J. Nonlinear Funct. Anal. 2025 (2025) 15 https://doi.org/10.23952/jnfa.2025.15

MATHRES

Journal of Nonlinear Functional Analysis

Available online at http://jnfa.mathres.org

LOCAL CONVERGENCE OF A LINE SEARCH FILTER SEQUENTIAL ADAPTIVE CUBIC REGULARIZATION ALGORITHM FOR NONLINEAR CONSTRAINED OPTIMIZATION

QINGHUI GAO[1,2]

[1]Faculty of Education, Henan Normal University, Xinxiang 453007, China
[2]School of Mathematics and Statistics, Henan Normal University, Xinxiang 453007, China

Abstract. An adaptive regularisation algorithm using cubics (ARC) in association with the line search filter technique for solving nonlinear equality constrained programming is proposed in this paper. In each iteration, the trial step providing sufficient descent is generated by solving a corresponding ARC subproblem. Then, the step size is decided by backtracking line search together with the filter technique to obtain the next iteration point. The advantage of this method is that resolving ARC subproblem many times to determine a new iteration point can be avoided and hence the expensive computation can be lessened. And the difficult decisions in regard to the choice of penalty parameters in the merit functions can be avoided by using the filter technique. Second order correction steps are introduced in the proposed algorithm to overcome Maratos effect. Convergence analysis demonstrates that fast local convergence can be achieved under some mild assumptions. The preliminary numerical results are reported.

Keywords. Cubic regularization; Line search filter; Local convergence; Nonlinear constrained optimization.

2020 MSC. 49M37, 65K05, 65K10, 90C30.

1. INTRODUCTION

Trust region methods and line search methods are two common strategies to guarantee the global convergence of algorithms for various nonlinear optimization problems [1–4]. Adaptive regularization algorithms using Cubics (ARC) were initially proposed in [5] for solving unconstrained optimization problems. ARC has been demonstrated to have excellent convergence properties and iteration complexity [6]. For some recent methods (variations) based on ARC for unconstrained optimization, we refer to [7–17] and some methods [18–22] were designed to solve or use ARC subproblems more efficiently. Meanwhile, some impressive algorithms based on ARC framework were also employed to solve constrained optimization problems; see, e.g.,

E-mail address: gaoqinghui@htu.edu.cn

Received January 29, 2025; Accepted April 10, 2025.

[23–26]. However, like trust region methods, we have to resolve ARC subproblem many times to determine a new iteration in ARC methods. This can increase computation cost.

Based on the method in [27], where line search filter techniques were used to avoid the difficulty of choosing an appropriate penalty parameter in merit functions methods, we propose another line search filter sequential adaptive cubic regularization algorithm for nonlinear constrained optimization with global convergence and focus on analysing local convergence. In each iteration, we use composite step methods to compute the trial step. To this end, we compute a normal step and a tangential step. The normal step is responsible for reducing the constraint violation. The tangential step is used to provide sufficient reduction of the model. It is computed by solving a standard ARC subproblem which is transformed from the ARC subproblem with linearized constraints via reduced Hessian methods, which is different from [27] where the projective matrix is used. After the trial step is computed, the step size is decided by backtracking line search together with filter technique to obtain the next iteration point. The advantage of this method is that resolving ARC subproblem many times to determine a new iteration point can be avoided and hence the expensive computation can be lessened. Furthermore, the difficult decisions in regard to the choice of penalty parameters in the merit functions can be avoided by using filter technique. In addition, ARC mechanism has a potential advantage that it can avoid difficulties resulting from the incompatibility of the intersection of linearized constraints with trust-region bounds in trust region methods for solving constrained optimization. Since filter methods may also suffer from the Maratos effect [28], we introduce second order correction steps to overcome that. Convergence analysis shows that fast local convergence can be achieved under some mild assumptions.

The remainder of this paper is organized as follows. In Section 2, we present the development of the algorithm in detail. The local convergence analysis is reported in Section 3. Numerical results are reported in Section 4. The concluding remarks are presented 5, the last section.

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. 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, and by $o(t_k)$ a sequence$\{v_k\}$ satisfying $\|v_k\| \leq \beta_k t_k$ for some positive sequence $\{\beta_k\}$ with $\beta_k \to 0 (k \to \infty)$.

2. Development of the Algorithm

Consider the nonlinear constrained optimization problem

$$\begin{aligned} \underset{x \in \mathbb{R}^n}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & c(x) = 0, \end{aligned} \tag{2.1}$$

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. Define the Lagrangian function for this problem by $\mathscr{L}(x,\lambda) := f(x) - \lambda^T c(x)$. We use $A(x)$ to denote the Jacobian matrix of $c(x)$, namely,

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

where $c_i(x)$ is the ith of vector $c(x)$. The first-order optimality optimization conditions of equality constrained problem (2.1) are that there exist an x^* and a vector of Lagrange multipliers λ^*

for which $\nabla_x \mathscr{L}(x^*, \lambda^*) = 0$ and $c(x^*) = 0$. Analogy to the idea of sequential quadratic programming (SQP) methods, we consider the following cubic regularization subproblem with linearized constraints in each iteration

$$\begin{aligned} \underset{d\in\mathbb{R}^n}{\text{minimize}} \quad & m_k(d) \overset{\text{def}}{=} f_k + g_k^T d + \frac{1}{2} d^T H_k d + \frac{1}{3}\sigma_k \|d\|^3 \\ \text{subject to} \quad & A_k d + c_k = 0. \end{aligned} \tag{2.2}$$

Here $f_k = f(x_k)$, $g_k = \nabla f(x_k)$, $A_k = A(x_k)$, $c_k = c(x_k)$, H_k denotes the Hessian $\nabla_{xx}\mathscr{L}(x_k, \lambda_k)$ or its approximation, and $\sigma_k \in \mathbb{R}^+$ is an adaptive parameter. In the iteration x_k of the algorithm, in order to calculate the next iteration x_{k+1}, we decompose the overall step as the sum of normal step and tangential step in each iteration using the idea of the composite step approach [**?**]. The overall step is decomposed into $d_k = n_k + t_k$, where n_k is called as a normal step which provides feasibility, and t_k is called as a tangential step which is used to obtain sufficient reduction of the objective function's model, i.e., $A_k n_k + c_k = 0$. Assuming that A_k has full column rank. We can compute the normal step n_k as following:

$$n_k = -A_k^T (A_k A_k^T)^{-1} c_k. \tag{2.3}$$

Moreover, to ensure sufficient decrease in model function, we also assume that

$$\|n_k\| \le \beta_1 \min\left\{1, \frac{\beta_2}{\sqrt{\sigma_k}^{\beta_3}}\right\} \frac{1}{\sqrt{\sigma_k}}, \tag{2.4}$$

where $\beta_1 \in (0,1], \beta_2 > 0$ and $\beta_3 \in (0,1)$. If (2.4) does not hold, the algorithm goes to a feasibility restoration phase, which is discussed later. First, we consider the case that (2.4) holds. Having obtained the normal step, we turn to the tangential step. Let the column of N_k be an orthonormal basis for the null-space of A_k. The tangential step t_k is approximately expressed as

$$t_k = N_k t_k^N, \tag{2.5}$$

where t_k^N is the solution(or its approximation) of the following problem

$$\text{minimize} \quad f(x_k) + (g_k^N)^T t^N + \frac{1}{2}(t^N)^T H_k^N t^N + \frac{1}{3}\sigma_k \|t^N\|^3, \tag{2.6}$$

where

$$g_k^N := N_k^T (g_k + H_k n_k) \text{ and } H_k^N := N_k^T H_k N_k. \tag{2.7}$$

After computing t_k^N, we define from (2.5) that

$$\begin{aligned} m_k^t(t_k) \quad & := f(x_k) + (g_k^N)^T t_k^N + \frac{1}{2}(t_k^N)^T H_k^N t_k^N + \frac{1}{3}\sigma_k \|t_k^N\|^3 \\ & = f(x_k) + (g_k + H_k n_k)^T t_k + \frac{1}{2} t_k^T H_k t_k + \frac{1}{3}\sigma_k \|t_k\|^3. \end{aligned} \tag{2.8}$$

To obtain the next iterate point x_{k+1}, we use a backtracking line search procedure combining filter method where a decreasing sequence of step sizes $\alpha_{k,l} \in (0,1] (l = 0,1,2,\ldots)$ where $\lim_{l\to\infty} \alpha_{k,l} = 0$ is tried until some acceptance rules are satisfied and the trial point is acceptable to the current filter. These acceptance rules are described in details as follows.

Let $h(x) := \|c(x)\|$. The same definition of a filter is the same as in [29] where the filter is defined as a set $\mathscr{F}_k \subseteq [0,\infty) \times \mathbb{R}$ containing all prohibited $(h(x_j), f(x_j))$ pairs in iteration k. At the beginning, we can set $\mathscr{F}_0 = \{(h,f) \in \mathbb{R}^2 : h > h(x_0)\}$. A trial point x is acceptable for the filter $\mathscr{F}_k$ only if

$$h(x) \le (1-\gamma_h)h(x_j) \text{ or } f(x) \le f(x_j) - \gamma_f h(x_j) \tag{2.9}$$

holds for all $(h(x_j), f(x_j))$ in $\mathscr{F}_k$ with fixed constants $\gamma_h, \gamma_f \in (0,1)$. This measure gives a mechanism whereby a sequence $\{x_k\}$ is forced towards feasibility if $\{f(x_k)\}$ is monotonically decreasing and bounded below. Nevertheless, this type of sequence $\{x_k\}$ could still be accepted even if it converges to a nonoptimal point. In order to avoid this case, the following condition

$$f(x) \le f(x_k) + \mu[m_k(d_k) - m_k(0)] \tag{2.10}$$

is employed to be the acceptance criterion whenever the following conditions

$$m_k(0) - m_k(d_k) \ge 0 \text{ and } (m_k(0) - m_k(d_k))^\tau \left(\alpha_{k,l}\sqrt{\sigma_k}\right)^{\tau-1} > \kappa_h (h(x_k))^\phi \tag{2.11}$$

follows for the current trial step, where $\kappa_h > 0,\ 0 < \phi < \frac{1}{2}$ and $\tau \ge 1$ are all fixed constants. It remains to consider another problem called cycling that may occur between two points that satisfy one of the conditions in (2.9) and fail the other. Fortunately, filter method can avoid this. Define $\mathscr{F}_k \subseteq [0,+\infty] \times \mathbb{R}$ including all prohibited (h,f)-pairs in iteration k. Thus a trail point x is acceptable to the filter if $(h(x), f(x)) \notin \mathscr{F}_k$. At first, initialize the filter as $\mathscr{F}_0 = \{(h,f) \in \mathbb{R}^2 : h \ge h_{\max}\}$ for some $h_{\max} > h(x_0)$. If the accepted trial step does not satisfy condition (2.11), the filter is augmented for a new iteration by employing the update formula

$$\mathscr{F}_{k+1} := \mathscr{F}_k \cup \{(h,f) \in \mathbb{R}^2 : h \ge (1-\gamma_h)h(x_k) \text{ and } f \ge f(x_k) - \gamma_f h(x_k)\} \tag{2.12}$$

if the accepted trial step size does not satisfy condition (2.11).

To deal with the situation where (2.4) does not hold and to ensure that the backtracking line search procedure terminates after a finite number of trial steps, we employ a feasibility restoration phase. It could be any iterative algorithm with the goal of finding a less infeasible point, and different methods could be used at different stages of the optimization procedure. In order to detect the situation where no admissible step size can be found and the feasibility restoration phase has to be invoked, we set a threshold for the trial step sizes. If the current trial step $\alpha_{k,l}$ is still large enough that the switching conditions (2.11) holds for some $\alpha < \alpha_{k,l}$, we will not switch to the feasibility restoration phase because there is still the possibility that a shorter step length might be accepted by the condition (2.10). Otherwise, the decision whether to switch to the feasibility restoration phase depends on (2.9). From (2.11) and the linear approximations of $f(x_k + \alpha d_k)$ and $h(x_k + \alpha d_k)$, we set a threshold $\alpha_k^{\min}$ in each iteration k as follows. If $-g_k^T d_k - \frac{1}{2} d_k^T H_k d_k - \frac{1}{3}\sigma_k \|d_k\|^3 > 0$ and $-g_k^T d_k > 0$,

$$\alpha_k^{\min} = \mu_\alpha \cdot \begin{cases} \min\left\{\gamma_h, \dfrac{\gamma_f h(x_k)}{-g_k^T d_k}, \dfrac{\kappa_h [h(x_k)]^\phi \sqrt{\sigma_k}^{1-\tau}}{(-g_k^T d_k)^\tau}\right\} & \text{if } d_k^T H_k d_k \ge 0, \\ \min\left\{\gamma_h, \dfrac{\gamma_f h(x_k)}{-g_k^T d_k}, \dfrac{\kappa_h [h(x_k)]^\phi \sqrt{\sigma_k}^{1-\tau}}{(-g_k^T d_k - \frac{1}{2} d_k^T H_k d_k - \frac{1}{3}\sigma_k \|d_k\|^3)^\tau}\right\} & \text{otherwise.} \end{cases} \tag{2.13a}$$

Otherwise,

$$\alpha_k^{\min} = \mu_\alpha \gamma_h. \tag{2.13b}$$

When $\alpha_{k,l} < \alpha_k^{\min}$, we can deduce that the algorithm turns to feasibility restoration phase. Following this rule, if (2.11) holds for a step size $\alpha \le \alpha_{k,l}$, we can ensure that the algorithm does not switch to the feasibility restoration phase and that the backtracking line search procedure is finite, namely, it eventually either delivers a new iterate point x_{k+1} or reverts to the feasibility restoration phase. Filter methods can also suffer from the Maratos effect which can result in poor local convergence. Therefore, we use a second order correction step w_k to overcome Maratos effect as follows.

$$w_k = -A_k^T(A_kA_k^T)^{-1}c(x_k+d_k). \tag{2.14}$$

Now we give the full description of our algorithm (see Algorithm 1).

Algorithm 1: Line search filter sequential adaptive cubic regularization algorithm

---

0. Initialization.
   (i) Given starting point x_0, constants $\sigma_{\max} > \sigma_{\min} > 0$, an initial $\sigma_0 > 0$ such that $\sigma_{\min} \le \sigma_0 \le \sigma_{\max}$, an initial symmetric matrix H_0.
   (ii) Set constants $0 < \gamma_0 < \gamma_1 \le 1 \le \gamma_2$, $0 < \eta_1 < \eta_2 < 1$, $\beta_1 \in (0,1]$, $\beta_2 > 0$; β_3, β, κ_h, γ_h, $\gamma_f \in (0,1)$; $0 < \phi < 0.5$, $\tau \ge 1$, $0 < \tau_1 \le \tau_2 < 1$.
   (iii) Set the filter $\mathscr{F}_0 = \{(h,f) \in \mathbb{R}^2 : h \ge h_{\max} > h(x_0)\}$ and the iteration counter $k = 0$.
1. Compute $f_k, g_k, h(x_k), A_k, N_k$.
2. Stop if x_k is a stationary point of optimization problem (2.1).
3. Compute n_k by (2.3). If (2.4) holds, compute t_k by (2.6) and set $d_k = n_k + t_k$. Otherwise, go to Step 7.
4. Line search filter with second order correction steps.
   4.1 Compute $\alpha_k^{\min}$ by (2.13). Set $\alpha_k^0 = 1$ and $l = 0$.
   4.2 If $\alpha_{k,l} < \alpha_k^{\min}$, go to step 7. Otherwise, compute a second order correction step w_k by (2.14) and set the new trial point $\bar{x}_{k+1} = x_k + \alpha_{k,l}d_k + \alpha_{k,l}^2 w_k$.
   4.3 If $(h(\bar{x}_{k+1}), f(\bar{x}_{k+1})) \in \mathscr{F}_k$, reject the trial step size and go to Step 4.7.
   4.4 If (2.11) holds for $\alpha_{k,l}$ and d_k, go to step 4.5. Otherwise go to Step 4.6.
   4.5 If $f(\bar{x}_{k+1}) \le f(x_k) + \mu[m_k(d_k) - m_k(0)]$, set $\alpha_k = \alpha_{k,l}$, $x_{k+1} = \bar{x}_{k+1}$ and go to Step 5. Otherwise, go to Step 4.7.
   4.6 If $h(\bar{x}_{k+1}) \le (1-\gamma_h)h(x_k)$ or $f(\bar{x}_{k+1}) \le f(x_k) - \gamma_f h(x_k)$ holds, set $\alpha_k = \alpha_{k,l}$, $x_{k+1} = \bar{x}_{k+1}$, augment the filter using (2.12) and go to Step 5.
   4.7 Choose $\alpha_{k,l} \in [\tau_1\alpha_{k,l}, \tau_2\alpha_{k,l}]$, set $l = l+1$ and go back to Step 4.2.
5. If $m_k(0) - m_k(d_k) > 0$, compute $\rho_k = \frac{f(x_k)-f(x_{k+1})}{m_k(0)-m_k(d_k)}$ and set
$$\sigma_{k+1} \in \begin{cases} [\min\{\sigma_k, \sigma_{\max}\}, \min\{\gamma_2\sigma_k, \sigma_{\max}\}) & \text{if } \rho_k < \eta_1, \\ [\min\{\gamma_1\sigma_k, \sigma_{\max}\}, \min\{\sigma_k, \sigma_{\max}\}) & \text{if } \rho_k \in [\eta_1,\ \eta_2), \\ (\min\{\gamma_0\sigma_k, \sigma_{\max}\}, \min\{\gamma_1\sigma_k, \sigma_{\max}\}] & \text{if } \rho_k \ge \eta_2. \end{cases}$$
   Otherwise, set $\sigma_{k+1} \in [\min\{\sigma_k, \sigma_{\max}\}, \min\{\gamma_2\sigma_k, \sigma_{\max}\})$.
6. Compute H_{k+1}. Set $k = k+1$ and go to Step 1.
7. Feasibility restoration phase.
   7.1 Compute a new iterate point x_{k+1} by decreasing $h(x)$ so that x_{k+1} satisfies both (2.9) and $(h(x_{k+1}), f(x_{k+1})) \notin \mathscr{F}_k$.
   7.2 Determine σ_{k+1}.
   7.3 Augment the filter (for x_k) and go to Step 6.

It can be verified that the global convergence analysis in [27] is still valid with modification of introducing second order correction steps. In this paper, we focus on the local convergence analysis.

3. Local Convergence Rate

Assumptions L. Algorithm 1 generates an infinite sequence x_k of iterate points that remain in a closed, bounded domain $X \subset \mathbb{R}^n$, regularisation parameter $\sigma_{\min} \leq \sigma_k \leq \sigma_{\max}$ for some constants $\sigma_{\min}, \sigma_{\max} > 0$.
(L1) $x_k \to x_*$ and x_* is a KKT point of problem (2.1).
(L2) H_k is semipositive definite on the null space of the Jocabian A_k^T for each k.
(L3) There exist positive constants M_A and M_H such that, for all k,

$$\sigma_{\min}(A_k) \geq M_A, \ \|H_k\| \leq M_H, \tag{3.1}$$

where $\sigma_{\min}(A_k)$ is the smallest singular value of A_k.
(L4) There exists a positive constant r such that $\frac{1}{2}p^T \nabla^2_{xx}\mathscr{L}(x_*, \lambda_*)p > 2r\|p\|^2$ if $A(x_*)^T p = 0$.
(L5) $(\hat{H}_k - W_k)d_k = o(\|d_k\|)$, where $\hat{H}_k = H_k + \sigma_k\|d_k\|I$ and

$$W_k := \nabla^2_{xx}\mathscr{L}(x_k, \lambda_*) = \nabla^2 f(x_k) + \sum_{i=1}^m \lambda_*^{(i)} \nabla^2 c^{(i)}(x_k).$$

From (2.3) and Condition (L3), one finds that there exists a constant $\kappa_n > 0$ such that

$$\|n_k\| \leq \kappa_n h(x_k). \tag{3.2}$$

Moreover, under Condition (L3), we can define

$$\lambda_k := -(A_k A_k^T)^{-1} A_k g_k. \tag{3.3}$$

First we give some preliminary results, Lemm1–Lemma 9. The main local convergence results is presented in Theorem 1.

Lemma 3.1. *Suppose that $k \notin \mathscr{R}_{\text{inc}}$, and t_k^c is the Cauchy step for (2.8), Then, for all $k \geq 0$,*

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

Proof. From (2.8) and the definition of the Cauchy step t_k^c, one finds that

$$t_k^c = -\beta_k^c(g_k + H_k n_k) \text{ and } \beta_k^c = \arg\min_{\beta \in \mathbb{R}^+} m_k^t(-\beta(g_k + H_k n_k)). \tag{3.4}$$

Due to $\beta \geq 0$, the Cauchy-Schwarz inequality and (3.4) indicate that

$$\begin{aligned}
& f(x_k) - m_k^t(t_k) \\
\geq \ & f(x_k) - m_k^t(-\beta(g_k + H_k n_k)) \\
= \ & \beta\|g_k + H_k n_k\|^2 - \frac{1}{2}\beta^2(g_k + H_k n_k)^T H_k (g_k + H_k n_k) - \frac{1}{3}\beta^3 \sigma_k \|g_k + H_k n_k\|^3 \\
\overset{(2.7)}{\geq} \ & \beta\|g_k^N\|^2(1 - \frac{1}{2}\beta\|H_k\| - \frac{1}{3}\beta^2 \sigma_k \|g_k^N\|).
\end{aligned} \tag{3.5}$$

Let

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

For $\beta\in[0,\hat{\beta}_k]$, it follows that $f(x_k)\geq m_k^t(t_k)$ provided that $1-\frac{1}{2}\beta\|H_k\|-\frac{1}{3}\beta^2\sigma_k\|g_k^N\|\geq 0$ and $\beta\geq 0$. Denote

$$\theta_k := \frac{1}{\sqrt{2}\max\left\{1+\|H_k\|,2\sqrt{\sigma_k\|g_k^N\|}\right\}}. \tag{3.6}$$

Combining

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

and

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

it follows that $0<\theta_k\leq\hat{\beta}_k$. Replacing β in (3.5) with θ_k indicates that

$$f(x_k)-m_k^t(t_k) \geq \frac{\|g_k^N\|^2(1-\frac{1}{2}\theta_k\|H_k\|-\frac{1}{3}\theta_k^2\sigma_k\|g_k^N\|)}{\sqrt{2}\max\left\{1+\|H_k\|,2\sqrt{\sigma_k\|g_k^N\|}\right\}}. \tag{3.7}$$

From the definition of θ_k in (3.6), we can deduce that $\theta_k\|H_k\|\leq 1$ and $\theta_k^2\sigma_k\|g_k^N\|\leq 1$. Therefore, the numerator of (3.7) is bounded below by $\frac{1}{6}\|g_k^N\|^2$. This result together with (3.7) yields that (3.10) is true. □

Lemma 3.2. *Suppose that (L2) holds. Then*

$$\|t_k\|\leq\sqrt{3}\sqrt{\frac{\|g_k^N\|}{\sigma_k}},\ k\geq 0 \tag{3.8}$$

holds.

Proof. Suppose that

$$\|t_k\|>\sqrt{3}\sqrt{\frac{\|g_k^N\|}{\sigma_k}} \tag{3.9}$$

for $k\geq 0$. From (L2), one finds that

$$m_k^t(t_k)-f(x_k) = (g_k+H_kn_k)^Tt_k+\frac{1}{2}t_k^TH_kt_k+\frac{1}{3}\sigma_k\|t_k\|^3$$

$$\overset{(2.7)}{\geq} \quad -\|t_k\|\|g_k^N\| + \frac{1}{3}\sigma_k\|t_k\|^3.$$

Combining (3.9) and $\frac{1}{3}\sigma_k\|t_k\|^3 - \|t_k\|\|g_k^N\| > 0$, one finds that $m_k^t(t_k) - f(x_k) > 0$. However, the above result contradicts (3.10). Thus the claim is true. □

Lemma 3.3. *Suppose that Assumptions L hold. Then, for all $k \geq 0$,*

$$h(x_k) = o(\|\alpha_k d_k\|^2) = o(\|d_k\|^2) \tag{3.10}$$

if x_k satisfies (2.11).

Proof. From (L2), (2.4) and (3.2), we can deduce that there exists a constant $M > 0$ such that, for all k, $\|g_k^N\| \leq M$ holds. Therefore, together with (3.8), one finds that

$$\|d_k\| \leq \|n_k\| + \|t_k\| \leq \min\left\{\sqrt{3M} + \beta_1, \sqrt{3M} + \beta_1\beta_2\sigma_k^{-\frac{\beta_3}{2}}\right\}\sigma_k^{-\frac{1}{2}}.$$

It follows from this result, (2.11), and $1 - \tau \leq 0$ that

$$\begin{aligned}
h(x_k) \quad & < \kappa_h^{-\frac{1}{\phi}}(-\alpha_k g_k^T d_k - \frac{1}{2}\alpha_k^2 d_k^T H_k d_k - \frac{1}{3}\alpha_k^3\sigma_k\|d_k\|^3)^{\frac{\tau}{\phi}}(\alpha_k\sqrt{\sigma_k})^{\frac{\tau-1}{\phi}} \\
& \leq \kappa_h^{-\frac{1}{\phi}}\|g_k + \frac{1}{2}\alpha_k H_k d_k\|^{\frac{\tau}{\phi}}\|\alpha_k d_k\|^{\frac{\tau}{\phi}}(\alpha_k\sqrt{\sigma_k})^{\frac{\tau-1}{\phi}} \\
& \overset{(3.1)}{\leq} \kappa_h^{-\frac{1}{\phi}}\left(M_g + \frac{1}{2}M_H(\sqrt{3M} + \beta_1)\sigma_{\min}^{-\frac{1}{2}}\right)^{\frac{\tau}{\phi}}(\sqrt{3M} + \beta_1)^{\frac{\tau-1}{\phi}}\|\alpha_k d_k\|^{\frac{1}{\phi}},
\end{aligned}$$

where $M_g \overset{\text{def}}{=} \max_{x\in X}\|\nabla f(x)\|$. Due to the last inequality, $0 < \phi < \frac{1}{2}$ gives that

$$h(x_k) = O(\|d_k\|^{\frac{1}{\phi}}) = o(\|d_k\|^2).$$

Therefore, the conclusion (3.10) follows. □

To prove the local convergence result, we introduce the exact penalty function

$$\Psi_\nu(x) = f(x) + \nu h(x) \tag{3.11}$$

only as a technical device. Moreover, the following model of the penalty function

$$q_\nu(x_k; d) = f(x_k) + g_k^T d + \frac{1}{2}d^T H_k d + \frac{1}{3}\sigma_k\|d\|^3 + \nu\|A_k d + c_k\|, \tag{3.12}$$

is employed, where g_k, H_k, A_k, c_k are defined same as in (2.2). From (2.14), it is easy to verify that

$$w_k = O(\|d_k\|^2). \tag{3.13}$$

Lemma 3.4. *Suppose that Assumptions L hold. Then $c(x_k + d_k + w_k) = o(\|d_k\|^2)$.*

Proof. Using the continuity of $c(x)$, one finds that

$$\begin{aligned}
c(x_k + d_k + w_k) \quad & = \quad c(x_k + d_k) + A(x_k + d_k)w_k + O(\|w_k\|^2) \\
& \overset{(2.14)}{=} \quad -(A_k - A(x_k + d_k))w_k + O(\|w_k\|^2) \\
& = \quad -(A_k - (A_k + O(\|d_k\|)))w_k + O(\|w_k\|^2) \\
& = \quad O(\|d_k\|\|w_k\|) + O(\|w_k\|^2)
\end{aligned}$$

$$\overset{(3.13)}{=} \quad o(\|d_k\|^2).$$

Thus the lemma is true. □

Lemma 3.5. *Suppose that the Assumptions L hold. Then, there exists an integer P_0 such that, for all $k \geq P_0$,*

$$q_\nu(x_k;0) - q_\nu(x_k;d_k) \geq r\|d_k\|^2 \geq 0, \tag{3.14}$$

with $\nu \geq \nu_0 \overset{\text{def}}{=} \|\lambda_k\| + (\frac{1}{2}M_H + r)\kappa_n$.

Proof. Since t_k^N is a solution of (2.6), one finds that

$$(N_k^T H_k N_k + \sigma_k \|t_k^N\| I)t_k^N = -N_k^T(g_k + H_k n_k),$$

and then $(t_k^N)^T N_k^T H_k N_k t_k^N + \sigma_k \|t_k^N\|(t_k^N)^T t_k^N = -(t_k^N)^T N_k^T(g_k + H_k n_k)$. Due to $t_k = N_k t_k^N$, we can see that

$$-(g_k + H_k n_k)^T t_k - \frac{1}{2}t_k^T H_k t_k = \frac{1}{2}(t_k^N)^T N_k^T H_k N_k t_k^N + \sigma_k \|t_k^N\|(t_k^N)^T t_k^N.$$

Using (3.2) and $h(x_k) \to 0$ if $k \to \infty$, it follows that $n_k \to 0$ if $k \to \infty$. Furthermore, one finds that $\sigma_k \geq \sigma_{\min} > 0$. As a result, we can deduce that there exists an integer P_0 such that, for all $k > P_0$, $\|n_k\| < 1$ and (2.4) hold. Since $f(x)$ and $c(x)$ are twice continuously differentiable, we see that $\mathscr{L}(x,\lambda)$ is twice continuously differentiable. As a result, W_k can be sufficiently close to $\nabla_{xx}^2 \mathscr{L}(x_*,\lambda_*)$ for all $k \geq P_0$. Thus (L4) implies that $\frac{1}{2}p^T W_k p > \frac{3}{2}r\|p\|^2$ if $A(x_k)p = 0$. Condition (L5) implies that, for all $k \geq P_0$, if $A(x_k)p = 0$, H_k can be sufficiently close to $\nabla_{xx}^2 \mathscr{L}(x_*,\lambda_*)$ such that $\frac{1}{2}p^T H_k p > r\|p\|^2$. Using the above results, one finds that

$$\begin{aligned}
& q_\nu(x_k;0) - q_\nu(x_k;d_k) \\
= \;& -(g_k + H_k n_k)^T t_k - \frac{1}{2}t_k^T H_k t_k - \frac{1}{2}n_k^T H_k n_k + \nu\|c_k\| - \lambda_k^T c_k + o(\|d_k\|^2) \\
\geq \;& \frac{1}{2}(t_k^N)^T N_k^T H_k N_k t_k^N - \frac{1}{2}n_k^T H_k n_k + (\nu - \|\lambda_k\|)\|c_k\| + o(\|d_k\|^2) \\
\geq \;& r\|t_k\|^2 + (\nu - \|\lambda_k\|)\|c_k\| - \frac{1}{2}n_k^T H_k n_k + o(\|d_k\|^2) \\
\geq \;& r\|t_k\|^2 + \frac{(\nu - \|\lambda_k\|)}{\kappa_n}\|n_k\| - \frac{1}{2}M_H\|n_k\|^2 + o(\|d_k\|^2) \\
\geq \;& r\|d_k\|^2 + o(\|d_k\|^2) \\
\geq \;& \frac{r}{2}\|d_k\|^2.
\end{aligned}$$

for $k \geq P_0$. Hence, the desired claim immediately follows. □

Lemma 3.6. *Suppose that Assumptions L hold. Then there exists an integer $P_1 \geq P_0$ (P_0 is from Lemma 3.5) such that, for all $k \geq P_1$,*

$$\Psi_\nu(x_k) - \Psi_\nu(x_k + d_k + w_k) \geq \varsigma(q_\nu(x_k;0) - q_\nu(x_k;d_k)), \tag{3.15}$$

with $\nu \geq \nu_0$ and $\varsigma \in (0,1)$.

Proof. Combining (3.11) and (3.12), we obtain that

$$\Psi_\nu(x_k) - \Psi_\nu(x_k + d_k + w_k) - \varsigma(q_\nu(x_k;0) - q_\nu(x_k;d_k))$$

$$
\begin{aligned}
&= f_k - f(x_k+d_k+w_k) + \nu\|c_k\| - \varsigma(-g_k^T d_k - \frac{1}{2}d_k^T H_k d_k - \frac{1}{3}\sigma_k\|d_k\|^3 + \nu\|c_k\|) \\
&= -g_k^T d_k - \frac{1}{2}d_k^T \nabla^2 f_k d_k - g(x_k+d_k)^T w_k + \nu\|c_k\| \\
&\quad -\varsigma(-g_k^T d_k - \frac{1}{2}d_k^T H_k d_k - \frac{1}{3}\sigma_k\|d_k\|^3 + \nu\|c_k\|) + o(\|d_k\|^2). \qquad (3.16)
\end{aligned}
$$

From (2.14), one finds that

$$
\begin{aligned}
g(x_k+d_k)^T w_k &= (g_k + O(\|d_k\|))^T w_k \\
&\overset{(2.14)}{=} g_k^T w_k + o(\|d_k\|^2) \\
&\overset{(2.14),(3.3)}{=} \lambda_k^T c(x_k+d_k) + o(\|d_k\|^2) \\
&= \frac{1}{2}\sum_{i=1}^{m} \lambda_k^{(i)} d_k^T \nabla^2 c_k^{(i)} d_k + o(\|d_k\|^2).
\end{aligned}
$$

Due to $\nu \geq \nu_0 = \|\lambda_k\| + (\frac{1}{2}M_H + r)\kappa_n$, it follows from the above equality and (3.16) that

$$
\begin{aligned}
& \Psi_\nu(x_k) - \Psi_\nu(x_k+d_k+w_k) - \varsigma(q_\nu(x_k;0) - q_\nu(x_k;d_k)) \\
=\ & -g_k^T d_k - \frac{1}{2}d_k^T H_k d_k + \nu\|c_k\| + \frac{1}{2}d_k^T(H_k - \nabla_{xx}^2\mathscr{L}(x_k,\lambda_k))d_k \\
& -\varsigma(-g_k^T d_k - \frac{1}{2}d_k^T H_k d_k + \nu\|c_k\|) + o(\|d_k\|^2) \\
\overset{(L5)}{=}\ & (1-\varsigma)(-g_k^T d_k - \frac{1}{2}d_k^T H_k d_k + \nu\|c_k\|) + o(\|d_k\|^2) \\
\overset{(3.14)}{\geq}\ & \frac{(1-\varsigma)r}{2}\|d_k\|^2 + o(\|d_k\|^2).
\end{aligned}
$$

Hence, we can obtain that there exists an integer $P_1 \geq P_0$ such that, for all $k \geq P_1$, $(1-\varsigma)r\|d_k\|^2 + o(\|d_k\|^2) \geq 0$. Therefore, the claim (3.15) is true. □

Lemma 3.7. *Suppose that Assumptions L hold. Then there exists an integer $P_2 \geq P_1$ (P_1 is from Lemma 3.6) such that for $k \geq P_2$,*

$$
f(x_k+d_k+w_k) \leq f(x_k) + \mu(g_k^T d_k + \frac{1}{2}d_k^T H_k d_k + \frac{1}{3}\sigma_k\|d_k\|^3)
$$

whenever (2.11) holds.

Proof. Lemma 3.4, Lemma 3.5 and $\mu \in (0, 1/2)$ indicate that there exists P_1 such that

$$
\Psi_{\nu_0}(x_k) - \Psi_{\nu_0}(x_k+d_k+w_k) \geq (\frac{1}{2}+\mu)(q_{\nu_0}(x_k;0) - q_{\nu_0}(x_k;d_k)) \qquad (3.17)
$$

follows for every $k \geq P_1$. Then, from (3.11) and (3.17), we can deduce that

$$
\begin{aligned}
& f(x_k) - f(x_k+d_k+w_k) + \mu(g_k^T d_k + \frac{1}{2}d_k^T H_k d_k + \frac{1}{3}\sigma_k\|d_k\|^3) \\
=\ & \Psi_{\nu_0}(x_k) - \Psi_{\nu_0}(x_k+d_k+w_k) - \nu_0(h(x_k) - h(x_k+d_k+w_k)) \\
& +\mu(g_k^T d_k + \frac{1}{2}d_k^T H_k d_k + \frac{1}{3}\sigma_k\|d_k\|^3)
\end{aligned}
$$

$$\begin{aligned}
&\overset{(3.10)}{\geq} \quad (\frac{1}{2}+\mu)(q_{\nu_0}(x_k;0)-q_{\nu_0}(x_k;d_k))+\mu(g_k^Td_k+\frac{1}{2}d_k^TH_kd_k)+o(\|d_k\|^2)\\
&= \quad \frac{1}{2}(q_{\nu_0}(x_k;0)-q_{\nu_0}(x_k;d_k))+\mu\nu_0\|c_k\|+o(\|d_k\|^2)\\
&\overset{(3.14)}{\geq} \quad \frac{1}{4}r\|d_k\|^2+o(\|d_k\|^2).
\end{aligned}$$

Thus there exists an integer $P_2 \geq P_1$ such that $\frac{1}{2}r\|d_k\|^2+o(\|d_k\|^2) \geq 0$ for all $k \geq P_2$ and the desired conclusion holds. □

Lemma 3.8. *Let Assumptions L hold. Then there exist constants $\nu_1,\ \nu_2,\ \nu_3 > 0$ with*

$$2\gamma_h\nu_2 < (1+\gamma_h)(\nu_2-\nu_1)-2\gamma_f, \tag{3.18}$$

$$2\nu_3 \geq (1+\gamma_h)\nu_1+(1+\gamma_h)\nu_2, \tag{3.19}$$

such that ν_i is greater than ν_0 for $i=1,2,3$. Furthermore, for $k \geq P_2$ with P_2 in Lemma 3.7, we have that

$$\Psi_{\nu_i}(x_k)-\Psi_{\nu_i}(x_k+d_k+\omega_k) \geq \frac{1+\gamma_h}{2}(q_{\nu_i}(x_k;0)-q_{\nu_i}(x_k;d_k)) \overset{(3.14)}{\geq} 0 \tag{3.20}$$

for $i=1,2,3$ and all choices of w_k:

$$\omega_k = w_k, \tag{3.21}$$

$$\omega_k = w_k+d_{k+1}+w_{k+1}, \tag{3.22}$$

$$\omega_k = w_k+d_{k+1}+w_{k+1}+d_{k+2}+w_{k+2}, \tag{3.23}$$

$$\omega_k = w_k+d_{k+1}+w_{k+1}+d_{k+2}+w_{k+2}+d_{k+3}+w_{k+3} \tag{3.24}$$

as long as $x_{l+1} = x_l+d_l+w_l$ for $l \in \{k,...k+j\}$ with $j \in \{-1,0,1,2\}$, respectively.

Proof. Choose $\nu_1 \geq \nu_0$ with ν_0 in Lemma 3.5, and define

$$\nu_2 = \frac{1+\gamma_h}{1-\gamma_h}\nu_1+\frac{3\gamma_f}{1-\gamma_h},\ \nu_3 = (1-\gamma_h)\nu_2-\gamma_f. \tag{3.25}$$

Then, $\nu_2,\nu_3 \geq \nu_1 \geq \nu_0$, (3.18), and (3.19) hold immediately. Due to $\frac{1+\gamma_h}{2} < 1$, Lemma 3.6 and Lemma 3.5 imply that (3.20) holds for ω_k in (3.21) when $k \geq P_2$ with P_2 in Lemma 3.7. From ω_k in (3.22), i.e., $\omega_k = w_k+d_{k+1}+w_{k+1}$, as long as x_{k+1} is taken as $x_k+d_k+w_k$ in Algorithm 1, we can obtain that

$$\begin{aligned}
& \quad \Psi_{\nu_i}(x_k)-\Psi_{\nu_i}(x_k+d_k+\omega_k)\\
&= \quad \Psi_{\nu_i}(x_k)-\Psi_{\nu_i}(x_k+d_k+w_k+d_{k+1}+w_{k+1})\\
&= \quad \Psi_{\nu_i}(x_k)-\Psi_{\nu_i}(x_k+d_k+w_k)+\Psi_{\nu_i}(x_k+d_k+w_k)\\
& \quad -\Psi_{\nu_i}(x_k+d_k+w_k+d_{k+1}+w_{k+1})\\
&= \quad \Psi_{\nu_i}(x_k)-\Psi_{\nu_i}(x_k+d_k+w_k)+\Psi_{\nu_i}(x_{k+1})-\Psi_{\nu_i}(x_{k+1}+d_{k+1}+w_{k+1})\\
&\geq \quad \frac{1+\gamma_h}{2}(q_{\nu_i}(x_k;0)-q_{\nu_i}(x_k;d_k))+\Psi_{\nu_i}(x_{k+1})-\Psi_{\nu_i}(x_{k+1}+d_{k+1}+w_{k+1}).
\end{aligned}$$

Combining Lemma 3.6 and Lemma 3.5, one finds that

$$\Psi_{\nu_i}(x_{k+1})-\Psi_{\nu_i}(x_{k+1}+d_{k+1}+w_{k+1}) \geq \varsigma(q_{\nu_i}(x_{k+1};0)-q_{\nu_i}(x_{k+1};d_{k+1})) \geq 0$$

since $k+1>k\geq P_2\geq P_1\geq P_0$ and $\nu_1\geq\nu_0$. It follows from results above that (3.20) holds for ω_k in (3.22). Therefore, we can verify that (3.20) holds for ω_k in (3.23) and (3.24) in the same manner. □

For brevity, we denote some notations. Due to $x_k\to x_*$ $(k\to\infty)$, we denote that

$$M=\{x_k:k\geq P_2 \text{ and } \Psi_{\nu_i}(x_k)\leq\Psi_{\nu_i}(x_*)+\kappa\}, \tag{3.26}$$

where $\kappa>0$ is chosen so that, for all $x_k\in M$, $(h(x_k),f(x_k))\notin\mathscr{F}_{P_2}$. This is possible because $H_{P_2}=\min\{h:(h,f)\in\mathscr{F}_{P_2}\}>0$ and $\max\{h(x_k):x_k\in M\}\to 0$ when $k\to\infty$ (see Subsection 14.5 in [30]). Furthermore, for $k\in\mathbb{N}$, we denote

$$\mathscr{G}_k=\{(h,f):h\geq(1-\gamma_h)h(x_k) \text{ and } f\geq f(x_k)-\gamma_f h(x_k)\} \tag{3.27}$$

as the filter building blocks, and index sets $I_{k_1}^{k_2}=\{l=k_1,...,k_2-1:l\in\mathscr{A}\}$ for $k_1<k_2$. Using (2.12) and the definition of $\mathscr{A}$, one finds that

$$\mathscr{F}_{k_2}=\mathscr{F}_{k_1}\cup\bigcup_{l\in I_{k_1}^{k_2}}\mathscr{G}_l. \tag{3.28}$$

holds for $k_1\leq k_2$.

Lemma 3.9. *Let Assumptions L hold. Let $l\geq P_2$ (P_2 is from Lemma 3.7) with $h(x_l)>0$. Then*

$$if\ \Psi_{\nu_2}(x_l)-\Psi_{\nu_2}(x)\geq\frac{1+\gamma_h}{2}(q_{\nu_2}(x_l;0)-q_{\nu_2}(x_l;d_l))\ for\ a\ given\ x\in\mathbb{R}^n,$$

then,

$$(h(x),f(x))\notin\mathscr{G}_l, \tag{3.29}$$

where ν_2 is from Lemma 3.8.

$$If\ \Psi_{\nu_2}(x_{P_3})-\Psi_{\nu_2}(x_k)\geq\frac{1+\gamma_h}{2}(q_{\nu_2}(x_{P_3};d_{P_3})-q_{\nu_2}(x_{P_3};d_{P_3}))\ for\ all\ x_k\in M,$$

then,

$$(h(x_k),f(x_k))\notin\mathscr{F}_{P_3}, \tag{3.30}$$

where P_3 is the first iteration after P_2 with $x_{P_3}\in M$.

Proof. Due to $\nu_1\geq\nu_0$, Lemma 3.8 implies that $q_{\nu_2}(x_l;0)-q_{\nu_2}(x_l;d_l)\geq 0$. Combining (3.12) and $A_ld_l+c_l=0$, one finds that

$$\begin{aligned}\Psi_{\nu_2}(x_l)-\Psi_{\nu_2}(x) &\geq\frac{1+\gamma_h}{2}(q_{\nu_2}(x_l;0)-q_{\nu_2}(x_l;d_l))\\ &=\frac{1+\gamma_h}{2}(q_{\nu_1}(x_l;0)-q_{\nu_1}(x_l;d_l)+(\nu_2-\nu_1)h(x_l))\\ &\overset{(3.14)}{\geq}\frac{1+\gamma_h}{2}(\nu_2-\nu_1)h(x_l).\end{aligned} \tag{3.31}$$

If $f(x)<f(x_l)-\gamma_f h(x_l)$, the desired results holds. Otherwise, from $h(x_l)>0$, one finds that

$$\begin{aligned}h(x_l)-h(x) &\overset{(3.11),(3.31)}{\geq}\frac{1+\gamma_h}{2\nu_2}(\nu_2-\nu_1)h(x_l)+\frac{f(x)-f(x_l)}{\nu_2}\\ &\geq\frac{1+\gamma_h}{2\nu_2}(\nu_2-\nu_1)h(x_l)-\frac{\gamma_f}{\nu_2}h(x_l)\end{aligned}$$

$$\overset{(3.18)}{\geq} \quad \gamma_h h(x_l).$$

Then the proof of (3.29) is complete.

(3.26) and $x_k \in M$ indicate that $(h(x_k), f(x_k)) \notin \mathscr{F}_{P_2}$. Choose any $l \in I_{P_2}^{P_3}$. It follows from (3.31) that

$$\Psi_{\nu_2}(x_{P_3}) - \Psi_{\nu_2}(x_k) \geq \frac{1+\gamma_h}{2}(\nu_2 - \nu_1)h(x_{P_3}). \tag{3.32}$$

Because x_{P_3} is the first point after P_2 with $x_{P_3} \in M$, the definition of M in (3.26) together with $l \leq P_3$ implies that

$$\begin{aligned} \Psi\nu_3(x_l) &\overset{(3.26)}{\geq} \Psi\nu_3(x_{P_3}) \\ &= \Psi\nu_2(x_{P_3}) + (\nu_3 - \nu_2)h(x_{P_3}) \\ &\overset{(3.32)}{\geq} \Psi\nu_2(x_k) + (\nu_3 - \frac{1+\gamma_h}{2}\nu_1 - \frac{1-\gamma_h}{2}\nu_2)h(x_{P_3}) \\ &\overset{(3.19)}{\geq} \Psi_{\nu_2}(x_k). \end{aligned} \tag{3.33}$$

If $f(x_k) < f(x_l) - \gamma_f h(x_l)$, it follows that $(h(x_k), f(x_k)) \notin \mathscr{G}_l$ directly. Otherwise, $f(x_k) \geq f(x_l) - \gamma_f h(x_l)$ together with (3.11) yields that

$$h(x_k) \overset{(3.33)}{<} \frac{1}{\nu_2}(f(x_l) + \nu_3 h(x_l) - f(x_k)) \leq \frac{\nu_3 + \gamma_f}{\nu_2} h(x_l) \overset{(3.25)}{=} (1-\gamma_h)h(x_l),$$

which implies that $(h(x_k), f(x_k)) \notin \mathscr{G}_l$. Then, (3.28) implies that (3.30) holds. □

Theorem 3.10. *Let Assumptions L hold. Then, for k sufficiently large, $x_{k+1} = x_k + d_k$ or $x_{k+1} = x_k + d_k + w_k$ are taken, and x_k converges to x_* superlinearly.*

Proof. Firstly, we show that, for $k \geq P_3 + 2$, the following statements follow

(i$_k$) $\Psi_{\nu_i}(x_l) - \Psi_{\nu_i}(x_k) \geq \frac{1+\gamma_h}{2}(q_{\nu_i}(x_l;0) - q_{\nu_i}(x_l;d_l))$ for $i \in \{2,3\}$ and $P_3 \leq l \leq k-2$,

(ii$_k$) $x_k \in M$,

(iii$_k$) $x_k = x_{k-1} + d_{k-1} + w_{k-1}$.

Next, we prove that, for $k = P_3 + 2$, these statements hold. Assume that $x_{P_3} + d_{P_3}$ is not acceptable for the filter. In this case, let $\bar{x}_{P_3+1} = x_{P_3} + d_{P_3} + w_{P_3}$. It follows from (3.20) and (3.21) with $i = 3$, $k = P_3$, $x_{P_3} \in M$ that

$$\Psi_{\nu_3}(\bar{x}_{P_3+1}) \leq \Psi_{\nu_3}(x_{P_3}) \leq \Psi_{\nu_3}(x_*) + \kappa.$$

Then, the definition of M indicates that

$$\bar{x}_{P_3+1} \in M. \tag{3.34}$$

Using (3.20) with $i = 2$. (3.30) and (3.34), we see that $(h(x_{P_3+1}), f(x_{P_3+1})) \notin \mathscr{F}_{P_3}$. Meanwhile, if (2.11) is satisfied, Lemma 3.7 implies that $x_{P_3+1} = x_{P_3} + d_{P_3} + w_{P_3}$ is taken in the algorithm. If (2.11) does not hold, then $h(x_{P_3}) > 0$. From (2.11) for $i = 2$, $k = P_3$, (3.21) and (3.29) for $l = P_3$, one finds that

$$h(\bar{x}_{P_3+1}) \leq (1-\gamma_h)h(x_{P_3}) \text{ or } f(\bar{x}_{P_3+1}) \leq f(x_{P_3}) - \gamma_f h(x_{P_3}).$$

Therefore, $x_{P_3+1} = \bar{x}_{P_3+1} = x_{P_3} + d_{P_3} + w_{P_3}$. We now consider iteration $P_3 + 1$. It follows from (2.11) for $k = P_3$ and (3.22) that

$$\Psi_{v_i}(x_{P_3}) - \Psi_{v_i}(x_{P_3+1} + d_{P_3+1} + w_{P_3+1}) \geq \frac{1+\gamma_h}{2}(q_{v_i}(x_{P_3};0) - q_{v_i}(x_{P_3};d_{P_3})) \tag{3.35}$$

for $i = 2,3$. This result indicates that

$$\bar{x}_{P_3+2} \stackrel{\text{def}}{=} x_{P_3+1} + d_{P_3+1} + w_{P_3+1} \in M. \tag{3.36}$$

Using (3.30), (3.35), and (3.36), one finds that $(h(\bar{x}_{P_3+2}), f(\bar{x}_{P_3+2})) \notin \mathscr{F}_{P_3}$. If $P_3 \notin I_{P_3}^{P_3-1}$, from (3.28), $(h(\bar{x}_{P_3+2}), f(\bar{x}_{P_3+2})) \notin \mathscr{F}_{P_3+1}$ follows. Otherwise, (3.35) for $i = 2$ and (3.29) indicate that $(h(\bar{x}_{P_3+2}), f(\bar{x}_{P_3+2})) \notin \mathscr{G}_{P_3}$. Then, this result with (3.28) implies that $(h(\bar{x}_{P_3+2}), f(\bar{x}_{P_3+2})) \notin \mathscr{F}_{P_3+1}$. Therefore, $\bar{x}_{P_3+2}$ is accepted in step 4.3. Similarly, one finds that $\bar{x}_{P_3+2}$ is accepted in Step 4.5 or Step 4.6. Thus $x_{P_3+2} = \bar{x}_{P_3+2}$, which proves $(\mathrm{i}_{P_3+2}) - (\mathrm{iii}_{P_3+2})$. For all $P_3 + 2 \leq l \leq k$ with some $k \geq P_3 + 2$, we assume that $(\mathrm{i}_l) - (\mathrm{iii}_l)$ hold. Define $\bar{x}_{k+1} = x_k + d_k + w_k$. From (3.20) with (3.22) for $i = 2,3$, one finds that

$$\Psi_{v_i}(x_{k-1}) - \Psi_{v_i}(\bar{x}_{k+1}) \geq \frac{1+\gamma_h}{2}(q_{v_i}(x_{k-1};0) - q_{v_i}(x_{k-1};d_{k-1})) \geq 0. \tag{3.37}$$

Then, we choose l with $P_3 \leq l < k-1$ and consider the following two cases.

Case (1). If $k = P_3 + 2$, then $l = P_3$. Combining (3.20) and (3.24), we can deduce that, for $i = 2,3$,

$$\Psi_{v_i}(x_l) - \Psi_{v_i}(\bar{x}_{k+1}) \geq \frac{1+\gamma_h}{2}(q_{v_i}(x_l;0) - q_{v_i}(x_l;d_l)) \geq 0. \tag{3.38}$$

Case (2). If $k > P_3 + 2$, (3.37) implies that $\Psi_{v_i}(x_{k-1}) \geq \Psi_{v_i}(\bar{x}_{k+1})$. Hence, one finds that (3.38) follows from (i_{k-1}).

In either case, $\Psi_{v_3}(\bar{x}_{k+1}) \leq \Psi_{v_3}(x_{P_3})$ follows directly from (3.38). Therefore, $x_{P_3} \in M$ implies that

$$\bar{x}_{k+1} \in M. \tag{3.39}$$

Furthermore, it follows from (3.30), (3.38), and (3.39) when $i = 2$ and $l = P_3$ that

$$(h(\bar{x}_{k+1}), f(\bar{x}_{k+1})) \notin \mathscr{F}_{P_3}.$$

Moreover, it follows from (3.38) when $i = 2$ and $l = P_3$, (3.39) and (3.30) that $(h(\bar{x}_{k+1}), f(\bar{x}_{k+1})) \notin \mathscr{F}_{P_3}$. For $l \in I_{P_3}^k$, we have from (3.37) and (3.38) with (3.29) that $(h(\bar{x}_{k+1}), f(\bar{x}_{k+1})) \notin \mathscr{G}_l$. Therefore, (3.28) implies that $\bar{x}_{k+1}$ is not denied in Step 4.3. Thus we can verify that $\bar{x}_{k+1}$ is accepted in Step 4.5 or Step 4.6 in the same manner. Hence, $x_{k+1} = \bar{x}_{k+1}$ indicates that $(\mathrm{i}_{k+1}) - (\mathrm{iii}_{k+1})$.

Next, if we replace B_k by $\hat{H}_k$ in [31], (L5) can ensure that $\{x_k\} \to x_*$ with a superlinear rate. $\square$

4. Numerical Results

In this section, we present the numerical results of Algorithm 1 which were performed on a desktop with Intel(R) Core(TM) i9-12900K 3.20 GHz. Numerical testing was implemented as a MATLAB code and run under MATLAB (R2022a).

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

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

is satisfied. The numerical results are reported in Table 1. The test problems are from CUTEst collection [32]. 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.

TABLE 1 Numerical results of Algorithm 1.

Problem	Dimension		NIT	NF	NC	NG	Res
	n	m					
AIRCRFTA	8	5	2	3	3	2	1.5932e−08
ARGTRIG	200	200	3	3	4	4	6.8423e−07
AUG2D	703	300	3	4	4	3	9.0601e−12
BDVALUE	102	100	1	2	2	1	2.7664e−07
BOOTH	2	2	2	2	3	2	8.8818e−16
BRATU3D	125	27	2	2	3	2	1.3946e-12
BROYDN3D	500	500	2	2	3	2	1.3946e−12
BT1	2	1	6	8	7	6	6.2509e−10
BT2	3	1	6	6	7	6	5.7545e−12
BT3	5	3	3	3	4	3	5.0337e−08
BT4	3	2	5	5	6	5	2.6930e−12
BT5	3	2	4	4	5	4	3.2447e−09
BT6	5	2	11	12	12	11	3.3946e−08
BT7	5	3	7	8	8	7	6.8003e−14
BT8	5	2	11	12	12	11	3.3717e−07
BT9	4	2	5	5	6	5	7.5925e−09
BT10	2	2	2	2	3	2	2.0895e−09
BT11	5	3	7	8	8	7	1.1500e−09
BT12	5	3	4	5	5	4	1.3465e−09
BYRDSPHR	3	2	7	7	8	7	3.5170e−13
CLUSTER	2	2	5	9	6	5	3.2963e−08
DTOC4	299	198	2	3	3	2	3.9938e−10
DTOC5	19	9	0	1	1	1	0.0000e+00
GENHS28	10	8	4	4	5	4	8.9137e−07
GOTTFR	2	3	2	2	3	2	2.1823e−10
HAGER1	1001	500	3	4	4	3	9.4429e−07
HAGER2	1001	500	3	4	4	3	7.3961e−07
HAGER3	1001	500	2	3	3	2	6.3420e−07
HATFLDF	3	3	3	4	4	3	5.2096e−07
HATFLDG	25	25	2	2	3	2	4.0736e−12
HEART8	8	8	2	2	3	2	1.1043e−11

Table 1 continued

Problem	Dimension n	Dimension m	NIT	NF	NC	NG	Res
HIMMELBA	2	2	1	1	2	1	5.3134e−07
HIMMELBC	2	2	2	2	3	2	2.8422e−13
HIMMELBE	3	3	1	1	2	1	7.4308e−07
HS06	2	1	5	11	6	5	3.0642e−13
HS07	2	1	5	5	6	5	4.5113e-10
HS08	2	2	2	2	3	2	2.4788e−13
HS09	2	1	9	53	10	9	2.8111e−07
HS26	3	1	18	19	19	18	6.9635e−07
HS27	3	1	13	14	14	13	4.5717e−09
HS28	3	1	3	4	4	3	1.9369e−08
HS39	4	2	8	9	9	8	6.4001e−07
HS40	4	3	3	4	4	3	1.7066e−10
HS42	4	2	5	5	6	5	1.0657e−11
HS46	5	2	18	19	19	18	4.8371e−07
HS47	5	3	16	17	17	16	6.9164e−07
HS48	5	2	3	4	4	3	5.4963e−11
HS49	5	2	16	17	17	16	4.8668e−07
HS50	5	3	9	10	10	9	4.4526e−13
HS51	5	3	2	3	3	2	1.4423e−08
HS52	5	3	2	3	3	2	3.1612e−10
HS56	7	4	10	11	11	10	5.5067e−07
HS61	3	2	6	6	7	7	6.3198e−07
HS77	5	2	10	11	11	10	2.3628e−08
HS78	5	3	8	9	9	8	2.7821e−07
HS79	5	3	5	5	6	5	6.7966e−10
HS100LNP	7	2	6	8	7	6	4.6236e−09
HS111LNP	10	3	20	20	21	20	2.7174e−07
HYPCIR	2	2	1	1	2	1	5.4209e−07
INTEGREQ	5	5	3	4	4	3	3.8263e−07
MARATOS	2	1	3	4	4	3	1.5309e−08
MWRIGHT	5	3	6	9	7	6	2.4979e−10
ORTHREGB	27	6	7	8	8	8	1.1322e−08
POWELLSQ	2	2	1	1	2	1	8.1584e−07
RECIPE	3	3	2	2	3	2	3.0306e−15
S235	3	1	13	15	14	13	1.0455e−08
S252	3	1	10	11	11	10	6.0873e−09
S265	4	2	1	1	2	1	3.9252e−17
S269	5	3	29	29	30	29	8.5895e−07
S316	2	1	2	2	3	3	4.8916e−08
S317	2	1	5	5	6	6	3.3780e−12
S318	2	1	5	5	6	6	3.2496e−11

Table 1 continued

Problem	Dimension		NIT	NF	NC	NG	Res
	n	m					
S319	2	1	7	7	8	8	6.3140e−08
S320	2	1	25	46	26	26	9.1824e−07
S321	2	1	17	35	18	18	2.9056e−07
S335	3	2	9	8	10	19	8.1612e−11
S336	3	2	5	5	6	5	2.3466e−08
S338	3	2	6	6	7	6	1.8903e−09
S344	3	1	9	10	10	9	5.7431e−07
S373	9	6	10	9	11	10	8.4754e−11
S378	10	3	17	19	18	17	1.9417e−10
S394	20	12	9	14	10	9	4.7112e−11
S395	50	1	8	9	9	8	6.1161e−12
ZANGWIL3	3	3	2	2	3	2	1.2098e−20

5. Conclusion

We introduced an adaptive regularization with cubics algorithm in association with backtracking line search filter technique for solving the nonlinear equality constrained programming, where second-order correction steps were introduced. After the new step is computed, we employed line search filter techniques to generate the next iteration point, which 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. And the cubic regularization presents a way to avoid difficulties resulting from the incompatibility of the intersection of linearized constraints with trust-region bounds in constrained optimization. Compared with algorithms based on line search frameworks, which require that the Lagrangian Hessian or its approximation is uniformly positive definite on the null space of the Jacobian of constraints to guarantee the descent property of the search directions, the proposed algorithm only requires semi-positive definiteness of Lagrangian Hessian to ensure the global convergence. In the local convergence analysis we demonstrated that Maratos effect can be avoided and fast local convergence rate can be achieved. The preliminary numerical results were also reported.

Acknowledgments
The author is grateful to the anonymous referees and the editor for their time and consideration on this paper. The author gratefully acknowledges the supports of Key Scientific Research Project for Colleges and Universities in Henan Province (23B110002), Horizontal Project: Research on the Algorithm for Predicting the Service Life of the Filter Element (5201019160013).

References

[1] R. H. Byrd, R. B. Schnabel, G. A. Schultz, A Trust region algorithm for nonlinearly constrained optimization, SIAM J. Numer. Anal. 24 (1987) 1152-1170.

[2] J. Dennis, M. Heinkenschloss, N. Vicente, On the convergence theory of trust-region-based algorithms for equality-constrained optimization, SIAM J. Optim. 7 (1997) 927-950.

[3] J. E. Dennis, M. El-Alem, M. C. Maciel, A global convergence theory for general trust-region-based algorithms for equality constrained optimization, SIAM J. Optim. 7 (1995) 177–207.

[4] R. Fletcher, L.M. Gould, L. Nicholas, T. Sven, L. Philippe, A. Wachter, Global convergence of a trust-region sqp-filter algorithm for general nonlinear programming, SIAM J. Optim. 13 (2000) 635-659.

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

[6] 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 (2011) 295–319.

[7] S. Bellavia, G. Gurioli, B. Morini, Adaptive cubic regularization methods with dynamic inexact Hessian information and applications to finite-sum minimization, IMA J. Numer. Anal. 41 (2021) 764–799.

[8] H. Benson, D. Shanno, Cubic regularization in symmetric rank-1 quasi-Newton methods, Math. Program. Comput. 10 (2018) 457–486.

[9] E. Bergou, D. Youssef, G. Serge, A line-search algorithm inspired by the adaptive cubic regularization framework and complexity analysis, J. Optim. Theory Appl. 178 (2018) 885–913.

[10] T. Bianconcini, M. Sciandrone, A cubic regularization algorithm for unconstrained optimization using line search and nonmonotone techniques, Optim. Methods Softw. 31 (2016) 1008–1035.

[11] C. Cartis, N. Gould, P. Toint, A concise second-order complexity analysis for unconstrained optimization using high-order regularized models, Optim. Methods Softw. 35 (2020) 243–256.

[12] T. Dehghan Niri, M. Heydari, M. M. Hosseini, An improvement of adaptive cubic regularization method for unconstrained optimization problems, Int. J. Comput. Math. 98 (2021) 271–287.

[13] J.-P. Dussault, $\mathrm{ARC_q}$: a new adaptive regularization by cubics, Optim. Methods Softw. 33 (2018) 322–335.

[14] J. M. Martínez, M. Raydan, Cubic-regularization counterpart of a variable-norm trust-region method for unconstrained minimization, J. Global Optim. 68 (2017) 367–385.

[15] S. Park, S. H. Jung, P. M. Pardalos, Combining stochastic adaptive cubic regularization with negative curvature for nonconvex optimization, J. Optim. Theory Appl. 184 (2020) 953–971.

[16] L. Sha, Z. Wei, L. Li, A trust region algorithm with adaptive cubic regularization methods for nonsmooth convex minimization, Comput. Optim. Appl. 51 (2012) 551–573.

[17] T. Zhao, H. Liu, Z. Liu, New subspace minimization conjugate gradient methods based on regularization model for unconstrained optimization, Numer. Algor. 87 (2021) 1501–1534.

[18] E. Bergou, Y. Diouane, S. Gratton, On the use of the energy norm in trust-region and adaptive cubic regularization subproblems, Comput. Optim. Appl. 68 (2017) 533–554.

[19] N. I. M. Gould, M. Porcelli, P. L. Toint, Updating the regularization parameter in the adaptive cubic regularization algorithm, Comput. Optim. Appl. 53 (2012) 1–22.

[20] Y. Hsia, R.-L. Sheu, Y.-X. Yuan, Theory and application of p-regularized subproblems for $p > 2$, Optim. Methods Softw. 32 (2017) 1059–1077.

[21] R. Jiang, M.-C. Yue, Z. Zhou, An accelerated first-order method with complexity analysis for solving cubic regularization subproblems, Comput. Optim. Appl. 79 (2021) 471–506.

[22] F. Lieder, Solving large-scale cubic regularization by a generalized eigenvalue problem, SIAM J. Optim. 30 (2020) 3345–3358.

[23] N. Agarwal, N. Boumal, B. Bullins, C. Cartis, Adaptive regularization with cubics on manifolds, Math. Program. 188 (2021) 85–134.

[24] C. Cartis, N. I. Gould, P. L. Toint, Universal regularization methods: varying the power, the smoothness and the accuracy, SIAM J. Optim. 29 (2019) 595–615.

[25] C. Cartis, N. I. M. Gould, P. L. Toint, An adaptive cubic regularization algorithm for nonconvex optimization with convex constraints and its function-evaluation complexity, IMA J. Numer. Anal. 32 (2012) 1662–1695.

[26] Y. Pei, S. Song, D. Zhu, A sequential adaptive regularisation using cubics algorithm for solving nonlinear equality constrained optimization, Comput. Optim. Appl. 84 (2023) 1005–1033.

[27] Y. Pei, J. Wang, S. Song, Q. Gao, D. Zhu, A line search filter sequential adaptive cubic regularisation algorithm for nonlinearly constrained optimization, (2024). https://arxiv.org/abs/2409.04784

[28] R. Fletcher, S. Leyffer, Nonlinear programming without a penalty function. Math. Program. 91 (1999), 239-269.

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

[30] A. R. Conn, N. I. M. Gould, P. L. Toint, Trust-region methods, MPS/SIAM Series on Optimization, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Programming Society (MPS), Philadelphia, 2000.

[31] P. T. Boggs, J. W. Tolle, P. Wang, On the local convergence of quasi-newton methods for constrained optimization, SIAM J. Control Optim. 20 (1982) 161–171.

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