J. Nonlinear Funct. Anal. 2021 (2021), Article ID 28 https://doi.org/10.23952/jnfa.2021.28

MATHRES Journal of Nonlinear Functional Analysis
Available online at http://jnfa.mathres.org

A DERIVATIVE-FREE CUBIC REGULARIZATION METHOD FOR NONLINEAR SYSTEMS OF EQUATIONS

XIAOJIN HUANG

School of Mathematics and Statistics, Yulin Normal University, Yulin 537000, China

Abstract. We propose a derivative-free cubic regularization method for solving nonlinear systems of equations without available derivatives. The novel feature of the method is that, based on locally interpolation models, the search direction in each iteration is allowed to be a solution of a model-based cubic regularization approximation formulated by the special structure of equations that ensures a significant improvement. We present a set of wild conditions that the search direction must be satisfied so that the global convergence of the method for solving the nonlinear equations is guaranteed. The global convergence and the fast local convergence rate of the proposed method are established, and numerical experiments are provided to illustrate the reliability of the proposed method.

Keywords. Adaptive cubic regularization; Derivative-free method; Global convergence; Lagrange interpolation; Nonlinear programming.

1. INTRODUCTION

In this paper, we consider iterative methods for finding the solutions of the following nonlinear equations

$$F(x) = 0, \tag{1.1}$$

where $F(x) = (f_1(x), f_2(x), \ldots, f_n(x))^\top : \mathbb{R}^n \to \mathbb{R}^n$, and the Jacobian of F is not available or requires a prohibitive amount of storage. (1.1) has been extensively investigated as a unified mathematical model in industrial and applied mathematics; see, e.g., [1, 2, 3, 4, 5] and the references therein. Recently, various algorithms have been introduced for solving problem (1.1). For the algorithms with their fast local superlinear convergence properties, such as, the Gauss-Newton method, the Newton's method, the trust region method, and quasi-Newton method, we refer to [6, 7, 8, 9, 10, 11, 12, 13] and the references therein. However, many real-world applications which can be modeled as the nonlinear equation have a theoretical and computational difficulties, that is, the objective functions are of the black-box type so that first order derivatives are not available [14, 15]. This fact has led to the improvement of derivative-free methods in

E-mail address: xjhuangikun@163.com.

Received March 9, 2021; Accepted September 5, 2021.

the literautre. In some practical applications, the derivative-free methods are an important task in order to solve the problem; see, e.g., [12, 16] and the references therein. In [17], Conn et al. proposed an algorithm for optimizing a nonlinear function without constraints and available first-order derivatives based on approximating the objective function by a quadratic polynomial interpolation model and using this model within a trust region framework, and then extended this algorithm to solve certain constrained optimization problems.

This approach built quadratic functions for individual constraints and combined them into a penalty function later. This approach is more efficient for constrained derivative-free optimization problems than combining the constraints into a penalty function first and building an interpolation model for the unconstrained problem. In [18], the global convergence of a trust region derivative-free method for the solution of nonlinear programming problems was extended to cover a wide class of derivative-free methods. However, their focus is only on global convergence without the analysis for local rates of convergence. Recently, Zhang et al. [19] took full advantage of the least-squares problem structure by constructing individual models interpolating each function to least-squares objective rather than to the objective itself whose approach is similar to [17]. The algorithm was proved to be locally convergent in [20]. The algorithm proposed in this paper belongs to model-based derivative-free methods. Adaptive regularized methods have been recently studied as an alternative to classical globalization techniques for nonlinear constrained and unconstrained optimization; see, e.g., [21, 22, 23]. To the best of our knowledge, the use of a cubic overestimator of the objective function as a regularization technique for the computation of the step was first considered by Griewank [24] as a means for constructing the affine-invariant variants of the Newton's method which are globally convergent.

Recently, Cartis, Gould, and Toint [25] considered a similar idea that they proposed an adaptive regularization algorithm by using Cubics (ARC) for unconstrained optimization. At each iteration of the approach, an approximate global minimizer of a local cubic regularization of the objective function is determined for ensuring a significant improvement in the objective as long as the Hessian of under mild assumptions. The ARC iterates show excellent global and fast convergence, see, e.g., [25, 26] for special methods. The goal of this paper is to connect the approximation technique, the particularly polynomial interpolation, with the adaptive cubic regularization algorithm for solving unconstrained problem (2.1). Taking advantage of the derivative-free approximation technique, we present an adaptive cubic regularization algorithm, which is a combination of the adaptive cubic regularization methods and the derivative-free methods. We first build the corresponding quadratic models to approximate each component function f_i, $i = 1, \ldots, n$ by means of the polynomial interpolation, which is only based on the objective function values computed at sampling points to form the least-square problem. Define our local cubic model for the polynomial interpolating least-square problem with a cubic term, called regularization weight. Updating the coefficient of the regularization weight of the local cubic model at each iteration, an improved step is determined, which is the approximate global minimizer of the model, and this gives an estimate of the decrease in the cubic model, accurately predicting the behavior of the interpolating least-square problem along these steps. The updating of the interpolating radius is connected with the coefficient of the regularization weight implicitly. The global convergence and the fast local convergence rate of the proposed algorithm are established under some conditions. All these clearly display that the adaptive

cubic regularization approach is efficient for (2.1). We finally present some numerical results to show the effectiveness of the main results.

The outline of the paper is as follows. Section 2 gives basic definitions of the interpolation models and introduces the derivative-free adaptive cubic regularization algorithm. Section 3 recalls some basic properties of the interpolation models and shows the global convergence of this new algorithm. The local convergent results are established in Section 4. The numerical performances of our algorithm are provided in Section 5, the last section.

As a general rule for notations, $\|\cdot\|$ is the 2-norm for a vector, and the induced 2-norm for a matrix. x_i denotes the ith component of a vector $x \in \mathbb{R}^n$. The subscript k denotes an index for a sequence.

2. PRELIMINARIES

A typical method for solving (1.1) involves solving the following optimization problem

$$\min_{x\in\mathbb{R}^n} \Phi(x) \stackrel{\text{def}}{=} \frac{1}{2}\|F(x)\|^2 = \frac{1}{2}\sum_{i=1}^{n} f_i^2(x), \tag{2.1}$$

where $\|\cdot\|$ denotes the Euclidian norm.

Now, we construct the local models for our derivative-free adaptive cubic regularization algorithm. Suppose that we have a Λ-poised set $Y \subseteq \mathscr{B}(z;\Delta)$ with $C_{n+1}^1 \le |Y| \le C_{n+2}^2$. For each $i = 1,\ldots,n$, let $q_i(x) \in \mathscr{P}_n^2$ be polynomial interpolating function of $f_i(x)$ on Y. Replacing $f_i(x)$ by $q_i(x)$ in the objective function, we obtain

$$Q(x) = \frac{1}{2}\sum_{i=1}^{n} q_i^2(x), \tag{2.2}$$

which is a polynomial of degree four normally. We define our local cubic model $\phi(x,s)$ around x by

$$\phi(x,s) = Q(x) + g_Q(x)^\top s + \frac{1}{2}s^\top B_Q(x)s + \frac{1}{3}\sigma\|s\|^3, \tag{2.3}$$

where $g_Q(x) = J(x)^\top q(x)$ with $q(x) = (q_1(x), q_2(x), \ldots, q_n(x))^\top$ and $J(x) = (\nabla q_1(x), \nabla q_2(x), \ldots,$ and $\nabla q_n(x))^\top$ denotes the Jacobian of $q(\cdot)$. $B_Q(x)$ is an approximation to the local Hessian of $Q(x)$. In particular, let $B_Q(x) = J(x)^\top J(x)$. Here $\sigma > 0$ is a cubic parameter. For constructing this algorithm, we employ (2.3) as an approximation to $Q(x)$ in each iteration of our algorithm. The cubic term of the model, called regularization weight, is rather closely linked to the size of the third derivative, and σ performs a double task. Namely, it may account not only for the discrepancy between (2.2) and its second order Taylor expansion, but also for the difference between the exact and the approximate Hessian of (2.2). Besides, if the reduction of $\Phi(x)$ is desirable, the value of σ is decreased; otherwise it is increased. The global convergence and local superlinear convergence of the new algorithm are established under reasonable assumptions.

We now derive a class of derivative-free algorithm based on the cubic regularization. Throughout the algorithm stated below, we fix the number of points in the sampling set, i.e., $|Y_k| = N_p$, for all $k \ge 0$, where $N_p \in [C_{n+1}^1, C_{n+2}^2]$ is an integer constant. We denote the resulting iterations by x_k, where k is the iteration number.

Now, we give the details of the derivative-free algorithm based on the cubic regularization for the solution of least-squares (2.1).

Algorithm A. A derivative-free adaptive cubic regularization algorithm.

Step 0. Choose an initial point x_0 and $\sigma_0 > 0$, $0 < \Delta_0 \leq 1$, and $N_p \geq C_{n+1}^1 = n+1$. Choose an initial set of interpolation points Y_0, where Y_0 is Λ-poised in $\mathscr{B}(x_0;\Delta_0)$ with $x_0 \in Y_0$. Choose $\varepsilon \in (0,1)$ and $\beta > 0$, $\kappa_\Delta > 0$, $0 < \eta_1 \leq \eta_2 < 1$, $1 < \gamma_1 < \gamma_2$, and $\gamma_3 \in (0,1)$. Set $k = 0$ and $j = 0$.

Step 1. Construct the polynomial interpolating model Q of Φ on Y_k, and obtain the information $g_{Q,k} \overset{\text{def}}{=} g_Q(x_k)$ and $B_{Q,k} \overset{\text{def}}{=} B_Q(x_k)$ with $x_k \in Y_k$.

Step 2. If $\|g_{Q,k}\| \leq \varepsilon$, let $\Delta_k^{(0)} = \Delta_k$. Possibly modify Y_k as needed to make sure Y_k is Λ-poised in $\mathscr{B}(x_k;\Delta_k)$, where $\Delta_k = \min\{\Delta_k^{(0)}, \beta\|g_{Q,k}\|\}$, and $g_{Q,k}$ is recalculated with the new Y_k since Y_k has changed. Determine the corresponding interpolation model. It was shown in [18, Lemma 7.3] that, unless x_k is a first-order stationary point, this can be achieved in a finite number of steps.

Step 3. Compute s_k

$$\phi_k(s_k) \leq \phi_k(s_k^c), \tag{2.4}$$

with

$$s_k^c = -\alpha_k^c g_{Q,k} \text{ and } \alpha_k^c = \arg\min_{\alpha \in \mathbb{R}_+} \phi_k(-\alpha g_{Q,k}),$$

where $\phi_k(s_k) = \phi(x_k, s_k)$ with $\phi(\cdot,\cdot)$ being defined by (2.3).

Step 4. Let $\Delta_{k,j} = \Delta_k$. If

$$\Delta_{k,j} > \kappa_\Delta \min\{\|s_k\|, \|g_{Q,k}\|\}, \tag{2.5}$$

set $\Delta_{k,j+1} = \gamma_3 \Delta_{k,j}$, update Y_k such that Y_k is Λ-poised in $\mathscr{B}(x_k;\Delta_k)$ with $\Delta_k = \Delta_{k,j+1}$ and $x_k \in Y_k$, set $j = j+1$, and go to Step 1.

Step 5. Compute $\Phi(x_k + s_k)$ and ρ_k, where the ρ_k is defined by

$$\rho_k = \frac{\Phi(x_k) - \Phi(x_k + s_k)}{\phi_k(0) - \phi_k(s_k)}. \tag{2.6}$$

Set

$$x_{k+1} = \begin{cases} x_k + s_k, & \text{if } \rho_k \geq \eta_1, \\ x_k, & \text{if otherwise,} \end{cases} \tag{2.7}$$

and

$$\sigma_{k+1} \in \begin{cases} (0, \sigma_k], & \text{if } \rho_k > \eta_2, & \text{[very successful iteration]} \\ [\sigma_k, \gamma_1 \sigma_k], & \text{if } \eta_1 \leq \rho_k \leq \eta_2, & \text{[successful iteration]} \\ [\gamma_1 \sigma_k, \gamma_2 \sigma_k], & \text{otherwise .} & \text{[unsuccessful iteration]} \end{cases} \tag{2.8}$$

Step 6. If $\rho_k \geq \eta_1$, then let $\Delta_{k+1} = \Delta_k$, the interpolating points set Y_k is updated to take into consideration the new point x_{k+1} to form Y_{k+1} being Λ-poised in $\mathscr{B}(x_{k+1};\Delta_{k+1})$ with $x_{k+1} \in Y_{k+1}$. Set $k = k+1$, and go to Step 1. Or go to Step 1 directly.

Since finding a global minimizer of the model $\phi_k(s)$ may not be essential in practise (it might be prohibitively expensive from a computational point of view), we relax this requirement by letting s_k be an approximation to such a minimizer. Initially, we only require that s_k ensures that the decrease in the model is at least as good as that provided by a suitable Cauchy point obtained by globally minimizing $\phi_k(s)$ along the current negative gradient direction of (2.2) in Step 3. The issue $\Delta_k \leq \kappa_\Delta \|s_k\|$ can be achieved by Step 4, which is important in the convergent analysis. The step s_k is accepted and new iteration x_{k+1} is set to be $x_k + s_k$ whenever the predicted model decrease $\phi_k(0) - \phi_k(s_k)$ is realized by the actual decrease in the objective, $\Phi(x_k) - \Phi(x_{k+1})$. This

is measured by computing the ratio ρ_k in (2.6) and requiring ρ_k to be greater than a prescribed positive constant η_1. Since the current parameter σ_k has resulted in a successful step, there is no pressing reason to increase it. Indeed, there may be benefit in decreasing it if good agreement between the model and the function is observed. Without loss of generality, we use $\phi_k(s) = \phi(x_k, s)$ with $\phi(\cdot,\cdot)$ defined by (2.3) in the following analysis.

3. Model Properties and Global convergence

In this section, we recall some basic properties of the interpolation models and show the global convergence for this new algorithm.

The following lemma shows that, under our assumptions of Λ-poisedness (by either definition) of the sampling set Y, an interpolating polynomial on Y would be at least a local fully linear model. This lemma follows directly from [27, Theorem 5.4].

Lemma 3.1. *Given any $\Delta > 0$, $z \in \mathbb{R}^n$, and $Y = \{y^0, y^1, \ldots, y^p\} \subseteq \mathscr{B}(z;\Delta)$ Λ-poised in $\mathscr{B}(z;\Delta)$ with $C^1_{n+1} \le |Y| \le C^2_{n+1}$, let $q(\cdot) \in \mathscr{P}^2_n$ be an interpolating polynomial of f on Y, i.e.,*

$$q(y^i) = f(y^i),\ i = 1, \ldots, |Y|.$$

If $f : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable and ∇f is Lipschitz continuous with Lipschitz constant L in an open set containing $\mathscr{B}(0;\Delta)$, then

$$\|\nabla f(z+s) - \nabla q(z+s)\| \le \hat{\kappa}_{eg}(n,\Lambda)(\|\nabla^2 q\| + L)\Delta,$$

and

$$|f(z+s) - q(z+s)| \le \hat{\kappa}_{ef}(n,\Lambda)(\|\nabla^2 q\| + L)\Delta^2,$$

where $\hat{\kappa}_{eg}$ and $\hat{\kappa}_{ef}$ are positive constants depending on n and Λ.

Give a Λ-poised set $Y \subseteq \mathscr{B}(z;\Delta)$ with $C^1_{n+1} \le |Y| \le C^2_{n+2}$, for each $i = 1, \ldots, n$. Suppose that $q_i(x) \in \mathscr{P}^2_n$ is a polynomial interpolating model of $f_i(x)$ on Y. Then, based on above lemma, there exist positive constants κ_{eg} and κ_{ef} such that, for any $s \in \mathscr{B}(0;\Delta)$,

$$\|\nabla f_i(z+s) - \nabla q_i(z+s)\| \le \kappa_{eg}\Delta, \tag{3.1}$$

$$|f_i(z+s) - q_i(z+s)| \le \kappa_{ef}\Delta^2,$$

for all $i = 1, \ldots, n$, where κ_{ef} and κ_{ef} are positive constants depending only on n, Λ, F, and $\max\{\|\nabla^2 q_i\|,\ i = 1, \ldots, n\}$. In particular, κ_{eg} and κ_{ef} depend either on z nor Δ.

Now, define $conv(L_{enl}(x_0))$ to be the convex hull of $L_{enl}(x_0)$ with

$$L_{enl}(x_0) = \bigcup_{x \in L(x_0)} \mathscr{B}(x;\Delta_0) \text{ and } L(x_0) = \{x \in \mathbb{R}^n : \Phi(x) \le \Phi(x_0)\},$$

where Φ is defined in (2.1). In this paper, we also need the following two assumptions.

Assumption 3.2. Suppose that F is twice continuously differentiable and the level set $L(x_0) = \{x \in \mathbb{R}^n : \Phi(x) \le \Phi(x_0)\}$ is bounded.

Assumption 3.3. There exists a constant κ_H, which is independent of the iteration number k in the Algorithm A, such that if $q_i,\ i = 1, \ldots, n$ is the polynomial interpolating model of f_i on a Λ-poised sampling set Y_k constructed as in the Algorithm A, then $\|\nabla^2 q_i\| \le \kappa_H$, for $i = 1, \ldots, n$.

Recall that F is assumed twice continuously differentiable, and none of their first-order or second-order derivatives is explicitly available. Let $\nabla F(x)$ be the Jacobian matrix of $F(x)$. Then, the gradient $\nabla\Phi(x)$ and the Hessian matrix $\nabla^2\Phi(x)$ are given respectively by

$$\nabla\Phi(x) = \nabla F(x)^\top F(x),\ \nabla^2\Phi(x) = \nabla F(x)^\top \nabla F(x) + \sum_{i=1}^{n} f_i(x)\nabla^2 f_i(x),$$

where $\nabla F(x) = (\nabla f_1(x), \nabla f_2(x), \ldots, \nabla f_n(x))^\top$.

Based on Assumption 3.2 and 3.3, the following lemmas reveal some bound results for the considered functions, whose proofs are similar to the auxiliary Lemma 3.2 and Lemma 3.3 in [19]. They are omitted here.

Lemma 3.4. *Under Assumption 3.2, there exist positive constants L_0, L_1, and L_2 such that*

$$\begin{aligned}
&\|F(x)\| \le L_0,\\
&\|F(x) - F(y)\| \le L_1\|x-y\|,\\
&\|\nabla F(x)\| \le L_1,\\
&\|\nabla F(x) - \nabla F(y)\| \le L_2\|x-y\|,\\
&\|\nabla^2 f_i(x)\| \le L_2,\ i = 1,\ldots,n,
\end{aligned}$$

for any $x, y \in conv(L_{enl}(x_0))$.

Lemma 3.5. *Under Assumptions 3.2 and 3.3, there exist constants κ_{nq} and κ_{ng}, independent of k, such that if q_i, $i = 1,\ldots,n$ is the polynomial interpolating model of f_i on a Λ-poised sampling set Y_k constructed as in Algorithm A, then*

$$|q_i(y)| \le \kappa_{nq} \text{ and } \|\nabla q_i(y)\| \le \kappa_{ng},\ \text{ for all } i = 1,\ldots,n. \tag{3.2}$$

In addition, there exists a constant κ_J such that

$$\|J(y)\| \le \kappa_J, \tag{3.3}$$

where $y \in Y_k$, and $Y_k \subseteq L_{enl}(x_0)$.

Next, we establish the global convergence of Algorithm A. We need the following lemma to guarantee a lower bound on the decrease in Φ predicted from the cubic model. It can be obtain from [26, Lemma 2.1], we omit the proof here.

Lemma 3.6. *Let s_k be the solution of (2.4). Then*

$$\phi_k(0) - \phi_k(s_k) \ge \frac{\|g_{Q,k}\|}{6}\min\left\{\frac{\|g_{Q,k}\|}{\|B_{Q,k}\|}, \frac{\sqrt{3}}{4}\sqrt{\frac{\|g_{Q,k}\|}{\sigma_k}}\right\}, \tag{3.4}$$

where $g_{Q,k}$ and $B_{Q,k}$ are defined by the Step 1 of Algorithm A.

For Step 2, we have the following lemma. It was proved in [18, Lemma 5.1].

Lemma 3.7. *If $\|\nabla\Phi(x_k)\| \neq 0$, the Step 2 in Algorithm A terminates in a finite number of improvement steps.*

Next, we obtain a useful bound on the step. The proof of the following lemma is omitted since it is similar to the one presented in [26, Lemma 2.2].

Lemma 3.8. *Let Assumptions 3.2 and 3.3 hold, and the step s_k satisfy (2.4), then*

$$\|s_k\| \leq 3\sqrt{\frac{\|g_{Q,k}\|}{\sigma_k}}, \ k \geq 0. \tag{3.5}$$

Let us assume that some iterative solver is used on each (major) iteration k to approximately minimize $\phi_k(s)$ for the Step 3 of Algorithm A. We first express the derivative of the cubic model $\phi_k(s)$ in (2.3) as

$$\nabla_s \phi_k(s) = g_{Q,k} + B_{Q,k}s + \sigma_k \|s\| s.$$

Set the termination criterion for its inner iterations i to be

$$\|\nabla_s \phi_k(s_{k,i})\| \leq \kappa_\theta \min\{1, \|s_{k,i}\|\} \cdot \|g_{Q,k}\|, \tag{3.6}$$

where $s_{k,i}$ are the inner iterates generated by the solver, and κ_θ is any constant in $(0,1)$. Condition (3.6) is always satisfied by any minimizer $s_{k,i}$ of $\phi_k(s)$ since $\nabla_s \phi_k(s_{k,i}) = 0$. Thus condition (3.6) can always be achieved by an iterative solver, and the worst that could happen is to iterate until an exact minimizer of $\phi_k(s)$ is found. We hope in practice to terminate well before this inevitable outcome. From (3.6), we have the following.

Assumption 3.9. $\|\nabla_s \phi_k(s_k)\| \leq \kappa_\theta \min\{1, \|s_k\|\} \cdot \|g_{Q,k}\|$, where $s_k \overset{\text{def}}{=} s_{k,i}$ with i being the last inner iteration.

It now remains to show that the loop between Steps 1 and 4 can be finitely terminated. The following Lemma is similar to the Lemma 4.1 of [28]. For the completeness of this paper, we list the proof here.

Lemma 3.10. *Under the conditions of Lemma 3.5, and Assumption 3.9, for all $k \geq 0$,*

$$\|s_k\| \geq \frac{(1-\kappa_\theta)\varepsilon}{\kappa_J^2 + 3n^{\frac{1}{4}}\sqrt{\sigma_k \kappa_J \kappa_{nq}}}, \tag{3.7}$$

and there exists a $\kappa(\sigma) > 0$ such that the loop between Steps 1 and 3 terminates in at most

$$\left\lceil \frac{\log \kappa(\sigma_k) + \log \varepsilon}{\log \gamma_3} \right\rceil_+, \tag{3.8}$$

iterations, where $\lceil a \rceil_+$ denotes the maximum of zero and the first integer larger than or equal to a.

Proof. From the algorithm, we see that if it has not terminated, then

$$\|g_{Q,k}\| > \varepsilon. \tag{3.9}$$

From Assumption 3.9, we see that, for all $k \geq 0$,

$$\begin{aligned} \kappa_\theta \min\{1, \|s_k\|\} \cdot \|g_{Q,k}\| &\geq \|\nabla_s \phi_k(s_k)\| \\ &\geq \|g_{Q,k} + B_{Q,k}s_k + (\sigma_k \|s_k\|)s_k\| \\ &\geq \|g_{Q,k}\| - \|B_{Q,k}s_k + (\sigma_k \|s_k\|)s_k\|. \end{aligned}$$

It follows from (3.9) that

$$\begin{aligned} \|B_{Q,k}s_k + (\sigma_k \|s_k\|)s_k\| &\geq \|g_{Q,k}\| - \kappa_\theta \min\{1, \|s_k\|\} \cdot \|g_{Q,k}\| \\ &\geq (1 - \kappa_\theta \min\{1, \|s_k\|\}) \cdot \|g_{Q,k}\| \\ &\geq (1-\kappa_\theta)\varepsilon. \end{aligned}$$

Taking this bound, and using (3.2), (3.3), and (3.5), we conclude from Lemma 3.4 that

$$\begin{aligned}(1-\kappa_\theta)\varepsilon &\le \kappa_J^2\|s_k\| + \sigma_k\|s_k\|^2 \\ &\le \left(\kappa_J^2 + \sigma_k\cdot 3\sqrt{\frac{\|g_{Q,k}\|}{\sigma_k}}\right)\|s_k\| \\ &= \left(\kappa_J^2 + 3\sqrt{\sigma_k\|g_{Q,k}\|}\right)\|s_k\| \\ &\le \left(\kappa_J^2 + 3n^{\frac{1}{4}}\sqrt{\sigma_k\kappa_J\kappa_{nq}}\right)\|s_k\|,\end{aligned}$$

which proves (3.7). Define

$$\mu(\sigma_k) \stackrel{\text{def}}{=} \frac{1-\kappa_\theta}{\kappa_J^2 + 3n^{\frac{1}{4}}\sqrt{\sigma_k\kappa_J\kappa_{nq}}}.$$

Then, we may use this lower bound to deduce that the loop between Steps 1 and 4 terminates as soon as (2.5) is violated, which must happen if j is large enough to ensure that

$$\begin{aligned}\Delta_{k,j} = \gamma_3^j\Delta_k &\le \gamma_3^j \\ &\le \kappa_\Delta\min\{\mu(\sigma_k),1\}\varepsilon \\ &\le \kappa_\Delta\min\{\|s_k\|,\|g_{Q,k}\|\},\end{aligned}$$

where we used $\Delta_k \le 1$ to have the first inequality, and (3.9) and (3.7) to derive the last inequality. This implies that j never exceeds

$$\left\lceil \frac{\log[\kappa_\Delta\min\{\mu(\sigma_k),1\}] + \log\varepsilon}{\log\gamma_3}\right\rceil_+,$$

which in turn yields (3.8) with $\kappa(\sigma_k) \stackrel{\text{def}}{=} \kappa_\Delta\min\{\mu(\sigma_k),1\}$. □

Since the loop between Steps 1 and 4 always terminates finitely, $\Delta_k \le \kappa_\Delta\|s_k\|$ holds for all $k \ge 0$.

Next, we give an auxiliary lemma.

Lemma 3.11. *Let Assumptions 3.2, 3.3, and 3.9 hold, and*

$$\sqrt{\sigma_k\|g_{Q,k}\|} > \frac{216}{(1-\eta_2)\sqrt{3}}\left(\sqrt{n}L_0\kappa_{eg}\kappa_\Delta + L_0L_2 + L_1^2 + \kappa_J^2\right) \stackrel{\text{def}}{=} \kappa_{LB}. \tag{3.10}$$

Then the iteration k of the Algorithm A is successful with $\rho_k \ge \eta_2$, and

$$\sigma_{k+1} \le \sigma_k, \text{ for all } k \text{ sufficiently large.} \tag{3.11}$$

Proof. From (3.10), we have $\sqrt{\sigma_k\|g_{Q,k}\|} \ge \kappa_J^2 \ge \|J(x_k)^\top J(x_k)\| = \|B_{Q,k}\|$. Thus, (3.4) becomes

$$\phi_k(s_k) - \phi_k(0) \le -\frac{\sqrt{3}}{24}\frac{\|g_{Q,k}\|^{\frac{3}{2}}}{\sqrt{\sigma_k}} < 0. \tag{3.12}$$

It then follows from (2.6) and (2.8) that

$$\rho_k > \eta_2 \Longleftrightarrow r_k \stackrel{\text{def}}{=} \Phi(x_k+s_k) - \Phi(x_k) - \eta_2[\phi_k(s_k) - \phi_k(0)] < 0.$$

Note that, for $k \ge 0$, r_k can be rewritten as

$$r_k = \Phi(x_k+s_k) - \phi_k(s_k) + (1-\eta_2)[\phi_k(s_k) - \phi_k(0)].$$

We then develop the first term on the right-hand side of this expression by using a Taylor expression of $\Phi(x_k+s_k)$, given that, for $k \geq 0$,

$$\Phi(x_k+s_k) - \phi_k(s_k) = \nabla\Phi(\xi_{1,k})^\top s_k - g_{Q,k}^\top s_k - \frac{1}{2} s_k^\top B_{Q,k} s_k - \frac{1}{3}\sigma_k \|s_k\|^3, \tag{3.13}$$

for some $\xi_{1,k} \in (x_k, x_k+s_k)$. Observe that

$$\begin{aligned}
\|\nabla\Phi(\xi_{1,k}) - g_{Q,k}\| &= \|F(\xi_{1,k})^\top \nabla F(\xi_{1,k}) - q(x_k)^\top J(x_k)\| \\
&= \|F(\xi_{1,k})^\top \nabla F(\xi_{1,k}) - F(x_k)^\top J(x_k)\| \\
&= \|F(\xi_{1,k})^\top \nabla F(\xi_{1,k}) - F(x_k)^\top J(x_k) + F(x_k)^\top \nabla F(x_k) \\
&\quad - F(x_k)^\top \nabla F(x_k) + F(\xi_{1,k})^\top \nabla F(x_k) - F(\xi_{1,k})^\top \nabla F(x_k)\| \\
&= \Big\| F(x_k)^\top \left(\nabla F(x_k) - J(x_k)\right) + F(\xi_{1,k})^\top \left(\nabla F(\xi_{1,k}) - \nabla F(x_k)\right) \\
&\quad + \left(F(\xi_{1,k}) - F(x_k)\right)^\top \nabla F(x_k) \Big\| \\
&\leq \left(\sqrt{n} L_0 \kappa_{eg} \kappa_\Delta + L_0 L_2 + L_1^2\right) \|s_k\| \\
&= \kappa_L \|s_k\|,
\end{aligned} \tag{3.14}$$

where $\kappa_L \stackrel{\text{def}}{=} \sqrt{n} L_0 \kappa_{eg} \kappa_\Delta + L_0 L_2 + L_1^2$, $\|\xi_{1,k} - x_k\| \leq \|s_k\|$, and the last inequality we used (3.2), and the fact that Y_k is Λ-poised in $\mathscr{B}(x_k;\Delta_k)$ with $x_k \in Y_k$ and $\Delta_k \leq \kappa_\Delta \|s_k\|$. Thus, the Cauchy-Schwarz inequality, (3.3), and (3.13) yield that, for $k \geq 0$,

$$\Phi(x_k+s_k) - \phi_k(s_k) \leq \kappa_L \|s_k\|^2.$$

In view of (3.12), (3.13), and (3.14), Lemma 3.8 provides the following upper bound for r_k, namely,

$$r_k \leq \frac{\|g_{Q,k}\|}{\sigma_k} \left[9\kappa_L - \frac{\sqrt{3}(1-\eta_2)}{24} \sqrt{\sigma_k \|g_{Q,k}\|} \right],$$

which together with (3.10) implies $r_k < 0$. Thus iteration k is successful, and (3.11) follows from (2.8). This completes the proof. □

We next consider what happens when the number of successful iterations is finite.

Lemma 3.12. *Let Assumptions 3.2–3.9 hold. If the number of successful iterations is finite, then*

$$\lim_{k \to +\infty} \|\nabla\Phi(x_k)\| = 0. \tag{3.15}$$

Proof. Let us consider the iterations that come after the last successful iteration. After the last successful iterate is computed, indexed by, say k_0, the construction of the algorithm implies that $x_{k_0+1} = x_{k_0+i} \stackrel{\text{def}}{=} x_*$ for all $i \geq 1$. Since there are no more successful iterations, i.e., $\rho_k < \eta_1$, for all sufficiently large $k \geq k_0+1$, it follows that σ_k increases by at least γ_1, and $\sigma_k \to \infty$, as $k \to \infty$, $k \geq k_0+1$. We then note that $\lim_{k\to\infty} \|g_{Q,k}\| = 0$. If $\|g_{Q,k}\|$ was bounded away from

zero, k would be a very successful iteration by Lemma 3.11, which yield a contradiction. From Lemma 3.8, we also have

$$\|s_k\| \leq 3\sqrt{\frac{\|g_{Q,k}\|}{\sigma_k}} \to 0, \text{ as } k \to \infty,\ k \geq k_0+1.$$

Now, for each $k \geq k_0+1$, we have

$$\|\nabla\Phi(x_k) - g_{Q,k}\| \leq \sqrt{n}\kappa_{eg}L_0\Delta_k \leq \sqrt{n}\kappa_{eg}\kappa_\Delta L_0\|s_k\|.$$

Observe that

$$\|\nabla\Phi(x_k)\| \leq \|\nabla\Phi(x_k) - g_{Q,k}\| + \|g_{Q,k}\|.$$

The two terms on the right-hand side converge to zero from the analysis above, and (3.15) follows. □

The next lemma gives an upper bound on σ_k when $g_{Q,k}$ is bounded away from zero. We now show that if the model gradient $\|g_{Q,k}\|$ converges to zero on a subsequence, so does the true gradient $\|\nabla\Phi(x_k)\|$.

Lemma 3.13. *For any subsequence $\{k_j\}$ such that*

$$\liminf_{j\to+\infty} \|g_{Q,k_j}\| = 0, \tag{3.16}$$

it holds that

$$\liminf_{j\to+\infty} \|\nabla\Phi(x_{k_j})\| = 0. \tag{3.17}$$

Proof. From (3.16) and the mechanism of Step 3, we see that $Y_{k_j} \subseteq \mathscr{B}(x_{k_j};\Delta_{k_j}) \subseteq \mathscr{B}(x_{k_j};\beta\|g_{Q,k_j}\|)$ is Λ-poised for all large j. From Lemmas 3.4 and 3.5, and (3.1), we have

$$\|\nabla f_i(x_{k_j}) - \nabla q_i(x_{k_j})\| \ \leq\ \kappa_{eg}\Delta_{k_j} \ \leq\ \kappa_{eg}\beta\|g_{Q,k_j}\|.$$

As a consequence,

$$\begin{aligned}\|\nabla\Phi(x_{k_j})\| &\leq \|\nabla\Phi(x_{k_j}) - g_{Q,k_j}\| + \|g_{Q,k_j}\| \\ &\leq (\sqrt{n}L_0\kappa_{eg}\beta + 1)\|g_{Q,k_j}\|,\end{aligned}$$

for all j sufficiently large. Due to $\|g_{Q,k_j}\| \to 0$, we obtain that (3.17) holds. □

We denote the index of all successful iterations of the Algorithm A by

$$\mathscr{S} \overset{\text{def}}{=} \{k \geq 0 :\ k \text{ successful or very successful in the sense of (2.8)}\}. \tag{3.18}$$

Theorem 3.14. *Let Assumptions 3.2, 3.3, and 3.9 hold. Then*

$$\liminf_{k\to+\infty} \|\nabla\Phi(x_k)\| = 0. \tag{3.19}$$

Proof. If there are finitely many successful iterations, then Lemma 3.12 implies that (3.19) holds. Now let us assume that infinitely successful iterations occur. Suppose that (3.16) does not hold, i.e., there exists a $\varepsilon > 0$ such that

$$\|g_{Q,k}\| > \varepsilon, \tag{3.20}$$

for all large k. We first prove that σ_k is bounded above, that is,

$$\sigma_k \leq \max\left\{\sigma_0, \frac{\gamma_2\kappa_{LB}^2}{\varepsilon}\right\}, \text{ for all } k = 0,\ldots,j. \tag{3.21}$$

For any $k \in \{0,\dots,j\}$, due to $\|g_{Q,k}\| > \varepsilon$, we conclude from Lemma 3.11 the implication

$$\sigma_k \geq \frac{\kappa_{LB}^2}{\varepsilon} \Longrightarrow \sigma_{k+1} \leq \sigma_k. \tag{3.22}$$

Thus, when $\sigma_0 \leq \gamma_2 \kappa_{LB}^2/\varepsilon$, (3.22) implies $\sigma_k \leq \gamma_2 \kappa_{LB}^2/\varepsilon$, $\forall\, k \in \{0,\dots,j\}$, where the factor γ_2 is introduced for the case when σ_k is less than $\kappa_{LB}^2/\varepsilon$ and the iteration k is not very successful. Letting $k = 0$ in (3.22) gives (3.21) when $\sigma_0 \geq \gamma_2 \kappa_{LB}^2/\varepsilon$. Next, we derive a contradiction to (3.21). Since $\Phi(x_k)$ is bounded below by zero, it follows from the algorithm, (2.6), (3.3), and (3.20) that

$$\begin{aligned}
+\infty &\geq \sum_{k=0}^{\infty} [\Phi(x_k) - \Phi(x_{k+1})] \\
&\geq \sum_{k\in\mathscr{S}} [\Phi(x_k) - \Phi(x_{k+1})] \\
&\geq \sum_{k\in\mathscr{S}} \eta_1 [\phi_k(0) - \phi_k(s_k)] \\
&\geq \sum_{k\in\mathscr{S}} \frac{\eta_1 \varepsilon}{6} \min\left\{ \frac{\varepsilon}{\kappa_j^2}, \frac{\sqrt{3}}{4}\sqrt{\frac{\varepsilon}{\sigma_k}} \right\}.
\end{aligned}$$

Hence, $\sigma_k \to \infty$, $k \in \mathscr{S}$, as $k \to \infty$. Now, it contradicts (3.21). Hence, (3.16) holds. Then (3.19) follows from Lemma 3.13. This completes the proof. □

Corollary 3.15. *Let Assumptions 3.2, 3.3, and 3.9 hold. Then*

$$\lim_{k\to+\infty} \|\nabla\Phi(x_k)\| = 0. \tag{3.23}$$

Proof. If there are finitely many successful iterations, then Lemma 3.12 implies that (3.23) holds. Now, we assume that $\mathscr{S}$ is infinite. Suppose, for the purpose of establishing a contradiction, that there exists a subsequence $\{k_i\} \subseteq \mathscr{S}$ such that

$$\|\nabla\Phi(x_k)\| \geq \varepsilon_0, \tag{3.24}$$

for some $\varepsilon_0 > 0$, and for all i (we can ignore the other iterates since x_k does not change during such iterations). From Lemma 3.13, we obtain that $\|g_{Q,k}\| \geq \varepsilon$, for some $\varepsilon > 0$, and for all i sufficiently large. Without loss of generality, we pick ε such that

$$\varepsilon \leq \frac{\varepsilon_0}{2(2+\kappa_{eg}\beta L_0)}. \tag{3.25}$$

The proof of Theorem 3.14 then ensures the existence for each k_i in the subsequence of a first iteration $l_i > k_i$ such that $\|g_{Q,l_i}\| \leq \varepsilon$. By removing elements from $\{k_i\}$, without loss of generality and without a change of notation, we obtain that there exists another subsequence indexed by $\{l_i\}$ such that

$$\|g_{Q,k}\| \geq \varepsilon, \text{ for } k_i \leq k < l_i, \text{ and } \|g_{Q,k_j}\| < \varepsilon, \tag{3.26}$$

for sufficiently large i, with inequality (3.24) being retained.

Let $\mathscr{K} \stackrel{\text{def}}{=} \{k \in \mathscr{S} : k_i \leq k < l_i\}$, where $\{k_i\}$ and $\{l_i\}$ are defined above. Since $k \subseteq \mathscr{S}$, it follows from (2.6), (2.7), (3.4), Assumptions 3.2–3.9, and (3.26) that

$$\Phi(x_k) - \Phi(x_{k+1}) \geq \frac{\eta_1 \varepsilon}{6} \min \left\{ \frac{\varepsilon}{\kappa_f^2}, \frac{\sqrt{3}}{4} \sqrt{\frac{\|g_{Q,k}\|}{\sigma_k}} \right\}, \; k \in \mathscr{K}. \tag{3.27}$$

Since $\{\Phi(x_k)\}$ is monotonically decreasing and bounded from below by zero, it is convergent. Hence, the left-hand side of (3.27) converges to zero as $k \to \infty$. Thus, (3.27) implies

$$\sqrt{\frac{\|g_{Q,k}\|}{\sigma_k}} \to 0, \; k \to \infty, \; k \in \mathscr{K}.$$

Due to the limit above, the bound (3.27) asymptotically becomes

$$\Phi(x_k) - \Phi(x_{k+1}) \geq \frac{\sqrt{3}\eta_1 \varepsilon}{24} \sqrt{\frac{\|g_{Q,k}\|}{\sigma_k}}, \text{ for all } k \in \mathscr{K} \text{ sufficiently large.}$$

Together with the relation above, Lemma 3.8 and the definition of $\mathscr{K}$ provide the bound

$$\Phi(x_k) - \Phi(x_{k+1}) \geq \frac{\sqrt{3}\eta_1 \varepsilon}{72} \|s_k\|, \text{ for all } k_i \leq k < l_i, \; k \in \mathscr{K}, \; i \text{ sufficiently large.}$$

Summing up the inequality above over k with $k_i \leq k < l_i$, and employing (2.7) and the triangle inequality, we obtain

$$\begin{aligned} \frac{72}{\sqrt{3}\eta_1 \varepsilon} [\Phi(x_{k_i}) - \Phi(x_{l_i})] &\geq \sum_{k=k_i, k \in \mathscr{S}}^{l_i - 1} \|s_k\| \\ &= \sum_{k=k_i, k \in \mathscr{S}}^{l_i - 1} \|x_k - x_{k+1}\| \\ &\geq \|x_{k_i} - x_{l_i}\|, \end{aligned} \tag{3.28}$$

for all i sufficiently large. Since $\{\Phi(x_k)\}$ is convergent, $\{\Phi(x_{k_i}) - \Phi(x_{l_i})\}$ converges to zero as $i \to \infty$. Thus (3.28) implies that $\|x_{k_i} - x_{l_i}\|$ converges to zero as $i \to \infty$. Now,

$$\|\nabla \Phi(x_{k_i})\| \leq \|\nabla \Phi(x_{k_i}) - \nabla \Phi(x_{l_i})\| + \|\nabla \Phi(x_{l_i}) - g_{Q,l_i}\| + \|g_{Q,l_i}\|.$$

The first term of the right-hand side tends to zero because of Lemma 3.4 and is thus bounded by ε for i sufficiently large. The third term is bounded by ε from (3.26). For the second term, we use the fact that, from (3.25) and the mechanism of the criticality step (Step 3 of Algorithm A) at iteration l_i, the set of interpolation points Y_{l_i} is Λ-poised in $\mathscr{B}(x_{l_i}; \Delta_{l_i})$, where $\Delta_{l_i} \leq \beta \|g_{Q,l_i}\|$. As a consequence, we obtain from these bounds and (3.26) that

$$\|\nabla \Phi(x_{k_i})\| \leq (2 + \kappa_{eg} \beta L_0) \varepsilon \leq \frac{1}{2} \varepsilon_0,$$

for i large enough, which contradicts (3.24). Hence, our initial assumption must be false and the theorem follows. □

4. Local Convergence

In this section, we discuss the local convergence properties of the Algorithm A under some suitable conditions. In this section, we have the following additional assumptions.

Assumption 4.1. Assume that $F(x_*) = 0$, where $x_* \in \mathbb{R}^n$ is the solution of (2.1).

Assumption 4.2. Assume that $\nabla F(x_*)$ is nonsingular.

We begin the discussion with the following lemma.

Lemma 4.3. *Under Assumptions 3.2–4.2, we have that*

$$\lim_{k\to\infty} \|F(x_k)\| = 0, \text{ and } \lim_{k\to\infty} \|g_{Q,k}\| = 0. \tag{4.1}$$

In addition, there exists constants $L_3 > 0$ and $L_4 > 0$ such that

$$\|J(x_k) - \nabla F(x_k)\| \le L_3 \|F(x_k)\|, \tag{4.2}$$

and $\|\nabla^2 \Phi(x_k) - B_{Q,k}\| \le L_4 \|F(x_k)\|$.

Proof. Assumption 4.2 implies $\nabla F(x_k)$ is nonsingular for all k sufficiently large. Thus, with Lemma 3.4, the eigenvalues of $\nabla F(x_k)$ are bounded. From Corollary 3.15, and Assumption 4.2, we have $\lambda_{\min}^{F_k} \|F(x_k)\| \le \|\nabla F(x_k)^\top F(x_k)\| = \|\nabla \Phi(x_k)\| \to 0$, $k \to \infty$, where $\lambda_{\min}^{F_k}$ is the leftmost absolute eigenvalue of $\nabla F(x_k)$. It now follows from Lemma 3.5 that

$$\|g_{Q,k}\| = \|g_Q(x_k)\| = \|J(x_k)^\top q(x_k)\| = \|J(x_k)^\top F(x_k)\| \le \kappa_J \|F(x_k)\|. \tag{4.3}$$

Hence, we have that (4.1) holds. Now, by (4.2), it follows from the Step 3 of the Algorithm A that Y_k is Λ-poised in $\mathscr{B}(x_k; \beta \|g_{Q,k}\|)$. By (3.1), (4.3), Assumptions 3.2, 3.3, and Lemma 3.4, we have

$$\begin{aligned} \|J(x_k) - \nabla F(x_k)\| &\le \sqrt{n}\kappa_{eg}\Delta_k \\ &\le \sqrt{n}\kappa_{eg}\beta \|g_{Q,k}\| \\ &\le n\kappa_{eg}\beta\kappa_{ng}\|F(x_k)\| \\ &\le L_3\|F(x_k)\|, \end{aligned}$$

where $L_3 \overset{\text{def}}{=} n\kappa_{eg}\beta\kappa_{ng}$ and β is defined in the Step 0 of the algorithm. Furthermore, from Lemma 3.4 and the inequality above, we obtain

$$\begin{aligned} & \|\nabla^2 \Phi(x_k) - B_{Q,k}\| \\ = \; & \|\nabla F(x_k)^\top \nabla F(x_k) + \sum_{i=1}^n f_i(x_k)\nabla^2 f_i(x_k) - J(x_k)^\top J(x_k)\| \\ \le \; & \|\nabla F(x_k)^\top \nabla F(x_k) - J(x_k)^\top J(x_k)\| + \sum_{i=1}^n |f_i(x_k)| \cdot \|\nabla^2 f_i(x_k)\| \\ \le \; & \|\nabla F(x_k)^\top \nabla F(x_k) - \nabla F(x_k)^\top J(x_k) + \nabla F(x_k)^\top J(x_k) - J(x_k)^\top J(x_k)\| + L_2 \sum_{i=1}^n |f_i(x_k)| \\ \le \; & \|\nabla F(x_k)\| \cdot \|\nabla F(x_k) - J(x_k)\| + \|J(x_k)\| \cdot \|\nabla F(x_k) - J(x_k)\| + L_2 \|F(x_k)\|_1 \\ \le \; & L_1 L_3 \|F(x_k)\| + \kappa_J L_3 \|F(x_k)\| + \sqrt{n} L_2 \|F(x_k)\| \\ = \; & L_4 \|F(x_k)\|, \end{aligned}$$

where $L_4 \overset{\text{def}}{=} L_1L_3 + \kappa_J L_3 + \sqrt{n}L_2$ and $\|\cdot\|_1$ denotes the 1-norm, i.e., $\|x\|_1 \overset{\text{def}}{=} \sum_{i=1}^{n} |x_i|$. □

To be able to guarantee convergence properties for the proposed algorithm, further requirements are needed to be placed on s_k. The strongest conditions are, of course, the first and second order (necessary) optimality conditions that s_k satisfies provided it is the (exact) global minimizer of $\phi_k(s)$ over $\mathbb{R}^n$. The choice of s_k, however, may be in general prohibitively expensive from a computational point of view, and thus, for most practical purposes, inefficient. Thus, the conditions we require on s_k in what follows, are some derivations of first- and second- order optimality when s_k is computed in each subspace including $g_{Q,k}$ (not only the previous results still hold, but we can prove further convergence properties).

In what follows, we require that s_k satisfies

$$g_{Q,k}^\top s_k + s_k^\top B_{Q,k} s_k + \sigma_k \|s_k\|^3 = 0, \ k \geq 0, \tag{4.4}$$

and

$$s_k^\top B_{Q,k} s_k + \sigma_k \|s_k\|^3 \geq 0, \ k \geq 0. \tag{4.5}$$

The next lemma presents some suitable choices for s_k that achieve (4.4) and (4.5) from [25]. We omit its proof here.

Lemma 4.4. *Suppose that s_k is the global minimizer of $\phi_k(s)$, for $s_k \in \mathscr{L}_k$, where $\mathscr{L}_k$ is a subsequence of $\mathbb{R}^n$. Then s_k satisfies (4.4) and (4.5). Furthermore, letting A_k denote any orthogonal matrix whose columns form a basis of $\mathscr{L}_k$, we have that $A_k^\top B_{Q,k} A_k + \sigma_k \|s_k\| I$ is positive semi-definite, $s \in \mathbb{R}^n$. Then s_k achieves (4.4) and (4.5).*

Cauchy point (2) satisfies (4.4) and (4.5) since it globally minimizes ϕ_k over the subsequence generated by $-g_{Q,k}$. To improve the properties and performance, it may be necessary to minimize ϕ_k over (increasingly) larger subspaces (that each contain $g_{Q,k}$ so that (2.4) can still be achieved). The following lemma gives a lower bound on the model decrease when (4.4) and (4.5) are satisfied.

Lemma 4.5. *If s_k satisfies (4.4), then $\phi_k(0) - \phi_k(s_k) \geq \frac{1}{2} s_k^\top B_{Q,k} s_k$.*

Requiring that s_k satisfies (4.4) may not necessarily implies (3.4), unless $s_k = -g_{Q,k}$. Nevertheless, when minimizing ϕ_k globally over successive subspaces, condition (3.4) can be easily ensured by including $g_{Q,k}$ in each of the subspaces. This is the approach we take in our implementation of the Algorithm A, where the subspaces generated by Lanczos method naturally include the gradient. Thus, we assume the Cauchy condition (3.4) still holds.

The following theorem shows that all iterations are ultimately very successful provided some further assumption on the level of resemblance between the approximate interpolating Hessians $B_{Q,k}$ and the true Hessians $\nabla^2 \Phi(x_k)$ holds as the iterates converge to a local minimizer.

Theorem 4.6. *Let Assumptions 3.2–4.2 hold, and let s_k satisfy (4.4) where $\nabla F(x_*)$ is nonsingular. Then there exists $R_{\min} > 0$ such that*

$$\|s_k\| \leq \frac{\kappa_J}{R_{\min}} \|F(x_k)\|, \text{ for all } k \text{ sufficiently large.} \tag{4.6}$$

Furthermore, all iterations are eventually very successful, and σ_k is bounded from above.

Proof. Since $\nabla F(x_*)$ is nonsingular, then it is $\nabla F(x_k)$ for all k sufficiently large. From Lemma 3.4, we have that $\nabla^2 f_i(x_k),\ i = 1,\ldots,n$ is bounded from above. Together with the second term of (3), we obtain that $\nabla^2\Phi(x_k)$ is positive definite. In particular, there exists a positive constant $R_{\min}$ such that

$$\frac{s_k^\top \nabla^2\Phi(x_k)s_k}{\|s_k\|^2} > 2R_{\min}, \text{ for all } k \text{ sufficiently large.}$$

Thus, thanks to Lemma 4.3 and Lemma 3.4, we obtain that, for all sufficiently large k,

$$\begin{aligned} 2R_{\min}\|s_k\|^2 &\leq s_k^\top \nabla^2\Phi(x_k)s_k \\ &= s_k^\top(\nabla^2\Phi(x_k) - B_{Q,k})s_k + s_k^\top B_{Q,k}s_k \\ &\leq R_{\min}\|s_k\|^2 + \|J(x_k)s_k\|^2. \end{aligned}$$

Moreover, since s_k satisfies (4.4), we conclude from Assumptions 3.2 and 3.3 that

$$\|J(x_k)s_k\|^2 = s_k^\top B_{Q,k}s_k \leq -g_{Q,k}^\top s_k \leq \|g_{Q,k}\|\cdot\|s_k\| \leq \kappa_J\|F(x_k)\|\cdot\|s_k\|.$$

With the two relations above, we obtain bound (4.6). To prove iteration k is a very successful iteration, we need to prove the inequality holds. From the Taylor mean theorem, we have

$$\begin{aligned} &\Phi(x_k+s_k) - \phi_k(s_k) \\ &= \Phi(x_k) + \nabla\Phi(x_k)^\top s_k + \tfrac{1}{2}s_k^\top\nabla\Phi(\xi_{2,k})s_k \\ &\quad - \left(Q(x_k) + g_{Q,k}^\top s_k + \frac{1}{2}s_k^\top B_{Q,k}s_k + + \frac{1}{3}\sigma_k\|s_k\|^3\right) \\ &= F(x_k)^\top(\nabla F(x_k) - J(x_k))s_k + \frac{1}{2}s_k^\top(\nabla^2\Phi(\xi_{2,k}) - B_{Q,k})s_k - \frac{1}{3}\sigma_k\|s_k\|^3, \end{aligned}$$

where $\xi_{2,k} \in (x_k, x_k+s_k)$. Observe that the first term of the last equality above can be bounded as following

$$F(x_k)^\top(\nabla F(x_k) - J(x_k))s_k \leq \sqrt{n}\kappa_{eg}\kappa_\Delta\|F(x_k)\|\cdot\|s_k\|^2,\ k \geq 0. \tag{4.7}$$

From the second term, we have

$$\|\nabla^2\Phi(\xi_{2,k}) - B_{Q,k}\| \leq \|\nabla^2\Phi(x_k) - \nabla^2\Phi(\xi_{2,k})\| + \|\nabla^2\Phi(x_k) - B_{Q,k}\|,\ k \geq 0.$$

The first term is bounded as following

$$\begin{aligned}
& \|\nabla^2\Phi(x_k)-\nabla^2\Phi(\xi_{2,k})\| \\
= & \left\|\nabla F(x_k)^\top\nabla F(x_k)+\sum_{i=1}^n f_i(x_k)\nabla^2 f_i(x_k)-\nabla F(\xi_{2,k})^\top\nabla F(\xi_{2,k})-\sum_{i=1}^n f_i(\xi_{2,k})\nabla^2 f_i(\xi_{2,k})\right\| \\
\le & \|\nabla F(x_k)^\top\nabla F(x_k)-\nabla F(\xi_{2,k})^\top\nabla F(\xi_{2,k})\|+\left\|\sum_{i=1}^n f_i(x_k)\nabla^2 f_i(x_k)\right\|+\left\|\sum_{i=1}^n f_i(\xi_{2,k})\nabla^2 f_i(\xi_{2,k})\right\| \\
\le & \|\nabla F(x_k)^\top\nabla F(x_k)-\nabla F(x_k)^\top\nabla F(\xi_{2,k})+\nabla F(x_k)^\top\nabla F(\xi_{2,k})-\nabla F(\xi_{2,k})^\top\nabla F(\xi_{2,k})\| \\
& +L_2\sum_{i=1}^n|f_i(x_k)|+L_2\sum_{i=1}^n|f_i(\xi_{2,k})| \\
\le & \|\nabla F(x_k)\|\cdot\|\nabla F(x_k)-\nabla F(\xi_{2,k})\|+\|\nabla F(\xi_{2,k})\|\cdot\|\nabla F(x_k)-\nabla F(\xi_{2,k})\| \\
& +L_2\|F(x_k)\|_1+L_2\sum_{i=1}^n|f_i(\xi_{2,k})-f_i(x_k)+f_i(x_k)| \\
\le & 2L_1L_2\|\xi_{2,k}-x_k\|+L_2\|F(x_k)\|_1+L_2\sum_{i=1}^n|f_i(x_k)|+L_2\sum_{i=1}^n|f_i(\xi_{2,k})-f_i(x_k)| \\
\le & 2L_1L_2\|\xi_{2,k}-x_k\|+2\sqrt{n}L_2\|F(x_k)\|+nL_1L_2\|\xi_{2,k}-x_k\| \\
\le & (2+n)L_1L_2\|\xi_{2,k}-x_k\|+2\sqrt{n}L_2\|F(x_k)\|.
\end{aligned}$$

Since $\xi_{2,k}\in(x_k,x_k+s_k)$, we have $\|\xi_{2,k}-x_k\|\le\|s_k\|$, which together with (4.6) and Lemma 4.3, gives $\|\xi_{2,k}-x_k\|\to 0$. This, together with the continuity of $f_i, i=1,\dots,n$ and Lemma 4.3, implies $\|[\nabla^2\Phi(x_k)-\nabla^2\Phi(\xi_{2,k})\|\to 0$, as $k\to\infty$. It now follows from the limit above and Lemma 4.3 that

$$\|\nabla^2\Phi(\xi_{2,k})-B_{Q,k}\|\to 0,\ k\to\infty. \tag{4.8}$$

Relation (4.4) implies

$$\begin{aligned}
\phi_k(0)-\phi_k(s_k) = & -g_{Q,k}^\top s_k-\tfrac{1}{2}s_k^\top B_{Q,k}s_k-\tfrac{1}{3}\sigma_k\|s_k\|^3 \\
= & s_k^\top B_{Q,k}s_k+\sigma_k\|s_k\|^3-\frac{1}{2}s_k^\top B_{Q,k}s_k-\frac{1}{3}\sigma_k\|s_k\|^3 \\
\ge & \frac{1}{2}s_k^\top B_{Q,k}s_k \\
\ge & \frac{1}{2}\|J(x_k)s_k\|^2 \\
\ge & \frac{1}{2}R_{\min}\|s_k\|^2.
\end{aligned}$$

It now follows from Assumptions 4.1 and 4.2, (4.7), (4.8), and the inequality above that

$$\begin{aligned}
r_k = & F(x_k)^\top(\nabla F(x_k)-J(x_k))s_k+\frac{1}{2}s_k^\top(\nabla^2\Phi(\xi_{2,k})-B_{Q,k})s_k-\frac{1}{3}\sigma_k\|s_k\|^3 \\
& +(1-\eta_2)[\phi_k(s_k)-\phi_k(0)] \\
\le & \frac{\|s_k\|^2}{2}\left(2\sqrt{2}\kappa_{eg}\kappa_\Delta\|F(x_k)\|+\|[\nabla^2\Phi(x_k)-\nabla^2\Phi(\xi_{2,k})\|-(1-\eta_2)R_{\min}\right) \\
< & 0,
\end{aligned}$$

for all k sufficiently large. Hence, all iterations are eventually very successful. Since σ_k is not allowed to increase on the very successful steps of the Algorithm A, and every k sufficiently large is very successful, σ_k is bounded from above. □

Lemma 4.7. *Let Assumptions 3.2–4.2 hold. Then, for each $k \in \mathscr{S}$, with $\mathscr{S}$ defined in (3.18), there exists an increasing positive single variable real function $L(\sigma_k) > 0$ with the real variable σ_k such that*

$$\left[1 - \left(\sqrt{n}\kappa_{eg}\kappa_{\Delta}\|F(x_{k+1})\| + \kappa_{\theta}\right)\right] \|g_{Q,k+1}\| \leq L(\sigma_k)\|F(x_k)\|^2, \tag{4.9}$$

for all sufficiently large $k \in \mathscr{S}$.

Proof. Letting $k \in \mathscr{S}$, we have $x_{k+1} = x_k + s_k$. From Assumption 3.9, we have

$$\begin{aligned}\|g_{Q,k+1}\| &\leq \|g_{Q,k+1} - \nabla\Phi(x_k+s_k)\| + \|\nabla\Phi(x_k+s_k) - \nabla_s\phi_k(s_k)\| + \|\nabla_s\phi_k(s_k)\| \\ &\leq \|g_{Q,k+1} - \nabla\Phi(x_k+s_k)\| + \|\nabla\Phi(x_k+s_k) - \nabla_s\phi_k(s_k)\| + \kappa_{\theta}\min\{1,\|s_k\|\}\cdot\|g_{Q,k}\|. \end{aligned} \tag{4.10}$$

Since Y_{k+1} is Λ-poised, we conclude that

$$\begin{aligned}\|g_{Q,k+1} - \nabla\Phi(x_k+s_k)\| &\leq \|F(x_{k+1})\|\cdot\|J(x_{k+1}) - \nabla F(x_{k+1})\| \\ &\leq \sqrt{n}\kappa_{eg}\kappa_{\Delta}\|F(x_{k+1})\|\cdot\|g_{Q,k+1}\|. \end{aligned} \tag{4.11}$$

From Taylor's theorem and Lemma 4.3, we obtain

$$\begin{aligned}&\|\nabla\Phi(x_k+s_k) - \nabla_s\phi_k(s_k)\| \\ &\leq \|\nabla\Phi(x_k) - g_{Q,k}\| + \left\|\int_0^1 \left[\nabla^2\Phi(x_k+\omega s_k) - B_{Q,k}\right] s_k d\omega\right\| + \sigma_k\|s_k\|^2 \\ &\leq \|F(x_k)\|\cdot\|\nabla F(x_k) - J(x_k)\| + \sigma_k\|s_k\|^2 + \left\|\left[\nabla^2\Phi(x_k) - B_{Q,k}\right] s_k\right\| \\ &\quad + \left\|\int_0^1 \left[\nabla^2\Phi(x_k+\omega s_k) - \nabla^2\Phi(x_k)\right] d\omega\right\| \|s_k\| \\ &\leq L_3\|F(x_k)\|^2 + \sigma_k\|s_k\|^2 + L_4\|F(x_k)\|\cdot\|s_k\| \\ &\quad + \|s_k\|\int_0^1 \left[(2+n)L_1L_2\|\omega s_k\| + 2\sqrt{n}L_2\|F(x_k)\|\right] d\omega \\ &\leq L_3\|F(x_k)\|^2 + \sigma_k\|s_k\|^2 + L_4\|F(x_k)\|\cdot\|s_k\| \\ &\quad + \frac{(2+n)L_1L_2}{2}\|s_k\|^2 + 2\sqrt{n}L_2\|F(x_k)\|\cdot\|s_k\|. \end{aligned} \tag{4.12}$$

From (3.1), Lemma 3.4, and Lemma 4.3, we have

$$\begin{aligned}\|g_{Q,k}\| &\leq \|g_{Q,k} - \nabla\Phi(x_k)\| + \|\nabla\Phi(x_k) - \nabla\Phi(x_k+s_k)\| + \|\nabla\Phi(x_k+s_k) - g_{Q,k+1}\| + \|g_{Q,k+1}\| \\ &\leq L_3\|F(x_k)\|^2 + L_2\|F(x_k)\|\cdot\|s_k\| + L_1^2\|s_k\| + \sqrt{n}\kappa_{eg}\kappa_{\Delta}\|F(x_{k+1})\|\cdot\|s_k\| + \|g_{Q,k+1}\| \end{aligned}$$

which implies that, for all $k \geq 0$,

$$\begin{aligned}\kappa_{\theta}\min\{1,\|s_k\|\}\cdot\|g_{Q,k}\| \leq{}& \kappa_{\theta}L_3\|F(x_k)\|^2 + \kappa_{\theta}L_2\|F(x_k)\|\cdot\|s_k\| + \kappa_{\theta}L_1^2\|s_k\|^2 \\ &+ \sqrt{n}\kappa_{\theta}\kappa_{eg}\kappa_{\Delta}\|F(x_{k+1})\|\cdot\|s_k\| + \kappa_{\theta}\|g_{Q,k+1}\|. \end{aligned} \tag{4.13}$$

Therefore, substituting (4.11), (4.12), (4.13) into (4.10), we obtain that, for all $k \geq 0$,

$$
\begin{aligned}
&\|g_{Q,k+1}\| \leq \sqrt{n}\kappa_{eg}\kappa_{\Delta}\|F(x_{k+1})\| \cdot \|g_{Q,k+1}\| + L_3\|F(x_k)\|^2 + \sigma_k\|s_k\|^2 + L_4\|F(x_k)\| \cdot \|s_k\| \\
&+\frac{(2+n)L_1L_2}{2}\|s_k\|^2 + 2\sqrt{n}L_2\|F(x_k)\| \cdot \|s_k\| + \kappa_{\theta}L_3\|F(x_k)\|^2 + \kappa_{\theta}L_2\|F(x_k)\| \cdot \|s_k\| \\
&+\kappa_{\theta}L_1^2\|s_k\|^2 + \kappa_{\theta}\sqrt{n}\kappa_{eg}\kappa_{\Delta}\|F(x_{k+1})\| \cdot \|s_k\| + \kappa_{\theta}\|g_{Q,k+1}\| \\
&= \left(\sqrt{n}\kappa_{eg}\kappa_{\Delta}\|F(x_{k+1})\| + \kappa_{\theta}\right) \cdot \|g_{Q,k+1}\| + \left[\frac{(2+n)L_1L_2}{2} + \sigma_k + \kappa_{\theta}L_1^2\right]\|s_k\|^2 \\
&+\left[L_4 + 2\sqrt{n}L_2 + \kappa_{\theta}L_2\right]\|F(x_k)\| \cdot \|s_k\| + (1+\kappa_{\theta})L_3\|F(x_k)\|^2 \\
&+\sqrt{n}\kappa_{\theta}\kappa_{eg}\kappa_{\Delta}\|F(x_{k+1})\| \cdot \|s_k\| \\
&\leq \left(\sqrt{n}\kappa_{eg}\kappa_{\Delta}\|F(x_{k+1})\| + \kappa_{\theta}\right) \cdot \|g_{Q,k+1}\| + \left[\frac{(2+n)L_1L_2}{2} + \sigma_k + \kappa_{\theta}L_1^2\right]\|s_k\|^2 \\
&+\left[L_4 + 2\sqrt{n}L_2 + \kappa_{\theta}L_2\right]\|F(x_k)\| \cdot \|s_k\| + (1+\kappa_{\theta})L_3\|F(x_k)\|^2 \\
&+\sqrt{n}\kappa_{\theta}\kappa_{eg}\kappa_{\Delta}\left[\|F(x_k)\| + L_1\|s_k\|\right] \cdot \|s_k\| \\
&= \left(\sqrt{n}\kappa_{eg}\kappa_{\Delta}\|F(x_{k+1})\| + \kappa_{\theta}\right)\|g_{Q,k+1}\| + \left[\sqrt{n}\kappa_{\theta}\kappa_{eg}\kappa_{\Delta}L_1 + \frac{(2+n)L_1L_2}{2} + \sigma_k + \kappa_{\theta}L_1^2\right]\|s_k\|^2 \\
&+\left[\sqrt{n}\kappa_{\theta}\kappa_{eg}\kappa_{\Delta} + L_4 + 2\sqrt{n}L_2 + \kappa_{\theta}L_2\right]\|F(x_k)\| \cdot \|s_k\| + (1+\kappa_{\theta})L_3\|F(x_k)\|^2 \\
&\leq \left(\sqrt{n}\kappa_{eg}\kappa_{\Delta}\|F(x_{k+1})\| + \kappa_{\theta}\right) \cdot \|g_{Q,k+1}\| \\
&+\frac{\sqrt{n}\kappa_{ng}}{R_{min}}\left[\sqrt{n}\kappa_{\theta}\kappa_{eg}\kappa_{\Delta}L_1 + \frac{(2+n)L_1L_2}{2} + \sigma_k + \kappa_{\theta}L_1^2\right]\|F(x_k)\|^2 \\
&+\frac{n\kappa_{ng}^2}{R_{min}^2}\left[\sqrt{n}\kappa_{\theta}\kappa_{eg}\kappa_{\Delta} + L_4 + 2\sqrt{n}L_2 + \kappa_{\theta}L_2\right]\|F(x_k)\|^2 + (1+\kappa_{\theta})L_3\|F(x_k)\|^2 \\
&= \left(\sqrt{n}\kappa_{eg}\kappa_{\Delta}\|F(x_{k+1})\| + \kappa_{\theta}\right) \cdot \|g_{Q,k+1}\| + L(\sigma_k)\|F(x_k)\|^2, \qquad (4.14)
\end{aligned}
$$

where $L(\sigma_k)$ is defined as following

$$
\begin{aligned}
L(\sigma_k) = &\frac{\sqrt{n}\kappa_{ng}}{R_{min}}\left[\sqrt{n}\kappa_{\theta}\kappa_{eg}\kappa_{\Delta}L_1 + \frac{(2+n)L_1L_2}{2} + \sigma_k + \kappa_{\theta}L_1^2\right] + (1+\kappa_{\theta})L_3 \\
&+\tfrac{n\kappa_{ng}^2}{R_{min}^2}\left[\sqrt{n}\kappa_{\theta}\kappa_{eg}\kappa_{\Delta} + L_4 + 2\sqrt{n}L_2 + \kappa_{\theta}L_2\right].
\end{aligned}
$$

Hence, (4.9) follows from (4.14). □

Corollary 4.8. *If the conditions of Theorem 4.6 hold, then*

$$
\frac{\|F(x_{k+1})\|}{\|F(x_k)\|} \to 0, \text{ as } k \to \infty, \qquad (4.15)
$$

and

$$
\frac{\|x_{x+1} - x_*\|}{\|x_k - x_*\|} \to 0, \text{ as } k \to \infty. \qquad (4.16)
$$

Furthermore,

$$
\|x_{x+1} - x_*\| = O(\|x_k - x_*\|^2), \text{ as } k \to \infty. \qquad (4.17)
$$

Proof. Since Y_{k+1} is Λ-poised in $B(x_{k+1};\Delta_{k+1})$, we conclude from (3.1), and the Step 3 of Algorithm A that

$$\begin{aligned}\|\nabla\Phi(x_{k+1})\| &\leq \|g_{Q,k+1}-\nabla\Phi(x_{k+1})\|+\|g_{Q,k+1}\|\\ &\leq \sqrt{n}\kappa_{eg}\beta\|F(x_{k+1})\|\cdot\|g_{Q,k+1}\|+\|g_{Q,k+1}\|\\ &\leq \left(1+\sqrt{n}\kappa_{eg}\beta\|F(x_{k+1})\|\right)\|g_{Q,k+1}\|.\end{aligned}$$

Moreover, as Theorem 4.6 gives that all iterates are eventually very successful, and σ_k is bounded above, say by some $\sigma_{\max}$, (4.9) holds for all k sufficiently large. Thus

$$\left[1-\left(\sqrt{n}\kappa_{eg}\kappa_{\Delta}\|F(x_{k+1})\|+\kappa_{\theta}\right)\right]\|g_{Q,k+1}\|\leq L_{\sigma_{\max}}\|F(x_k)\|^2,$$

for all k sufficiently large, where $L_{\sigma_{\max}}=L(\sigma_{\max})$. Since $\nabla F(x_*)$ is nonsingular, then it is $\nabla F(x_{k+1})$ for all k sufficiently large, and there exists a positive constant γ such that $\|\nabla F(x_{k+1})^{-1}\|\leq\gamma$. It follows that

$$\begin{aligned}\frac{\|F(x_{k+1})\|}{\|F(x_k)\|} &= \frac{\|\nabla F(x_{k+1})^{-1}\nabla\Phi(x_{k+1})\|}{\|F(x_k)\|}\\ &\leq \frac{\gamma\|\nabla\Phi(x_{k+1})\|}{\|F(x_k)\|}\\ &\leq \frac{\gamma(1+\sqrt{n}\kappa_{eg}\beta\|F(x_{k+1})\|)\|g_{Q,k+1}\|}{\|F(x_k)\|}\\ &\leq \frac{\gamma(1+\sqrt{n}\kappa_{eg}\beta\|F(x_{k+1})\|)L_{\sigma_{\max}}\|F(x_k)\|^2}{[1-(\sqrt{n}\kappa_{eg}\kappa_{\Delta}\|F(x_{k+1})\|+\kappa_{\theta})]\|F(x_k)\|}\\ &= \frac{\gamma(1+\sqrt{n}\kappa_{eg}\beta\|F(x_{k+1})\|)L_{\sigma_{\max}}}{[1-(\sqrt{n}\kappa_{eg}\kappa_{\Delta}\|F(x_{k+1})\|+\kappa_{\theta})]}\|F(x_k)\| \qquad (4.18)\end{aligned}$$

for all k sufficiently large. The right-hand side of (4.18) tends to zero as $k\to\infty$ due to $\|F(x_k)\|\to 0$, as $k\to\infty$, under Assumption 4.2. Thus, (4.15) holds. From the standard Taylor theorem of $F(x_{k+1})$ and $F(x_k)$ around x_*, and $F(x_*)=0$ with full rank $\nabla F(x_*)$, we have

$$\begin{aligned}F(x_{k+1}) &= F(x_*)+\nabla F(\xi_{3,k})(x_{k+1}-x_*),\\ F(x_k) &= F(x_*)+\nabla F(\xi_{4,k})(x_k-x_*),\end{aligned}$$

for all k sufficiently large, where $\xi_{3,k}\in(x_{k+1},x_*)$, and $\xi_{4,k}\in(x_k,x_*)$. From Lemma 3.4 and Assumption 4.1, we obtain

$$\begin{aligned}\|F(x_{k+1})\| &\leq L_1\|x_{k+1}-x_*\|,\\ \|F(x_k)\| &\leq L_1\|x_k-x_*\|.\end{aligned}$$

Thus,

$$\begin{aligned}\frac{\|F(x_{k+1})\|}{\|x_{k+1}-x_*\|} &\leq C_1,\\ \frac{\|F(x_k)\|}{\|x_k-x_*\|} &\leq C_2,\end{aligned}$$

where C_1 and C_2 are some constants. Hence,

$$\lim_{k\to\infty}\frac{\|x_{k+1}-x_*\|}{\|x_k-x_*\|}=\lim_{k\to\infty}\frac{\|F(x_k)\|}{\|x_k-x_*\|}\frac{\|x_{k+1}-x_*\|}{\|F(x_{k+1})\|}\frac{\|F(x_{k+1})\|}{\|F(x_k)\|}=0$$

and (4.16) holds as $k\to 0$. So, (4.17) follows. □

Corollary 4.8 holds in the case when the stronger conditions of Theorem 4.6 are satisfied. Particularly, we require the stronger condition Assumption 4.2, then, the limit $F(x_k)\to 0$ is guaranteed to hold.

5. Numerical Results

In this section, we report the results of the computational experiments performed in order to assess the effectiveness of the Algorithm A. All of our experiments were performed on a single processor of a desktop computer with Intel(R) Core (TM) i3-4150U @3.50 GHz. And all programs are written in Matlab code, and the parameters in Algorithm A are chosen as follows:

$$\gamma_1=1.5,\ \gamma_2=3,\ \gamma_3=0.1,$$
$$\eta_1=0.3,\ \eta_2=0.8,\ \sigma_0=1,$$
$$\Delta_0=0.4,\ \kappa_\Delta=0.9,\ \beta=0.5.$$

Now, we test the algorithm with the above parameters. In our algorithm, we terminated the iteration when

$$\|g_{Q,k}\|\le 10^{-5},$$

or iteration number exceeds 200, and any run exceeding this is flagged as a failure.

The numerical results are listed in the following table. The table shows the names of problems that are denoted by "name", the number of its variables, which is denoted by "n", the approximate optimal value "$\|F\|$" at the approximate solution"x_*", the value of "$\|g_Q\|$" at the approximate optimal solution "x_*", and the number of outer iteration which is denoted by "k". We compare the text problem with two different sampling points. The left-hand side of table 1 shows the numerical results of sampling points with the natural choice of $2n+1$, and the other side shows the numerical results of sampling points with $(n+1)(n+2)/2$. The set of the test problems mainly consist of [29]. The remaining problems were collected from a variety of places during the development of the Algorithm A code. The test problem HS40 is the problem 40 in Hock and Schittkowski [29]. SC206, SC207, and SC208 are the Problems 206, 207, and 208 in Schittkowski [30]. The rest problems are from [29].

We see that, for each n, the outer iteration occurs in the $N_p=2n+1$ column being very encouraging. We can observe that the number of outer iteration of the sampling set with more interpolating points is almost less than the other one. It's reasonable since the latter can approximate the original problem more accurately with more interpolating points in the sampling set. The algorithm we propose with $N_p=2n+1$ uses small sampling set that can construct a quadratic model with less computation numbers while the number of the whole outer iteration is increasing. In the case of the other sampling set, to construct a quadratic model needs more computation numbers while the number of the whole outer iteration is less than the former one. We can't say which one is better, it depends. And here we show that the proposed algorithm is effective and promising.

TABLE 1. Numerical Results

Problem	dim	$2n+1$			$(n+1)(n+2)/2$		
		k	$\|g_Q\|$	$\|F\|$	k	$\|g_Q\|$	$\|F\|$
HS40	4	7	6.779569e-006	1.805985e-005	7	9.606303e-006	2.014731e-005
SC206	2	5	5.964599e-006	4.291315e-007	4	6.877607e-006	7.967604e-007
SC207	2	7	6.513801e-006	3.026469e-006	6	9.848525e-006	1.500385e-006
SC208	2	95	7.838677e-006	5.308939e-006	20	1.808174e-006	8.318998e-007
Ferraris and Tronconi	2	43	4.317743e-006	1.723028e-005	17	9.064597e-006	9.640606e-006
Extended Powell singual function	4	60	3.567228e-006	1.314412e-004	57	8.841074e-006	1.789611e-004
Extended Powell singual function	8	59	5.841614e-006	1.899750e-004	9	8.210961e-006	1.212916e-004
Broyden Tridiagonal function	5	59	6.667181e-006	5.262892e-006	4	5.094843e-006	9.565117e-007
Discrete boundary value function	5	3	5.800534e-007	2.733637e-007	3	1.136031e-006	4.211831e-007
Discrete boundary value function	10	4	3.439812e-007	1.545640e-007	3	3.421256e-007	1.615803e-007
Extended Tridiagonal 1 function	4	11	6.124970e-006	4.606700e-005	8	4.329193e-006	5.105501e-005
Extended BD1 function	4	99	7.001012e-006	9.211647e-006	failure		
POWER function (CUTE)	5	4	6.735499e-006	1.087452e-006	4	6.938087e-006	1.119569e-006
Extended Powell function	4	4	2.208470e-007	1.404933e-007	4	3.384554e-007	1.422355e-007
Box three-dimensional function	3	12	1.959474e-007	2.906171e-006	12	1.893713e-007	1.683440e-006

Acknowledgements

The author gratefully acknowledges the partial supports of the Guangxi Natural Science Foundation (2018JJA110105), the Open Project of Guangxi Colleges and Universities Key Laboratory of Complex System Optimization and Big Data Processing (2017CSOBDP0104), the Foundation of Yulin Normal University (G2019ZK12), the Project of Improving Youth Scientific Researching Ability of Guangxi (2019KY0587).

REFERENCES

[1] S. Zeng, S. Migórski, A.A. Khan, Nonlinear quasi-hemivariational inequalities: existence and optimal control, SIAM J. Control Optim. 59 (2021), 1246-1274.

[2] Z Liu, D. Motreanu, S. Zeng, Generalized penalty and regularization method for differential variational-hemivariational inequalities, SIAM J. Optim. 31 (2021), 1158-1183.

[3] A.J. Wood, B.F. Wollenberg, Power Generations, Operations, and Control, Wiley, New York, 1996.

[4] S. Migórski, A.A. Khan, S. Zeng, Inverse problems for nonlinear quasi-hemivariational inequalities with application to mixed boundary value problems, Inverse Probl. 36 (2020), 024006.

[5] S. Migórski, M. Sofonea, S. Zeng, Well-posedness of history-dependent sweeping processes, SIAM J. Math. Anal. 51 (2021), 1082-1107.

[6] J.Y. Fan, Y.X. Yuan, On the quadratic convergence of the Levenberg-Marquardt method without nonsingularity assumption, Computing 74 (2005), 23-39.

[7] M. Marazzi, J. Nocedal, Wedge trust region methods for derivative-free optimizationm Math. Program. 91 (2002), 289-305.

[8] J.J. Moré, S.M. Wild, Benchmarking derivative-free optimization algorithms, SIAM J. Optim. 20 (2009), 172-191.

[9] M.J.D. Powell, On trust region methods for unconstrained minimization without derivatives, Math. Program. 97 (2003), 605-623.

[10] M.J.D. Powell, Least Frobenius norm updating of quadratic models that satisfy interpolation conditions, Math. Program. Ser. B 100 (2004), 183-215.

[11] M.J.D. Powell, Development of NEWUOA for unconstrained minimization without derivatives, IMA J. Numer. Anal. 28 (2008), 649-664.

[12] J.R. Sharma, H. Arora, An efficient derivative free iterative method for solving systems of nonlinear equations, Appl. Anal. Discrete Math. 7 (2013), 390-403.

[13] P.L. Toint, Nonlinear step size control, trust regions and regularization for unconstrained optimization, Optim. Methods Softw. 28 (2013), 82-95.

[14] M.T. Darvishi, A. Barati, Super cubic iterative methods to solve systems of nonlinear equations, Appl. Math. Comput. 188 (2007), 1678-1685,
[15] M. Frontini, E. Sormani, Third-order methods from quadrature formulae for solving systems of nonlinear equations, Appl. Math. Comput. 149 (2004), 771-782.
[16] M. Grau-Sánchez, A. Grau, M. Noguera, Frozen divided difference scheme for solving systems of nonlinear equations, J. Comput. Appl. Math. 235 (2011), 1739-1743.
[17] A.R. Conn, K. Scheinber, Ph.L. Toint, A derivative-free algorithm in practice, In: Proceedings of the 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, AIAA, Reston, 1998.
[18] A.R. Conn, K. Scheinber, L.N. Vicente, Global convergence of general derivative-free trust-region algorithms to fisrt- and second-order critical points, SIAM J. Optim. 20 (2009), 387-415.
[19] H. Zhang, A.R. Conn, K. Scheinber, A derivative-free algorithm for least-squares minimization, SIAM J. Optim. 20 (2010), 3555-3576.
[20] H. Zhang, A.R. Conn, On the local convergence of a derivative-free algorithm for least-squares minimization, Comput. Optim. Appl. 51 (2012), 481-507.
[21] S. Bellavia, Convergence of a regularized euclidean residual algorithm for nonlinear least-squares, SIAM J. Numer. Anal. 48 (2010), 1-29.
[22] H.Y. Benson, D.F. Shanno, Interior-Point methods for nonconvex nonlinear programming: cubic regularization, Comput. Optim. Appl. 58 (2014), 323-346.
[23] S. Bellavia, B. Morini, Strong local convergence properties of adaptive regularized methods for nonlinear least squares, IMA J. Numer. Anal. 35 (2015), 947-968.
[24] A. Griewank, The modification of Newton's method for unconstrained optimization by bounding cubic terms, Technical Report NA/12, University of Cambridge, 1981
[25] C. Cartes, N.I.M. Gould, Ph.L. Toint, Adaptive cubic regularisation methods for unconstrained optimization. Part I: motivation, convergence and numerical results, Math. Program. 127 (2011), 245-295.
[26] C. Cartis, N.I.M. Gould, Ph.L. Toint, Adaptive cubic regularisation methods for unconstrained optimization. Part II: worst-case function- and derivative-evaluation complexity, Math. Program. Ser. A. 130 (2011), 295-319.
[27] A.R. Conn, K. Scheinber, L.N. Vicente, Introduction to Derivative-Free Optimization, MPS/SIAM Ser. Optim. 8, SIAM, Philadelphia, 2009.
[28] C. Cartis, N.I.M. Gould, Ph.L. Toint, On the oracle complexity of fisrt-order and derivative-free algorithms for smooth nonconvex minimization, SIAM J. Optim. 22 (2012), 66-86.
[29] W. Hock, K. Schittkowski, Test Examples for Nonlinear Programming Codes. Lecture Notes in Economics and Mathematical Systems, 187. Springer-Verlag, New York, 1981.
[30] K. Schittkowski, More test examples for nonlinear mathematical programming codes, Lecture Notes in Economics and Mathematics System, Vol. 282, Springer-Verlag, Berlin, Heidelberg, 1987.