\sqrt{\sigma_k}}\right\}, \tag{19}$$

for all k.

Proof From the definition of $m^N(x, n)$, the normal step (3) and (16), we have that

$$\begin{aligned}
\Delta m_k^N &= \|c(x_k)\| - \|c(x_k) + A_k n_k\| \\
&= \|c(x_k)\| - \|c(x_k) + \alpha_k(A_k n_k^c + c(x_k)) - \alpha_k c(x_k)\| \\
&= \|c(x_k)\| - (1 - \alpha_k)\|c(x_k)\| \\
&= \alpha_k\|c(x_k)\|.
\end{aligned} \tag{20}$$

We note that $0 < \alpha_k \leq 1$ from (4). Thus the first inequality in (19) can be obtained directly from (20).

From (20) and (4), we can also obtain that

$$\Delta m_k^N = \alpha_k\|c(x_k)\| \geq \min\left\{\|c(x_k)\|, \frac{\theta\|c(x_k)\|}{\|n_k^c\|\sqrt{\sigma_k}}\right\}.$$

Then (19) holds from (16).

The following result shows that the normal step is bounded because of the total reduction of the normal model is bounded.

Lemma 2 *Suppose that assumptions (G1)-(G3) hold. Then there is a constant* $\kappa_{\rm mn} > 0$ *such that*

$$\kappa_{\rm mn}\|n_k\| \leq \Delta m_k^N \leq \|c(x_k)\|, \tag{21}$$

for all k.

Proof The result is obviously true if $c(x_k) = 0$, since (17) and (3) imply that $n_k = 0$. So now assume that $c(x_k) \neq 0$. Lemma 1 gives that

$$\Delta m_k^N \geq \min\left\{\|c(x_k)\|, \frac{\theta}{\kappa_{\rm n}\sqrt{\sigma_k}}\right\}. \tag{22}$$

If

$$\sigma_k \geq \left(\frac{\kappa_{\rm n}\|c(x_k)\|}{\theta}\right)^{-2},$$

it follows from (22) and that

$$\Delta m_k^N \geq \frac{\theta}{\kappa_{\rm n}\sqrt{\sigma_k}} = \frac{\kappa_{\rm mn}}{\sqrt{\sigma_k}} \geq \kappa_{\rm mn}\|n_k\|, \tag{23}$$

where $\kappa_{\rm mn} = \frac{\theta}{\kappa_{\rm n}}$.

If

$$\sigma_k < \left(\frac{\kappa_{\rm n}\|c(x_k)\|}{\theta}\right)^{-2},$$

we have from (22) that

$$\Delta m_k^N \geq \|c(x_k)\|. \tag{24}$$

Since $\alpha_k \leq 1$ from (4), (16) ensures that

$$\|n_k\| = |\alpha_k|\|n_k^c\| \leq \|n_k^c\| \leq \kappa_{\rm n}\|c(x_k)\|. \tag{25}$$

(25) and (24) imply that

$$\Delta m_k^N \geq \frac{1}{\kappa_{\rm n}}\|n_k\| \geq \frac{\theta}{\kappa_{\rm n}}\|n_k\| = \kappa_{\rm mn}\|n_k\|, \tag{26}$$

since $0 < \theta \leq 1$. Therefore, when both (23) and (26) hold, the lemma is true since (19) gives that $\Delta m_k^N \leq \|c(x_k)\|$.

We now turn to the tangential step. The Cauchy step t_k^c for the subproblem (6) is then the minimizer of the model (8) in the steepest-descent direction $t_k^s = -N_k g_k^N$. That is

$$t_k^c = -\beta_k t_k^s, \quad \text{and} \quad \beta_k = \arg\min_{\beta\in\mathbb{R}^+} m^H(x_k, B_k, -\beta t_k^s).$$

Then we can obtain the lower bound of the decrease brought by tangential step t_k.

Lemma 3 *Suppose that t_k is computed by* (5)*. Then*

$$\Delta m_k^H \geq \kappa_{\mathrm{mt}} \|g_k^N\| \min \left\{ \frac{\|g_k^N\|}{1+\|B_k^N\|}, \frac{1}{2}\sqrt{\frac{\|g_k^N\|}{\sigma_k}} \right\}, \tag{27}$$

for all $k \geq 0$, where $\kappa_{\mathrm{mt}} = \frac{1}{6\sqrt{2}}$.

Proof For any $\beta \geq 0$, using the Cauchy-Schwarz inequality, we have that

$$\begin{aligned}
\Delta m_k^H &= m^H(x_k, B_k, 0) - m^H(x_k, B_k, t_k) \\
&= m^H(x_k, B_k, 0) - m^H(x_k, B_k, N_k t_k^N) \\
&\geq m^H(x_k, B_k, 0) - m^H(x_k, B_k, t_k^c) \\
&\geq m^H(x_k, B_k, 0) - m^H(x_k, B_k, -\beta N_k g_k^N) \\
&= \beta \|g_k^N\|^2 - \frac{1}{2}\beta^2 \langle g_k^N, B_k^N g_k^N \rangle - \frac{1}{3}\beta^3 \sigma_k \|N_k g_k^N\|^3 \\
&\geq \beta \|g_k^N\|^2 \left(1 - \frac{1}{2}\beta \|B_k^N\| - \frac{1}{3}\beta^2 \sigma_k \|g_k^N\|\right).
\end{aligned} \tag{28}$$

Let

$$\begin{aligned}
\overline{\beta}_k &= \frac{3}{2\sigma_k \|g_k^N\|} \left(-\frac{1}{2}\|B_k^N\| + \sqrt{\frac{1}{4}\|B_k^N\|^2 + \frac{4}{3}\sigma_k \|g_k^N\|} \right) \\
&= 2\left(\frac{1}{2}\|B_k^N\| + \sqrt{\frac{1}{4}\|B_k^N\|^2 + \frac{4}{3}\sigma_k \|g_k^N\|} \right)^{-1}.
\end{aligned}$$

Then for all $\beta \in [0, \overline{\beta}_k]$,

$$m^H(x_k, B_k, 0) \geq m^H(x_k, B_k, t_k)$$

because $1 - \frac{1}{2}\beta\|B_k^N\| - \frac{1}{3}\beta^2 \sigma_k \|g_k^N\| \geq 0$.

Let

$$\theta_k \stackrel{\text{def}}{=} \left(\sqrt{2} \max \left\{ 1 + \|B_k^N\|, 2\sqrt{\sigma_k \|g_k^N\|} \right\} \right)^{-1}. \tag{29}$$

From the inequalities

$$\begin{aligned}
&\sqrt{\frac{1}{4}\|B_k^N\|^2 + \frac{4}{3}\sigma_k \|g_k^N\|} \\
&\leq \frac{1}{2}\|B_k^N\| + \frac{2}{\sqrt{3}}\sqrt{\sigma_k \|g_k^N\|} \\
&\leq 2 \max \left\{ \frac{1}{2}\|B_k^N\|, \frac{2}{\sqrt{3}}\sqrt{\sigma_k \|g_k^N\|} \right\} \\
&\leq \sqrt{2} \max \left\{ 1 + \|B_k^N\|, 2\sqrt{\sigma_k \|g_k^N\|} \right\},
\end{aligned}$$

and

$$\frac{1}{2}\|B_k^N\| \leq \sqrt{2}\,\max\left\{1+\|B_k^N\|, 2\sqrt{\sigma_k\|g_k^N\|}\right\},$$

it follows that $0 < \theta_k \leq \overline{\beta}_k$.

Thus replace β in (28) with θ_k, we obtain that

$$\begin{aligned} & m^H(x_k, B_k, 0) - m^H(x_k, B_k, t_k) \\ \geq & \frac{\|g_k^N\|^2(1-\frac{1}{2}\theta_k\|B_k^N\| - \frac{1}{3}\theta_k^2\sigma_k\|g_k^N\|)}{\sqrt{2}\,\max\left\{1+\|B_k^N\|, 2\sqrt{\sigma_k\|g_k^N\|}\right\}}. \end{aligned} \tag{30}$$

It now follows from the definition (29) of θ_k that $\theta_k\|B_k^N\| \leq 1$ and $\theta_k^2\sigma_k\|g_k^N\| \cdot \|N_k\|^3 \leq 1$, so that the numerator of (30) is bounded below by $\frac{1}{6}\|g_k^N\|^2$. This and (30) imply the inequality in (27).

The following lemma gives a useful bound for the tangential step t_k.

Lemma 4 *Suppose that Assumption (G5) holds. Then the tangential step*

$$\|t_k\| \leq \sqrt{3}\sqrt{\frac{\|g_k^N\|}{\sigma_k}} \tag{31}$$

for all k.

Proof Suppose that

$$\|t_k\| > \sqrt{3}\sqrt{\frac{\|g_k^N\|}{\sigma_k}}, \tag{32}$$

for $k \geq 0$. In view of $g_k^N = N_k^T(g_k + B_k n_k)$ and N_k is a orthogonal basis for the null-space of A_k,

$$\begin{aligned} & m^H(x_k, B_k, t_k) - m^H(x_k, B_k, 0) \\ = & (g_k + B_k n_k)^T t_k + \frac{1}{2} t_k^T B_k t_k + \frac{1}{3}\sigma_k\|t_k\|^3 \\ \overset{(G5)}{\geq} & (g_k + B_k n_k)^T t_k + \frac{1}{3}\sigma_k\|t_k\|^3 \\ \geq & -\|t_k\|\|g_k^N\| + \frac{1}{3}\sigma_k\|t_k\|^3 \\ \overset{(32)}{>} & 0. \end{aligned}$$

But this contradicts with (27). Hence the assumption (32) must be false. The lemma is true.

Assumption (G2) implies that

$$f(x_k) \geq f_{\min}, \tag{33}$$

for all k and some constant $f_{\min}$.

For convenience, define the auxiliary function

$$\Phi(x,\mu) = \frac{\phi(x,\mu) - f_{\min}}{\mu}.$$

The important results of $\Phi(x,\mu)$ are as follows.

Lemma 5 *Suppose that $\{k_i\}$, $i \geq 0$, are the indices of the successful iterations resulting from the algorithm. Then*

$$\Phi(x_{k_i},\mu_{k_i}) \geq 0, \tag{34}$$

and for any $j > 0$,

$$\Phi(x_{k_{i+j}},\mu_{k_{i+j}}) \leq \Phi(x_{k_i},\mu_{k_i}) - \eta_1 \frac{\Delta m_{k_i}}{\mu_{k_i}}. \tag{35}$$

Proof Due to the fact that $\mu_k > 0$ for all k, and $\phi(x_k,\mu_k) \geq f_{\min}$ from (33), (34) holds. Next, we prove the second conclusion. Because k_i and k_{i+1} are consecutive successful iterations, we have that

$$x_l = x_{k_{i+1}}$$

for $k_i + 1 \leq l \leq k_{i+1}$. Then it follows from $\rho_{k_i} \geq \eta_1$ that

$$\phi(x_{k_{i+1}},\mu_{k_i}) - f_{\min} = \phi(x_{k_i+1},\mu_{k_i}) - f_{\min} \leq \phi(x_{k_i},\mu_{k_i}) - f_{\min} - \eta_1 \Delta m_{k_i},$$

and thus

$$\Phi(x_{k_{i+1}},\mu_{k_i}) \leq \Phi(x_{k_i},\mu_{k_i}) - \eta_1 \frac{\Delta m_{k_i}}{\mu_{k_i}}.$$

From the definition of Φ, and the fact that $\mu_{k_i} \leq \mu_{k_{i+1}}$ and (33) give

$$\Phi(x_{k_{i+1}},\mu_{k_i}) - \Phi(x_{k_{i+1}},\mu_{k_{i+1}}) = (f(x_{k_{i+1}}) - f_{\min}) \left(\frac{1}{\mu_{k_i}} - \frac{1}{\mu_{k_{i+1}}} \right) \geq 0.$$

Combining the above two inequalities give that (35) holds for $j = 1$, which then implies that $\Phi(x_{k_{i+1}},\mu_{k_{i+1}}) \leq \Phi(x_{k_i},\mu_{k_i})$. Therefore, (35) is hold for any $j > 0$.

We next investigate the error between the penalty function and its overall model at the new iterate $x_k + p_k$.

Lemma 6 *Suppose that assumptions (G1)-(G3) and* (31) *hold. Then*

$$\phi(x_k + p_k, \mu_k) - m(x_k, B_k, \mu_k, p_k) \leq \kappa_0(1 + \mu_k)\|p_k\|^2 \tag{36}$$
$$\leq \kappa_{\mathrm{m}}(1 + \mu_k)\frac{1}{\sigma_k}, \tag{37}$$

for some constants κ_0 *and* κ_{m}.

Proof A Taylor expansion of $f(x_k + p_k)$ and the Lipschitz continuity of $g(x)$ give that

$$f(x_k + p_k) - f(x_k) - g_k^T p_k \leq \kappa_{\mathrm{fx}}\|p_k\|^2,$$

for some Lipschitz constant κ_{fx}. Similarly, the lipschitz continuity of $A(x)$ implies that

$$\|c(x_k + p_k)\| - \|c(x_k) + A_k p_k\| \leq \|c(x_k + p_k) - A_k p_k - c(x_k)\| \leq \kappa_{\mathrm{cx}}\|p_k\|^2,$$

for another Lipschitz constant κ_{cx}. The above inequalities with the bound

$$p_k^T B_k p_k \overset{\text{(G3)}}{\leq} \kappa_B\|p_k\|^2,$$

yield that

$$\begin{aligned}
&\phi(x_k + p_k, \mu_k) - m(x_k, B_k, \mu_k, p_k) \\
&= f(x_k + p_k) - f(x_k) - g_k^T p_k - \frac{1}{2}p_k^T B_k p_k - \frac{1}{3}\sigma_k\|p_k\|^3 \\
&\leq (\kappa_{\mathrm{fx}} + \frac{1}{2}\kappa_B + \mu_k\kappa_{\mathrm{cx}})\|p_k\|^2 \\
&\leq \kappa_0(1 + \mu_k)\|p_k\|^2
\end{aligned}$$

with $\kappa_0 = \max\{\kappa_{\mathrm{fx}} + \frac{1}{2}\kappa_B, \kappa_{\mathrm{cx}}\}$, i.e., (36) holds.

From the assumptions (G1)-(G3) and (31), there exists a constant M_g so that for all k

$$\|g_k^N\| \leq M_g.$$

Then we can have from (31) that

$$\|p_k\| \leq \|n_k\| + \|t_k\| \leq (1 + \sqrt{3M_g})\frac{1}{\sqrt{\sigma_k}}. \tag{38}$$

Using this bound and (36), we have that (37) holds with $\kappa_m = \kappa_0(1 + \sqrt{3M_g})^2$.

3.2 Feasibility

The following theorem shows the feasibility of the algorithm. The index set of successful iteration of the algorithm is denoted as

$$\mathcal{S} \stackrel{\text{def}}{=} \{k \geq 0 : \ k \text{ successful or very successful in the sense of (13)}\}.$$

Theorem 1 *Suppose that assumptions (G1) and (G2) hold. Then*

$$\lim_{k\to\infty} c(x_k) = 0.$$

Proof Since the result is obviously if $c(x_l) = 0$ for all sufficiently large l, we consider any infeasibel iteration x_l such that $\|c(x_l)\| > 0$. First, we shall prove that the algorithm can not stop at such a point. The Lipschitz continuity of $\|c(x_l)\|$ from assumption (G1) gives that

$$|\|c(x)\| - \|c(x_l)\|| \leq L_1\|x - x_l\|, \tag{39}$$

for some $L_1 > 0$ and all x. Thus (39) certainly holds for all x in an open ball

$$\mathcal{O}_l \stackrel{\text{def}}{=} \left\{x \mid \|x - x_l\| < \frac{\|c(x_l)\|}{2L_1}\right\}. \tag{40}$$

We then get that

$$|\|c(x)\| - \|c(x_l)\|| < \frac{1}{2}\|c(x_l)\|$$

from (39) and (40) and thus that

$$\|c(x)\| > \frac{1}{2}\|c(x_l)\|, \tag{41}$$

which ensures that the bound of $c(x)$ is far away from zero for arbitrary $x \in \mathcal{O}_l$. (10) and (41) guarantee that

$$\begin{aligned} \Delta m_k \ &\geq \nu\mu_k \min\left\{\|c(x_k)\|, \tfrac{\theta}{\kappa_{\rm n}}\tfrac{1}{\sqrt{\sigma_k}}\right\} \\ &\geq \nu\mu_k \min\left\{\tfrac{1}{2}\|c(x_l)\|, \tfrac{\theta}{\kappa_{\rm n}}\tfrac{1}{\sqrt{\sigma_k}}\right\}, \end{aligned} \tag{42}$$

for any $x_k \in \mathcal{O}_l$.

Let

$$\sigma_l^0 = \max\left\{\left(\frac{\kappa_{\rm n}\|c(x_l)\|}{2\theta}\right)^{-2}, \left(\frac{(1-\eta_2)\nu\theta\mu_0}{\kappa_{\rm m}\kappa_{\rm n}(1+\mu_0)}\right)^{-2}\right\}. \tag{43}$$

If $\sigma_k \geq \sigma_l^0$, (10), (19) and (43) imply that

$$\Delta m_k \geq \nu\mu_k \frac{\theta}{\kappa_{\rm n}\sqrt{\sigma_k}},$$

which combines with (37) to give

$$\begin{aligned} 1-\rho_k &= \frac{\phi(x_k+p_k,\mu_k)-m(x_k,B_k,\mu_k,p_k)}{\Delta m_k} \\ &\le \frac{\kappa_{\rm m}\kappa_{\rm n}(1+\mu_k)}{\nu\theta\mu_k\sqrt{\sigma_k}} \\ &\le 1-\eta_2. \end{aligned}$$

Hence, so long as $\sigma_k \ge \sigma_l^0$, a very successful iteration will occur from any $x_k \in \mathcal{O}_l$.

Next, we show that there must be a first successful iterate $x_{k_j}, k_j > k_0 \stackrel{\rm def}{=} l$, not in $\mathcal{O}_l$. Suppose that all iterates $x_{k_i} \in \mathcal{O}_l$. Then

$$\sigma_{k_i} \le \sigma^0 \stackrel{\rm def}{=} \gamma_1 \max\{\sigma_l, \sigma_l^0\}. \tag{44}$$

Since iteration $k_i \in \mathcal{S}$ is successful, the bounds (35), (42), (44) shows that

$$\begin{aligned} \Phi(x_{k_{i+1}},\mu_{k_{i+1}}) &\le \Phi(x_{k_i},\mu_{k_i}) - \eta_1\frac{\Delta m_{k_i}}{\mu_{k_i}} \\ &\le \Phi(x_{k_i},\mu_{k_i}) - \eta_1\nu\min\left\{\frac{1}{2}\|c(x_l)\|, \frac{\theta}{\kappa_{\rm n}\sqrt{\sigma^0}}\right\}. \end{aligned}$$

By summing the first j of these successful iterations, we get

$$\begin{aligned} \Phi(x_{k_{i+j}},\mu_{k_{i+j}}) &\le \Phi(x_{k_i},\mu_{k_i}) - \eta_1\frac{\Delta m_{k_i}}{\mu_{k_i}} \\ &\le \Phi(x_{k_i},\mu_{k_i}) - j\eta_1\nu\min\left\{\frac{1}{2}\|c(x_l)\|, \frac{\theta}{\kappa_{\rm n}\sqrt{\sigma^0}}\right\}. \end{aligned}$$

Since the right side of this inequality can be made arbitrarily negative by increasing j, $\Phi(x_{k_{i+j}},\mu_{k_{i+j}})$ must eventually be negative, this is contradiction with (34). Our assumption that $x_{k_i}, k_i \in \mathcal{S} \ge l$ is remain in $\mathcal{O}_l$ must be incorrect, and thus there must be a first iterate $x_{k_j}, j > 0$, not in $\mathcal{O}_l$.

Consider the iterates between x_{k_j} and $x_{k_0} = x_l$. (35) and (42) imply that

$$\Phi(x_{k_k},\mu_{k_k}) \le \Phi(x_{k_i},\mu_{k_i}) - \eta_1\nu\min\left[\frac{1}{2}\|c(x_l)\|, \frac{\theta}{\kappa_{\rm n}\sqrt{\sigma_{k_i}}}\right], \tag{45}$$

for any $i < k \le j$, since $x_{k_i} \in \mathcal{O}_l$. If

$$\sigma_{k_i} \le \left(\frac{\kappa_{\rm n}\|c(x_l)\|}{2\theta}\right)^{-2},$$

(35) and (45) with $k = j$ imply that

$$\Phi(x_{k_j},\mu_{k_j}) \le \Phi(x_{k_i},\mu_{k_i}) - \frac{1}{2}\eta_1\nu\|c(x_l)\| \le \Phi(x_l,\mu_l) - \frac{1}{2}\eta_1\nu\|c(x_l)\|. \tag{46}$$

On the other hand, if

$$\sigma_{k_i} < \left(\frac{\kappa_{\rm n}\|c(x_l)\|}{2\theta}\right)^{-2},$$

for all $0 \le i \le j$.

$$\Phi(x_{k_{i+1}}, \mu_{k_{i+1}}) \le \Phi(x_{k_i}, \mu_{k_i}) - \frac{\eta_1\nu\theta}{\kappa_{\rm n}\sqrt{\sigma_{k_i}}}, \tag{47}$$

for each $0 \le i \le j$. Then by summing (47),

$$\Phi(x_{k_j}, \mu_{k_j}) \le \Phi(x_l, \mu_l) - \frac{\eta_1\nu\theta}{\kappa_{\rm n}}\sum_{i=0}^{j-1}\frac{1}{\sqrt{\sigma_{k_i}}}. \tag{48}$$

But as $x_{k_j} \notin \mathcal{O}_l$, (38) and (40) imply that

$$\frac{\|c(x_l)\|}{2L_1} \le \|x_{k_j} - x_l\| \le \sum_{i=0}^{j-1}\|x_{k_{i+1}} - x_{k_i}\| \le 4(1+\sqrt{3M_g})\sum_{i=0}^{j-1}\frac{1}{\sqrt{\sigma_{k_i}}}, \tag{49}$$

and then (48) gives

$$\Phi(x_{k_j}, \mu_{k_j}) \le \Phi(x_l, \mu_l) - \frac{\eta_1\nu\theta\|c(x_l)\|}{8(1+\sqrt{3M_g})L_1\kappa_{\rm n}}. \tag{50}$$

Thus in all cases (46) and (50) ensure that

$$\Phi(x_{k_j}, \mu_{k_j}) \le \Phi(x_l, \mu_l) - \kappa_c\|c(x_l)\|, \tag{51}$$

where $\kappa_c = \frac{1}{2}\eta_1\nu\max\left\{1, \frac{\theta}{4(1+\sqrt{3M_g})L_1\kappa_{\rm n}}\right\}$. (51) proves that $\|c(x_l)\|$ cannot be bounded away from zero since l is arbitrary and $\{\Phi(x_{k_j}, \mu_{k_j})\}$ is decreasing and bounded from below from Lemma 5.

3.3 Optimality

The feasibility of the algorithm has been proved in Theorem 1. Then we turn to the optimality condition.

Lemma 7 *Suppose that assumptions (G1)-(G3) hold and that* (21) *holds for all k sufficiently large. Then there are constants $\mu_{\max} > 0$ and $\kappa_{\rm bdm} \ge 0$, and an index k_1 such that*

$$\Delta m_k \ge \Delta m_k^H + (\mu_k - \kappa_{\rm mk})\Delta m_k^N. \tag{52}$$

and

$$\mu_k = \mu_{\max}, \tag{53}$$

for all $k \ge k_1$.

Proof (10) implies that μ_k satisfies

$$\Delta m_k \geq \nu\mu_k \Delta m_k^N, \tag{54}$$

Since assumptions (G3) and (G1) imply that g_k and B_k are bounded, (21) gives that $n_k \leq \Delta m_k^N/\kappa_{\text{mn}}$ and $\Delta m_k^N \leq \|c(x_k)\|$, and $c(x_k)$ is bounded by (G1), thus we have that

$$\Delta m_k^F \geq -\kappa_{\text{mk}}\Delta m_k^N,$$

for some constant $\kappa_{\text{mk}} > 0$. Then (54) immediately gives (52), from which we obtain that

$$\Delta m_k \geq (\mu_k - \kappa_{\text{mk}})\Delta m_k^N,$$

due to $\Delta m_k^H > 0$.

Hence (54) is evidently satisfied for all $\mu_k \geq \frac{\kappa_{\text{mk}}}{1-\nu}$. In view of (14), every increase of an insufficient μ_{k-1} must be by at least $\max\{(\tau_1 - 1)\mu_0, \tau_2\}$, and therefore the penalty parameter can only be increased a finite number of times. The required value $\mu_{\max}$ is at most the first value of μ_k that exceeds $\frac{\kappa_{\text{mk}}}{1-\nu}$, which proves the lemma.

Lemma 8 *Suppose that assumptions (G1)-(G3) hold and that* (21) *holds for all k sufficiently large. Then there is a constant $\kappa_{\text{btm}} > 0$ and an index k_1 such that*

$$\Delta m_k \geq \kappa_{\text{btm}}\Delta m_k^H, \tag{55}$$

for all $k \geq k_1$.

Proof Consider $k \geq k_1$ as in Theorem 7, in which case $\mu_k = \mu_{\max}$, and let

$$\kappa_{\text{btm}} \stackrel{\text{def}}{=} \frac{1}{2}\min\left\{1, \frac{\nu\mu_{\max}}{\kappa_{\text{mk}} - \mu_{\max}}\right\}.$$

If

$$-\frac{1}{2}\Delta m_k^H \leq (\mu_{\max} - \kappa_{\text{mk}})\Delta m_k^N.$$

In this event (52) implies that $\Delta m_k \geq \frac{1}{2}\Delta m_k^H$, thus proves that (55).

If

$$-\frac{1}{2}\Delta m_k^H > (\mu_{\max} - \kappa_{\text{mk}})\Delta m_k^N.$$

Since both Δm_k^H and Δm_k^N are nonnegative, it must be that $\mu_{\max} \leq \kappa_{\text{mk}}$, in which case (10) and (11) imply that

$$\Delta m_k > \frac{\nu\mu_{\max}}{2(\kappa_{\text{mk}} - \mu_{\max})}\Delta m_k^H.$$

Thus the lemma is proved.

Lemma 9 *Assume that* (19) *holds. Suppose that x_k is not a first-order critical point for* (1)*. Then there exists a $\sigma_k^0 > 0$ such that $\rho_k \geq \eta_2$ if $\sigma_k \geq \sigma_k^0$.*

Proof Consider two cases, $c(x_k) \neq 0$ and $c(x_k) = 0$.

Firstly, suppose that $c(x_k) \neq 0$. Then (10) and (19) imply that

$$\Delta m_k \geq \nu\mu_k \min\left\{\|c(x_k)\|, \frac{\theta}{\kappa_{\rm n}\sqrt{\sigma_k}}\right\}. \tag{56}$$

Let

$$\sigma_k^0 = \max\left\{\left(\frac{\kappa_{\rm m}\kappa_{\rm n}(1+\mu_{-1})}{\nu\theta\mu_{-1}(1-\eta_2)}\right)^2, \left(\frac{\theta}{\kappa_{\rm n}\|c(x_k)\|}\right)^2\right\}. \tag{57}$$

If $\sigma_k \geq \sigma_k^0$, (56) and (57) show that

$$\Delta m_k \geq \nu\mu_k \frac{\theta}{\kappa_{\rm n}\sqrt{\sigma_k}},$$

which combines (37) and the fact $\{\sigma_k\}$ is nondecreasing to give

$$\begin{aligned} 1-\rho_k &= \frac{\phi(x_k+p_k,\mu_k) - m(x_k,B_k,\mu_k,p_k)}{\Delta m_k} \\ &\leq \frac{\kappa_{\rm m}\kappa_{\rm n}(1+\mu_k)}{\nu\theta\mu_k\sqrt{\sigma_k}} \leq \frac{\kappa_{\rm m}\kappa_{\rm n}(1+\mu_{-1})}{\nu\theta\mu_{-1}\sqrt{\sigma_k}} \\ &\overset{(57)}{\leq} 1-\eta_2, \end{aligned}$$

which shows that $\rho_k \geq \eta_2$ and $\sigma_{k+1} \leq \sigma_k$.

Now suppose that $c(x_k) = 0$. Then, we have that $n_k = 0$ and so $\Delta m_k = Deltam_k^H$. It then follows from Lemma 3 that

$$\Delta m_k \geq \kappa_{\rm mt}\|g_k^N\|^{\frac{3}{2}} \min\left\{\frac{\sqrt{\|g_k^N\|}}{1+\|B_k^N\|}, \frac{1}{2\sqrt{\sigma_k}}\right\}. \tag{58}$$

Let l_k be the index of the last iteration such that $c(x_{l_k}) \neq 0$. Then

$$\mu_k \leq \bar{\mu} \overset{\rm def}{=} \max\{\tau_1\mu_{l_k}, \mu_{l_k}+\tau_2\}. \tag{59}$$

Note that $\|g_k^N\| \neq 0$ since x_k is not a first-order critical point and $c(x_k) = 0$.

Let

$$\sigma_k^0 = \max\left\{\left(\frac{1+\kappa_B}{2\sqrt{\|g_k^N\|}}\right)^2, \left(\frac{2\kappa_{\rm m}(1+\bar{\mu})}{\kappa_{\rm mt}(1-\eta_2)\|g_k^N\|^{\frac{3}{2}}}\right)^2\right\}. \tag{60}$$

Then, if $\sigma_k \geq \sigma_k^0$, we can obtain that

$$
\begin{aligned}
1-\rho_k &= \frac{\phi(x_k+p_k,\mu_k)-m(x_k,B_k,\mu_k,p_k)}{\Delta m_k} \\
&\overset{(37),(58)}{\leq} \frac{2\kappa_{\mathrm{m}}(1+\mu_k)}{\kappa_{\mathrm{mt}}\|g_k^N\|^{\frac{3}{2}}\sqrt{\sigma_k}} \overset{(59),(60)}{\leq} \frac{2\kappa_{\mathrm{m}}(1+\bar{\mu})}{\kappa_{\mathrm{mt}}\|g_k^N\|^{\frac{3}{2}}\sqrt{\sigma_k}} \\
&\leq 1-\eta_2.
\end{aligned}
$$

Then, we also obtain that $\rho_k \geq \eta_2$ and $\sigma_{k+1} \leq \sigma_k$.

The following results are given based on the above conclusions.

Theorem 2 *Suppose that* (27)*, assumptions (G1)-(G5) hold. Then*

$$\liminf_{k\to\infty} \|N_k^T g_k\| = 0.$$

Proof Suppose that $N_k^T g_k$ is bounded away from zero, that is

$$\|N_k^T g_k\| \geq 2\epsilon > 0 \tag{61}$$

for some constant $\epsilon > 0$ and for all k. $N_k^T B_k n_k$ tends to zero from Theorem 1 and assumption (G3). Thus it follows from (61) and the definition of g_k^N that

$$\|g_k^N\| \geq \epsilon > 0 \tag{62}$$

for all $k \geq k_2$.
The assumption (G3), (55) and (27) imply that

$$
\begin{aligned}
\Delta m_k &\geq \kappa_{\mathrm{btm}}\kappa_{\mathrm{mt}}\epsilon \min\left\{\frac{\|g_k^N\|}{\kappa_{\mathrm{Bn}}}, \frac{1}{2}\sqrt{\frac{\|g_k^N\|}{\sigma_k}}\right\} &(63)\\
&\geq \kappa_{\mathrm{btm}}\kappa_{\mathrm{mt}}\epsilon \min\left\{\frac{\epsilon}{\kappa_{\mathrm{Bn}}}, \frac{1}{2}\sqrt{\frac{\epsilon}{\sigma_k}}\right\}, &(64)
\end{aligned}
$$

for all $k \geq k_3 \overset{\mathrm{def}}{=} \max\{k_1, k_2\}$, where $\kappa_{\mathrm{Bn}} = \kappa_B + 1$. Let,

$$\sigma^0 \overset{\mathrm{def}}{=} \max\left\{\left(\frac{2\sqrt{\epsilon}}{\kappa_{\mathrm{Bn}}}\right)^{-2}, \left(\frac{(1-\eta_2)\kappa_{\mathrm{btm}}\kappa_{\mathrm{mt}}\epsilon^{\frac{3}{2}}}{2\kappa_{\mathrm{m}}(1+\mu_{\max})}\right)^{-2}\right\}. \tag{65}$$

And suppose that $\sigma_k \geq \sigma^0$, then (65) gives

$$\frac{\epsilon}{\kappa_{\mathrm{Bn}}} \geq \frac{1}{2}\sqrt{\frac{\epsilon}{\sigma_k}},$$

and hence (64) gives

$$\Delta m_k \geq \frac{\kappa_{\mathrm{btm}}\kappa_{\mathrm{mt}}\epsilon^{\frac{3}{2}}}{2}\frac{1}{\sqrt{\sigma_k}}.$$

Combining this with (37), we obtain that

$$1-\rho_k = \frac{\phi(x_k+p_k,\mu_k)-m(x_k,B_k,\mu_k,p_k)}{\Delta m_k} \le \frac{2\kappa_{\rm m}(1+\mu_{\max})}{\kappa_{\rm btm}\kappa_{\rm mt}\epsilon^{\frac{3}{2}}}\frac{1}{\sqrt{\sigma_k}}.$$

The fact that $\sigma_k \ge \sigma^0$ and (65) imply that $\rho_k \ge \eta_2$. Thus the iteration will be very successful whenever $k \ge k_3$ and $\sigma_k \ge \sigma^0$.

We now prove that

$$\sigma_k < \gamma_2\sigma^0, \tag{66}$$

for all $k \ge k_3$.

Suppose that (66) is false. Let k be the first iteration such that $\sigma_{k+1} \ge \gamma_2\sigma^0$. And $\gamma_2\sigma_k \ge \sigma_{k+1}$ follows from (15), therefore $\sigma_k \ge \sigma^0$. But the iteration k must have been very successful as we concluded at the end of the last paragraph, thus $\sigma_k \ge \sigma_{k+1}$. This contradicts the assumption that k is the first iteration such that $\sigma_{k+1} \ge \gamma_2\sigma^0$. Hence, the inequality (66) must hold. So from (64) and (66), we obtain that

$$\Delta m_k \ge s_3 \stackrel{\rm def}{=} \kappa_{\rm btm}\kappa_{\rm mt}\epsilon \min\left\{\frac{\epsilon}{\kappa_{\rm Bn}}, \frac{1}{2}\sqrt{\frac{\epsilon}{\gamma_1\sigma^0}}\right\}.$$

For all $k \in \mathcal{S} \ge k_3$,

$$\phi(x_k,\mu_{\max}) - \phi(x_{k+1},\mu_{\max}) \ge \eta_1\Delta m_k \ge \eta_1 s_3,$$

summing the above over the set of successful iteration between iteration k_3 and k, and letting τ_k be the number of such successful iteration, we deduce that

$$\phi(x_k,\mu_{\max}) \le \phi(x_{k_3},\mu_{\max}) - \tau_k\eta_1 s_3.$$

Since assumption (G1) implies that $\{\phi(x,\mu_{\max})\}$ is bounded from below, we obtain that τ_k is bounded. But this is impossible, since this would imply that there is an index l such that $x_k = x_l$ for all $k \ge l$ after which the cubic regularisation parameter σ_k must converge to infinity, which is not compatible with Lemma 9. Thus our initial hypothesis (61) is false and the conclusion of this theorem holds.

A stronger conclusion can obtain under the following assumption.
(**G6**) $N_k^T g_k$ is uniformly continuous on the sequence of iterates $\{x_k\}$, namely, $\|N_{t_i}^T g_{t_i} - N_{l_i}^T g_{l_i}\| \to 0$, whenever $\|x_{t_i} - x_{l_i}\| \to 0$, $i \to \infty$, where $\{x_{t_i}\}$ and $\{x_{l_i}\}$ are subsequence of $\{x_k\}$.

Theorem 3 *Suppose that assumptions(G1)-(G6) hold. Then*

$$\lim_{k\to\infty} \|N_k^T g_k\| = 0.$$

Proof Assume that there is an infinite subsequence $\{t_i\} \subset \mathcal{S}$ such that

$$\|N_{t_i}^T g_{t_i}\| \geq 3\epsilon, \tag{67}$$

for some $\epsilon > 0$ and for all i. Theorem 7 ensures that for each t_i, there is a first successful iteration $l_i > t_i$ such that $\|N_{l_i}^T g_{l_i}\| < 2\epsilon$. Thus $\{l_i\} \subset \mathcal{S}$ and for all i and $t_i \leq k < l_i$,

$$\|N_k^T g_k\| \geq 2\epsilon, \text{ and } \|N_{l_i}^T g_{l_i}\| < 2\epsilon. \tag{68}$$

Similar to (62), we have that

$$\|g_k^N\| \geq \epsilon > 0. \tag{69}$$

Let $\mathcal{K} \stackrel{\text{def}}{=} \{k \in \mathcal{S} : t_i \leq k < l_i\}$, where $\{t_i\}$ and $\{l_i\}$ were defined above. Since $\mathcal{K} \in \mathcal{S}$, it follows from (27), Assumption (G3), and (55) that

$$\Delta m_k \geq \kappa_{\text{btm}} \kappa_{\text{mt}} \epsilon \min\left\{\frac{\epsilon}{\kappa_{\text{Bn}}}, \frac{1}{2}\sqrt{\frac{\|g_k^N\|}{\sigma_k}}\right\}, \; k \in \mathcal{K}. \tag{70}$$

From (35), (69), (70), we have for all $k \in \mathcal{K}$ that,

$$\Phi(x_k, \mu_k) - \Phi(x_{k+1}, \mu_{k+1}) \geq \frac{\eta_1 \kappa_{\text{btm}} \kappa_{\text{mt}} \epsilon^{\frac{3}{2}}}{\mu_k} \min\left\{\frac{\epsilon}{\kappa_{\text{Bn}}}, \frac{1}{2\sqrt{\sigma_k}}\right\}. \tag{71}$$

Since $\{\Phi(x_k, \mu_k)\}$ is monotonically decreasing and bounded from below, (71) implies that

$$\frac{1}{2\sqrt{\sigma_k}} \to 0, \quad k \to \infty, \quad k \in \mathcal{K}. \tag{72}$$

From (72), (71) gives that

$$\Phi(x_k, \mu_k) - \Phi(x_{k+1}, \mu_{k+1}) \stackrel{(53)}{\geq} \frac{\eta_1 \kappa_{\text{btm}} \kappa_{\text{mt}} \epsilon^{\frac{3}{2}}}{2\mu_{\max}\sqrt{\sigma_k}}, \tag{73}$$

for all sufficiently large $k \in \mathcal{K}$. From (68) and (72), we have that (38) is satisfied when $\mathcal{I} := \mathcal{K}$, and thus

$$\Phi(x_k, \mu_k) - \Phi(x_{k+1}, \mu_{k+1}) \geq \frac{\eta_1 \kappa_{\text{btm}} \kappa_{\text{mt}} \epsilon^{\frac{3}{2}}}{2\mu_{\max}(1 + \sqrt{3}M_g)} \|p_k\|,$$

for all $t_i \leq k < l_i$, $k \in \mathcal{S}$, i sufficiently large. Summing the above between iteration t_i and l_i, and employing the triangle inequality, we deduce that,

$$\begin{aligned} \frac{2\mu_{\max}(1 + \sqrt{3}M_g)}{\eta_1 \kappa_{\text{btm}} \kappa_{\text{mt}} \epsilon^{\frac{3}{2}}} [\Phi(x_{t_i}, \mu_{t_i}) - \Phi(x_{l_i}, \mu_{l_i})] &\geq \sum_{k=t_i, k \in \mathcal{S}}^{l_i - 1} \|p_k\| \\ &= \sum_{k=t_i, k \in \mathcal{S}}^{l_i - 1} \|x_{k+1} - x_k\| \\ &= \|x_{l_i} - x_{t_i}\|, \end{aligned} \tag{74}$$

for all i sufficiently large. Since $\{\Phi(x_k, \mu_k)\}$ is convergent, thus

$$\lim_{i\to\infty} \|\Phi(x_{t_i}, \mu_{t_i}) - \Phi(x_{l_i}, \mu_{l_i})\| = 0,$$

and then

$$\lim_{i\to\infty} \|x_{l_i} - x_{t_i}\| = 0.$$

From assumption (G6),

$$\lim_{i\to\infty} \|N_{l_i}^T g_{l_i} - N_{t_i}^T g_{t_i}\| = 0, \tag{75}$$

but from (67) and (68),

$$\|N_{l_i}^T g_{l_i} - N_{t_i}^T g_{t_i}\| \geq \|N_{t_i}^T g_{t_i}\| - \|N_{l_i}^T g_{l_i}\| \geq \epsilon, \tag{76}$$

this is in contradiction with (75). Then the theorem is proved.

4 Numerical Results

In this section, we present the numerical results of Algorithm 1 which have been performed on a desktop with Intel(R) Core(TM) i7-6700 CPU @ 3.40GHz. Algorithm 1 is implemented as a MATLAB code and run under MATLAB(R2018a).

In the implementation, the regularisation parameter is updated by the following rules

$$\sigma_{k+1} = \begin{cases} \max\{\min\{\sigma_k, \|N_k^T g_k\|\}, \epsilon_m\} & \text{if } \rho_k \geq \eta_2, \\ \sigma_k & \text{if } \rho_k \in [\eta_1, \eta_2), \\ 2\sigma_k & \text{if } \rho_k < \eta_1, \end{cases} \tag{77}$$

where the $\epsilon_m = 10^{-16}, \eta_1 = 0.01, \eta_2 = 0.9$.

Other parameters are set as follows.

$$\nu = 10^{-4}, \tau_1 = 2, \tau_2 = 1.$$

The computation terminates when $\|c(x_k)\| \leq \epsilon$ and $\|N_k^T g_k\| \leq \epsilon = 10^{-6}$ are both satisfied. The test problems are from CUTEst collection [17] and the results are reported in Table 1, where NF and NC are the numbers of computation of the objective function and the numbers of computation of the constraints, respectively.

Table 1: Numerical results of the Algorithm 1

Problem	Dimension		NF,NC	$\|c(x_k)\|$	$\|N_k^T g_k\|$
	n	m			

Table 1 continued

Problem	Dimension n	Dimension m	NF,NC	$\|c(x_k)\|$	$\|N_k^T g_k\|$
AIRCRFTA	8	5	2	1.5932e-08	0
ARGTRIG	1000	1000	3	5.9215e-12	0
ARTIF	1002	1002	11	1.1690e-07	0
BDVALUE	1002	1000	2	6.4590e-08	0
BOOTH	2	2	4	0	0
BROYDN3D	1000	1000	11	2.4647e-07	0
BT1	2	1	5	2.8036e-08	0
BT2	3	1	14	6.5725e-14	4.2043e-15
BT3	5	3	14	0	5.9346e-08
BT4	3	2	8	2.2062e-12	2.7148e-07
BT5	3	2	5	8.2439e-07	3.2411e-09
BT6	5	2	12	8.9752e-07	6.3805e-13
BT7	5	3	13	4.5025e-12	3.7409e-08
BT8	5	2	11	3.3717e-07	1.5543e-15
BT9	4	2	13	2.5720e-08	1.3700e-07
BT10	2	2	6	4.4039e-09	0
BT11	5	3	10	5.6118e-12	3.6854e-08
BT12	5	3	8	6.4062e-08	9.0762e-10
BYRDSPHR	3	2	10	7.7791e-11	4.6885e-13
CLUSTER	2	2	10	4.8201e-08	0
GENHS28	10	8	5	1.6653e-16	2.5622e-13
GOTTFR	2	3	7	1.1392e-09	0
HATFLDF	3	3	5	8.7862e-07	0
HIMMELBA	2	2	5	0	0
HIMMELBC	2	2	5	2.5104e-09	0
HIMMELBE	3	3	4	0	0
HS06	2	1	10	4.1300e-13	5.8965e-08
HS07	2	1	7	1.7626e-09	1.4652e-08
HS08	2	2	5	2.5873e-07	0
HS09	2	1	9	1.7764e-15	5.4561e-07
HS26	3	1	18	5.4108e-12	7.7635e-07
HS27	3	1	20	2.4577e-08	4.9757e-07
HS28	3	1	3	0	1.9907e-06
HS39	4	2	8	1.0916e-08	1.6396e-07
HS40	4	3	3	1.7206e-07	3.1065e-07
HS42	4	2	3	5.6563e-09	8.6790e-09
HS46	5	2	18	7.6410e-07	2.4454e-08
HS47	5	3	16	2.8813e-07	5.2673e-07
HS48	5	2	4	9.9301e-16	4.4598e-13
HS49	5	2	16	8.8818e-16	3.5702e-07
HS50	5	3	9	8.8818e-16	1.1203e-16
HS51	5	3	2	0	1.4494e-11
HS52	5	3	5	3.9252e-16	2.4474e-10

Table 1 continued

Problem	Dimension		NF,NC	$\|c(x_k)\|$	$\|N_k^T g_k\|$
	n	m			
HS56	7	4	10	4.8508e-07	2.4045e-12
HS61	3	2	7	6.7055e-08	4.3475e-09
HS77	5	2	9	7.3125e-11	9.0627e-10
HS78	5	3	8	2.8435e-07	6.5241e-14
HS79	5	3	4	8.1313e-09	3.9293e-09
HYPCIR	2	2	4	1.9989e-10	0
MARATOS	2	1	5	3.0114e-12	1.0900e-11
MWRIGHT	5	3	7	5.3368e-14	1.9160e-13
POWELLBS	2	2	12	7.8350e-07	0
RECIPE	3	3	8	1.7347e-19	0
S216	2	1	8	4.1389e-13	6.9983e-10
S219	4	2	27	1.1443e-09	1.4894e-12
S235	3	1	12	5.4818e-09	1.8671e-08
S252	3	1	20	1.7635e-07	3.8376e-09
S254	3	2	5	4.1170e-07	1.1735e-08
S269	5	3	9	1.1102e-16	2.0190e-07
S316	2	1	3	3.9328e-07	0
S317	2	1	5	9.3695e-08	9.0262e-07
S318	2	1	5	4.8702e-09	8.7080e-08
S319	2	1	6	1.6644e-12	3.0013e-11
S320	2	1	8	7.1054e-15	6.3238e-13
S321	2	1	9	7.8773e-12	2.9117e-09
S322	2	1	16	1.4133e-11	5.0594e-07
S336	3	2	8	2.2204e-15	1.5721e-13
S338	3	2	7	3.5279e-11	5.6554e-12
S344	3	1	13	7.3896e-13	9.4816e-07
S345	3	1	22	5.3824e-13	5.5569e-07
S373	9	6	11	2.5657e-10	1.1464e-10
S378	10	3	27	7.4865e-09	8.0702e-08
S394	20	1	20	1.4158e-11	8.4186e-13
S395	50	1	27	8.6207e-11	3.7018e-12
ZANGWIL3	3	3	2	1.4606e-14	0

5 Conclusion

We have introduced a sequential adaptive regularisation using cubics (SARC) algorithm for solving nonlinear equality constrained programming. In the presented algorithm, a specific form of the step is constructed based on composite-step method. More precisely, the trial step is decomposed to be the sum of a normal and tangential step, where the former has a specific form depending on the Jacobian and the constraint function and the latter belongs to the null space of the transposed Jacobian and is calculated by solving a standard ARC subproblem. The exact penalty function is used as the merit function to decide wether the trial step is accepted. Global convergence is proved under some suitable assumptions and preliminary numerical results are given.

Acknowledgements The authors gratefully acknowledge the supports of the National Natural Science Foundation of China (12071133), Key Scientific Research Project for Colleges and Universities in Henan Province (21A110012).

Conflict of interest

The authors declare that they have no conflict of interest.

Data availability statements

There is no data generated or analysed in this paper.

References

1. Agarwal, N., Boumal, N., Bullins, B., Cartis, C.: Adaptive regularization with cubics on manifolds. Math. Program. **188**(1, Ser. A), 85–134 (2021)
2. Bellavia, S., Gurioli, G., Morini, B.: Adaptive cubic regularization methods with dynamic inexact Hessian information and applications to finite-sum minimization. IMA J. Numer. Anal. **41**(1), 764–799 (2021)
3. Benson, H., Shanno, D.: Cubic regularization in symmetric rank-1 quasi-Newton methods. Mathematical Programming Computation **10**(4), 457–486 (2018)
4. Benson, H.Y., Shanno, D.F.: Interior-point methods for nonconvex nonlinear programming: cubic regularization. Comput. Optim. Appl. **58**(2), 323–346 (2014)
5. Bergou, E., Diouane, Y., Gratton, S.: On the use of the energy norm in trust-region and adaptive cubic regularization subproblems. Comput. Optim. Appl. **68**(3), 533–554 (2017)
6. Bergou, E., Youssef, D., Serge, G.: A line-search algorithm inspired by the adaptive cubic regularization framework and complexity analysis. Journal of Optimization Theory and Applications **178**(3), 885–913 (2018)
7. Bianconcini, T., Sciandrone, M.: A cubic regularization algorithm for unconstrained optimization using line search and nonmonotone techniques. Optim. Methods Softw. **31**(5), 1008–1035 (2016)
8. Birgin, E., Martínez, J.: On regularization and active-set methods with complexity for constrained optimization. SIAM Journal on Optimization **28**(2), 1367–1395 (2018)
9. Cartis, C., Gould, N., Toint, P.: Adaptive cubic regularisation methods for unconstrained optimization. Part I: motivation, convergence and numerical results. Mathematical Programming **127**(2), 245–295 (2011)
10. Cartis, C., Gould, N., Toint, P.: Adaptive cubic regularisation methods for unconstrained optimization. Part II: worst-case function- and derivative-evaluation complexity. Mathematical Programming **130**(2), 295–319 (2011)
11. Cartis, C., Gould, N., Toint, P.: A concise second-order complexity analysis for unconstrained optimization using high-order regularized models. Optimization Methods & Software **35**(2), 243–256 (2020)
12. Cartis, C., Gould, N.I., Toint, P.L.: Universal regularization methods: varying the power, the smoothness and the accuracy. SIAM J. Optim. **29**(1), 595–615 (2019)

13. Cartis, C., Gould, N.I.M., Toint, P.L.: An adaptive cubic regularization algorithm for nonconvex optimization with convex constraints and its function-evaluation complexity. IMA J. Numer. Anal. **32**(4), 1662–1695 (2012)
14. Cartis, C., Gould, N.I.M., Toint, P.L.: On the evaluation complexity of constrained nonlinear least-squares and general constrained nonlinear optimization using second-order methods. SIAM J. Numer. Anal. **53**(2), 836–851 (2015)
15. Dehghan Niri, T., Heydari, M., Hosseini, M.M.: An improvement of adaptive cubic regularization method for unconstrained optimization problems. Int. J. Comput. Math. **98**(2), 271–287 (2021)
16. Dussault, J.P.: $\mathrm{ARC_q}$: a new adaptive regularization by cubics. Optim. Methods Softw. **33**(2), 322–335 (2018)
17. Gould, N.I.M., Orban, D., Toint, P.L.: Cutest: a constrained and unconstrained testing environment with safe threads for mathematical optimization. Comput. Optim. Appl. **60**(3), 545–557 (2015)
18. Gould, N.I.M., Porcelli, M., Toint, P.L.: Updating the regularization parameter in the adaptive cubic regularization algorithm. Comput. Optim. Appl. **53**(1), 1–22 (2012)
19. Hsia, Y., Sheu, R.L., Yuan, Y.x.: Theory and application of p-regularized subproblems for $p > 2$. Optim. Methods Softw. **32**(5), 1059–1077 (2017)
20. Jiang, R., Yue, M.C., Zhou, Z.: An accelerated first-order method with complexity analysis for solving cubic regularization subproblems. Comput. Optim. Appl. **79**(2), 471–506 (2021)
21. Lieder, F.: Solving large-scale cubic regularization by a generalized eigenvalue problem. SIAM J. Optim. **30**(4), 3345–3358 (2020)
22. Lubkoll, L., Schiela, A., Weiser, M.: An affine covariant composite step method for optimization with PDEs as equality constraints. Optim. Methods Softw. **32**(5), 1132–1161 (2017)
23. Martínez, J.M.: On high-order model regularization for constrained optimization. SIAM J. Optim. **27**(4), 2447–2458 (2017)
24. Martínez, J.M., Raydan, M.: Cubic-regularization counterpart of a variable-norm trust-region method for unconstrained minimization. J. Global Optim. **68**(2), 367–385 (2017)
25. Park, S., Jung, S.H., Pardalos, P.M.: Combining stochastic adaptive cubic regularization with negative curvature for nonconvex optimization. J. Optim. Theory Appl. **184**(3), 953–971 (2020)
26. Sha, L., Wei, Z., Li, L.: A trust region algorithm with adaptive cubic regularization methods for nonsmooth convex minimization. Computational Optimization and Applications. An International Journal **51**(2), 551–573 (2012)
27. Vardi, A.: A trust region algorithm for equality constrained minimization: convergence properties and implementation. SIAM J. Numer. Anal. **22**(3), 575–591 (1985)
28. Zhao, T., Liu, H., Liu, Z.: New subspace minimization conjugate gradient methods based on regularization model for unconstrained optimization. Numer. Algorithms **87**(4), 1501–1534 (2021)