Journal of Computational Mathematics
Vol.xx, No.x, 200x, 1–28.

http://www.global-sci.org/jcm
doi:??

A LINE SEARCH FILTER SEQUENTIAL ADAPTIVE CUBIC REGULARISATION ALGORITHM FOR NONLINEARLY CONSTRAINED OPTIMIZATION*

Yonggang Pei
Henan Provincial Engineering Laboratory for Big Data Statistical Analysis and Optimal Control, College of Mathematics and Information Science, Henan Normal university, Xinxiang 453007, China
Email: peiyg@163.com
Shaofang Song
School of Mathematics and Statistics, Southwest University, Chongqing 400715, China
Email: ssf731713505@163.com
Yanyan Chen
Henan Provincial Engineering Laboratory for Big Data Statistical Analysis and Optimal Control, College of Mathematics and Information Science, Henan Normal university, Xinxiang 453007, China
Email: cyy1610403053@163.com
Detong Zhu
Mathematics and Science College, Shanghai Normal University, Shanghai 200234, China
Email: dtzhu@shnu.edu.cn

Abstract

In this paper, a sequential adaptive regularization algorithm using cubics (ARC) is presented to solve nonlinear equality constrained optimization. It is motivated by the idea of handling constraints in sequential quadratic programming methods. In each iteration, we decompose the new step into the sum of the normal step and the tangential step by using composite step approaches. Using a projective matrix, we transform the constrained ARC subproblem into a standard ARC subproblem which generates the tangential step. After the new step is computed, we employ line search filter techniques to generate the next iteration point. Line search filter techniques enable the algorithm to avoid the difficulty of choosing an appropriate penalty parameter in merit functions and the possibility of solving ARC subproblem many times in one iteration in ARC framework. Global convergence is analyzed under some mild assumptions. Preliminary numerical results and comparison are reported.

Mathematics subject classification: 49M37, 65K05, 65K10, 90C30.
Key words: Nonlinear optimization, Cubic regularization, Global convergence, Line search filter, Sequential quadratic programming.

1. Introduction

Optimization algorithms play an important role in many fields, such as artificial intelligence, machine learning, signal processing, modeling design, transportation analysis, industry, structural engineering, economics, etc [7, 8]. Sequential quadratic programming (SQP) methods, which generate steps by solving quadratic subproblems, are among the most effective methods

* Received xxx / Revised version received xxx / Accepted xxx /

for nonlinearly constrained optimization. SQP methods have shown their strength when solving problems with significant nonlinearities in the constraints [1, 11, 18, 20] since it was first proposed by Wilson in [22].

SQP methods are often embedded in two fundamental strategies, line search and trust-region, to solve nonlinearly constrained optimization. Recently, a third alternative strategy, the adaptive regularization method using cubics (ARC), is presented. ARC was proposed by Cartis et al. [4] for solving unconstrained optimization. It can be viewed as an adaptive version of the cubic regularization of the Newton's method which was proposed by Griewank [14] and its global convergence rates were first established by Nesterov and Polyak [17]. Benson and Shanno [3] also described the development of cubic regularization methods. Recently, Bellavia, Gurioli, Morini, and Toint proposed an adaptive regularization method for nonconvex optimization using inexact function values and randomly perturbed derivatives [2]. ARC has shown its attractive convergence properties and promising numerical experiments performance [5] for solving unconstrained optimization. It can also be viewed as a non-standard trust region method while it offers an easy way to avoid difficulties resulting from the incompatibility of the intersection of linearized constraints with trust-region bounds in constrained optimization.

In this paper, we consider how to extend ARC to solve nonlinearly constrained optimization by referring to the idea of SQP methods, and propose a penalty-free sequential adaptive cubic regularization algorithm. In each iteration, two problems should be addressed. One is the computation of a new step. The other is the decision of a new iteration point by using the new step. To obtain a new step, we need to deal with an ARC subproblem with linearized constraints. Composite step approaches are utilized to compute the new step which is decomposed into the sum of the normal step and the tangential step [7]. The normal step is computed firstly and aims to reduce the constraint violation degree while it satisfies the linearized constraints. The tangential step is used to present sufficient decrease of the model. Using a projective matrix, we can transform the constrained ARC subproblem into a standard ARC subproblem which generates the tangential step. This projective matrix allows us to avoid the computational difficulties caused by poor choice of basis of null space of the Jacobian of the constraints in reduced Hessian methods.

After the new step is computed, we employ line search filter techniques in [21] to generate the next iteration point. Line search filter techniques can help us avoid the difficulty of choosing an appropriate penalty parameter in merit functions and the possibility of solving ARC subproblem many times in one iteration in ARC framework. So, the proposed algorithm is also called as line search filter sequential adaptive regularization algorithm using cubics (LsFSARC). The adaptive parameter in ARC is adjusted by the ratio of the reduction of the objective function to the reduction of the model. This is different from the standard ARC methods, where the reduction ratio is used to both update the adaptive parameter and decide the acceptance of the trial step. Global convergence is analyzed under some mild assumptions.

The following part of this paper is developed as follows. In Section 2, the computation of search direction is described and the filter sequential ARC algorithm combining line search for equality constrained optimization is developed as shown in Algorithm 2.1. The analysis of the global convergence to first-order critical point is presented in Section 3. Preliminary numerical results comparison are presented in Section 4 and conclusion is reported in Section 5.

Notation: Throughout this paper, we denote the transpose of a vector v by v^T and denote the transpose of a matrix A by A^T. Norms $\|\cdot\|$ denote the Euclidean norm and its compatible matrix norm. $|\mathcal{A}|$ denotes the number of elements in a set $\mathcal{A}$. Finally, we denote by $O(t_k)$ a sequence $\{v_k\}$ satisfying $\|v_k\| \leq \beta t_k$ for a constant $\beta > 0$ independent of k.

2. Line search filter sequential ARC algorithm

In this section, we focus on constructing an ARC algorithm combining line search filter technique to solve the nonlinear equality constrained optimization problem

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \qquad f(x) \tag{2.1a}$$

$$\text{subject to} \qquad c(x) = 0, \tag{2.1b}$$

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 with $m \leq n$. First, the process of computing the trial step is presented. Then, we give the acceptance mechanism for the trial step. The whole algorithm is reported in the end of this section.

Consider the constrained optimization problem (2.1). Let $A(x)$ denote the Jacobian matrix of $c(x)$, namely,

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

Assume that $A(x)$ has full row rank. We can define a projective matrix

$$P(x) := I - A(x)^T (A(x)A(x)^T)^{-1} A(x), \tag{2.2}$$

which is a projection onto the null space of $A(x)$. This projective matrix allows us to avoid the computational difficulties caused by poor choice of basis of null space of the Jacobian of the constraints in reduced Hessian methods.

Meanwhile, we use the same definition of Lagrange function in [19]

$$\ell(x) := f(x) - \lambda(x)^T c(x), \tag{2.3}$$

where $\lambda(x)$ is a projective version of the multiplier vector

$$\lambda(x) := (A(x)A(x)^T)^{-1} A(x) g(x) \in \mathbb{R}^m \tag{2.4}$$

with $g(x) := \nabla f(x)$ denoting the gradient of the objective function $f(x)$.

Hence, the KKT conditions can be expressed as

$$g(x) + A(x)y = 0, \quad c(x) = 0$$

for some $y \in \mathbb{R}^m$. Equivalently, the KKT conditions can be written as

$$P(x)g(x) = 0, \quad c(x) = 0. \tag{2.5}$$

To obtain search directions, we need to deal with the following subproblem which is similar to SQP methods in iteration k,

$$\underset{d \in \mathbb{R}^n}{\text{minimize}} \quad f_k + g_k^T d + \frac{1}{2} d^T H_k d + \frac{1}{3} \sigma_k \|d\|^3 \tag{2.6a}$$

$$\text{subject to} \quad A_k d + c(x_k) = 0, \tag{2.6b}$$

where $f_k := f(x_k)$, $g_k := \nabla f(x_k)$, $A_k := A(x_k)$, $c(x_k) := c(x_k)$, H_k denotes the Hessian of Lagrange function $\nabla_{xx}\ell(x_k)$ or its approximation and $\sigma_k \in \mathbb{R}^+$ is an adaptive parameter in ARC. In addition, we assume that A_k has full row rank for all k.

Instead of solving subproblem (2.6) directly, we decompose the overall step via composite methods as follows.

$$d_k = n_k + t_k,$$

where n_k is called as a normal step which is used to satisfy feasibility condition, and t_k is called as a tangential step for ensuring sufficient decrease of the function's model.

First, we can compute n_k by

$$n_k = -A_k^T (A_k A_k^T)^{-1} c(x_k). \tag{2.7}$$

Moreover, to ensure sufficient reduction in model function, we also require that the following condition

$$\|n_k\| \leq \beta_1 \min\left\{1, \frac{\beta_2}{\sqrt{\sigma_k}^{\beta_3}}\right\} \frac{1}{\sqrt{\sigma_k}}, \tag{2.8}$$

where fixed constants $\beta_1, \beta_2 > 0$ and $\beta_3 \in (0, 1)$.

However, (2.8) may not hold. So we distinguish two cases depending on that whether (2.8) holds. First, we consider the case (2.8) holds. The tangential step t_k is computed as

$$t_k = P_k u_k, \tag{2.9}$$

where $P_k := P(x_k)$ and u_k is the solution(or its approximation) of the following problem

$$\text{minimize} \quad f(x_k) + (P_k g_k)^T u + \frac{1}{2} u^T (P_k H_k P_k) u + \frac{1}{3} \sigma_k \|P_k u\|^3. \tag{2.10}$$

The above problem is constructed by using reduced Hessian methods.

After computing u_k, from (2.9) and the definition of P_k, we can define

$$m_k^t(t_k) := f(x_k) + g_k^T t_k + \frac{1}{2} t_k^T H_k t_k + \frac{1}{3} \sigma_k \|t_k\|^3. \tag{2.11}$$

Next, we discuss the mechanism of the acceptance of the trial point $x_k(\alpha_{k,l})$.

For the purpose of obtaining the next iteration x_{k+1}, we need to determine a step size α_k so that $x_{k+1} = x_k + \alpha_k d_k$. To this end, we use a backtracking line search procedure combining filter method where a decreasing sequence of step sizes $\alpha_{k,l} \in (0, 1] (l = 0, 1, 2, ...)$ with $\lim_{l \to \infty} \alpha_{k,l} = 0$ is tried until some acceptance rules are satisfied and the trial point is accepted by the current filter. For brevity, the trial point $x_k(\alpha_{k,l})$ is denoted by

$$x_k(\alpha_{k,l}) := x_k + \alpha_{k,l} d_k.$$

These acceptance rules are reported in detail as follows.

Define the constraint violation measure

$$h(x) := \|c(x)\|.$$

We use the same definition of a filter in [21] where the filter is defined as a set $\mathcal{F}_k \subseteq [0, \infty) \times \mathbb{R}$ containing all prohibited $(h(x_j), \ell(x_j))$ pairs in iteration k. At the beginning, we can set $\mathcal{F}_0 = \{(h, \ell) \in \mathbb{R}^2 : h > h(x_0)\}$.

A trial point $x_k(\alpha_{k,l})$ can be accepted only if it provides satisfying improvement of the infeasibility measure $h(x)$ or $\ell(x)$, i.e.,

$$h(x_k(\alpha_{k,l})) \leq (1 - \gamma_h)h(x_j) \quad \text{or} \quad \ell(x_k(\alpha_{k,l})) \leq \ell(x_j) - \gamma_\ell h(x_j) \tag{2.12}$$

holds for all $(h(x_j), \ell(x_j))$ in $\mathcal{F}_k$ with fixed constants $\gamma_h, \gamma_\ell \in (0, 1)$.

A trial point $x_k(\alpha_{k,l})$ is called to be acceptable to the filter $\mathcal{F}_k$ if

$$(\ell(x_k(\alpha_{k,l})), h(x_k(\alpha_{k,l}))) \notin \mathcal{F}_k.$$

This criterion indicates that, provided that $\{\ell(x_k)\}$ is monotonically decreasing and bounded below, a sequence $\{x_k\}$ is forced towards feasibility. However, this type of sequence $\{x_k\}$ could still be accepted even if it converges to a nonoptimal point. In order to prevent this from happening, define a model

$$m_k(\alpha) := \alpha g_k^T t_k + \frac{1}{2}\alpha^2 t_k^T H_k t_k + \frac{1}{3}\alpha^3 \sigma_k \|t_k\|^3 - \alpha(\nabla\lambda_k^T d_k)^T c(x_k), \tag{2.13}$$

where $\nabla\lambda_k := \nabla\lambda(x_k)$ denotes the gradient of $\lambda(x)$ at x_k. The following condition

$$\ell(x_k(\alpha_{k,l})) \leq \ell(x_k) + \mu m_k(\alpha_{k,l}) \tag{2.14}$$

is employed to be the acceptance criterion whenever the following switching conditions

$$m_k(\alpha_{k,l}) < 0 \quad \text{and} \quad (-m_k(\alpha_{k,l}))^\omega(\alpha_{k,l}\sqrt{\sigma_k})^{\omega-1} > \kappa_h(h(x_k))^\varsigma \tag{2.15}$$

hold for the current trial step size $\alpha_{k,l}$, where $0 < \mu < 1$, $\kappa_h > 0$, $\omega \geq 1$ and $\varsigma > 2$ are fixed constants.

In order to tackle the situation where no acceptable step can be found and the feasibility restoration procedure has to be started, we set a threshold

$$\alpha_k^{\min} := \begin{cases} \mu_\alpha \min\left\{\gamma_h, \frac{\gamma_l h(x_k)}{-g_k^T t_k + (\nabla\lambda_k^T d_k)^T c(x_k)}, \frac{\kappa_h [h(x_k)]^\phi \sigma_{k_j}^{1-\tau}}{[-g_k^T t_k + (\nabla\lambda_k^T d_k)^T c(x_k)]^\tau}\right\} & \text{if } \delta_k > 0, \\ \mu_\alpha \gamma_h & \text{otherwise,} \end{cases} \tag{2.16}$$

with $\delta_k := -g_k^T t_k + (\nabla\lambda_k^T d_k)^T c(x_k)$ and fixed constants $\mu_\alpha \in (0, 1]$, $\phi > 2, \tau \geq 1$. The algorithm goes to feasibility restoration procedure if $\alpha_{k,l} < \alpha_k^{\min}$.

Next we consider the case (2.8) does not hold. In this case, we use the same strategy as in [21]. That is, the algorithm relies on the feasibility restoration procedure, whose purpose is to generate a new iterate $x_{k+1} = x_k + r_k$ which is acceptable for $\mathcal{F}_k$ and satisfies (2.8), where r_k

is a solution of the following problem

$$\min_{x\in\mathbb{R}^n} h^2(x) \tag{2.17}$$

from x_k. For convenience, we denote the set $\mathcal{A} := \{k \mid x_k \text{ is added to the filter}\}$. One can see that $\mathcal{F}_k \subsetneqq \mathcal{F}_{k+1} \iff k \in \mathcal{A}$. Let $\mathcal{A}_{\text{inc}}$ be the set of all indices of those iterations in which the feasibility of restoration procedure is invoked when (2.8) does not hold.

Now, we are ready to state the linear search filter sequential adaptive regularisation algorithm with cubics for solving problem (2.1) as shown in Algorithm 2.1.

Algorithm 2.1 LsFSARC for nonlinear equality constrained optimization.

Step 0. Initialization.
(i) Given starting point x_0, an initial $\sigma_0 > 0$ such that $\sigma_{\min} \le \sigma_0$, an initial symmetric matrix H_0.
(ii) Set constants $1 < \gamma_1 \le \gamma_2$, $0 < \eta_1 < \eta_2 < 1$, $\beta_1 \in (0, 1]$, $\beta_2 > 0$, β_3, β, κ_h, γ_h, $\gamma_\ell \in (0, 1)$, $\varsigma > 2$, $\omega \ge 1$, $0 < \omega_1 \le \omega_2 < 1$.
(iii) Set the filter $\mathcal{F}_0 = \{(h, \ell) \in \mathbb{R}^2 : h \ge h_{\max} > h(x_0)\}$ and the iteration counter $k = 0$.
Step 1. Compute $f_k, g_k, h(x_k), A_k, \lambda_k := \lambda(x_k), H_k, P_k$.
Step 2. Stop if x_k is a stationary point of optimization problem (2.1), i.e., if it satisfies the KKT conditions (2.5).
Step 3. Compute n_k by (2.7). If (2.8) holds, compute t_k by (2.10) and set $d_k = n_k + t_k$. Otherwise, go to step 13.
Step 4. Compute $\alpha_k^{\min}$ with $\alpha_k^{\min}$ defined by (2.16). Set $\alpha_{k,0} = 1$ and $l = 0$.
Step 5. If $\alpha_{k,l} < \alpha_k^{\min}$, go to step 13. Otherwise, compute the new trial point $x_k(\alpha_{k,l}) = x_k + \alpha_{k,l} d_k$.
Step 6. If $(h(x_k(\alpha_{k,l})), \ell(x_k(\alpha_{k,l}))) \in \mathcal{F}_k$, reject $\alpha_{k,l}$ and go to step 10.
Step 7. If (2.15) holds, go to step 8. Otherwise go to step 9.
Step 8. If (2.14) holds, set $\alpha_k = \alpha_{k,l}$, $x_{k+1} = x_k(\alpha_k) = x_k + \alpha_k d_k$ and go to step 11. Otherwise, go to step 10.
Step 9. If (2.12) holds, set $\alpha_k = \alpha_{k,l}$, $x_{k+1} = x_k(\alpha_k) = x_k + \alpha_k d_k$, add x_k to the filter $\mathcal{F}_k$ and go to step 11. Otherwise, go to step 10.
Step 10. Choose $\alpha_{k,l+1} \in [\omega_1 \alpha_{k,l}, \omega_2 \alpha_{k,l}]$, set $l = l + 1$ and go back to step 5.
Step 11. If $m_k(\alpha_k) < 0$, compute

$$\rho_k := \frac{\ell(x_k + \alpha_k d_k) - \ell(x_k)}{m_k(\alpha_k)} \tag{2.18}$$

and set

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

Otherwise, set $\sigma_{k+1} \in (\gamma_1\sigma_k, \gamma_2\sigma_k]$.
Step 12. Set $k = k + 1$ and go to step 1.
Step 13. Feasibility restoration procedure.
13.1 Compute a new iterate point x_{k+1} by decreasing $h(x)$ for which x_{k+1} satisfies both (2.12) and $(h(x_{k+1}), \ell(x_{k+1})) \notin \mathcal{F}_k$.
13.2 Determine σ_{k+1} and go to step 12.

3. Global convergence

Assumptions G. Let $\{x_k\}$ be the sequence produced by Algorithm 2.1, where restoration iteration terminates successfully and the algorithm does not stop at a KKT point.
(G1) The iterations $\{x_k\} \subset X$, where X is a closed, bounded domain with $X \subset \mathbb{R}^n$.
(G2) $f(x)$ and $c(x)$ are differentiable on X, and $\nabla f(x)$ and $\nabla c(x)$ are Lipschitz-continuous over X.
(G3) There exists a constant $M_H > 0$ so that $\|H_k\| \leq M_H$ for all k.
(G4) H_k is semipositive definite on the null space of the Jocabian A_k for each k.
(G5) There exist constants $\delta_h, \kappa_n > 0$ such that if $h(x_k) \leq \delta_h$,

$$k \notin \mathcal{A}_{\text{inc}} \quad \text{and} \quad \|n_k\| \leq \kappa_n h(x_k). \tag{3.1}$$

(G6) There exists a constant $M_A > 0$ such that

$$\varrho_{\min}(A_k) \geq M_A$$

for $k \notin \mathcal{A}_{\text{inc}}$, where $\varrho_{\min}$ is the smallest singular value of A_k.

Using (G1), we can deduce that $\{\ell(x_k)\}$ is bounded below and $\{h(x_k)\}$ is bounded. Hence, there exist constants $\ell_{\min}$ and $h_{\max} > 0$ such that $\ell_{\min} \leq \ell(x_k)$ and $0 \leq h(x_k) \leq h_{\max}$ for all k.

From (G3), we can conclude that

$$\|P_k H_k P_k\| \leq \|P_k\|^2 \|H_k\| \leq M_H. \tag{3.2}$$

Moreover, from (G1) and (G6), one finds that there exist constants $M_\lambda, M_\lambda^{'} > 0$ such that for all k

$$\|\lambda_k\| \leq M_\lambda, \quad \|\nabla \lambda_k\| \leq M_\lambda^{'}. \tag{3.3}$$

3.1. Preliminary results

The next lemma provides the reduction in $f(x)$ predicted from the subproblem (2.10).

Lemma 3.1. *Suppose that $k \notin \mathcal{A}_{\text{inc}}$, and the step t_k^c is the Cauchy step for* (2.11). *Then*

$$f(x_k) - m_k^t(t_k) \geq \frac{\|P_k g_k\|}{6\sqrt{2}} \min\left\{ \frac{\|P_k g_k\|}{1 + \|H_k\|}, \frac{1}{2}\sqrt{\frac{\|P_k g_k\|}{\sigma_k}} \right\} \tag{3.4}$$

for all $k \geq 0$.

Proof. We can rewrite (2.11) as

$$m_k^t(t_k) = f(x_k) + (P_k g_k)^T t_k + \frac{1}{2} t_k^T H_k t_k + \frac{1}{3}\sigma_k \|t_k\|^3 \tag{3.5}$$

because of the definition of P_k. Hence, the Cauchy step t_k^c for (2.11) is

$$t_k^c = -\beta_k^c P_k g_k \;\; \text{and} \;\; \beta_k^c = \arg\min_{\beta \in \mathbb{R}^+} m_k^t(-\beta P_k g_k). \tag{3.6}$$

For any $\beta \geq 0$, combining the Cauchy-Schwarz and (3.6), one finds that

$$\begin{aligned} & f(x_k) - m_k^t(t_k) \\ \geq\ & f(x_k) - m_k^t(-\beta P_k g_k) \\ =\ & \beta \|P_k g_k\|^2 - \frac{1}{2}\beta^2 (P_k g_k)^T H_k P_k g_k - \frac{1}{3}\beta^3 \sigma_k \|P_k g_k\|^3 \\ \geq\ & \beta \|P_k g_k\|^2 (1 - \frac{1}{2}\beta \|H_k\| - \frac{1}{3}\beta^2 \sigma_k \|P_k g_k\|). \end{aligned} \tag{3.7}$$

Denote

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

Then for $\beta \in [0, \hat{\beta}_k]$, when $1 - \frac{1}{2}\beta\|H_k\| - \frac{1}{3}\beta^2 \sigma_k \|P_k g_k\| \geq 0$, it follows that $f(x_k) \geq m_k^t(t_k)$. Let

$$\theta_k := \frac{1}{\sqrt{2}\ \max\left\{ 1 + \|H_k\|, 2\sqrt{\sigma_k \|P_k g_k\|} \right\}}. \tag{3.8}$$

By employing the inequalities

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

and

$$\frac{1}{2}\|H_k\| \leq \sqrt{2}\ \max\left\{ 1 + \|H_k\|, 2\sqrt{\sigma_k \|P_k g_k\|} \right\},$$

it follows that $0 < \theta_k \leq \hat{\beta}_k$. Thus replace β in (3.7) with θ_k, we obtain that

$$\begin{aligned} & f(x_k) - m_k^t(t_k) \\ \geq\ & \frac{\|P_k g_k\|^2 (1 - \frac{1}{2}\theta_k \|H_k\| - \frac{1}{3}\theta_k^2 \sigma_k \|P_k g_k\|)}{\sqrt{2}\ \max\left\{ 1 + \|H_k\|, 2\sqrt{\sigma_k \|P_k g_k\|} \right\}}. \end{aligned} \tag{3.9}$$

Combining the definition (3.8) of θ_k, it follows that $\theta_k \|H_k\| \leq 1$ and $\theta_k^2 \sigma_k \|P_k g_k\| \leq 1$. Hence, the numerator of (3.9) is bounded below by $\frac{1}{6}\|P_kg_k\|^2$, which together with (3.9), implies that (3.4) holds.

The following result gives a critical bound on the tangential step, which plays a role in following analysis.

Lemma 3.2. *Suppose that (G4) holds. Then the tangential step satisfies*

$$\|t_k\| \le \sqrt{3}\sqrt{\frac{\|P_kg_k\|}{\sigma_k}}, \quad k \ge 0. \tag{3.10}$$

Proof. Suppose that

$$\|t_k\| > \sqrt{3}\sqrt{\frac{\|P_kg_k\|}{\sigma_k}} \tag{3.11}$$

for $k \ge 0$. Therefore, from (G4) and (3.5), we have

$$\begin{aligned} & m_k^t(t_k) - f(x_k) \\ = \;& (P_kg_k)^T t_k + \frac{1}{2}t_k^T H_k t_k + \frac{1}{3}\sigma_k\|t_k\|^3 \\ \ge \;& -\|t_k\|\|P_kg_k\| + \frac{1}{3}\sigma_k\|t_k\|^3. \end{aligned}$$

Due to (3.11), $\frac{1}{3}\sigma_k\|t_k\|^3 - \|t_k\|\|P_kg_k\| > 0$. Then $m_k^t(t_k) - f(x_k) > 0$, which contradicts (3.4). Hence the announced claim follows.

3.2. Feasibility

Lemma 3.3. *Suppose that Assumptions G hold. Suppose also that $|\mathcal{A}| < \infty$. Then*

$$\lim_{k\to\infty} h(x_k) = 0. \tag{3.12}$$

Proof. Due to $|\mathcal{A}| < \infty$, one finds that there exists an integer $K_0 \in \mathbb{N}$ so that x_k is not acceptable for the filter for all $k > K_0$. Note that $k \notin \mathcal{A}_{\text{inc}} \subseteq \mathcal{A}$, which implies that both (2.15) and (2.14) hold for α_k, for all $k > K_0$. At the same time, from the assumptions (G1)-(G3), and (G6), we get that

$$\|P_kg_k\| \le M_{pg}$$

for all k, where M_{pg} is a constant. Hence, along with (3.10), it follows that

$$\begin{aligned} & \|d_k\| && (3.13) \\ \le \;& \|n_k\| + \|t_k\| \\ \le \;& \min\left\{\sqrt{3M_{pg}} + \beta_1, \sqrt{3M_{pg}} + \beta_1\beta_2\sigma_k^{-\frac{\beta_3}{2}}\right\}\sigma_k^{-\frac{1}{2}}. && (3.14) \end{aligned}$$

Then we distinguish two cases, where $k \notin \mathcal{A}$.

Case 1 ($\omega > 1$). Using (2.15) and Assumptions G, one finds that

$$\begin{aligned}
& \kappa_h(h(x_k))^{\varsigma} \\
< \;& (-m_k(\alpha_k))^{\omega}\alpha_k^{1-\omega}\sqrt{\sigma_k}^{\omega-1} \\
= \;& [-g_k^T t_k - \frac{1}{2}\alpha t_k^T H_k t_k - \frac{1}{3}\alpha^2\sigma_k\|t_k\|^3 + (\nabla\lambda_k^T d_k)^T c(x_k)]^{\omega}\alpha_k\sqrt{\sigma_k}^{\omega-1} \\
\leq \;& \left(\|g_k\|\|t_k\| + \frac{1}{2}\|H_k\|\|t_k\|^2 + \frac{1}{3}\sigma_k\|t_k\|^3 + \|\nabla\lambda_k\|\|d_k\|\|c(x_k)\|\right)^{\omega}\alpha_k\sqrt{\sigma_k}^{\omega-1} \\
\overset{(3.13)}{\leq} \;& \Bigg(\|g_k\|\sqrt{3M_{pg}} + \frac{\sqrt{3M_{pg}}}{2}\|H_k\|\|t_k\| + M_{pg}\sqrt{3M_{pg}} \\
& + \left(\sqrt{3M_{pg}} + \beta_1\right)\|\nabla\lambda_k\|\|c(x_k)\|\Bigg)^{\omega}\alpha_k\sigma_k^{-\frac{1}{2}} \\
\overset{(3.3)}{\leq} \;& \left(\sqrt{3M_{pg}}M_g + \frac{3M_{pg}}{2}M_H\sigma_{\min}^{-\frac{1}{2}} + M_{pg}\sqrt{3M_{pg}} + \left(\sqrt{3M_{pg}} + \beta_1\right)M_{\lambda}^{'}M_c\right)^{\omega}\alpha_k\sigma_k^{-\frac{1}{2}} \\
= \;& \alpha_k\overline{M}^{\omega}\sigma_k^{-\frac{1}{2}},
\end{aligned}$$

where $M_c = \max_{x\in X}\|c(x)\|$ and

$$\overline{M} = \sqrt{3M_{pg}}M_g + \frac{3M_{pg}}{2}M_H\sigma_{\min}^{-1/2} + M_{pg}\sqrt{3M_{pg}} + \left(\sqrt{3M_{pg}} + \beta_1\right)M_{\lambda}^{'}M_c.$$

Combining this result and $1 - 1/\omega > 0$, we can see that

$$(h(x_k))^{\frac{\varsigma}{\omega}} > \left(\frac{\kappa_h}{\overline{M}^{\omega}}\right)^{1-1/\omega}(\alpha_k\frac{1}{\sqrt{\sigma_k}})^{1/\omega-1}(h(x_k))^{\varsigma}. \tag{3.15}$$

Calling upon (2.14), it holds that

$$\begin{aligned}
& \ell(x_k) - \ell(x_{k+1}) \\
\geq \;& -\mu m_k(\alpha_k) \\
\overset{(2.15)}{\geq} \;& \mu\kappa_h^{\frac{1}{\omega}}(\alpha_k\frac{1}{\sqrt{\sigma_k}})^{1-\frac{1}{\omega}}(h(x_k))^{\frac{\varsigma}{\omega}} \\
\overset{(3.15)}{>} \;& \mu\overline{M}^{1-\omega}\kappa_h(h(x_k))^{\varsigma}.
\end{aligned}$$

Case 2 ($\omega = 1$). (2.15) and (2.14) together indicate that

$$\ell(x_k) - \ell(x_{k+1}) \geq \mu\kappa_h(h(x_k))^{\varsigma}.$$

No matter what case happens, we can get that

$$\ell(x_k) - \ell(x_{k+1}) \geq \widetilde{M}(h(x_k))^{\varsigma} \geq 0, \tag{3.16}$$

where $\widetilde{M} > 0$ is a constant. Following (G1), it follows that $\ell(x_k)$ is bounded below, and from (3.16), $\ell(x_k)$ is also monotonically decreasing for all $k > K_0$. Hence, (3.16) implies that (3.12) holds as k tends to infinity.

From Lemma 1 of [12] and its corollary, the following results hold as well.

Lemma 3.4. *Let Assumptions G hold. Suppose also that* $|\mathcal{A}| = \infty$*. Then there exists a subsequence* $\{k_i\} \subseteq \mathcal{A}$ *such that*

$$\lim_{i\to\infty} h(x_{k_i}) = 0. \tag{3.17}$$

Theorem 3.1. *Let Assumptions G hold. Then*

$$\lim_{k\to\infty} h(x_k) = 0. \tag{3.18}$$

Proof. From Lemma 3.3 and Lemma 3.4, we can show that the sequence $h(x_k) \to 0$ using the idea of Lemma 8 in [15].

3.3. Optimality

Lemma 3.5. *Suppose that Assumptions G hold,* $k \notin \mathcal{A}_{\text{inc}}$*, and that* (3.4) *holds. Suppose furthermore that*

$$\|P_k g_k\| \geq \epsilon \tag{3.19}$$

for a constant $\epsilon > 0$ *independent of* k*, and that*

$$\sigma_k \geq \Delta_1 := \frac{(1+M_H)^2}{4\epsilon}, \tag{3.20}$$

$$h(x_k) \leq \frac{\epsilon\sqrt{\epsilon}}{24\sqrt{2}M'_\lambda(\sqrt{3M_{pg}}+\beta_1)}. \tag{3.21}$$

Then

$$-m_k(\alpha) \geq \alpha\frac{\epsilon\sqrt{\epsilon}}{24\sqrt{2}}\frac{1}{\sqrt{\sigma_k}} \tag{3.22}$$

for all $\alpha \in (0,1]$.

Proof. From Lemma 3.1, (3.19) and (3.20), it follows that

$$-g_k^T t_k - \frac{1}{2}t_k^T H_k t_k - \frac{1}{3}\sigma_k\|t_k\|^3 \geq \frac{\epsilon\sqrt{\epsilon}}{12\sqrt{2}}\frac{1}{\sqrt{\sigma_k}}. \tag{3.23}$$

Calling upon (G4), we have that

$$\begin{aligned} & -m_k(\alpha) - \alpha(\nabla\lambda_k^T d_k)^T c(x_k) \qquad (3.24)\\ = \quad & -\alpha g_k^T t_k - \frac{1}{2}\alpha^2 t_k^T H_k t_k - \frac{1}{3}\alpha^3\sigma_k\|t_k\|^3 \\ \overset{0<\alpha\leq 1}{\geq} \quad & -\alpha g_k^T t_k - \frac{1}{2}\alpha t_k^T H_k t_k - \frac{1}{3}\alpha\sigma_k\|t_k\|^3 \end{aligned}$$

$$\stackrel{(3.23)}{\geq} \quad \alpha \frac{\epsilon\sqrt{\epsilon}}{12\sqrt{2}} \frac{1}{\sqrt{\sigma_k}}. \tag{3.25}$$

Thus, it follows that

$$\begin{aligned}
& -m_k(\alpha) & (3.26)\\
\geq \quad & \alpha \left(\frac{\epsilon\sqrt{\epsilon}}{12\sqrt{2}} \frac{1}{\sqrt{\sigma_k}} + (\nabla \lambda_k^T d_k)^T c(x_k) \right) \\
\geq \quad & \alpha \left(\frac{\epsilon\sqrt{\epsilon}}{12\sqrt{2}} \frac{1}{\sqrt{\sigma_k}} - \|\nabla \lambda_k\| \|d_k\| c(x_k) \right) \\
\stackrel{(3.13)}{\geq} \quad & \alpha \frac{1}{\sqrt{\sigma_k}} \left(\frac{\epsilon\sqrt{\epsilon}}{12\sqrt{2}} - (\sqrt{3M_{pg}} + \beta_1) \|\nabla \lambda_k\| c(x_k) \right) \\
\geq \quad & \alpha \frac{\epsilon\sqrt{\epsilon}}{24\sqrt{2}} \frac{1}{\sqrt{\sigma_k}}, & (3.27)
\end{aligned}$$

as announced.

Lemma 3.6. *Suppose that (G1) holds. Then there exist constants M_h and $M_\ell > 1$ for which*

$$\ell(x_k + \alpha d_k) - \ell(x_k) - m_k(\alpha) \leq M_\ell \alpha^2 \sigma_k^{-1}, \tag{3.28a}$$
$$h(x_k + \alpha d_k) - (1 - \alpha) h(x_k) \leq M_h \alpha^2 \|d_k\|^2, \tag{3.28b}$$

for all $k \notin \mathcal{A}_{\text{inc}}$ and $\alpha \in (0, 1]$.

Proof. It is easy to follow that (3.28b) holds from Taylor expansions.

From (2.2), (2.7) and (2.9), it follows that

$$A_k d_k + c(x_k) = 0. \tag{3.29}$$

Combining this result and the definition of $\ell(x)$ in (2.3), we can have that

$$\begin{aligned}
& \ell(x_k + \alpha d_k) - \ell(x_k) \\
= \quad & f(x_k + \alpha d_k) - \lambda(x_k + \alpha d_k)^T c(x_k + \alpha d_k) - f_k + \lambda_k^T c(x_k) \\
= \quad & \alpha g_k^T d_k + O(\alpha^2 \|d_k\|^2) - [\lambda_k + \alpha \nabla \lambda(x_k)^T d_k + O(\alpha^2 \|d_k\|^2)]^T \\
& [c(x_k) + \alpha A_k d_k + O(\alpha^2 \|d_k\|^2)] + \lambda_k^T c(x_k) \\
= \quad & \alpha g_k^T d_k - (\lambda_k + \alpha \nabla \lambda_k^T d_k)^T (1 - \alpha) c(x_k) + \lambda_k^T c(x_k) + O(\alpha^2 \|d_k\|^2) \\
= \quad & \alpha g_k^T (n_k + t_k) - (1 - \alpha) \lambda_k^T c(x_k) - \alpha (\nabla \lambda_k^T d_k)^T c(x_k) + \lambda_k^T c(x_k) + O(\alpha^2 \|d_k\|^2) \\
\stackrel{(2.4),(2.7)}{=} \quad & \alpha g_k^T t_k - \alpha (\nabla \lambda_k^T d_k)^T c(x_k) + O(\alpha^2 \|d_k\|^2).
\end{aligned}$$

Therefore, it follows (3.13) that

$$\begin{aligned}
& \ell(x_k + \alpha d_k) - \ell(x_k) - m_k(\alpha) \\
= \quad & -\frac{1}{2} \alpha^2 t_k H_k t_k - \frac{1}{3} \alpha^3 \sigma_k \|t_k\|^3 + O(\alpha^2 \|d_k\|^2) \\
\stackrel{(G4)}{\leq} \quad & \frac{1}{2} \alpha^2 t_k H_k t_k + + O(\alpha^2 \|d_k\|^2)
\end{aligned}$$

$$= \quad O(\alpha^2(\frac{1}{\sqrt{\sigma_k}})^2).$$

Hence, the (3.28a) holds.

Lemma 3.7. *Let Assumptions G,* (3.19) *and* (3.21) *hold,* $k \notin \mathcal{A}_{\rm inc}$. *Assume also that*

$$\sigma_k \geq \Delta_2 := \max\left\{\Delta_1, \left(\frac{24\sqrt{2}M_\ell}{(1-\eta_2)\epsilon\sqrt{\epsilon}}\right)^2\right\}. \tag{3.30}$$

Then $\rho_k \geq \eta_2$ *for all* $\alpha \in (0, 1]$.

Proof. Calling upon (3.19), (3.21) and (3.30), it follows that Lemma 3.1 and Lemma 3.5 hold. As a result, we can deduce that

$$-m_k(\alpha) \geq \alpha \frac{\epsilon\sqrt{\epsilon}}{24\sqrt{2}} \frac{1}{\sqrt{\sigma_k}}.$$

Combining this result, (3.28a), the definition of ρ_k and (3.30), one finds that

$$\begin{aligned} & 1-\rho_k \\ = \quad & \frac{\ell(x_k+\alpha d_k)-\ell(x_k)-m_k(\alpha)}{-m_k(\alpha)} \\ \leq \quad & \frac{24\sqrt{2}M_\ell\alpha^2}{\alpha\epsilon\sqrt{\epsilon}\sqrt{\sigma_k}} \\ \leq \quad & 1-\eta_2. \end{aligned}$$

Therefore, the claim is true.

Lemma 3.8. *Let Assumptions G hold and let* $\{x_{k_i}\}$ *be a sequence with* $k_i \notin \mathcal{A}_{\rm inc}$. *Assume furthermore that* (3.22) *follows for a constant* $\epsilon > 0$ *independent of* k_i *and for all* $\alpha \in (0, 1]$. *Then there exists a constant*

$$\bar{\alpha} = \frac{(1-\mu)\epsilon\sqrt{\epsilon}\sqrt{\sigma_{\min}}}{24\sqrt{2}M_\ell} > 0$$

such that

$$\ell(x_{k_i}+\alpha d_{k_i}) - \ell(x_{k_i}) \leq \mu m_{k_i}(\alpha) \tag{3.31}$$

for all k_i *and* $\alpha \leq \bar{\alpha}$.

Proof. Using (3.28a) in Lemma 3.6, one finds that

$$\begin{aligned} & \ell(x_{k_i}+\alpha d_{k_i}) - \ell(x_{k_i}) - m_{k_i}(\alpha) \\ \leq \quad & M_\ell\alpha^2\sigma_{k_i}^{-1} \\ \leq \quad & M_\ell\frac{(1-\mu)\epsilon\sqrt{\epsilon}\sqrt{\sigma_{\min}}}{24\sqrt{2}M_\ell}\alpha\sigma_{k_i}^{-1} \end{aligned}$$

$$\begin{aligned} &\leq \frac{\alpha\epsilon\sqrt{\epsilon}}{24\sqrt{2}}(1-\mu)\sigma_{k_i}^{-\frac{1}{2}} \\ &\overset{(3.22)}{\leq} -(1-\mu)m_{k_i}(\alpha) \end{aligned}$$

for $\alpha \in (0, \bar{\alpha}]$, which indicates that (3.31) follows.

Lemma 3.9. *Suppose that Assumptions G hold and let $\{x_{k_i}\}$ be a sequence with $k_i \notin \mathcal{A}_{\text{inc}}$. Suppose that (3.22) holds for all $\alpha \in (0,1]$ and for a constant $\epsilon > 0$ independent of k_i. Then there exist constants $\nu_1, \nu_2 > 0$ for which*

$$(h(x_{k_i} + \alpha d_{k_i}), \ell(x_{k_i} + \alpha d_{k_i})) \notin \mathcal{F}_{k_i}$$

for all k_i and $\alpha \leq \min\{\nu_1, \nu_2 h(x_{k_i})\}$.

Proof. From the mechanism of Algorithm 2.1, we know that

$$(h(x_{k_i}), \ell(x_{k_i})) \notin \mathcal{F}_{k_i}. \tag{3.32}$$

Using (3.28a) and (3.22), it follows that

$$\begin{aligned} &\ell(x_{k_i} + \alpha d_{k_i}) - \ell(x_{k_i}) \\ \leq\ & m_{k_i}(\alpha) + M_\ell \alpha^2 \sigma_{k_i}^{-1} \\ \leq\ & -\frac{\alpha\epsilon\sqrt{\epsilon}}{24\sqrt{2}}\sigma_{k_i}^{-\frac{1}{2}} + M_\ell \alpha^2 \sigma_{k_i}^{-1}. \end{aligned}$$

Note that for $\alpha \leq \nu_1 := \frac{\epsilon\sqrt{\epsilon}\sqrt{\sigma_{\min}}}{24\sqrt{2}M_\ell}$, we can get that

$$\ell(x_{k_i} + \alpha d_{k_i}) \leq \ell(x_{k_i}). \tag{3.33}$$

Similarly, it follows from (3.28b) that

$$h(x_{k_i} + \alpha d_{k_i}) \leq h(x_{k_i}) \tag{3.34}$$

for $\alpha \leq \nu_2 h(x_{k_i})$, where $\nu_2 := \frac{\sigma_{\min}}{M_h(\sqrt{3M_{pg}}+\beta_1)^2}$.

Combining (3.32)-(3.34) and the initialization of the filter, we can deduce that $(h(x_{k_i} + \alpha d_{k_i}), \ell(x_{k_i} + \alpha d_{k_i})) \notin \mathcal{F}_{k_i}$.

Lemma 3.10. *Suppose that Assumptions G hold and $|\mathcal{A}| < \infty$. Assume also that (3.19) follows for all k. Then there exists a constant $\sigma_{\max} > 0$ independent of k so that $\sigma_k \leq \sigma_{\max}$ for all k.*

Proof. Due to $|\mathcal{A}| < \infty$, Lemma 3.3 indicates that (3.12) follows. Let $K_1 \geq K_0$ be given, which is sufficiently large enough so that $k \notin \mathcal{A}_{\text{inc}}$ follows for all $k \geq K_1$.

To obtain a contradiction, we assume that the iteration j is the first iteration after K_1 such that

$$\sigma_j \geq \gamma_2 \Delta_3 \tag{3.35}$$

with $\Delta_3 := \max\{\Delta_2, \sigma_{K_1}\}$. (3.35) implies that $\sigma_j \geq \gamma_2\sigma_{K_1}$. This result guarantees that $j \geq K_1 + 1$. Hence, one finds that $j - 1 \geq K_1$ and we can deduce that $j - 1 \notin \mathcal{A}_{\text{inc}}$. Calling upon step 11 of Algorithm 2.1, it follows from (3.35) that

$$\sigma_{j-1} \geq \Delta_3 \geq \Delta_2.$$

Thus, the result follows from Lemma 3.7.

Moreover, it follows from $\Delta_2 \geq \Delta_1$, (3.21) and (3.19) that Lemma 3.5 is applicable. Therefore, we can deduce that Lemma 3.9 is applicable, which indicates that $x_{j-1} + \alpha d_{j-1}$ is accepted by the filter, for all $\alpha \leq \min\{\nu_1, \nu_2 h(x_{j-1})\}$. It follows from this result, $\rho_{j-1} \geq \eta_2$ and the mechanism of the Algorithm 2.1 that

$$\sigma_{j-1} \geq \sigma_j \geq \gamma_2 \max\{\Delta_2, \sigma_{K_1}\},$$

which contradicts the fact that j is the first iteration after K_1 such that (3.35) follows. Hence, for all $k \geq K_1$, one finds that $\sigma_k \leq \gamma_2\Delta_3$. If we define

$$\sigma_{\max} = \max\{\sigma_0, \ldots, \sigma_{K_1}, \gamma_2\Delta_3\},$$

the desired conclusion follows.

Lemma 3.11. *Suppose that Assumptions G hold. Then*

$$h(x_k) = 0 \Rightarrow -m_k(\alpha) > 0, \tag{3.36}$$

and for all k and $\alpha \in (0, 1]$,

$$\mathcal{H}_k := \min\{h : (h, \ell) \in \mathcal{F}_k\} > 0. \tag{3.37}$$

Proof. From (2.7), one finds that

$$\begin{aligned} & \|n_k\| \\ = \;& \|A_k^T(A_kA_k^T)^{-1}c(x_k)\| \\ \leq \;& \|A_k^T(A_kA_k^T)^{-1}\|h(x_k) \\ \overset{(G6)}{\leq} \;& \frac{1}{M_A}h(x_k). \end{aligned} \tag{3.38}$$

Hence, using (3.38), $h(x_k) = 0$ yields that $n_k = 0$ and $c(x_k) = 0$, which indicates that $d_k = t_k$.

Moreover, we know that $\|P_kg_k\| > 0$ or the algorithm would stop in step 2. It follows from (G4) that

$$\begin{aligned} & -m_k(\alpha) \\ = \;& -\alpha g_k^Tt_k - \frac{1}{2}\alpha^2t_k^TH_kt_k - \frac{1}{3}\alpha^3\sigma_k\|t_k\|^3 + \alpha(\nabla\lambda_k^Td_k)^Tc(x_k) \\ = \;& -\alpha g_k^Tt_k - \frac{1}{2}\alpha^2t_k^TH_kt_k - \frac{1}{3}\alpha^3\sigma_k\|t_k\|^3 \\ \overset{0<\alpha<1}{\geq} \;& \alpha\Big(-g_k^Tt_k - \frac{1}{2}t_k^TH_kt_k - \frac{1}{3}\sigma_k\|t_k\|^3\Big) \end{aligned}$$

$$\overset{(3.4)}{\geq} \quad \frac{\alpha\|P_kg_k\|}{6\sqrt{2}}\min\left\{\frac{\|P_kg_k\|}{1+\|H_k\|}, \frac{1}{2}\sqrt{\frac{\|P_kg_k\|}{\sigma_k}}\right\} > 0.$$

Hence, (3.36) follows.

Next, we establish the second conclusion. Since $h_{\max} > 0$, for $k = 0$, calling upon the mechanism of Algorithm 2.1, one can show that the claim follows.

Let the claim holds for k. Provided that $h(x_k) > 0$, and x_k is accepted by the filter, we can get that $\mathcal{H}_{k+1} > 0$ in view of $\gamma_h \in (0, 1)$.

If $h(x_k) = 0$, it follows from (3.36) that $-m_k(\alpha) > 0$ for all $\alpha \in (0, 1]$. Hence, for all trial step sizes, we have that (2.15) holds. Thus, our algorithm always consider step 8. Furthermore, α_k satisfies (2.14). Hence, x_k is not accepted by the filter, which shows that $\mathcal{H}_{k+1} = \mathcal{H}_k > 0$, as announced.

Lemma 3.12. *Suppose that Assumptions G hold. Suppose also that $|\mathcal{A}| < \infty$. Then*

$$\lim_{k\to\infty} \|P_kg_k\| = 0.$$

Proof. To derive a contradiction, we assume that there exists a subsequence $\{x_{k_i}\}$ so that (3.19) follows, namely, $\|P_{k_i}g_{k_i}\| > \epsilon$ for all i.

Due to $|\mathcal{A}| < \infty$, for all $k_i \geq K_2$, there exists an integer $K_2 \geq K_1 \geq K_0$ so that $k_i \notin \mathcal{A}$. From the mechanism of Algorithm 2.1, one finds that (2.14) and (2.15) follow for all $k_i \geq K_2$. Consequently, the above results indicate that

$$\lim_{i\to\infty} m_{k_i}(\alpha_{k_i}) = 0 \tag{3.39}$$

because $\ell(x_{k_i})$ is monotonically decreasing and bounded below from (3.16).

Combining (3.4), Lemma 3.10, and $\|P_kg_k\| > \epsilon$, it follows that

$$\begin{aligned} & -g_{k_i}^T t_{k_i} - \frac{1}{2}t_{k_i}^T H_{k_i} t_{k_i} - \frac{1}{3}\sigma_k\|t_{k_i}\|^3 \\ \geq\; & \frac{\|P_{k_i}g_{k_i}\|}{6\sqrt{2}}\min\left\{\frac{\|P_{k_i}g_{k_i}\|}{1+\|H_{k_i}\|}, \frac{1}{2}\sqrt{\frac{\|P_{k_i}g_{k_i}\|}{\sigma_{k_i}}}\right\} \\ \geq\; & \tilde{\Delta}, \end{aligned} \tag{3.40}$$

where

$$\tilde{\Delta} := \frac{\epsilon}{6\sqrt{2}}\min\left\{\frac{\epsilon}{1+M_H}, \frac{\sqrt{\epsilon}}{2}\frac{1}{\sqrt{\sigma_{\max}}}\right\}.$$

Then, we can get that

$$\begin{aligned} & -m_{k_i}(\alpha_{k_i}) - \alpha_{k_i}(\nabla\lambda_{k_i}^T d_{k_i})^T c_{k_i} \\ =\; & -\alpha_{k_i} g_{k_i}^T t_{k_i} - \frac{1}{2}\alpha_{k_i}^2 t_{k_i}^T H_{k_i} t_{k_i} - \frac{1}{3}\alpha_{k_i}^3\sigma_{k_i}\|t_{k_i}\|^3 \\ \geq\; & \alpha_{k_i}\Big(- g_{k_i}^T t_{k_i} - \frac{1}{2}t_{k_i}^T H_{k_i} t_{k_i} - \frac{1}{3}\sigma_{k_i}\|t_{k_i}\|^3\Big) \\ \overset{(3.40)}{\geq}\; & \tilde{\Delta}\alpha_{k_i} \end{aligned}$$

for $k_i \geq K_2$. As a consequence, it follows that

$$-m_{k_i}(\alpha_{k_i}) \geq (\tilde{\Delta} + (\nabla \lambda_{k_i}^T d_{k_i})^T c_{k_i})\alpha_{k_i}. \tag{3.41}$$

Meanwhile, Lemma 3.3 provides that there exists an integer $K_3 \geq K_2$ so that $h(x_{k_i}) \leq \frac{\tilde{\Delta}\sqrt{\sigma_{\min}}}{2(\sqrt{3M_{pg}}+\beta_1)M_\lambda'}$ holds for all $k_i \geq K_3$. This result and (3.41) yield that

$$\begin{aligned} & -m_{k_i}(\alpha_{k_i}) \\ \geq \; & [\tilde{\Delta} - (\nabla \lambda_{k_i}^T d_{k_i})^T c_{k_i}]\alpha_{k_i} \\ \geq \; & [\tilde{\Delta} - \|\nabla \lambda_{k_i}\|\|d_{k_i}\| h(x_{k_i})]\alpha_{k_i} \\ \geq \; & \frac{1}{2}\tilde{\Delta}\alpha_{k_i} \end{aligned} \tag{3.42}$$

for $k_i \geq K_3$. The last inequality and (3.39) show that $\lim_{i\to\infty}\alpha_{k_i} = 0$.

In general, suppose that K_3 is large enough so that $\alpha_{k_i} < 1$. Hence, $\alpha_{k_i,0} = 1$ cannot be accepted. Moreover, it follows from $k_i \notin \mathcal{A}$ and $\alpha_{k_i,l_i} > \alpha_{k_i}$ that last rejected trial step size

$$\alpha_{k_i,l_i} \in [\alpha_{k_i}/\omega_2, \alpha_{k_i}/\omega_1] \tag{3.43}$$

satisfies (2.15). As a consequence, α_{k_i,l_i} can not be accepted since (2.14) does not hold, namely,

$$\ell(x_{k_i} + \alpha_{k_i,l_i} d_{k_i}) - \ell(x_{k_i}) > \mu[m_{k_i}(\alpha_{k_i,l_i})], \tag{3.44}$$

or it is not accepted by the current filter, namely,

$$(h(x_{k_i} + \alpha_{k_i,l_i} d_{k_i}), \ell(x_{k_i} + \alpha_{k_i,l_i} d_{k_i})) \in \mathcal{F}_{k_i} = \mathcal{F}_{K_2}. \tag{3.45}$$

Now, one provides that either (3.44) does not hold or $(h(x_{k_i}+\alpha_{k_i,l_i}d_{k_i}), \ell(x_{k_i}+\alpha_{k_i,l_i}d_{k_i})) \notin \mathcal{F}_{k_i}$ for sufficiently large k_i.

Consider (3.44). Combining (3.43) and $\lim_{i\to\infty}\alpha_{k,i} = 0$, we can get that $\lim_{i\to\infty}\alpha_{k_i,l_i} = 0$. Consequently, it follows from Lemma 3.8 that $\alpha_{k_i,l_i} \leq \bar{\alpha}$ for sufficiently large k_i, which indicates that (3.44) does not hold for those k_i.

Consider (3.45). Denote $\mathcal{H}_{K_2} = \min\{h : (h, \ell) \in \mathcal{F}_{K_2}\}$. Lemma 3.11 shows that $\mathcal{H}_{K_2} > 0$. Our assumptions together (3.28b) imply that

$$h(x_{k_i} + \alpha_{k_i,l_i} d_{k_i}) \leq (1 - \alpha_{k_i,l_i}) h(x_{k_i}) + M_h \alpha_{k_i,l_i}^2 (\sqrt{3M_{pg}} + \beta_1)^2 \sigma_{\min}^{-1}.$$

It follows from the above inequality, $\lim_{i\to\infty}\alpha_{k_i,l_i} = 0$ and $\lim_{i\to\infty} h(x_{k_i}) = 0$ that $h(x_{k_i} + \alpha_{k_i,l_i} d_{k_i}) < \mathcal{H}_{K_2}$ for k_i sufficiently large. which is contradiction with (3.45). Thus, the desired conclusion follows.

Then, we can get the following result from Lemma 3.1 in [10].

Lemma 3.13. *Suppose that (G1) and* (3.1) *hold. Then*

$$\|n_k\| \geq \frac{1}{\kappa_{hn}} h(x_k) \tag{3.46}$$

for a constant κ_{hn} independent of k.

Lemma 3.14. *Suppose that Assumptions G hold and let* $\{x_{k_i}\}$ *be a sequence with* $\|P_{k_i} g_{k_i}\| \geq \epsilon$ *for a constant* $\epsilon > 0$ *independent of* k_i. *Then there exists* $K \in \mathbb{N}$ *so that for all* $k_i \geq K$, $k_i \notin \mathcal{A}$.

Proof. Using Theorem 3.1, it follows that $\lim_{i\to\infty} h(x_{k_i}) = 0$, which indicates that $k_i \notin \mathcal{A}_{\rm inc}$ for $k_i \geq K_3$. Then we consider two cases.

Case 1 (there is a constant $\tilde{\kappa} > 0$ independent of k_i for which $\sigma_{k_i} \leq \tilde{\kappa}$ for all k_i).
By the same argument of (3.42), one finds that there is a constant $\bar{\Delta} > 0$ independent of k_i for which

$$-m_{k_i}(\alpha) \geq \bar{\Delta}\alpha \tag{3.47}$$

for k_i large enough and $\alpha \in (0, 1]$.

In general, suppose that (3.47) follows for all k_i. Combining Lemma 3.8 and Lemma 3.9, it follows that the constants $\bar{\alpha}, \nu_1, \nu_2 > 0$ are existent.

Choose $K \in \mathbb{N}$ with $K \geq K_3$ such that, for all $k_i \geq K$,

$$h(x_{k_i}) < \min\left\{\delta_h, \frac{\bar{\alpha}}{\nu_2}, \frac{\nu_1}{\nu_2}, \left(\frac{\bar{\Delta}^\omega \omega_1 \nu_2}{\kappa_h}\right)^{\frac{1}{\varsigma-2}}\right\}, \tag{3.48}$$

where ω_1 is from step 10 in Algorithm 2.1. The following part of this proof is in the same way as Lemma 10 in [21], which proves the claim.

Case 2 (there exist subsequences of $\{k_i\}$ such that σ_{k_i} tends to infinity).

Similarly, we can choose $K \in \mathbb{N}$ with $K \geq K_3$ for which

$$h(x_{k_j}) < \min\left\{\delta_h, \frac{\bar{\alpha}}{\nu_2}, \frac{\nu_1}{\nu_2}, \frac{\epsilon\sqrt{\epsilon}}{24\sqrt{2}M'_\lambda(\sqrt{3M_{pg}} + \beta_1)}, \left[\frac{(\epsilon\sqrt{\epsilon})^\omega \omega_1 \nu_2}{(24\sqrt{2})^\omega \kappa_{hn}\kappa_h}\right]^{\frac{1}{\varsigma-2}}\right\} \tag{3.49}$$

for all $k_j \geq K$. For the sake of brevity, let $\{k_j\}$ be a subsequence of $\{k_i\}$ for which

$$\lim_{j\to\infty} \frac{1}{\sqrt{\sigma_{k_j}}} = 0.$$

Hence, for k_j sufficiently large, we can get that $\sigma_{k_j} \geq \Delta_1$ for the constant $\Delta_1 > 0$ in (3.20). So (3.22) follows for those k_j, namely,

$$-m_{k_j}(\alpha) \geq \alpha \frac{\epsilon\sqrt{\epsilon}}{24\sqrt{2}} \frac{1}{\sqrt{\sigma_{k_j}}}. \tag{3.50}$$

Without loss of generality, suppose that (3.50) follows for those k_j.

From Lemma 3.11, for all $k_j \geq K$ with $h(x_{k_j}) = 0$, one finds that both (2.15) and (2.14) hold in iteration k_j. Hence, we can see that $k_j \notin \mathcal{A}$.

For those iterations $k_j \geq K$ with $h(x_{k_j}) > 0$, it follows from (3.49) that $k_j \notin \mathcal{A}_{\rm inc}$,

$$\frac{(24\sqrt{2})^\omega \kappa_{hn}\kappa_h [h(x_{k_j})]^{\varsigma-1}}{(\epsilon\sqrt{\epsilon})^\omega} < \omega_1 \nu_2 h(x_{k_j}) \tag{3.51}$$

and

$$\nu_2 h(x_{k_j}) < \min\{\bar{\alpha}, \nu_1\}. \tag{3.52}$$

For an arbitrary k_j such that $h(x_{k_j}) > 0$, we can define that

$$\zeta_{k_j} = \nu_2 h(x_{k_j}) \overset{(3.52)}{=} \min\{\bar{\alpha}, \nu_1, \nu_2 h(x_{k_j})\}. \tag{3.53}$$

Using Lemma 3.8 and Lemma 3.9, for $\alpha_{k_j,l} \leq \zeta_{k_j}$, we can conclude that both

$$(h(x_{k_j} + \alpha_{k_j,l_j} d_{k_j}), \ell(x_{k_j} + \alpha_{k_j,l_j} d_{k_j})) \notin \mathcal{F}_{k_j} \tag{3.54}$$

and

$$\ell(x_{k_j} + \alpha_{k_j,l_j} d_{k_j}) - \ell(x_{k_j}) \leq \mu[m_{k_j}(\alpha_{k_j,l_j})] \tag{3.55}$$

hold. Let $\alpha_{k_j,L}$ be the first trial step size such that (3.54) and (3.55) hold. The step 10 in Algorithm 2.1 provides that

$$\alpha \geq \omega_1 \zeta_{k_j} \overset{(3.53)}{=} \omega_1 \nu_2 h(x_{k_j}) \overset{(3.51)}{>} \frac{(24\sqrt{2})^\omega \kappa_{hn} \kappa_h [h(x_{k_j})]^{\varsigma - 1}}{(\epsilon\sqrt{\epsilon})^\omega}$$

for $\alpha \geq \alpha_{k_j,L}$. Moreover, (3.46) yields that

$$\frac{1}{\sqrt{\sigma_{k_j}}} \geq \|n_{k_j}\| \geq \frac{1}{\kappa_{hn}} h(x_{k_j}).$$

From above results, one finds that

$$\begin{aligned} & [-m_{k_j}(\alpha)]^\omega \left(\alpha \frac{1}{\sqrt{\sigma_{k_j}}}\right)^{1-\omega} \\ \overset{(3.50)}{\geq} & \left(\frac{\epsilon\sqrt{\epsilon}}{24\sqrt{2}}\right)^\omega \alpha (\sqrt{\sigma_{k_j}})^{-1} \\ > & \kappa_h [h(x_{k_j})]^\varsigma \end{aligned} \tag{3.56}$$

for $\alpha \geq \alpha_{k_j,L}$.

From (3.56), we know that the algorithm goes to step 8. Furthermore, (3.55) implies that (2.14) follows for $\alpha_{k_j,L}$. It remains to verify

$$\alpha_{k_j,L} \geq \alpha_{k_j}^{\min}.$$

Following (3.56), one finds that

$$\begin{aligned} & \alpha_{k_j,L} \\ \geq & \frac{\kappa_h [h(x_{k_j})]^\varsigma (\sqrt{\sigma_{k_j}})^{1-\omega}}{[-g_{k_j}^T t_{k_j} - \frac{1}{2}\alpha_{k_j,L} t_{k_j}^T H_{k_j} t_{k_j} - \frac{1}{3}\sigma_{k_j} \alpha_{k_j,L}^2 \|t_{k_j}\|^3 + (\nabla \lambda_{k_j}^T d_{k_j})^T c_{k_j}]^\omega} \\ \geq & \frac{\kappa_h [h(x_{k_j})]^\varsigma (\sqrt{\sigma_{k_j}})^{1-\omega}}{[-g_{k_j}^T t_{k_j} + (\nabla \lambda_{k_j}^T d_{k_j})^T c_{k_j}]^\omega}, \end{aligned}$$

which together (2.16) imply that $\alpha_{k_j,L} \geq \alpha_{k_j}^{\min}$.

Therefore, the algorithm does not go to the feasibility restoration procedure. Following this result and (3.54)-(3.56), one finds that $\alpha_{k_j,L}$ is the accepted step size α_{k_j}. Hence, for all $k_j \geq K$, $k_j \notin \mathcal{A}$ holds because of the mechanism of Algorithm 2.1.

Thus, no matter Case 1 or Case 2 happens, we know that the desired conclusion follows.

The following theorem gives global convergence conclusion.

Theorem 3.2. *Suppose that Assumptions G hold. Then*

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

Proof. To prove our claim, we consider two cases.
Case 1 ($|\mathcal{A}| < \infty$). Lemma 3.12 has proven this claim.
Case 2 (There exists a subsequence $\{x_{k_i}\}$ for which $k_i \in \mathcal{A}$).

To obtain a contradiction, assume that $\limsup_{k\to\infty}\|P_k g_k\| > 0$. Hence, there exist a subsequence $\{x_{k_{i_j}}\}$ so that $\|P_{k_{i_j}} g_{k_{i_j}}\| > \epsilon$ for a constant $\epsilon > 0$ independent of k_{i_j} and all k_{i_j}. Then, from Lemma 3.14, we know that there is an iteration k_{i_j} such that $k_{i_j} \notin \mathcal{A}$, which is in contradiction with the choice of $\{x_{k_{i_j}}\}$ such that $\lim_{i\to\infty}\|P_{k_i} g_{k_i}\| = 0$. Therefore, the required result holds.

4. Numerical Results

In this section, we present numerical results to show the efficiency of LsFSARC (Algorithm 2.1). We performed it with Intel(R) Core(TM) i5-6200U CPU @ 2.30GHz 2.40GHz. Numerical testing was implemented in MATLAB version 9.4.0.813654 (R2018a).

In our implementation, the parameters: $\varepsilon = 10^{-6}$, $\beta_1 = 0.1$, $\beta_2 = 100$, $\beta_3 = 0.01$, $\gamma_h = 10^{-5}$, $\kappa_h = 10^{-4}$, $\tau_2 = 2.01$, $\tau_1 = 2$, $\eta_1 = 0.01$, $\eta_2 = 0.9$. The algorithm terminates when

$$\text{Res} := \max\{\|P_k g_k\|, \|c(x_k)\|\} \leq \varepsilon$$

is satisfied. The numerical results are presented in Table 4.1. The test problems are from CUTEst collection [13]. n denotes the number of variables. m denotes the number of equality constraints. NF and NC are the numbers of computation of the objective function and constraint function, respectively. NIT denotes the numbers of iterations. The numbers of computation of the objective function's gradient is denoted by NG. The CPU times in Table 4.1 are counted in seconds.

For comparison, we include the corresponding results obtained from Algorithm 2.1 (Alg. 2.1) in [6] and Algorithm 2.2 (Alg. 2.2) in [16]. The comparison numerical results are reported in Table 4.2 and Table 4.3, respectively. The numerical results of LANCELOT are also from literatures [6] and [16], respectively. Furthermore, to display the performance based on the numerical results in Table 4.2 and Table 4.3 visually, we use the logarithmic performance profiles [9] (see Fig 4.1 and Fig 4.2). From Table 4.2, Table 4.3, Fig 4.1, and Fig 4.2, it can be seen that LsFSARC can be comparable with those algorithms for the given problems.

Table 4.1: Numerical results of the LsFSARC

Problem	Dimension		NIT	NF	NC	NG	Res	CPU-Time
	n	m						
AIRCRFTA	8	5	2	3	3	3	1.5932e−08	0.0164
ARGTRIG	200	200	3	3	4	4	6.8423e−07	3.8064
BDVALUE	102	100	2	3	3	3	9.5721e−10	0.1291
BOOTH	2	2	1	2	2	2	0.0000e+00	0.0074
BROYDN3D	500	500	4	5	5	5	1.0634e−09	18.2427
BT1	2	1	5	5	6	6	1.3889e−07	0.0234
BT2	3	1	9	9	10	10	7.2220e−10	0.0297
BT3	5	3	3	4	4	4	1.0934e−10	0.0084
BT4	3	2	5	5	1	6	2.5683e−07	0.0200
BT5	3	2	7	8	8	8	2.2452e−08	0.0147
BT6	5	2	12	15	13	13	4.2933e−07	0.0277
BT7	5	3	5	4	6	6	2.9464e−07	0.0334
BT8	5	2	6	6	7	7	3.5654e−07	0.0213
BT9	4	2	9	11	10	10	8.0670e−08	0.0198
BT10	2	2	2	2	3	3	2.0895e−09	0.0187
BT11	5	3	9	9	10	10	5.2740e−09	0.0349
BT12	5	3	8	8	9	9	3.5613e−07	0.0315
BYRDSPHR	3	2	8	8	9	9	3.5170e−13	0.0346
CLUSTER	2	2	4	7	5	5	3.7799e−10	0.0174
DTOC3	299	198	27	28	28	28	3.9893e−07	42.3079
DTOC4	299	198	3	4	4	4	2.3783e−07	17.4360
DTOC5	19	9	0	1	1	1	0.0000e+00	0.0009
GENHS28	10	8	5	5	6	6	3.3345e−08	0.0360
GOTTFR	2	3	5	8	6	6	2.1823e−10	0.0169
HAGER1	1001	500	16	16	17	17	5.8785e−07	102.1731
HAGER2	1001	500	12	12	13	13	4.6752e−07	71.8340
HAGER3	1001	500	10	10	11	11	2.9516e−07	52.7732
HATFLDF	3	3	3	3	4	4	1.7795e−08	0.0387
HATFLDG	25	25	2	2	3	3	4.0738e−12	0.0684
HEART8	8	8	5	6	6	6	1.8500e−07	0.2510
HIMMELBA	2	2	1	1	2	2	5.3134e−07	0.0108
HIMMELBC	2	2	2	2	3	3	2.8424e−13	0.0260
HIMMELBE	3	3	5	1	2	6	7.4308e−07	0.0121
HS06	2	1	13	13	14	14	6.0080e−08	0.0232
HS07	2	1	7	8	8	8	5.4175e−08	0.0273
HS08	2	2	2	2	3	3	2.5421e−13	0.0105
HS09	2	1	6	7	7	7	3.9241e−07	0.0112
HS26	3	1	9	10	10	10	1.2431e−08	0.0189
HS27	3	1	26	29	27	27	3.4457e−08	0.0447
HS28	3	1	5	7	6	6	1.9369e−08	0.0081
HS39	4	2	9	11	10	10	8.1036e−08	0.0350
HS40	4	3	19	19	20	120	9.5368e−07	0.0527

Table 4.1 continued

Problem	Dimension		NIT	NF	NC	NG	Res	CPU-Time
	n	m						
HS42	4	2	28	52	29	29	8.4021e−07	0.0657
HS46	5	2	12	13	13	13	8.8987e−07	0.0551
HS47	5	3	20	20	21	21	5.8020e−07	0.1012
HS48	5	2	4	5	5	5	8.6615e−08	0.0169
HS49	5	2	22	23	23	23	7.8456e−07	0.0975
HS50	5	3	12	15	13	13	1.3377e−07	0.0667
HS51	5	3	3	5	4	4	9.8047e−15	0.0143
HS52	5	3	6	6	7	7	3.1612e−10	0.0209
HS56	7	4	0	1	1	1	0.0000e+00	0.0043
HS61	3	2	6	6	7	7	6.3198e−07	0.0232
HS77	5	2	11	13	12	12	7.8820e−07	0.0267
HS78	5	3	12	14	13	13	5.3810e−07	0.0303
HS79	5	3	8	8	9	9	8.8824e−07	0.0353
HS100LNP	7	2	15	21	16	16	4.9734e−07	0.0540
HS111LNP	10	3	12	12	13	13	8.0141e−08	0.0587
HYPCIR	2	2	1	1	2	2	5.4209e−07	0.0118
INTEGREQ	5	5	1	1	2	2	3.8263e−07	0.0117
MARATOS	2	1	3	4	4	4	2.6776e−07	0.0072
MWRIGHT	5	3	8	10	9	9	6.1967e−09	0.0267
ORTHREGB	27	6	7	8	8	8	1.1322e−08	0.2326
POWELLSQ	2	2	1	1	2	2	8.1603e−07	0.0099
RECIPE	3	3	2	2	3	3	3.0307e−15	0.0118
S235	3	1	16	17	17	17	1.8935e−08	0.0230
S252	3	1	14	14	15	15	4.7686e−08	0.0267
S265	4	2	1	2	2	2	1.8081e−16	0.0073
S269	5	3	5	5	6	6	2.1785e−07	0.0201
S316	2	1	2	2	3	3	4.8916e−08	0.0094
S317	2	1	5	5	6	6	3.3780e−12	0.0143
S318	2	1	5	5	6	6	3.2496e−11	0.0143
S319	2	1	7	7	8	8	6.3140e−08	0.0151
S320	2	1	25	46	26	26	9.1824e−07	0.0498
S321	2	1	17	35	18	18	2.9056e−07	0.0406
S335	3	2	11	11	12	12	1.2615e−07	0.1254
S336	3	2	7	7	8	8	2.3466e−08	0.0211
S338	3	2	6	6	7	7	9.4251e−07	0.0209
S344	3	1	8	10	9	9	6.4195e−07	0.0146
S373	9	6	13	12	14	14	7.6678e−07	0.1131
S378	10	3	13	13	14	14	1.9417e−10	0.0707
S394	20	12	9	9	10	10	2.7574e−07	0.1158
S395	50	1	9	8	10	10	5.1312e−07	0.2480
ZANGWIL3	3	3	2	2	3	3	2.0128e−47	0.0208

Table 4.2: Comparison results 1

Problem	Dimension		Alg. 2.1 in [6]		LsFSARC		LANCELOT	
	n	m	NF	NC	NF	NC	NF	NC
AIRCRFTA	8	5	12	21	3	3	5	5
ARGTRIG	200	200	4	4	3	4	7	7
BDVALUE	102	100	4	6	3	3	2	2
BOOTH	2	2	2	2	2	2	3	3
BROYDN3D	500	500	14	14	5	5	6	6
BT1	2	1	8	9	5	6	48	48
BT2	3	1	11	11	9	10	22	22
BT3	5	3	6	6	4	4	16	16
BT4	3	2	9	9	5	1	28	28
BT5	3	2	5	5	8	8	16	16
BT6	5	2	14	14	15	13	26	26
BT7	5	3	16	18	4	6	48	48
BT8	5	2	12	19	6	7	25	25
BT9	4	2	13	15	11	10	20	20
BT10	2	2	7	7	2	3	19	19
BT11	5	3	8	8	9	10	19	19
BT12	5	3	7	8	8	9	21	21
BYRDSPHR	3	2	10	13	8	9	22	22
CLUSTER	2	2	8	8	7	5	10	10
DTOC3	299	198	4	4	28	28	26	26
DTOC4	299	198	3	3	4	4	17	17
DTOC5	19	9	4	4	1	1	14	14
GENHS28	10	8	3	3	5	6	11	11
GOTTFR	2	3	9	14	8	6	17	17
HAGER1	1001	500	2	2	16	17	13	13
HAGER2	1001	500	6	6	12	13	12	12
HAGER3	1001	500	14	14	10	11	13	13
HATFLDF	3	3	28	43	3	4	62	62
HATFLDG	25	25	8	9	2	3	16	16
HEART8	8	8	48	66	6	6	149	149
HIMMELBA	2	2	3	3	1	2	3	3
HIMMELBC	2	2	5	5	2	3	9	9
HIMMELBE	3	3	3	3	1	2	6	6
HS06	2	1	10	13	13	14	30	30
HS07	2	1	9	11	8	8	17	17
HS08	2	2	5	5	2	3	10	10
HS09	2	1	4	4	7	7	6	6
HS26	3	1	19	19	10	10	22	22
HS27	3	1	23	25	29	27	14	14
HS28	3	1	2	2	7	6	7	7
HS39	4	2	13	15	11	10	20	20
HS40	4	3	4	4	19	20	16	16

Table 4.2 continued

Problem	Dimension		Alg. 2.1 in [6]		LsFSARC		LANCELOT	
	n	m	NF	NC	NF	NC	NF	NC
HS42	4	2	4	4	52	29	13	13
HS46	5	2	17	17	13	13	19	19
HS47	5	3	18	18	20	21	21	21
HS48	5	2	3	3	5	5	8	8
HS49	5	2	15	15	23	23	18	18
HS50	5	3	9	9	15	13	11	11
HS51	5	3	2	2	5	4	7	7
HS52	5	3	3	3	6	7	13	13
HS56	7	4	9	9	1	1	18	18
HS61	3	2	6	6	6	7	16	16
HS77	5	2	16	16	13	12	23	23
HS78	5	3	5	5	14	13	12	12
HS79	5	3	5	5	8	9	11	11
HS100LNP	7	2	8	10	21	16	23	23
HS111LNP	10	3	12	13	12	13	69	69
HYPCIR	2	2	5	6	1	2	9	9
INTEGREQ	5	5	2	2	1	2	4	4
MARATOS	2	1	4	4	4	4	9	9
MWRIGHT	5	3	8	8	10	9	18	18
ORTHREGB	27	6	25	30	8	8	74	74
POWELLSQ	2	2	12	15	1	2	24	24
RECIPE	3	3	12	12	2	3	17	17
ZANGWIL3	3	3	6	6	2	3	8	8

Table 4.3: Comparison results 2

Problem	Dimension		Alg. 2.2 in [16]			LsFSARC			LANCELOT		
	n	m	NF	NC	NG	NF	NC	NG	NF	NC	NG
AIRCRFTA	8	5	3	3	3	3	3	3	5	5	5
BDVALUE	102	100	3	3	2	3	3	3	2	2	2
BOOTH	2	2	2	2	2	2	2	2	4	4	4
BT1	2	1	11	11	8	5	6	6	57	57	47
BT3	5	3	8	8	8	4	4	4	15	15	15
BT4	3	2	14	14	14	5	1	6	27	27	26
BT5	3	2	9	9	9	8	8	8	67	67	43
BT6	5	2	30	30	29	15	13	13	51	51	39
BT8	5	2	11	11	11	6	7	7	27	27	25
BT9	4	2	57	57	41	11	10	10	23	23	23
BT10	2	2	8	8	8	2	3	3	21	21	21
BT11	5	3	13	13	13	9	10	10	23	23	20
BT12	5	3	9	9	8	8	9	9	23	23	19
CLUSTER	2	2	8	8	8	7	5	5	13	13	10

Table 4.3 continued

Problem	Dimension		Alg. 2.2 in [16]			LsFSARC			LANCELOT		
	n	m	NF	NC	NG	NF	NC	NG	NF	NC	NG
GENHS28	10	8	9	9	9	5	6	6	10	10	10
GOTTFR	2	3	9	9	6	8	6	6	12	12	11
HATFLDG	25	25	25	25	7	2	3	3	24	24	20
HIMMELBA	2	2	2	2	2	1	2	2	3	3	3
HIMMELBC	2	2	7	7	6	2	3	3	9	9	8
HIMMELBE	3	3	3	3	3	1	2	6	4	4	4
HS06	2	1	14	14	11	13	14	14	58	58	42
HS07	2	1	12	12	12	8	8	8	24	24	19
HS08	2	2	6	6	5	2	3	3	11	11	10
HS09	2	1	7	7	7	7	7	7	11	11	11
HS26	3	1	36	36	26	10	10	10	33	33	31
HS28	3	1	10	10	9	7	6	6	4	4	4
HS39	4	2	57	57	41	11	10	10	674	674	630
HS40	4	3	7	7	7	19	20	20	15	15	14
HS42	4	2	11	11	9	52	29	29	13	13	13
HS46	5	2	29	29	27	13	13	13	28	28	25
HS48	5	2	13	13	10	5	5	5	4	4	4
HS49	5	2	27	27	22	23	23	23	25	25	25
HS50	5	3	25	25	15	15	13	13	19	19	19
HS51	5	3	10	10	9	5	4	4	3	3	3
HS52	5	3	8	8	7	6	7	7	11	11	11
HS61	3	2	13	13	11	6	7	7	18	18	17
HS77	5	2	29	29	26	13	12	12	35	35	30
HS78	5	3	9	9	9	14	13	13	26	26	15
HS79	5	3	13	13	13	8	9	9	12	12	12
HS100LNP	7	2	79	79	35	21	16	16	510	510	468
HYPCIR	2	2	6	6	5	1	2	2	7	7	7
INTEGREQ	5	5	2	2	2	1	2	2	3	3	3
MARATOS	2	1	5	5	5	4	4	4	9	9	9
ORTHREGB	27	6	7	7	7	8	8	8	140	140	116
RECIPE	3	3	12	12	12	2	3	3	43	43	37
ZANGWIL3	3	3	2	2	3	2	3	3	8	8	8

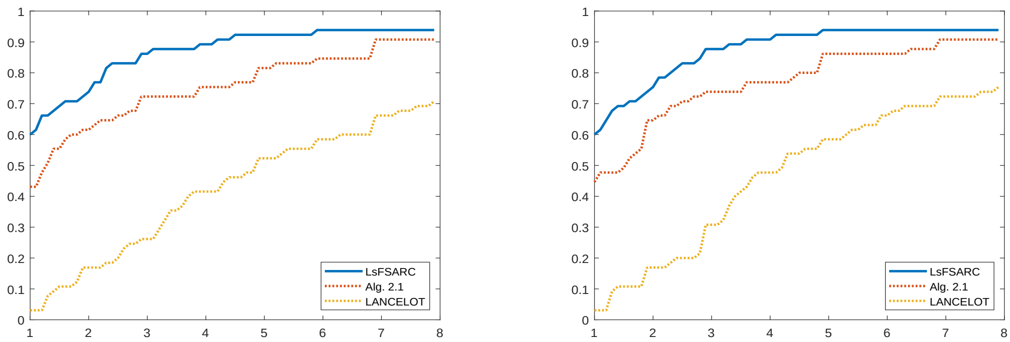

Fig. 4.1. Performance profiles based on NF (left) and NC (right)

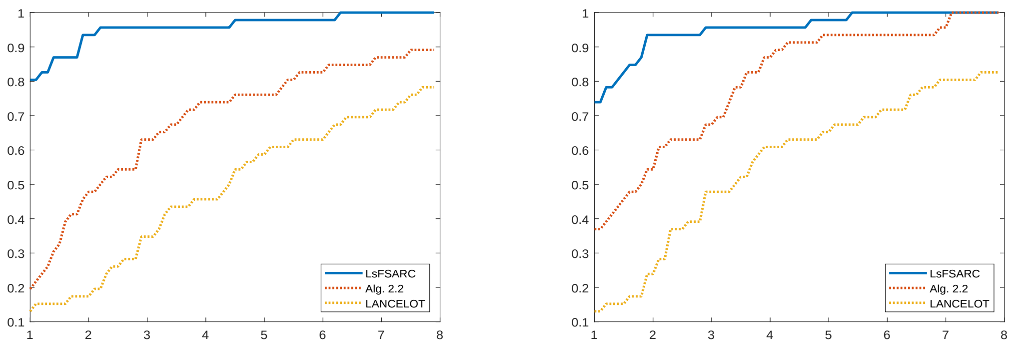

Fig. 4.2. Performance profiles based on NF (left) and NG (right)

5. Conclusion and discussion

We have introduced a line search filter sequential adaptive regularization algorithm using cubics (LsFSARC) to solve nonlinear equality constrained programming. It benefits from the idea of SQP methods. Composite step methods and projective matrices are used to obtain the new step which is decomposed into the sum of a normal step and a tangential step. The tangential step is computed by solving a standard ARC subproblem. The global convergence analysis is reported under some suitable assumptions. Preliminary numerical results and comparison results are presented to demonstrate the performance of LsFSARC. It can be observed that LsFSARC can be comparable with Algorithm 2.1 in [6] and Algorithm 2.2 in [16] for these test problems.

Compared with SQP algorithms where penalty function is employed as a merit function, LsFSARC is a penalty-free method and does not involve the calculation and update of penalty parameters. Naturally, the convergence analysis does not rely on the penalty parameters. Moreover, compared with two impressive and powerful penalty-free methods in [16, 21] which both require that the Lagrangian Hessian or its approximation is uniformly positive definite on the null space of the Jacobian of constraints for each k to guarantee the descent property of the search directions, LsFSARC only requires semipositive definiteness of Lagrangian Hessian to ensure the global convergence.

However, the convergence analysis is not complete since local convergence properties are not discussed. The proposed algorithm LsFSARC can also suffer Maratos effect. To avoid this and achieve fast convergence rate, we can introduce second-order corrections or other techniques in the algorithm. Moreover, we are working on the worst-case complexity bound for the number of iterations to find an ϵ-approximate KKT point.

Acknowledgments

The authors are very grateful to the editor and the referees, whose valuable suggestions and insightful comments helped to improve the paper's prestentation significantly. The authors gratefully acknowledge the supports of the National Natural Science Foundation of China (12071133).

References

[1] N. Andrei. *Modern numerical nonlinear optimization.* Springer-Verlag, 2022.

[2] S. Bellavia, G. Gurioli, B. Morini, P. Toint. Adaptive regularization for nonconvex optimization using inexact function values and randomly perturbed derivatives. *J. Complexity.* **68** (2022), 101591.

[3] H. Benson, D. Shanno. Interior-point methods for nonconvex nonlinear programming: cubic regularization. *Comput. Optim. Appl.* **58**:2 (2014), 323–346.

[4] C. Cartis, N. Gould, P. Toint. Adaptive cubic regularisation methods for unconstrained optimization. Part I: motivation, convergence and numerical results. *Math. Program.* **127**:2 (2011), 245–295.

[5] C. Cartis, N. Gould, P. Toint.. Adaptive cubic regularisation methods for unconstrained optimization. Part II: worst-case function- and derivative-evaluation complexity. *Math. Program.* **130**:2 (2011), 295–319.

[6] Z. Chen, Y. Dai, J. Liu. A penalty-free method with superlinear convergence for equality constrained optimization. *Comput. Optim. Appl.* **76**:3 (2020), 801–833.

[7] A. Conn, N. Gould, P. Toint. *Trust region methods.* Society for Industrial and Applied Mathematics, 2000.

[8] F. Curtis, K. Scheinberg. Adaptive stochastic optimization: A framework foranalyzing stochastic optimization algorithms. *IEEE Signal Proc. Mag.* **37**:5 (2020), 32–42.

[9] E. Dolan., J. Moré. Benchmarking optimization software with performance profiles. *Math. Program.* **91** (2002), 201–213.

[10] R. Fletcher, N. Gould, S. Leyffer, P. Toint, A. Wächter. Global convergence of a trust-region SQP-filter algorithm for general nonlinear programming. *SIAM J. Optim.* **13**:3 (2002), 635–659.

[11] R. Fletcher, S. Leyffer. Nonlinear programming without a penalty function. *Math. Program.* **91**:2 (2002), 239–269.

[12] R. Fletcher, S. Leyffer, P. Toint. On the global convergence of a filter-SQP algorithm. *SIAM J. Optim.* **13**:1 (2002), 44–59.

[13] N. Gould, D. Orban, P. Toint. Cutest: a constrained and unconstrained testing environment with safe threads for mathematical optimization. *Comput. Optim. Appl.* **60**:3 (2015), 545–557.

[14] A. Griewank. The modification of newton's method for unconstrained optimization by bounding cubic terms. Department of Applied Mathematics and Theoretical Physics, 1981.

[15] C. Gu, D. Zhu. A secant algorithm with line search filter method for nonlinear optimization. *Appl. Math. Model.* **35**:2 (2011), 879–894.

[16] X. Liu, Y. Yuan. A sequential quadratic programming method without a penalty function or a filter for nonlinear equality constrained optimization. *SIAM J. Optim.* **21**:2 (2011), 545-571.

[17] Y. Nesterov, B. Polyak. Cubic regularization of Newton method and its global performance. *Math. Program.* **108**:1 (2006), 177–205.

[18] J. Nocedal, S. Wright. *Numerical optimization (2nd ed)*. Springer-Verlag, 2006

[19] Y. Pei, D. Zhu. On the Global Convergence of a Projective Trust Region Algorithm for Nonlinear Equality Constrained Optimization. *Acta. Math. Sin.* **34**:12 (2018), 1804–1828.

[20] W. Sun, Y. Yuan. *Optimization theory and methods: nonlinear programming*. Springer-Verlag, 2006.

[21] A. Wächter, L. Biegler. Line search filter methods for nonlinearprogramming: motivation and global convergence. *SIAM J. Optim.* **16**:1 (2005), 1–31.

[22] R. Wilson. A simplicial algorithm for concave programming. PhD thesis, Harvard Business School, 1963.

[23] J. Zhang and D. Zhu. 1994. A projective quasi-Newton method for nonlinear optimization. *J. Comput. Appl. Math.* **53**:3 (1994), 291-307.