A Regularized Newton Method without Line Search for Unconstrained Optimization

Guidance

Associate Professor Nobuo YAMASHITA

Kenji UEDA

Department of Applied Mathematics and Physics

Graduate School of Informatics

Kyoto University

February 2009

Abstract

For unconstrained optimization, Newton-type methods have good convergence properties, and are used in practice. The Newton's method combined with a trust-region method (the TR-Newton method), the cubic regularization of Newton's method and the regularized Newton method with line search methods are such Newton-type methods. The TR-Newton method and the cubic regularization of Newton's method have to solve nonconvex subproblems at each iteration in order to get a search direction although these methods converge rapidly with fewer function evaluations. Thus their total computational times may become large. On the other hand, the regularized Newton method with line search methods gets its search direction by only solving linear equations. However, it may evaluate the objective function value many times in a line search step. Therefore, it is significant to construct a solution method whose behavior is similar to the TR-Newton method, and whose subproblems can be solved easily.

In this paper, we propose a regularized Newton method without line search. The proposed method controls a regularized parameter instead of a step size in order to guarantee the global convergence. We demonstrate that it is closely related to the TR-Newton method when the Hessian of the objective function is positive definite. Moreover, it does not solve nonconvex problems but linear equations as subproblems at each iteration. Thus, the proposed algorithm is regarded as a desired solution method mentioned above. We show that the proposed algorithm has the following convergence properties. (a) The proposed algorithm has global convergence under appropriate conditions. (b) It has superlinear rate of convergence under the local error bound condition. (c) Its global complexity bound, which is the first iteration k such that $\|\nabla f(x_k)\| \leq \epsilon$, is $O(\epsilon^{-2})$ when f is nonconvex, $O(\epsilon^{-\frac{5}{3}})$ when f is convex, and $O(\epsilon^{-1})$ when f is strongly convex. Moreover, we report numerical results that show that the proposed algorithm is competitive with the existing Newton-type methods, and hence it is very promising.

Contents

1 Introduction

In this paper, we consider the following unconstrained minimization problem.

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \ f(x), \tag{1.1}$$

where f is a twice continuously differentiable function from $\mathbb{R}^n$ into $\mathbb{R}$. Many solution methods for (1.1), such as the steepest descent method and the Newton's method, have been proposed [1, 2, 11, 14]. Usually, efficiencies of these solution methods are discussed from the following points of view [1, 2, 11, 14].

- Global convergence from an arbitrary initial point to a stationary point of f;
- Rate of convergence, such as the superlinear convergence and the quadratic convergence, in a neighborhood of a local optimal solution;
- Numerical results for benchmark problems such as CUTEr [7];
- The first iteration J_g satisfying $\|\nabla f(x_{J_g})\| \leq \epsilon$, or the first iteration J_f satisfying $f(x_{J_f}) - f^* \leq \epsilon$, where $\{x_k\}$ is a sequence generated by some algorithms, ϵ is a given positive constant and f^* is the optimal value of f.

The last item is important when we solve large-scale problems where an appropriate initial point is difficult to find and we want to estimate the computational time for a given accuracy of a solution in advance [3, 12, 13, 16, 17]. In this paper, J_g and J_f are referred to as global complexity bounds of the algorithm. In what follows, we discuss existing algorithms from the above four points of view, and then we explain a regularized Newton method proposed in this paper.

The steepest descent method is an iterative method which uses $-\nabla f(x_k)$ as a search direction. The steepest descent method has a global convergence and a linear rate of convergence under appropriate conditions. A convergence of the steepest descent method is generally slow as compared to that of the Newton-type methods. However, the steepest descent method is suitable for large-scale problems since it does not need to compute Hessian matrices of f. The global complexity bound of the steepest descent method is shown to be $J_g = O(\epsilon^{-2})$ when f is nonconvex, and $J_f = O(\epsilon^{-\frac{1}{2}})$ when f is convex [11].

The Newton's method uses Hessian matrices of f, and has a quadratic rate of convergence under appropriate conditions. Moreover, the Newton's method combined with a trust-region method [4] has global convergence. In what follows, we represent the TR-Newton method by the Newton's method with a trust-region method. For a current point x_k and a current trust-region Δ_k, the TR-Newton method adopts a search direction $\bar{d}_k(\Delta_k)$ as

$$\bar{d}_k(\Delta_k) \in \underset{\|d\| \leq \Delta_k}{\operatorname{argmin}} \left(f(x_k) + \nabla f(x_k)^T d + \frac{1}{2} d^T \nabla^2 f(x_k) d \right).$$

For large-scale problems with sparse Hessian matrices, the TR-Newton method can get a solution efficiently with the use of the sparsity. However, a complexity bound of the TR-Newton method remains unknown.

Recently, Nesterov and Polyak [13] proposed the cubic regularization of Newton's method. The cubic regularization of Newton's method has a global and quadratic convergence as well as the TR-Newton method. Moreover, the global complexity bound of the cubic regularization of Newton's method is shown to be $J_g = O(\epsilon^{-\frac{3}{2}})$ when f is nonconvex, and $J_f = O(\epsilon^{-\frac{1}{3}})$ when f is convex [12]. More recently, Cartis, Gould and Toint [3] extended the cubic regularization of Newton's method, called the adaptive cubic overestimation method, and they reported that the adaptive cubic overestimation method worked well as compared to the TR-Newton method in their numerical experiments. The cubic regularization of Newton's method uses a global minimizer of a cubic model function as the next iteration point. In order to get the minimizer, it solves certain nonlinear equations equivalent to minimizing the cubic model function. Since we do not know a computational complexity to solve the nonlinear equations, we cannot estimate the total computational complexity of the cubic regularization of Newton's method even if we know J_g or J_f.

When f is convex, the regularized Newton method [9, 10, 16, 17] is one of the efficient solution methods for (1.1). For a current point x_k, the regularized Newton method adopts a search direction d_k by

$$d_k = -(\nabla^2 f(x_k) + \mu_k I)^{-1} \nabla f(x_k),$$

where μ_k is a positive parameter. We call μ_k a regularized parameter. If f is convex, then a matrix $\nabla^2 f(x_k) + \mu_k I$ is positive definite, and hence d_k is a descent direction for f at x_k. Therefore, the regularized Newton method with an appropriate line search method, such as the Armijo's step size rule, has a global convergence property. Li, Fukushima, Qi and Yamashita [9] showed that the regularized Newton method, which sets the regularized parameter μ_k as $\mu_k = \|\nabla f(x_k)\|$, has a quadratic rate of convergence under the assumption that $\|\nabla f(x)\|$ provides a local error bound for (1.1) in a neighborhood of an optimal solution x^*. Moreover, Polyak [16] showed that the global complexity bound of the regularized Newton method, which also sets the regularized parameter μ_k as $\mu_k = \|\nabla f(x_k)\|$, is $J_g = O(\epsilon^{-4})$. Recently, Ueda and Yamashita [17] extended the regularized Newton method to the unconstrained nonconvex optimization. The extended regularized Newton method adopts the regularized parameter μ_k as

$$\mu_k = c_1 \min(0, -\lambda_{\min}(\nabla^2 f(x_k))) + c_2 \|\nabla f(x_k)\|^\delta,$$

where c_1, c_2 and δ are given positive constants, and $\lambda_{\min}(\nabla^2 f(x_k))$ is the minimum eigenvalue of $\nabla^2 f(x_k)$. Ueda and Yamashita [17] adopted the Armijo's step size rule as a line search method. They showed that the extended regularized Newton method has global convergence under appropriate conditions and superlinear convergence under the local error bound condition. Moreover, its global complexity bound is $J_g = O(\epsilon^{-2})$.

The TR-Newton method and the cubic regularization of Newton's method have to solve nonconvex subproblems at each iteration. A number of efficient solution methods for these subproblems have been proposed. However, a lot of computational complexities may be required to get an exact solution of the subproblem, and this complexity is unknown. On the other hand, the regularized Newton method with line search methods can get a search direction by only solving linear equations. However, it may evaluate the objective function value many times in a line search step. Therefore, it is desirable to construct a solution method whose behavior is similar to the TR-Newton method, and subproblems can be solved easily. In this paper, we proposed a regularized Newton method without line search. In order to guarantee the global convergence, it controls the regularized parameter μ_k. The proposed algorithm solves linear equations to get the search direction $d_k(\mu_k)$. As seen in the next section, the next iteration point $x_{k+1} = x_k + d_k(\mu_k)$ generated by the proposed algorithm coincides with the next iteration point $x_{k+1} = x_k + \bar{d}_k(\Delta_k)$ generated by the TR-Newton method with a certain trust-region Δ_k. Therefore, we expect that the proposed regularized Newton method behaves as well as the TR-Newton method. We show that the proposed algorithm has a global convergence property, and a superlinear convergence property under the local error bound condition. We also give global complexity bounds of the proposed algorithm. In particular, we show that the global complexity bounds are $J_g = O(\epsilon^{-2})$ when f is nonconvex, $J_g = O(\epsilon^{-\frac{5}{3}})$ and $J_f = O(\epsilon^{-\frac{2}{3}})$ when f is convex, and $J_g = O(\epsilon^{-1})$ and $J_f = O(\log \epsilon^{-1})$ when f is strongly convex.

This paper is organized as follows. In the next section, we propose a regularized Newton's method which controls the regularized parameter at each iteration. In Section 3, we show its global convergence. In Section 4, we establish superlinear convergence under the local error bound condition. In Section 5, we give the global complexity bounds of the proposed algorithm. Then, numerical results are presented and discussed in Section 6. Finally Section 7 concludes the paper.

Throughout the paper, we use the following notations. For a vector $x \in \mathbb{R}^n$, $\|x\|$ denotes the Euclidean norm defined by $\|x\| := \sqrt{x^T x}$. For a symmetric matrix $M \in \mathbb{R}^{n \times n}$, we denote the maximum eigenvalue and the minimum eigenvalue of M as $\lambda_{\max}(M)$ and $\lambda_{\min}(M)$, respectively. Then, $\|M\|$ denotes the ℓ_2 norm of M defined by $\|M\| := \sqrt{\lambda_{\max}(M^T M)}$. If M is symmetric positive semidefinite matrix, then $\|M\| = \lambda_{\max}(M)$. Furthermore, $M \succ (\succeq) 0$ denotes the positive (semi)definiteness of M, i.e., $\lambda_{\min}(M) > (\geq) 0$. $B(x, r)$ denotes a closed sphere with center x and radius r, i.e., $B(x, r) := \{y \in \mathbb{R}^n \mid \|y - x\| \leq r\}$. $\mathrm{dist}(x, S)$ denotes the distance between a vector $x \in \mathbb{R}^n$ and a set $S \subseteq \mathbb{R}^n$, i.e., $\mathrm{dist}(x, S) := \min_{y \in S} \|y - x\|$. For sets $S_1 \subseteq \mathbb{R}^n$ and $S_2 \subseteq \mathbb{R}^n$, $S_1 + S_2$ denotes the sum of S_1 and S_2 defined by $S_1 + S_2 := \{x + y \in \mathbb{R}^n \mid x \in S_1, y \in S_2\}$.

2 Proposed algorithm

In this section, we propose a regularized Newton method that controls the regularized parameter at each iteration. In what follows, x_k denotes the k-th iterative point, and g_k and H_k denotes the gradient $\nabla f(x_k)$ and the Hessian $\nabla^2 f(x_k)$, respectively.

For a given positive parameter ν_k, we consider a regularized parameter μ_k defined by

$$\mu_k := c\Lambda_k + \nu_k \|g_k\|^\delta, \tag{2.1}$$

where c and δ are given constants such that $c > 1$ and $\delta \geq 0$, and Λ_k is defined by

$$\Lambda_k := \max(0, -\lambda_{\min}(H_k)).$$

From the definition of Λ_k, the matrix $H_k + c\Lambda_k I$ is positive semidefinite even if f is nonconvex. Therefore, if $\|g_k\| \neq 0$, then $H_k + \mu_k I = H_k + c\Lambda_k I + \nu_k \|g_k\|^\delta I \succ 0$. Thus we can compute a vector $d_k(\nu_k)$ defined by

$$d_k(\nu_k) := -(H_k + c\Lambda_k I + \nu_k \|g_k\|^\delta I)^{-1} g_k. \tag{2.2}$$

The existing regularized Newton method uses a search direction $d_k(\nu)$ with ν_k fixed to a certain ν, and generates the next iterative point $x_{k+1} = x_k + t d_k(\nu)$ by controlling a step size t so that the objective function value decreases. In this paper, we propose to control ν_k in order to satisfy $f(x_{k+1}) < f(x_k)$ with $x_{k+1} = x_k + d_k(\nu_k)$.

In order to find an appropriate ν_k, we use the idea of updating trust-region Δ_k in the TR-Newton method. Let $m_k : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ be a model function of f at x_k defined by

$$m_k(d, \nu) := f(x_k) + g_k^T d + \frac{1}{2} d^T (H_k + c\Lambda_k I + \nu \|g_k\|^\delta I) d. \tag{2.3}$$

Note that $d_k(\nu_k)$ is a global minimizer of $m_k(\cdot, \nu_k)$ if $\|g_k\| \neq 0$. Let $\rho_k : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ be the ratio of the reduction of the objective function value to that of the model function value, i.e.,

$$\rho_k(d, \nu) := \frac{f(x_k) - f(x_k + d)}{f(x_k) - m_k(d, \nu)}.$$

If $\rho_k(d_k(\nu_k), \nu_k)$ is large, i.e., the reduction $f(x_k) - f(x_k + d_k(\nu_k))$ is sufficiently large as compared to the reduction of the model function, we adopt $d_k(\nu_k)$ and decrease the parameter ν_k. On the other hand, if $\rho_k(d_k(\nu_k), \nu_k)$ is small, i.e., the reduction $f(x_k) - f(x_k + d_k(\nu_k))$ is not large, we increase ν_k and compute $d_k(\nu_k)$ once again.

Based on the ideas, we propose the following algorithm. We call the proposed algorithm the adaptive regularized Newton method, because it uses an adaptive parameter ν.

The Adaptive Regularized Newton Method

Step 0 : Choose parameters $\nu_0, \nu_{\min}, \delta, c, \gamma_1, \gamma_2, \eta_1, \eta_2$ such that

$$\nu_0 \geq \nu_{\min} > 0,\ \delta \geq 0,\ c > 1,\ 1 < \gamma_1 \leq \gamma_2,\ 0 < \eta_1 \leq \eta_2 \leq 1.$$

Choose a starting point x_0. Set $k := 0$.

Step 1 : If the stopping criterion is satisfied, then terminate. Otherwise, go to Step 2.

Step 2 : **Step 2.0 :** Set $l_k := 1$ and $\bar{\nu}_{l_k} = \nu_k$.

Step 2.1 : Compute

$$d_k(\bar{\nu}_{l_k}) = -(H_k + c\Lambda_k I + \bar{\nu}_{l_k} \|g_k\|^\delta I)^{-1} g_k.$$

Step 2.2 : Compute

$$\rho_k(d_k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) = \frac{f(x_k) - f(x_k + d_k(\bar{\nu}_{l_k}))}{f(x_k) - m_k(d_k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k})}.$$

If $\rho_k(d_k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) < \eta_1$, then update $\bar{\nu}_{l_k+1} \in [\gamma_1 \bar{\nu}_{l_k}, \gamma_2 \bar{\nu}_{l_k}]$, set $l_k := l_k + 1$, and go to Step 2.1. Otherwise, go to Step 3.

Step 3 : If $\eta_2 > \rho_k(d_k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) \geq \eta_1$, then update $\nu_{k+1} \in [\bar{\nu}_{l_k}, \gamma_1 \bar{\nu}_{l_k}]$.
If $\rho_k(d_k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) \geq \eta_2$, then update $\nu_{k+1} \in [\nu_{\min}, \bar{\nu}_{l_k}]$.
Update $x_{k+1} = x_k + d_k(\bar{\nu}_{l_k})$. Set $k := k + 1$, and go to Step 1.

The proposed algorithm is closely related to the TR-Newton method as follows. Consider the case where H_k is positive definite. Then, since $\Lambda_k = 0$, the next iteration point x_{k+1} of the proposed algorithm lies on a trajectory Γ_k defined by

$$\Gamma_k := \left\{ x \in \mathbb{R}^n \mid x = x_k - (H_k + \nu I)^{-1} g_k, \ \nu \in (0, \infty) \right\}.$$

On the other hand, the next iteration point x_{k+1} of the TR-Newton method lie on a trajectory $\bar{\Gamma}_k$ defined by

$$\bar{\Gamma}_k := \left\{ x \in \mathbb{R}^n \;\middle|\; x = x_k + \bar{d}_k(\Delta), \ \bar{d}_k(\Delta) \in \operatorname*{argmin}_{\|d\| \leq \Delta} \left(f(x_k) + g_k^T d + \frac{1}{2} d^T H_k d \right), \ \Delta \in (0, \infty) \right\}.$$

In [4], it is shown that $\bar{d}_k(\Delta) \in \operatorname{argmin}_{\|d\| \leq \Delta} \left(f(x_k) + g_k^T d + \frac{1}{2} d^T H_k d \right)$ if and only if there exists $\lambda_k(\Delta)$ such that

$$\begin{aligned} (H_k + \lambda_k(\Delta) I) \bar{d}_k(\Delta) &= -g_k, \\ H_k + \lambda_k(\Delta) I &\succeq 0, \\ \lambda_k(\Delta) &\geq 0, \\ \lambda_k(\Delta)(\|\bar{d}_k(\Delta)\| - \Delta) &= 0. \end{aligned}$$

It then follows from the positive definiteness of H_k that

$$\bar{d}_k(\Delta) = \begin{cases} -H_k^{-1} g_k & \text{if } \|H_k^{-1} g_k\| \leq \Delta, \\ -(H_k + \lambda_k(\Delta) I)^{-1} g_k & \text{otherwise,} \end{cases}$$

where $\lambda_k(\Delta)$ is a positive constant such that $\|(H_k + \lambda_k(\Delta) I)^{-1} g_k\| = \Delta$. Therefore, the trajectory $\bar{\Gamma}_k$ can be written as

$$\begin{aligned} \bar{\Gamma}_k = &\{ x \in \mathbb{R}^n \mid x = x_k - (H_k + \lambda_k(\Delta) I)^{-1} g_k, \ \|(H_k + \lambda_k(\Delta) I)^{-1} g_k\| = \Delta, \ \Delta \in (0, \|H_k^{-1} g_k\|) \} \\ &\cup \{ x_k - H_k^{-1} g_k \}. \end{aligned}$$

Since $\lambda_k(\Delta)$ decreases monotonically on $(0, \|H_k^{-1} g_k\|)$, we have $\lim_{\Delta \to 0} \lambda_k(\Delta) = \infty$ and $\lim_{\Delta \to \|H_k^{-1} g_k\|} \lambda_k(\Delta) = 0$. Thus the trajectory Γ_k coincides with the trajectory $\bar{\Gamma}_k \setminus \{x_k - H_k^{-1} g_k\}$, and hence for a certain $\nu \in (0, \infty)$, there exists Δ such that $d_k(\nu) = \bar{d}_k(\Delta)$. From this fact, we expect that the proposed algorithm behaves as well as the TR-Newton method when H_k is positive definite. Figure 1 shows the search direction $d_k^{(\mathrm{SDM})}$ of the steepest descent method, the search direction $d_k^{(\mathrm{NM})}$ of the pure Newton's method, the search direction $d_k^{(\mathrm{RNM})}$ of the regularized Newton method, and the trajectory Γ_k of the proposed algorithm. The contour in Figure 1 is that of the quadratic model function $f(x_k) + g_k^T d + \frac{1}{2} d^T H_k d$.

On the other hand, when H_k is not positive definite, the behavior of the proposed algorithm may be different from that of the TR-Newton method. For example, consider the case where H_k is not positive semidefinite and $\|g_k\| = 0$. Then, $d_k(\nu)$ of the proposed algorithm is always 0 for any $\nu \in (0, \infty)$, while

$\bar{d}_k(\Delta)$ of the TR-Newton method is not 0. Therefore, the proposed algorithm do not necessarily have the same properties as the TR-Newton method.

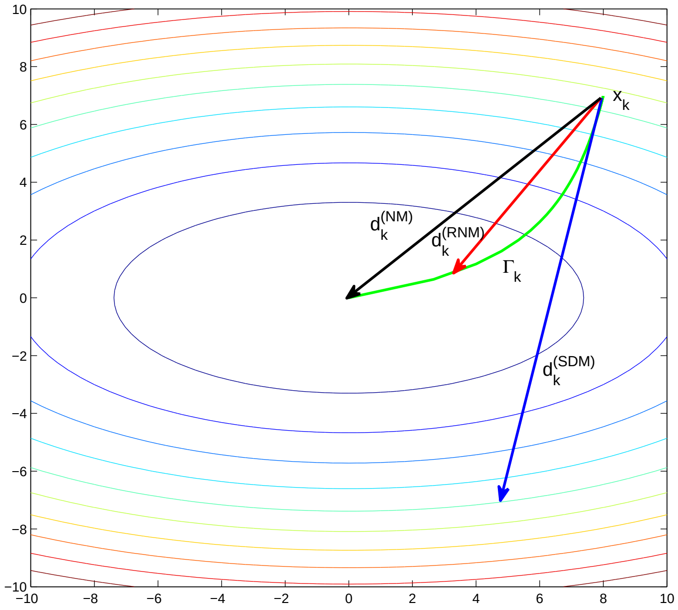

Figure 1: Image of relationship among $d_k^{(\mathrm{SDM})}, d_k^{(\mathrm{NM})}, d_k^{(\mathrm{RNM})}$ and Γ_k

In the remainder of this section, we show that the proposed algorithm is well-defined when $\|g_k\| \neq 0$.

Theorem 2.1. *If $\|g_k\| \neq 0$, then the proposed algorithm is well-defined, i.e., the number l_k of inner iterations is finite.*

Proof. Since f is twice continuously differentiable, we have from the definition of $d_k(\bar{\nu}_{l_k})$ that

$$\begin{aligned} f(x_k) - f(x_k + d_k(\bar{\nu}_{l_k})) &= -g_k^T d_k(\bar{\nu}_{l_k}) - O(\|d_k(\bar{\nu}_{l_k})\|^2) \\ &= g_k^T (H_k + c\Lambda_k I + \bar{\nu}_{l_k} \|g_k\|^\delta I)^{-1} g_k - O(\|d_k(\bar{\nu}_{l_k})\|^2). \end{aligned}$$

Moreover, from the definitions of $d_k(\bar{\nu}_{l_k})$ and $m_k(d_k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k})$, we have

$$\begin{aligned} f(x_k) - m_k(d_k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) &= -g_k^T d_k(\bar{\nu}_{l_k}) - \frac{1}{2} d_k(\bar{\nu}_{l_k})^T (H_k + c\Lambda_k I + \bar{\nu}_{l_k} \|g_k\|^\delta I) d_k(\bar{\nu}_{l_k}) \\ &= \frac{1}{2} g_k^T (H_k + c\Lambda_k I + \bar{\nu}_{l_k} \|g_k\|^\delta I)^{-1} g_k. \end{aligned}$$

It then follows from the definitions of $d_k(\bar{\nu}_{l_k})$ and $\rho_k(d_k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k})$ that

$$\begin{aligned} \rho_k(d_k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) &= \frac{g_k^T (H_k + c\Lambda_k I + \bar{\nu}_{l_k} \|g_k\|^\delta I)^{-1} g_k - O(\|d_k(\bar{\nu}_{l_k})\|^2)}{\frac{1}{2} g_k^T (H_k + c\Lambda_k I + \bar{\nu}_{l_k} \|g_k\|^\delta I)^{-1} g_k} \\ &= 2 - \frac{O\left(\left\|(H_k + c\Lambda_k I + \bar{\nu}_{l_k} \|g_k\|^\delta I)^{-1} g_k\right\|^2\right)}{\frac{1}{2} g_k^T (H_k + c\Lambda_k I + \bar{\nu}_{l_k} \|g_k\|^\delta I)^{-1} g_k} \\ &= 2 - \frac{O\left(\frac{1}{\bar{\nu}_{l_k}^2} \left\|\left(\frac{1}{\bar{\nu}_{l_k}} H_k + \frac{1}{\bar{\nu}_{l_k}} c\Lambda_k I + \|g_k\|^\delta I\right)^{-1} g_k\right\|^2\right)}{\frac{1}{2\bar{\nu}_{l_k}} g_k^T \left(\frac{1}{\bar{\nu}_{l_k}} H_k + \frac{1}{\bar{\nu}_{l_k}} c\Lambda_k I + \|g_k\|^\delta I\right)^{-1} g_k} \end{aligned} \tag{2.4}$$

From the updating rule of $\bar{\nu}_{l_k}$ in Step 2.2, we have $\bar{\nu}_{l_k} \to \infty$ as $l_k \to \infty$. Then, taking $l_k \to \infty$, the second term of the right-hand side of (2.4) goes to 0, and hence $\lim_{l_k \to \infty} \rho_k(d_k(\bar{\nu}_{l_k}) = 2 > \eta_1$. Therefore, the proposed algorithm is well-defined. □

In Sections 3 – 5, we will show global and superlinear convergence, and give the global complexity bounds. In the sections, for simplicity, we denote l_k and $\bar{\nu}_{l_k}$ of the last iteration in the inner loops of Steps 2.0 – 2.2 at each k as l_k^* and ν_k^*, respectively. We also denote $d_k(\nu_k^*)$, $m_k(d_k(\nu_k^*), \nu_k^*)$ and $\rho_k(d_k(\nu_k^*), \nu_k^*)$ as d_k^*, m_k^*, and ρ_k^*, respectively, i.e.,

$$d_k^* := d_k(\nu_k^*) = -(H_k + c\Lambda_k I + \nu_k^* I)^{-1} g_k, \tag{2.5}$$

$$m_k^* := m_k(d_k(\nu_k^*), \nu_k^*) = f(x_k) + g_k^T d_k^* + \frac{1}{2} {d_k^*}^T (H_k + c\Lambda_k I + \nu_k^* I) d_k^*, \tag{2.6}$$

$$\rho_k^* := \rho_k(d_k(\nu_k^*), \nu_k^*) = \frac{f(x_k) - f(x_k + d_k^*)}{f(x_k) - m_k^*}. \tag{2.7}$$

3 Global convergence

In this section, we investigate the global convergence property of the proposed algorithm. To this end, we need the following assumption.

Assumption 1. *There exists a compact set $\Omega \subseteq \mathbb{R}^n$ such that $\{x_k\} \subseteq \Omega$.*

Note that Assumption 1 holds if the level set of f at the initial point x_0 is compact.

First, we show the relationship between $\|d_k(\nu)\|$ and $\|g_k\|$.

Lemma 3.1. *Suppose that $\|g_k\| \neq 0$. Then, for any $\nu \in [\nu_{\min}, \infty)$,*

$$\|d_k(\nu)\| \leq \frac{\|g_k\|^{1-\delta}}{\nu}.$$

Proof. We have from (2.2) that

$$\begin{aligned}
\|d_k(\nu)\| &= \|(H_k + c\Lambda_k I + \nu \|g_k\|^\delta I)^{-1} g_k\| \\
&\leq \|(H_k + c\Lambda_k I + \nu \|g_k\|^\delta I)^{-1}\| \cdot \|g_k\| \\
&= \lambda_{\max}\Big((H_k + c\Lambda_k I + \nu \|g_k\|^\delta I)^{-1}\Big) \|g_k\| \\
&= \frac{\|g_k\|}{\lambda_{\min}(H_k + c\Lambda_k I + \nu \|g_k\|^\delta I)} \\
&\leq \frac{\|g_k\|^{1-\delta}}{\nu},
\end{aligned} \tag{3.1}$$

where the last inequality follows from the facts that $H_k + c\Lambda_k I$ is positive semidefinite and $\|g_k\| \neq 0$. □

Since the sequence $\{x_k\}$ is in the compact set Ω by Assumption 1, there exists $U_g > 0$ such that

$$\|g_k\| \leq U_g, \quad \forall k \geq 0. \tag{3.2}$$

The next lemma indicates that $\|d_k(\nu)\|$ is bounded above if $\|g_k\|$ is away from 0.

Lemma 3.2. *Suppose that Assumption 1 holds. Suppose also that there exists a constant $\epsilon > 0$ such that $\|g_k\| \geq \epsilon$. Then, for any $\nu \in [\nu_{\min}, \infty)$,*

$$\|d_k(\nu)\| \leq b(\epsilon),$$

where

$$b(\epsilon) := \max\left(\frac{U_g^{1-\delta}}{\nu_{\min}}, \frac{1}{\nu_{\min}\epsilon^{\delta-1}}\right).$$

Proof. When $\delta \leq 1$, it follows from Lemma 3.1, (3.2) and $\nu \geq \nu_{\min}$ that

$$\|d_k(\nu)\| \leq \frac{U_g^{1-\delta}}{\nu_{\min}}. \tag{3.3}$$

Meanwhile, when $\delta > 1$, it follows from Lemma 3.1, $\|g_k\| \geq \epsilon$ and $\nu \geq \nu_{\min}$

$$\|d_k(\nu)\| \leq \frac{1}{\nu_{\min}\epsilon^{\delta-1}}.$$

This completes the proof. □

When $\|g_k\| \geq \epsilon$ for all k, we have from Lemma 3.2 that

$$x_k + sd_k(\nu) \in \Omega + B(0, b(\epsilon)), \quad \forall s \in [0,1], \quad \forall k \geq 0.$$

Moreover, since $\Omega + B(0, b(\epsilon))$ is compact and f is twice continuously differentiable, there exists $U_H(\epsilon) > 0$ such that

$$\|\nabla^2 f(x)\| \leq U_H(\epsilon), \quad \forall x \in \Omega + B(0, b(\epsilon)). \tag{3.4}$$

Next, we show that the parameter ν_k^* in μ_k is bounded above when $\|g_k\| \geq \epsilon$ for all $k \geq 0$.

Lemma 3.3. *Suppose that Assumption 1 holds. Suppose also that there exists a constant $\epsilon > 0$ such that $\|g_k\| \geq \epsilon$ for all $k \geq 0$. Then,*

$$\nu_k^* \leq \nu_{\max}(\epsilon),$$

where

$$\nu_{\max}(\epsilon) := \max\left(\nu_0, \frac{\gamma_2 U_H(\epsilon)}{\epsilon^\delta}\right).$$

Proof. From Taylor's theorem, there exists $\tau \in (0, 1)$ such that

$$f(x_k + d_k(\nu)) = f(x_k) + g_k^T d_k(\nu) + \frac{1}{2} d_k(\nu)^T \nabla^2 f(x_k + \tau d_k(\nu)) d_k(\nu).$$

It then follows from the definition (2.3) of $m_k(d_k(\nu), \nu)$ that

$$\begin{aligned}
&f(x_k + d_k(\nu)) - m_k(d_k(\nu), \nu) \\
&= \frac{1}{2} d_k(\nu)^T \Big(\nabla^2 f(x_k + \tau d_k(\nu)) - (H_k + c\Lambda_k I + \nu\|g_k\|^\delta I)\Big) d_k(\nu) \\
&= \frac{1}{2} d_k(\nu)^T \Big(\nabla^2 f(x_k + \tau d_k(\nu)) - \nu\|g_k\|^\delta I\Big) d_k(\nu) - \frac{1}{2} d_k(\nu)^T (H_k + c\Lambda_k I) d_k(\nu) \\
&\leq \frac{1}{2}(U_H(\epsilon) - \nu\|g_k\|^\delta)\|d_k(\nu)\|^2 \\
&\leq \frac{1}{2}(U_H(\epsilon) - \nu\epsilon^\delta)\|d_k(\nu)\|^2,
\end{aligned} \tag{3.5}$$

where the first inequality follows from $H_k + c\Lambda_k \succeq 0$ and (3.4), and the last inequality follows from $\|g_k\| \geq \epsilon$. Now suppose that $\nu \geq U_H(\epsilon)/\epsilon^\delta$. Then, we have

$$f(x_k + d_k(\nu)) \leq m_k(d_k(\nu), \nu),$$

and hence

$$\rho_k(d_k(\nu), \nu) = \frac{f(x_k) - f(x_k + d_k(\nu))}{f(x_k) - m_k(d_k(\nu), \nu)} \geq 1.$$

Thus, if $\bar{\nu}_{l_k} \geq U_H(\epsilon)/\epsilon^\delta$, then inner loops of Step 2 must terminate. Therefore, ν_k^* must satisfy

$$\nu_k^* \leq \max\left(\nu_{k-1}^*, \left(\frac{U_H(\epsilon)}{\epsilon^\delta}\right)\gamma_2\right) \leq \cdots \leq \max\left(\nu_0, \left(\frac{U_H(\epsilon)}{\epsilon^\delta}\right)\gamma_2\right).$$

This completes the proof. □

Next, we give a lower bound of the reduction of the model function when $\|g_k\| \geq \epsilon$ for all $k \geq 0$.

Lemma 3.4. *Suppose that Assumption 1 holds. Suppose also that there exists a constant $\epsilon > 0$ such that $\|g_k\| \geq \epsilon$ for all $k \geq 0$. Then,*

$$f(x_k) - m_k^* \geq p(\epsilon)\epsilon^2,$$

where

$$p(\epsilon) := \frac{1}{2\Big((1+c)U_H(\epsilon) + \nu_{\max}(\epsilon)U_g^\delta\Big)}.$$

Proof. Since $H_k + c\Lambda_k I$ is positive semidefinite and $\|g_k\| \neq 0$, we have

$$\begin{aligned}\lambda_{\min}\Big((H_k + c\Lambda_k I + \nu_k^*\|g_k\|^\delta I)^{-1}\Big) &= \frac{1}{\lambda_{\max}(H_k + c\Lambda_k I + \nu\|g_k\|^\delta I)} \\ &= \frac{1}{\lambda_{\max}(H_k) + c\Lambda_k + \nu_k^*\|g_k\|^\delta}.\end{aligned}$$

It then follows from $\|g_k\| \geq \epsilon$, (3.2), (3.4) and Lemma 3.3 that

$$\lambda_{\min}\Big((H_k + c\Lambda_k I + \nu_k^*\|g_k\|^\delta I)^{-1}\Big) \geq \frac{1}{(1+c)U_H(\epsilon) + \nu_{\max}(\epsilon)U_g^\delta}. \tag{3.6}$$

Therefore, we have from the definition (2.5) of d_k^* and the definition (2.6) of m_k^* that

$$\begin{aligned}f(x_k) - m_k^* &= -g_k^T d_k^* - \frac{1}{2}d_k^{*T}(H_k + c\Lambda_k I + \nu_k^*\|g_k\|^\delta I)d_k^* \\ &= \frac{1}{2}g_k^T(H_k + c\Lambda_k I + \nu_k^*\|g_k\|^\delta I)^{-1}g_k \\ &\geq \frac{1}{2}\lambda_{\min}\Big((H_k + c\Lambda_k I + \nu_k^*\|g_k\|^\delta I)^{-1}\Big)\|g_k\|^2 \\ &\geq \frac{1}{2\Big((1+c)U_H(\epsilon) + \nu_{\max}(\epsilon)U_g^\delta\Big)}\|g_k\|^2 \qquad (3.7) \\ &\geq \frac{1}{2\Big((1+c)U_H(\epsilon) + \nu_{\max}(\epsilon)U_g^\delta\Big)}\epsilon^2,\end{aligned}$$

where the second inequality follows from (3.6), and the last inequality follows from $\|g_k\| \geq \epsilon$. □

By using the above lemma and the updating rule of x_k, we give a lower bound of the reduction $f(x_k) - f(x_{k+1})$ when $\|g_k\| \geq \epsilon$ for all $k \geq 0$.

Lemma 3.5. *Suppose that Assumption 1 holds. Suppose also that there exists a constant $\epsilon > 0$ such that $\|g_k\| \geq \epsilon$ for all $k \geq 0$. Then,*

$$f(x_k) - f(x_{k+1}) \geq \eta_1 p(\epsilon)\epsilon^2.$$

Proof. Since $\rho_k^* \geq \eta_1$, we have

$$f(x_k) - f(x_{k+1}) \geq \eta_1(f(x_k) - m_k^*) \geq \eta_1 p(\epsilon)\epsilon^2,$$

where the last inequality follows from Lemma 3.4. □

Now, we are at the position to prove the main theorem of this section.

Theorem 3.1. *Suppose that Assumption 1 holds. Then,*

$$\liminf_{k\to\infty}\|g_k\| = 0 \quad \text{or} \quad \|g_K\| = 0, \text{ for some } K \geq 0.$$

Proof. Suppose the contrary, i.e., there exists a constant ϵ such that $\|g_k\| \geq \epsilon$ for all $k \geq 0$. Then, we have from Lemma 3.5 that

$$f(x_0) - f(x_k) \geq \sum_{j=0}^{k-1}(f(x_j) - f(x_{j+1})) \geq \sum_{j=0}^{k-1} \eta_1 p(\epsilon)\epsilon^2 = \eta_1 p(\epsilon)\epsilon^2 k.$$

Taking $k \to \infty$, the right-hand side of the inequality goes to infinity, and hence $\lim_{k\to\infty} f(x_k) = -\infty$. This contradicts Assumption 1 and the continuity of f. Hence, we have $\liminf_{k\to\infty} \|g_k\| = 0$ or $\|g_K\| = 0$ for some $K \geq 0$. □

Remark 3.1. *Note that we can prove $\lim_{k\to\infty} \|g_k\| = 0$ in a way similar to the proof of [17, Theorem 3.1] by replacing the statement "If $\eta_2 > \rho_k(d_k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) \geq \eta_1$, then update $\nu_{k+1} \in [\bar{\nu}_{l_k}, \gamma_1\bar{\nu}_{l_k}]$. If $\rho_k(d_k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) \geq \eta_2$, then update $\nu_{k+1} \in [\nu_{\min}, \bar{\nu}_{l_k}]$." in Step 3 with "If $\rho_k(d_k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) \geq \eta_1$, then update $\nu_{k+1} = \nu_0$." However, this modification may increase the number of inner iterations.*

Remark 3.2. *The TR-Newton method has a global convergence property to a second-order critical point [4]. However, since $d_k(\bar{\nu}_{l_k}) = 0$ when $\|g_k\| = 0$, the proposed algorithm may not converge to a second-order critical point.*

4 Local convergence

In this section, we show that the proposed algorithm converges superlinearly when $\|\nabla f(x)\|$ provides a local error bound (see Assumption 2 (d) below). Note that the local error bound condition holds if the second-order sufficient optimality condition holds at x^*. But the converse is not true. Thus the local error bound condition is weaker than the second-order sufficient optimality condition. In order to prove the superlinear convergence, we use techniques similar to [17] where the regularized Newton method with Armijo's step size rule is shown to have a superlinear rate of convergence under the local error bound condition.

First, we make the following assumptions.

Assumption 2.

(a) $0 < \delta < 1$.

(b) *There exists a local optimal solution x^* of the problem (1.1).*

(c) $\nabla^2 f$ *is local Lipschitz continuous, i.e., there exist constants $b_1 \in (0, 1)$ and $\bar{L}_H > 0$ such that*

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \leq \bar{L}_H\|x - y\|, \quad \forall x, y \in B(x^*, b_1).$$

(d) $\|\nabla f(x)\|$ *provides a local error bound for the problem (1.1) on $B(x^*, b_1)$, i.e., there exists a constant $\kappa_1 > 0$ such that*

$$\kappa_1 \mathrm{dist}(x, X^*) \leq \|\nabla f(x)\|, \quad \forall x \in B(x^*, b_1),$$

where X^* is the local optimal solution set of (1.1).

Note that under Assumption 2 (c), the following inequality holds.

$$\|\nabla^2 f(y)(x - y) - (\nabla f(x) - \nabla f(y))\| \leq \frac{1}{2}\bar{L}_H\|x - y\|^2, \quad \forall x, y \in B(x^*, b_1). \tag{4.1}$$

Moreover, since f is twice continuously differentiable, there exists a positive constant $\bar{L}_g$ such that

$$\|\nabla f(x) - \nabla f(y)\| \leq \bar{L}_g\|x - y\|, \quad \forall x, y \in B(x^*, b_1). \tag{4.2}$$

In what follows, $\bar{x}_k$ denotes an arbitrary vector such that

$$\|x_k - \bar{x}_k\| = \mathrm{dist}(x_k, X^*), \quad \bar{x}_k \in X^*.$$

Since we consider the case where f is not necessarily convex, it is not always true that $\Lambda_k = 0$. Therefore, we now investigate the relationship between Λ_k and $\mathrm{dist}(x_k, X^*)$. To this end, we need the following property on a singular matrix.

Lemma 4.1. *Suppose that $M \in \mathbb{R}^{n\times n}$ is singular, then $\|I - M\| \geq 1$.*

Proof. It directly follows from [8, Corollary 5.6.16]. □

By using Lemma 4.1, we show the following key lemma for superlinear convergence.

Lemma 4.2. *Suppose that Assumption 2 holds. If $x_k \in B(x^*, b_1/2)$, then*

$$\Lambda_k \leq \bar{L}_H \mathrm{dist}(x_k, X^*).$$

Proof. When $H_k \succeq 0$, we have $\Lambda_k = 0$. Thus the desired inequality holds. Next, we assume $\lambda_{\min}(H_k) < 0$. Let $\bar{\lambda}_k^{(i)}$ be the i-th largest eigenvalue of $\nabla^2 f(\bar{x}_k)$. Since $\bar{x}_k \in X^*$, we have $\bar{\lambda}_k^{(i)} \geq 0$. Moreover, since $\nabla^2 f(\bar{x}_k)$ is a real symmetric matrix, $\nabla^2 f(\bar{x}_k)$ can be diagonalized by some orthogonal matrix $\bar{Q}_k$, i.e.,

$$\bar{Q}_k^T \nabla^2 f(\bar{x}_k) \bar{Q}_k = \mathrm{diag}(\bar{\lambda}_k^{(i)}),$$

where $\mathrm{diag}(\bar{\lambda}_k^{(i)})$ denotes the diagonal matrix whose (i,i) element is $\bar{\lambda}_k^{(i)}$. Then, we obtain

$$\begin{aligned}\lambda_{\min}(H_k)I - \bar{Q}_k^T H_k \bar{Q}_k &= \lambda_{\min}(H_k)I - \bar{Q}_k^T\Big(\nabla^2 f(\bar{x}_k) + (H_k - \nabla^2 f(\bar{x}_k))\Big)\bar{Q}_k \\ &= \lambda_{\min}(H_k)I - \mathrm{diag}(\bar{\lambda}_k^{(i)}) - \bar{Q}_k^T(H_k - \nabla^2 f(\bar{x}_k))\bar{Q}_k.\end{aligned}$$

Since $\bar{Q}_k^T H_k \bar{Q}_k$ has the eigenvalue $\lambda_{\min}(H_k)$, the matrix $\lambda_{\min}(H_k)I - \bar{Q}_k^T H_k \bar{Q}_k$ is singular. Thus $\lambda_{\min}(H_k)I - \mathrm{diag}(\bar{\lambda}_k^{(i)}) - \bar{Q}_k^T(H_k - \nabla^2 f(\bar{x}_k))\bar{Q}_k$ is also singular. On the other hand, $\lambda_{\min}(H_k)I - \mathrm{diag}(\bar{\lambda}_k^{(i)})$ is nonsingular because $\lambda_{\min}(H_k) < 0$ and $\bar{\lambda}_k^{(i)} \geq 0$.

Now let

$$M := \Big(\lambda_{\min}(H_k)I - \mathrm{diag}(\bar{\lambda}_k^{(i)})\Big)^{-1}\Big(\lambda_{\min}(H_k)I - \mathrm{diag}(\bar{\lambda}_k^{(i)}) - \bar{Q}_k^T(H_k - \nabla^2 f(\bar{x}_k))\bar{Q}_k\Big).$$

Then, M is singular. It then follows from Lemma 4.1 that

$$\begin{aligned}1 &\leq \|I - M\| \\ &= \left\| I - \left(I - \Big(\lambda_{\min}(H_k)I - \mathrm{diag}(\bar{\lambda}_k^{(i)})\Big)^{-1}\bar{Q}_k^T(H_k - \nabla^2 f(\bar{x}_k))\bar{Q}_k\right)\right\| \\ &= \left\|\Big(\lambda_{\min}(H_k)I - \mathrm{diag}(\bar{\lambda}_k^{(i)})\Big)^{-1}\bar{Q}_k^T(H_k - \nabla^2 f(\bar{x}_k))\bar{Q}_k\right\| \\ &\leq \left\|\Big(\lambda_{\min}(H_k)I - \mathrm{diag}(\bar{\lambda}_k^{(i)})\Big)^{-1}\right\| \cdot \|\bar{Q}_k^T(H_k - \nabla^2 f(\bar{x}_k))\bar{Q}_k\| \\ &= \left\|\Big(\lambda_{\min}(H_k)I - \mathrm{diag}(\bar{\lambda}_k^{(i)})\Big)^{-1}\right\| \cdot \|H_k - \nabla^2 f(\bar{x}_k)\|. \end{aligned} \tag{4.3}$$

We consider $\|(\lambda_{\min}(H_k)I - \mathrm{diag}(\bar{\lambda}_k^{(i)}))^{-1}\|$ and $\|H_k - \nabla^2 f(\bar{x}_k)\|$ separately. Since $\lambda_{\min}(H_k) < 0$ and $\bar{\lambda}_k^{(i)} \geq 0$, we have

$$\begin{aligned}\left\|\Big(\lambda_{\min}(H_k)I - \mathrm{diag}(\bar{\lambda}_k^{(i)})\Big)^{-1}\right\| &= \max_{1\leq i\leq n}\left|\lambda_{\min}(H_k) - \bar{\lambda}_k^{(i)}\right|^{-1} \\ &= \frac{1}{\min_{1\leq i\leq n}\left|\lambda_{\min}(H_k) - \bar{\lambda}_k^{(i)}\right|} \\ &\leq \frac{1}{|\lambda_{\min}(H_k)|} \\ &= \frac{1}{\Lambda_k}.\end{aligned} \tag{4.4}$$

Next, we consider $\|H_k - \nabla^2 f(\bar{x}_k)\|$. Since $x_k \in B(x^*, b_1/2)$, we have

$$\|\bar{x}_k - x^*\| \le \|\bar{x}_k - x_k\| + \|x_k - x^*\| \le \|x^* - x_k\| + \|x_k - x^*\| \le b_1,$$

and hence $\bar{x}_k \in B(x^*, b_1)$. It then follows from Assumption 2 (c) that

$$\|H_k - \nabla^2 f(\bar{x}_k)\| \le \bar{L}_H \|x_k - \bar{x}_k\| = \bar{L}_H \mathrm{dist}(x_k, X^*). \tag{4.5}$$

Therefore, we have from (4.3) – (4.5) that

$$1 \le \frac{\bar{L}_H \mathrm{dist}(x_k, X^*)}{\Lambda_k},$$

which is the desired inequality. □

Next, we show that $\|d_k(\nu)\| = O(\mathrm{dist}(x_k, X^*))$.

Lemma 4.3. *Suppose that Assumption 2 holds. If $x_k \in B(x^*, b_1/2)$, then*

$$\|d_k(\nu)\| \le \kappa_2 \mathrm{dist}(x_k, X^*), \quad \forall \nu \in [\nu_{\min}, \infty),$$

where

$$\kappa_2 := \frac{\bar{L}_H}{2\nu_{\min}\kappa_1^\delta} + \max\left(1, \frac{1}{c-1}\right).$$

Proof. First note that $\nabla f(\bar{x}_k) = 0$. From the definition (2.2) of $d_k(\nu)$ we have

$$\begin{aligned}
&\|d_k(\nu)\| \\
&= \left\|(H_k + c\Lambda_k I + \nu\|g_k\|^\delta I)^{-1} g_k\right\| \\
&= \left\|(H_k + c\Lambda_k I + \nu\|g_k\|^\delta I)^{-1}\Big(g_k - \nabla f(\bar{x}_k) - H_k(x_k - \bar{x}_k) + H_k(x_k - \bar{x}_k)\Big)\right\| \\
&\le \left\|(H_k + c\Lambda_k I + \nu\|g_k\|^\delta I)^{-1}\Big(g_k - \nabla f(\bar{x}_k) - H_k(x_k - \bar{x}_k)\Big)\right\| + \left\|(H_k + c\Lambda_k I + \nu\|g_k\|^\delta I)^{-1} H_k(x_k - \bar{x}_k)\right\| \\
&\le \left\|(H_k + c\Lambda_k I + \nu\|g_k\|^\delta I)^{-1}\right\| \|g_k - \nabla f(\bar{x}_k) - H_k(x_k - \bar{x}_k)\| + \left\|(H_k + c\Lambda_k I + \nu\|g_k\|^\delta I)^{-1} H_k\right\| \|x_k - \bar{x}_k\| \\
&\le \frac{\bar{L}_H}{2}\|x_k - \bar{x}_k\|^2 \left\|(H_k + c\Lambda_k I + \nu\|g_k\|^\delta I)^{-1}\right\| + \|x_k - \bar{x}_k\| \left\|(H_k + c\Lambda_k I + \nu\|g_k\|^\delta I)^{-1} H_k\right\| \\
&= \frac{\bar{L}_H}{2}\mathrm{dist}(x_k, X^*)^2 \left\|(H_k + c\Lambda_k I + \nu\|g_k\|^\delta I)^{-1}\right\| + \mathrm{dist}(x_k, X^*) \left\|(H_k + c\Lambda_k I + \nu\|g_k\|^\delta I)^{-1} H_k\right\|,
\end{aligned} \tag{4.6}$$

where the last inequality follows from (4.9). First, we consider $\|(H_k + c\Lambda_k I + \nu\|g_k\|^\delta I)^{-1}\|$. Since $x_k \in B(x^*, b_1/2)$, we have $\bar{x}_k \in B(x^*, b_1)$. It follows from $H_k + c\Lambda_k \succeq 0$, $\nu \ge \nu_{\min}$ and Assumption 2 (d) that

$$\begin{aligned}
\left\|(H_k + c\Lambda_k I + \nu\|g_k\|^\delta I)^{-1}\right\| &= \lambda_{\max}\Big((H_k + c\Lambda_k I + \nu\|g_k\|^\delta I)^{-1}\Big) \\
&= \frac{1}{\lambda_{\min}(H_k + c\Lambda_k I + \nu\|g_k\|^\delta I)} \\
&\le \frac{1}{\nu\|g_k\|^\delta} \\
&\le \frac{1}{\nu_{\min}\kappa_1^\delta \mathrm{dist}(x_k, X^*)^\delta}.
\end{aligned} \tag{4.7}$$

Next, we consider $\|(H_k + c\Lambda_k I + \nu\|g_k\|^\delta I)^{-1} H_k\|$. Let $\lambda_k^{(i)}$ be the i-th largest eigenvalue of H_k. Then, the eigenvalues of $(H_k + c\Lambda_k I + \nu\|g_k\|^\delta I)^{-1} H_k$ are given by

$$\frac{\lambda_k^{(i)}}{\lambda_k^{(i)} + c\Lambda_k + \nu\|g_k\|^\delta}, \quad 1 \le i \le n.$$

Now we consider two cases: (a) $\lambda_k^{(i)} \ge 0$ and (b) $\lambda_k^{(i)} < 0$.

Case (a): This case implies that

$$\frac{\left|\lambda_k^{(i)}\right|}{\left|\lambda_k^{(i)} + c\Lambda_k + \nu\|g_k\|^\delta\right|} \le 1.$$

Case (b): In this case, since $-\Lambda_k = \lambda_{\min}(H_k) \le \lambda_k^{(i)} < 0$, we have $\lambda_k^{(i)} - \lambda_{\min}(H_k) \ge 0$ and $|\lambda_k^{(i)}| \le |\lambda_{\min}(H_k)|$. Therefore, we have

$$\begin{aligned}\frac{\left|\lambda_k^{(i)}\right|}{\left|\lambda_k^{(i)} + c\Lambda_k + \nu\|g_k\|^\delta\right|} &= \frac{\left|\lambda_k^{(i)}\right|}{\left|(\lambda_k^{(i)} - \lambda_{\min}(H_k)) - (c-1)\lambda_{\min}(H_k) + \nu\|g_k\|^\delta\right|} \\ &\le \frac{|\lambda_{\min}(H_k)|}{\lambda_k^{(i)} - \lambda_{\min}(H_k) + (c-1)\left|\lambda_{\min}(H_k)\right| + \nu\|g_k\|^\delta} \\ &\le \frac{1}{c-1}.\end{aligned}$$

Thus we have

$$\frac{\left|\lambda_k^{(i)}\right|}{\left|\lambda_k^{(i)} + c\Lambda_k + \nu\|g_k\|^\delta\right|} \le \max\left(1, \frac{1}{c-1}\right), \quad 1 \le i \le n,$$

and hence

$$\left\|(H_k + c\Lambda_k I + \nu\|g_k\|^\delta I)^{-1} H_k\right\| \le \max\left(1, \frac{1}{c-1}\right). \tag{4.8}$$

From (4.6) – (4.8), we have

$$\begin{aligned}\|d_k(\nu)\| &\le \frac{\bar{L}_H}{2\nu_{\min}\kappa_1^\delta}\mathrm{dist}(x_k, X^*)^{2-\delta} + \max\left(1, \frac{1}{c-1}\right)\mathrm{dist}(x_k, X^*) \\ &\le \left(\frac{\bar{L}_H}{2\nu_{\min}\kappa_1^\delta} + \max\left(1, \frac{1}{c-1}\right)\right)\mathrm{dist}(x_k, X^*),\end{aligned}$$

which is the desired inequality. □

From the above lemma, we can show that the next iteration point $x_{k+1} = x_k + d_k(\nu) \in B(x^*, b_1)$ if x_k is sufficiently close to x^*.

Lemma 4.4. *Suppose that Assumption 2 holds. Let $b_2 := b_1/(\kappa_2+1)$. If $x_k \in B(x^*, b_2)$, then*

$$x_k + d_k(\nu) \in B(x^*, b_1), \quad \forall \nu \in [\nu_{\min}, \infty).$$

Proof. Since $b_2 \le b_1/2$, we have $x_k \in B(x^*, b_1/2)$. Therefore, we obtain

$$\begin{aligned}\|x_k + d_k(\nu) - x^*\| &\le \|x_k - x^*\| + \|d_k(\nu)\| \\ &\le \|x_k - x^*\| + \kappa_2 \mathrm{dist}(x_k, X^*) \\ &\le \|x_k - x^*\| + \kappa_2\|x_k - x^*\| \\ &\le (\kappa_2+1)b_2 = b_1,\end{aligned}$$

where the second inequality follows from Lemma 4.3. □

From Lemma 4.4 and the convexity of the set $B(x^*, b_1)$, we have

$$x_k + sd_k(\nu) \in B(x^*, b_1), \quad \forall s \in [0,1], \quad \forall \nu \in [\nu_{\min}, \infty)$$

if $x_k \in B(x^*, b_2)$. It then follows from Assumption 2 (c) that

$$\|\nabla^2 f(x_k + sd_k(\nu)) - H_k\| \le \bar{L}_H \|d_k(\nu)\|, \quad \forall s \in [0,1], \quad \forall \nu \in [\nu_{\min}, \infty). \tag{4.9}$$

Now, we show that $l_k^* = 1$ and $\nu_k^* \le \nu_{k-1}^*$ if x_k is sufficiently close to x^*.

Lemma 4.5. *Suppose that Assumption 2 holds. Let*

$$b_3 := \min\left(b_2, \left(\frac{\nu_{\min}\kappa_1^\delta}{\kappa_2 \bar{L}_H}\right)^{\frac{1}{1-\delta}}\right).$$

If $x_k \in B(x^*, b_3)$, then $l_k^* = 1$ and $\nu_k^* \le \nu_{k-1}^*$. In particular, if $x_0, x_1, \ldots, x_k \in B(x^*, b_3)$, then $\nu_k^* \le \nu_0$.

Proof. Since $c\Lambda_k \ge 0$, we have from (3.5) that

$$\begin{aligned} f(x_k + d_k(\nu)) - m_k(d_k(\nu), \nu) &\le \frac{1}{2} d_k(\nu)^T (\nabla^2 f(x_k + \tau(\nu) d_k(\nu)) - H_k - \nu \|g_k\|^\delta I) d_k(\nu) \\ &\le \frac{1}{2} (\|\nabla^2 f(x_k + \tau(\nu) d_k(\nu)) - H_k\| - \nu \|g_k\|^\delta) \|d_k(\nu)\|^2 \qquad (4.10) \\ &\le \frac{1}{2} (\bar{L}_H \|d_k(\nu)\| - \nu \|g_k\|^\delta) \|d_k(\nu)\|^2 \\ &\le \frac{1}{2} \left(\frac{\bar{L}_H \|d_k(\nu)\|}{\|g_k\|^\delta} - \nu\right) \|g_k\|^\delta \|d_k(\nu)\|^2. \end{aligned}$$

where the third inequality follows from (4.9). It then follows from Assumption 2 (d), Lemma 4.3 and $\nu \ge \nu_{\min}$ that

$$\begin{aligned} f(x_k + d_k(\nu)) - m_k(d_k(\nu), \nu) &\le \frac{1}{2} \left(\frac{\bar{L}_H \kappa_2}{\kappa_1^\delta} \mathrm{dist}(x_k, X^*)^{1-\delta} - \nu\right) \|g_k\|^\delta \|d_k(\nu)\|^2 \\ &\le \frac{1}{2} \left(\frac{\bar{L}_H \kappa_2}{\kappa_1^\delta} \|x_k - x^*\|^{1-\delta} - \nu_{\min}\right) \|g_k\|^\delta \|d_k(\nu)\|^2 \\ &\le 0, \end{aligned}$$

where the second inequality follows from $\nu \ge \nu_{\min}$, and the last inequality follows from $x_k \in B(x^*, b_3)$. Therefore, we have $\rho(d_k(\nu), \nu) \ge 1$, and hence $l_k^* = 1$ and $\nu_k^* \le \nu_{k-1}^*$. The second part of the Lemma directly follows from the updating rule of ν. □

Next, we show that $\mathrm{dist}(x_k, X^*)$ converges to 0 superlinearly, as long as $\{x_k\}$ lies in a neighborhood of x^*.

Lemma 4.6. *Suppose that Assumption 2 holds. If $x_0, x_1, \ldots, x_k, x_{k+1} \in B(x^*, b_3)$, then*

$$\mathrm{dist}(x_{k+1}, X^*) = O\left(\mathrm{dist}(x_k, X^*)^{1+\delta}\right).$$

Therefore, there exists a positive constant b_4 such that

$$\mathrm{dist}(x_k, X^*) \le b_4 \Rightarrow \mathrm{dist}(x_{k+1}, X^*) \le \frac{1}{2} \mathrm{dist}(x_k, X^*).$$

Proof. We have from Assumption 2 (d) that

$$\begin{aligned}
\operatorname{dist}(x_{k+1}, X^*) &\le \frac{1}{\kappa_1}\|g_{k+1}\| \\
&\le \frac{1}{\kappa_1}\|H_k d_k^* + g_k\| + \frac{\bar{L}_H}{2\kappa_1}\|d_k^*\|^2 \\
&= \frac{1}{\kappa_1}\left\|c\Lambda_k d_k^* + \nu_k^*\|g_k\|^\delta d_k^*\right\| + \frac{\bar{L}_H}{2\kappa_1}\|d_k^*\|^2 \\
&\le \frac{c\Lambda_k}{\kappa_1}\|d_k^*\| + \frac{\nu_k^*}{\kappa_1}\|g_k\|^\delta\|d_k^*\| + \frac{\bar{L}_H}{2\kappa_1}\|d_k^*\|^2 \\
&\le \frac{c\Lambda_k}{\kappa_1}\|d_k^*\| + \frac{\nu_0}{\kappa_1}\|g_k\|^\delta\|d_k^*\| + \frac{\bar{L}_H}{2\kappa_1}\|d_k^*\|^2, \qquad (4.11)
\end{aligned}$$

where the second inequality follows from (4.9), the first equality follows from the definition (2.5) of d_k^*, and the last inequality follows from Lemma 4.5. From (4.2), we have

$$\|g_k\|^\delta = \|g_k - \nabla f(\bar{x}_k)\|^\delta \le \bar{L}_g^\delta \operatorname{dist}(x_k, X^*)^\delta. \qquad (4.12)$$

Therefore, we obtain from (4.11), (4.12), Lemma 4.2 and Lemma 4.3 that

$$\begin{aligned}
\operatorname{dist}(x_{k+1}, X^*) &\le \frac{c\kappa_2\bar{L}_H}{\kappa_1}\operatorname{dist}(x_k, X^*)^2 + \frac{\nu_0\kappa_2\bar{L}_g^\delta}{\kappa_1}\operatorname{dist}(x_k, X^*)^{1+\delta} + \frac{\kappa_2^2\bar{L}_H}{2\kappa_1}\operatorname{dist}(x_k, X^*)^2 \\
&\le \frac{\kappa_2(2c\bar{L}_H + 2\nu_0\bar{L}_g^\delta + \kappa_2\bar{L}_H)}{2\kappa_1}\operatorname{dist}(x_k, X^*)^{1+\delta}.
\end{aligned}$$

□

Lemma 4.6 shows that $\{\operatorname{dist}(x_k, X^*)\}$ converges to 0 superlinearly if $x_k \in B(x^*, b_3)$ for all k. Now we give a sufficient condition for $x_k \in B(x^*, b_3)$ for all k.

Lemma 4.7. *Suppose that Assumption 2 holds. Let $b_5 := \min(b_3, b_4)$ and $b_6 := \frac{1}{1+2\kappa_2}b_5$. If $x_0 \in B(x^*, b_6)$, then $x_k \in B(x^*, b_5)$ for all k.*

Proof. We prove the lemma by induction. First we consider the case where $k = 0$. Since $b_6 < b_5 \le b_3 \le b_2 \le b_1/2$, we have $x_0 \in B(x^*, b_1/2)$. Therefore, from Lemma 4.3, we obtain

$$\begin{aligned}
\|x_1 - x^*\| &= \|x_0 + d_0^* - x^*\| \\
&\le \|x_0 - x^*\| + \|d_0^*\| \\
&\le \|x_0 - x^*\| + \kappa_2 \operatorname{dist}(x_0, X^*) \\
&\le (1+\kappa_2)\|x_0 - x^*\| \\
&\le (1+\kappa_2)b_6 \\
&\le \frac{1+\kappa_2}{1+2\kappa_2}b_5 \le b_5,
\end{aligned}$$

which shows that $x_1 \in B(x^*, b_5)$. Next, we consider the case where $k \ge 1$. Suppose that $x_j \in B(x^*, b_5),\ j = 1, \ldots, k$. It follows from Lemma 4.6 that

$$\operatorname{dist}(x_j, X^*) \le \frac{1}{2}\operatorname{dist}(x_{j-1}, X^*) \le \cdots \le \left(\frac{1}{2}\right)^j \operatorname{dist}(x_0, X^*) \le \left(\frac{1}{2}\right)^j \|x_0 - x^*\| \le \left(\frac{1}{2}\right)^j b_6.$$

Therefore,

$$\|d_j\| \le \kappa_2 \operatorname{dist}(x_j, X^*) \le \left(\frac{1}{2}\right)^j \kappa_2 b_6. \qquad (4.13)$$

Thus we obtain

$$\|x_{k+1} - x^*\| \le \|x_0 - x^*\| + \sum_{j=0}^{k} \|d_j^*\| \le (1 + 2\kappa_2)b_6 = b_5,$$

which shows that $x_{k+1} \in B(x^*, b_5)$. This completes the proof. □

By using Lemmas 4.6 and 4.7, we give the rate of convergence.

Theorem 4.1. *Suppose that Assumption 2 holds. Let $\{x_k\}$ be a sequence generated by the proposed algorithm with $x_0 \in B(x^*, b_6)$. Then, $\{\mathrm{dist}(x_k, X^*)\}$ converges to 0 at the rate of $1+\delta$. Moreover, $\{x_k\}$ converges to a local optimal solution $\hat{x} \in B(x^*, b_5)$.*

Proof. The first part of the theorem directly follows from Lemmas 4.6 and 4.7. Therefore, we only show the second part. For all integers $p > q \ge 0$, we obtain

$$\|x_p - x_q\| \le \sum_{j=q}^{p-1} \|d_j^*\| \le \kappa_2 b_6 \sum_{j=q}^{p-1} \left(\frac{1}{2}\right)^j \le \kappa_2 b_6 \sum_{j=q}^{\infty} \left(\frac{1}{2}\right)^j \le \kappa_2 b_6 \left(\frac{1}{2}\right)^{q-1},$$

where the second inequality follows from (4.13). Thus, $\{x_k\}$ is a Cauchy sequence, and hence it converges. □

Remark 4.1. *Note that in a way similar to the proof of [9, Theorem 3.2], we can prove that $\{x_k\}$ converges to $\hat{x}$ at the rate of $1 + \delta$.*

Remark 4.2. *We get a rapid convergence if we take a larger δ. However, we cannot guarantee the quadratic convergence since δ must be less than 1. Note that when the second-order sufficient condition holds at x^*, we can prove that the proposed algorithm with $\delta = 1$ has quadratic convergence.*

5 Global complexity bound

In this section, we estimate the global complexity bound of the proposed algorithm. We consider three cases (a) f is nonconvex, (b) f is convex and (c) f is strongly convex.

5.1 Nonconvex case

In this subsection, we consider the case where f is nonconvex. Throughout this subsection, we need the following assumptions in addition to Assumption 1.

Assumption 3.

(a) $\delta \le 1/2$.

(b) *Let $b_7 := U_g^{1-\delta}/\nu_{\min}$. $\nabla^2 f$ is Lipschitz continuous on $\Omega + B(0, b_7)$, i.e., there exists $L_H > 0$ such that*

$$\|\nabla^2 f(x) - \nabla^2 f(y)\| \le L_H \|x - y\|, \quad \forall x, y \in \Omega + B(0, b_7).$$

Under Assumption 1, the inequality (3.2) holds. Moreover, there exists $f_{\min}$ such that

$$f(x_k) \ge f_{\min}, \quad \forall k \ge 0.$$

From Assumptions 1 and 3 (a), the inequality (3.3) holds. Therefore, we have

$$x_k + s d_k(\nu) \in \Omega + B(0, b_7), \quad \forall s \in [0, 1], \quad \forall \nu \in [0, \infty), \quad \forall k \ge 0. \tag{5.1}$$

It then follows from Assumption 3 (b) that

$$\|\nabla^2 f(x_k + s d_k(\nu)) - H_k\| \le L_H \|d_k(\nu)\|, \quad \forall s \in [0,1], \quad \forall \nu \in [0,\infty), \quad \forall k \ge 0. \tag{5.2}$$

Moreover, since $\Omega + B(0, b_7)$ is compact and f is twice continuously differentiable, there exists $U_H > 0$ such that

$$\|\nabla^2 f(x)\| \le U_H, \quad \forall x \in \Omega + B(0, b_7). \tag{5.3}$$

The next lemma indicates that the parameter ν_k^* is bounded above by some positive constant independent of k.

Lemma 5.1. *Suppose that Assumptions 1 and 3 hold. Then,*

$$\nu_k^* \le \nu_{\max},$$

where

$$\nu_{\max} := \max\left(\nu_0, \gamma_2 \sqrt{L_H U_g^{1-2\delta}}\right).$$

Proof. From the inequalities (4.10) of Lemma 4.5 and (5.2), we have

$$\begin{aligned} f(x_k + d_k(\nu)) - m_k(d_k(\nu), \nu) &\le \frac{1}{2}(L_H \|d_k(\nu)\| - \nu \|g_k\|^\delta)\|d_k(\nu)\|^2 \\ &\le \frac{1}{2}\left(\frac{L_H \|g_k\|^{1-\delta}}{\nu} - \nu\|g_k\|^\delta\right)\|d_k(\nu)\|^2 \\ &\le \frac{1}{2\nu}\left(L_H U_g^{1-2\delta} - \nu^2\right)\|g_k\|^\delta \|d_k(\nu)\|^2, \end{aligned} \tag{5.4}$$

where the first inequality follows from (5.2), the second inequality follows from Lemma 3.1, and the last inequality follows from (3.2). Now we suppose that $\nu \ge \sqrt{L_H U_g^{1-2\delta}}$. Then, we have

$$f(x_k + d_k(\nu)) \le m_k(d_k(\nu), \nu),$$

and hence

$$\rho_k(d_k(\nu), \nu) = \frac{f(x_k) - f(x_k + d_k(\nu))}{f(x_k) - m_k(d_k(\nu), \nu)} \ge 1.$$

Therefore, from the updating rule of $\bar{\nu}_{l_k}$, ν_k^* must satisfy

$$\nu_k^* \le \max\left(\nu_{k-1}^*, \left(\sqrt{L_H U_g^{1-2\delta}}\right)\gamma_2\right) \le \cdots \le \max\left(\nu_0, \left(\sqrt{L_H U_g^{1-2\delta}}\right)\gamma_2\right).$$

This completes the proof. □

From the above lemma, we show that the number l_k^* of inner iterations at the k-th iteration is bounded above by some positive constant independent of k.

Theorem 5.1. *Suppose that Assumptions 1 and 3 hold. Then, for all k,*

$$l_k^* \le l_{\max},$$

where

$$l_{\max} := \left\lceil \log_{\gamma_1}\left(\frac{\nu_{\max}}{\nu_{\min}}\right) + 1 \right\rceil.$$

Proof. We have from Lemma 5.1 that $\nu_{\min} \leq \bar{\nu}_{l_k} \leq \nu_{\max}$. From the updating rule of ν, we have $\bar{\nu}_{l_k+1} \geq \gamma_1 \bar{\nu}_{l_k}$, and hence we obtain the desired inequality. □

Next, we give a lower bound of the reduction of the model function.

Lemma 5.2. *Suppose that Assumptions 1 and 3 hold. Then,*

$$f(x_k) - m_k^* \geq p_1 \|g_k\|^2,$$

where

$$p_1 := \frac{1}{2((1+c)U_H + \nu_{\max} U_g^\delta)}.$$

Proof. It directly follows from (5.3), Lemma 5.1 and the inequality (3.7) of Lemma 3.4. □

By using this lemma, we give a lower bound of the reduction $f(x_k) - f(x_{k+1})$.

Lemma 5.3. *Suppose that Assumptions 1 and 3 hold. Then,*

$$f(x_k) - f(x_{k+1}) \geq \eta_1 p_1 \|g_k\|^2.$$

Proof. In a way similar to the proof of Lemma 3.5, we obtain the desired inequality. □

Now, we obtain the following global complexity bound J_g.

Theorem 5.2. *Suppose that Assumptions 1 and 3 hold. Let $\{x_k\}$ be a sequence generated by the proposed algorithm. Let J_g be the first iteration such that $\|g_{J_g}\| \leq \epsilon$. Then,*

$$J_g \leq \frac{f(x_0) - f_{\min}}{\eta_1 p_1} \epsilon^{-2}.$$

Proof. It follows from Lemma 5.3 that

$$f(x_0) - f_{\min} \geq f(x_0) - f(x_k) \geq \sum_{j=0}^{k-1} (f(x_j) - f(x_{j+1})) \geq \eta_1 p_1 \sum_{j=0}^{k-1} \|g_j\|^2 \geq k \eta_1 p_1 \left(\min_{0 \leq j \leq k-1} \|g_j\| \right)^2.$$

Then, we have

$$\min_{0 \leq j \leq k-1} \|g_j\| \leq \left(\frac{f(x_0) - f_{\min}}{k \eta_1 p_1} \right)^{\frac{1}{2}},$$

and hence

$$k \geq \frac{f(x_0) - f_{\min}}{\eta_1 p_1} \epsilon^{-2}$$

implies $\min_{0 \leq j \leq k-1} \|g_j\| \leq \epsilon$. This completes the proof. □

The above global complexity bound is same as that of the steepest descent method. On the other hand, it can be reduced under the following additional assumption on the minimum eigenvalue of H_k.

Assumption 4. *There exist positive constants $\bar{\delta}$ and κ_3 such that*

$$\Lambda_k \leq \kappa_3 \|g_k\|^{\bar{\delta}}, \quad \forall k \geq 0.$$

Before we show the reduced complexity bound, we give sufficient conditions for Assumption 4.

Proposition 5.1.

(a) *Suppose that f is convex. Then, Assumption 4 holds for any $\bar{\delta}$ and κ_3.*

(b) *Suppose that Assumption 3 holds. Suppose also that f is analytic and $\nabla^2 f(x) \succeq 0$ for any x such that $\nabla f(x) = 0$. Then, Assumption 4 holds.*

Proof. The statement (a) directly follows from the fact that $\Lambda_k = 0, \forall k \geq 0$ when f is convex.

Next, we show (b). Let $X_1 := \{x \in \mathbb{R}^n \mid \|\nabla f(x)\| = 0\}$ and $X_2 := \{x \in \mathbb{R}^n \mid \|\nabla f(x)\| = 0,\ \nabla^2 f(x) \succeq 0\}$. In a way similar to the proof of Lemma 4.2, we can show that there exists $c_1 > 0$ such that

$$\Lambda_k \leq c_1 \mathrm{dist}(x_k, X_2),$$

when Assumption 3 holds. Moreover, it is shown in [15] that there exist $c_2 > 0$ and $\bar{\delta} > 0$ such that

$$\mathrm{dist}(x, X_1) \leq c_2 \|\nabla f(x)\|^{\bar{\delta}}, \quad \forall x \in \Omega,$$

when f is analytic. It then follows from $X_1 = X_2$ that

$$\Lambda_k \leq c_1 c_2 \|g_k\|^{\bar{\delta}},$$

and hence Assumption 4 holds. □

Remark 5.1. *If f is quasi-convex, then $\nabla^2 f(x) \succeq 0$ for any x such that $\nabla f(x) = 0$ [5]. Thus, an analytic quasi-convex function satisfies the assumptions of Proposition 5.1 (b).*

Now we show that the global complexity bound J_g is reduced to $O(\epsilon^{-\frac{2+\delta}{1+\delta}})$ under Assumption 4. To this end, we need the following assumption on δ.

Assumption 5. $\delta \leq \bar{\delta}$.

First, we give the relationship between $\|d_k^*\|$ and $\|g_k\|$.

Lemma 5.4. *Suppose that Assumptions 1 and 3 hold. Then,*

$$\|d_k^*\| \geq \frac{1}{(1+c)U_H + \nu_{\max} U_g^{\delta}} \|g_k\|.$$

Proof. From the definition (2.5) of d_k^*, we have

$$g_k = (H_k + c\Lambda_k I + \nu_k^* \|g_k\|^{\delta} I) d_k^*. \tag{5.5}$$

It then follows from (3.2), (5.3) and Lemma 5.1 that

$$\begin{aligned}\|g_k\| &= \|(H_k + c\Lambda_k I + \nu_k^* \|g_k\|^{\delta} I) d_k^*\| \\ &\leq \|H_k + c\Lambda_k I + \nu_k^* \|g_k\|^{\delta} I\| \cdot \|d_k^*\| \\ &\leq (U_H + cU_H + \nu_{\max} U_g^{\delta}) \|d_k^*\|.\end{aligned}$$

This completes the proof. □

Next, we show the following key lemma for the desired global complexity bound J_g.

Lemma 5.5. *Suppose that Assumptions 1, 3, 4 and 5 hold. Then,*

$$\|g_{k+1}\| \leq \kappa_4 \max\left(\|g_k\|^{\delta} \|d_k^*\|, \|d_k^*\|^2\right),$$

where

$$\kappa_4 := c\kappa_3 U_g^{\bar{\delta}-\delta} + \nu_{\max} + \frac{1}{2} L_H.$$

Proof. From (5.1) and Assumption 3 (b), we have

$$\|H_k d_k^* - (g_{k+1} - g_k)\| \le \frac{L_H}{2}\|d_k^*\|^2,$$

and hence

$$\|g_{k+1}\| \le \|H_k d_k^* + g_k\| + \frac{L_H}{2}\|d_k^*\|^2. \tag{5.6}$$

Moreover, we have from the definition (2.5) of d_k^* that

$$H_k d_k^* + g_k = -c\Lambda_k d_k^* - \nu_k^* \|g_k\|^\delta d_k^*.$$

It then follows from (5.6) that

$$\begin{aligned}
\|g_{k+1}\| &\le \|H_k d_k^* + g_k\| + \frac{L_H}{2}\|d_k^*\|^2 \\
&\le c\Lambda_k \|d_k^*\| + \nu_k^* \|g_k\|^\delta \|d_k^*\| + \frac{L_H}{2}\|d_k^*\|^2 \\
&\le c\kappa_3 \|g_k\|^{\bar\delta} \|d_k^*\| + \nu_{\max} \|g_k\|^\delta \|d_k^*\| + \frac{L_H}{2}\|d_k^*\|^2 \\
&= c\kappa_3 \|g_k\|^{\bar\delta-\delta} \|g_k\|^\delta \|d_k^*\| + \nu_{\max} \|g_k\|^\delta \|d_k^*\| + \frac{L_H}{2}\|d_k^*\|^2 \\
&\le c\kappa_3 U_g^{\bar\delta-\delta} \|g_k\|^\delta \|d_k^*\| + \nu_{\max} \|g_k\|^\delta \|d_k^*\| + \frac{L_H}{2}\|d_k^*\|^2 \\
&\le \left(c\kappa_3 U_g^{\bar\delta-\delta} + \nu_{\max} + \frac{L_H}{2} \right) \max\left(\|g_k\|^\delta \|d_k^*\|, \|d_k^*\|^2 \right),
\end{aligned}$$

where the third inequality follows from Assumption 4 and Lemma 5.1, and the fourth inequality follows from (3.2). □

By using Lemmas 5.4 and 5.5, we give a lower bound of the reduction of the model function.

Lemma 5.6. *Suppose that Assumptions 1, 3, 4 and 5 hold. Then,*

$$f(x_k) - m_k^* \ge p_2 \|g_{k+1}\|^{\frac{2+\bar\delta}{1+\delta}},$$

where

$$p_2 := \min\left(\frac{\nu_{\min}}{2\kappa_4^2}, \frac{\nu_{\min}}{2\kappa_4((1+c)U_H + \nu_{\max} U_g^\delta)}, \frac{\nu_{\min}^{\frac{1}{1-\delta}}}{2\kappa_4^{\frac{2-\delta}{2(1-\delta)}} U_g^{\frac{2-3\delta-\delta^2}{2(1+\delta)(1-\delta)}}} \right).$$

Proof. We have from the equality (5.5) of Lemma 5.4 and $H_k + c\Lambda_k I \succeq 0$ that

$$\begin{aligned}
f(x_k) - m_k^*(d_k^*) &= -g_k^T d_k^* - \frac{1}{2} {d_k^*}^T (H_k + c\Lambda_k I + \nu_k^* \|g_k\|^\delta I) d_k^* \\
&= \frac{1}{2} {d_k^*}^T (H_k + c\Lambda_k I + \nu_k^* \|g_k\|^\delta I) d_k^* \qquad (5.7) \\
&\ge \frac{1}{2} \nu_k^* \|g_k\|^\delta \|d_k^*\|^2 \\
&\ge \frac{1}{2} \nu_{\min} \|g_k\|^\delta \|d_k^*\|^2. \qquad (5.8)
\end{aligned}$$

In what follows, we consider two cases: (i) $\|d_k^*\|^2 \le \|g_k\|^\delta \|d_k^*\|$ and (ii) $\|d_k^*\|^2 \ge \|g_k\|^\delta \|d_k^*\|$.

Case (i): In this case, we have from Lemma 5.5 that

$$\|g_{k+1}\| \leq \kappa_4 \|g_k\|^\delta \|d_k^*\|, \tag{5.9}$$

and hence

$$\|d_k^*\| \geq \frac{1}{\kappa_4}\|g_k\|^{-\delta}\|g_{k+1}\|.$$

It then follows from (5.8) that

$$\begin{aligned} f(x_k) - m_k^* &\geq \frac{1}{2}\nu_{\min}\|g_k\|^\delta \left(\frac{1}{\kappa_4}\|g_k\|^{-\delta}\|g_{k+1}\|\right)^2 \\ &= \frac{\nu_{\min}}{2\kappa_4^2}\|g_k\|^{-\delta}\|g_{k+1}\|^2, \end{aligned} \tag{5.10}$$

where the last inequality follows from Lemma 5.1.

On the other hand, we have from (5.8), (5.9) and Lemma 5.4 that

$$\begin{aligned} f(x_k) - m_k^* &\geq \frac{\nu_{\min}}{2\kappa_4}\|d_k^*\| \cdot \|g_{k+1}\| \\ &\geq \frac{\nu_{\min}}{2\kappa_4((1+c)U_H + \nu_{\max}U_g^\delta)}\|g_k\| \cdot \|g_{k+1}\|. \end{aligned} \tag{5.11}$$

Now we consider two cases: (a) $\|g_{k+1}\| \geq \|g_k\|^\alpha$ and (b) $\|g_{k+1}\| \leq \|g_k\|^\alpha$, where α is an arbitrary positive constant.

Case (a): This case implies that

$$\|g_k\|^{-\delta} \geq \|g_{k+1}\|^{-\frac{\delta}{\alpha}}.$$

It then follows from (5.10) that

$$f(x_k) - m_k^* \geq \frac{\nu_{\min}}{2\kappa_4^2}\|g_{k+1}\|^{2-\frac{\delta}{\alpha}}. \tag{5.12}$$

Case (b): In this case, we have

$$\|g_k\| \geq \|g_{k+1}\|^{\frac{1}{\alpha}}.$$

It then follows from (5.11) that

$$f(x_k) - m_k^* \geq \frac{\nu_{\min}}{2\kappa_4((1+c)U_H + \nu_{\max}U_g^\delta)}\|g_{k+1}\|^{1+\frac{1}{\alpha}}. \tag{5.13}$$

Since α is an arbitrary positive constant, we choose $\alpha := 1+\delta$, which minimizes $\max(2-\frac{\delta}{\alpha}, 1+\frac{1}{\alpha})$. Then, we have

$$2 - \frac{\delta}{\alpha} = 1 + \frac{1}{\alpha} = \frac{2+\delta}{1+\delta}.$$

It then follows from (5.12) and (5.13) that

$$f(x_k) - m_k^* \geq \min\left(\frac{\nu_{\min}}{2\kappa_4^2}, \frac{\nu_{\min}}{2\kappa_4((1+c)U_H + \nu_{\max}U_g^\delta)}\right)\|g_{k+1}\|^{\frac{2+\delta}{1+\delta}}. \tag{5.14}$$

Case (ii): In this case, we have from Lemma 5.5 that

$$\|g_{k+1}\| \le \kappa_4 \|d_k^*\|^2. \tag{5.15}$$

It then follows from Lemma 3.1 that

$$\|g_{k+1}\| \le \kappa_4 \|d_k^*\|^2 \le \frac{\kappa_4}{(\nu_k^*)^2}\|g_k\|^{2(1-\delta)} \le \frac{\kappa_4}{\nu_{\min}^2}\|g_k\|^{2(1-\delta)}.$$

Thus we have

$$\|g_k\|^\delta \ge \left(\frac{\nu_{\min}^2}{\kappa_4}\|g_{k+1}\|\right)^{\frac{\delta}{2(1-\delta)}}. \tag{5.16}$$

From (5.8), (5.15) and (5.16), we have

$$\begin{aligned} f(x_k) - m_k^* &\ge \frac{\nu_{\min}}{2\kappa_4}\left(\frac{\nu_{\min}^2}{\kappa_4}\right)^{\frac{\delta}{2(1-\delta)}} \|g_{k+1}\|^{1+\frac{\delta}{2(1-\delta)}} \\ &= \frac{\nu_{\min}^{\frac{1}{1-\delta}}}{2\kappa_4^{\frac{2-\delta}{2(1-\delta)}}}\|g_{k+1}\|^{\frac{2+\delta}{1+\delta} - \frac{2-3\delta-\delta^2}{2(1+\delta)(1-\delta)}}. \end{aligned}$$

Since $\delta \in (0, \frac{1}{2}]$, we have

$$\frac{2-3\delta-\delta^2}{2(1+\delta)(1-\delta)} \ge 0.$$

Moreover, from (3.2), we have

$$\|g_{k+1}\| \le U_g.$$

Thus we obtain

$$f(x_k) - m_k^* \ge \frac{\nu_{\min}^{\frac{1}{1-\delta}}}{2\kappa_4^{\frac{2-\delta}{2(1-\delta)}} U_g^{\frac{2-3\delta-\delta^2}{2(1+\delta)(1-\delta)}}}\|g_{k+1}\|^{\frac{2+\delta}{1+\delta}}. \tag{5.17}$$

Therefore, we obtain from (5.14) and (5.17) that

$$f(x_k) - m_k^* \ge \min\left(\frac{\nu_{\min}}{2\kappa_4^2}, \frac{\nu_{\min}}{2\kappa_4((1+c)U_H + \nu_{\max}U_g^\delta)}, \frac{\nu_{\min}^{\frac{1}{1-\delta}}}{2\kappa_4^{\frac{2-\delta}{2(1-\delta)}} U_g^{\frac{2-3\delta-\delta^2}{2(1+\delta)(1-\delta)}}}\right)\|g_{k+1}\|^{\frac{2+\delta}{1+\delta}}.$$

This completes the proof. □

By using the above lemma, we give a lower bound of the reduction $f(x_k) - f(x_{k+1})$.

Lemma 5.7. *Suppose that Assumptions 1, 3, 4 and 5 hold. Then,*

$$f(x_k) - f(x_{k+1}) \ge \eta_1 p_2 \|g_{k+1}\|^{\frac{2+\delta}{1+\delta}}.$$

Proof. In a way similar to the proof of Lemma 3.5, we obtain the desired inequality. □

Finally, by using this lemma, we obtain the desired global complexity bound J_g.

Theorem 5.3. *Suppose that Assumptions 1, 3, 4 and 5 hold. Let $\{x_k\}$ be a sequence generated by the proposed algorithm. Let J_g be the first iteration such that $\|g_{J_g}\| \le \epsilon$. Then,*

$$J_g \le \frac{f(x_0) - f_{\min}}{\eta_1 p_2}\epsilon^{-\frac{2+\delta}{1+\delta}} + 1.$$

Proof. It directly follows from the proof of Theorem 5.2. □

Remark 5.2. *Under Assumption 4, the global complexity bound $O(\epsilon^{-\frac{2+\delta}{1+\delta}})$ of the proposed algorithm is better than $O(\epsilon^{-2})$ of the steepest descent method.*

5.2 Convex case

In this subsection, we consider the case where f is convex. We need the following assumptions instead of Assumption 3.

Assumption 6.

(a) $\delta \leq 1/2$.

(b) $\nabla^2 f$ *is Lipschitz continuous on* $\Omega + B(0, b_7)$ *with modulus* L_H.

(c) f *is convex.*

From Proposition 5.1 (a), Assumption 4 holds for any $\bar{\delta}$. Moreover, under Assumptions 1 and 6, Lemma 5.1, Theorems 5.1 and 5.3 hold. Thus we can directly get the following global complexity bound J_g.

Theorem 5.4. *Suppose that Assumptions 1 and 6 hold. Let* $\{x_k\}$ *be a sequence generated by the proposed algorithm. Let* J_g *be the first iteration such that* $\|g_{J_g}\| \leq \epsilon$*. Then,*

$$J_g \leq \frac{f(x_0) - f_{\min}}{\eta_1 p_2} \epsilon^{-\frac{2+\delta}{1+\delta}} + 1.$$

In particular, if $\delta = 1/2$*, then*

$$J_g \leq \frac{f(x_0) - f_{\min}}{\eta_1 p_2} \epsilon^{-\frac{5}{3}} + 1.$$

In what follows, we discuss the global complexity bound J_f. From Assumption 1 and Theorem 3.1, there exists a solution x^* of (1.1). Moreover, there exists $U_x > 0$ such that

$$\|x_k - x^*\| \leq U_x, \quad \forall k \geq 0. \tag{5.18}$$

First, we give the following technical lemma.

Lemma 5.8. *Let* β*,* γ *and* u *be positive parameters such that* $0 < \beta \leq 1$*,* $\gamma \geq 0$ *and* $u > 0$*. Then,*

$$(1 + \gamma\alpha)^\beta \geq 1 + \frac{(1 + \gamma u)^\beta - 1}{u} \alpha, \quad \forall \alpha \in [0, u]. \tag{5.19}$$

Proof. Let $h(t) := (1 + \gamma t)^\beta$. Since $0 < \beta \leq 1$ and $\gamma \geq 0$, we have

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} h(t) = -\frac{\beta(1-\beta)\gamma^2}{(1+\gamma t)^{2-\beta}} \leq 0, \quad \forall t \in [0, \infty)$$

Therefore, $h(t)$ is concave on $[0, u]$. Let $\alpha \in [0, u]$. Then, $\alpha/u \in [0, 1]$. It then follows from the concavity of h that

$$\begin{aligned} h(\alpha) &= h\left(\frac{\alpha}{u} u + \left(1 - \frac{\alpha}{u}\right) 0\right) \\ &\geq \frac{\alpha}{u} h(u) + \left(1 - \frac{\alpha}{u}\right) h(0) \\ &= 1 + \frac{(1 + \gamma u)^\beta - 1}{u} \alpha, \end{aligned}$$

which is the desired inequality. $\square$

By using Lemma 5.8, we obtain the global complexity bound J_f. Note that the proof technique is similar to [13, Theorem 6] where the global complexity bound J_f of the cubic regularization of Newton's method is given.

Theorem 5.5. *Suppose that Assumptions 1 and 6 hold. Let $\{x_k\}$ be a sequence generated by the proposed algorithm. Let J_f be the first iteration such that $f(x_{J_f}) - f(x^*) \le \epsilon$. Then,*

$$J_f = O\left(\epsilon^{-\frac{1}{1+\delta}}\right).$$

In particular, if $\delta = 1/2$, then

$$J_f = O\left(\epsilon^{-\frac{2}{3}}\right).$$

Proof. Since f is convex, we have from (5.18) that

$$f(x_{k+1}) - f(x^*) \le g_{k+1}^T(x_{k+1} - x^*) \le U_x\|g_{k+1}\|.$$

It then follows from Lemma 5.7 that

$$f(x_k) - f(x_{k+1}) \ge \frac{\eta_1 p_2}{U_x^{\frac{2+\delta}{1+\delta}}} \left(f(x_{k+1}) - f(x^*)\right)^{\frac{2+\delta}{1+\delta}}.$$

Denoting $\alpha_k := f(x_k) - f(x^*)$, $\beta := 1/(1+\delta)$ and $\gamma := \eta_1 p_2 / U_x^{\frac{2+\delta}{1+\delta}}$, we obtain

$$\alpha_k \ge \alpha_{k+1} + \gamma\alpha_{k+1}^{1+\beta}.$$

Then, we have

$$\begin{aligned}
\frac{1}{\alpha_{k+1}^\beta} - \frac{1}{\alpha_k^\beta} &\ge \frac{1}{\alpha_{k+1}^\beta} - \frac{1}{(\alpha_{k+1} + \gamma\alpha_{k+1}^{1+\beta})^\beta} \\
&= \frac{\alpha_{k+1}^\beta(1+\gamma\alpha_{k+1}^\beta)^\beta - \alpha_{k+1}^\beta}{\alpha_{k+1}^{2\beta}(1+\gamma\alpha_{k+1}^\beta)^\beta} \\
&= \frac{(1+\gamma\alpha_{k+1}^\beta)^\beta - 1}{\alpha_{k+1}^\beta(1+\gamma\alpha_{k+1}^\beta)^\beta}.
\end{aligned} \tag{5.20}$$

Since $\alpha_{k+1}^\beta \le \alpha_0^\beta$ and $\beta \le 1$, substituting $u := \alpha_0^\beta$ and $\alpha := \alpha_{k+1}^\beta$ into (5.19) of Lemma 5.8 yields

$$1 + \frac{(1+\gamma\alpha_0^\beta)^\beta - 1}{\alpha_0^\beta}\alpha_{k+1}^\beta \le (1+\gamma\alpha_{k+1}^\beta)^\beta \le (1+\gamma\alpha_0^\beta)^\beta.$$

It then follows from (5.20) that

$$\begin{aligned}
\frac{1}{\alpha_{k+1}^\beta} &\ge \frac{1}{\alpha_k^\beta} + \frac{(1+\gamma\alpha_0^\beta)^\beta - 1}{\alpha_0^\beta(1+\gamma\alpha_0^\beta)^\beta} \\
&\ge \frac{1}{\alpha_0^\beta} + \frac{(1+\gamma\alpha_0^\beta)^\beta - 1}{\alpha_0^\beta(1+\gamma\alpha_0^\beta)^\beta}(k+1) \\
&= \frac{(1+\gamma\alpha_0^\beta)^\beta + \left((1+\gamma\alpha_0^\beta)^\beta - 1\right)(k+1)}{\alpha_0^\beta(1+\gamma\alpha_0^\beta)^\beta},
\end{aligned}$$

and hence

$$\alpha_k \le \left(\frac{\alpha_0^\beta(1+\gamma\alpha_0^\beta)^\beta}{(1+\gamma\alpha_0^\beta)^\beta + \left((1+\gamma\alpha_0^\beta)^\beta - 1\right)k}\right)^{\frac{1}{\beta}}.$$

Therefore, $f(x_k) - f(x^*) = \alpha_k \le \epsilon$, provided that

$$k \ge \frac{\alpha_0^\beta(1+\gamma\alpha_0^\beta)^\beta\epsilon^{-\beta} - (1+\gamma\alpha_0^\beta)^\beta}{(1+\gamma\alpha_0^\beta)^\beta - 1}.$$

This completes the proof. □

Remark 5.3. *The global complexity bounds* $J_g = O(\epsilon^{-\frac{2+\delta}{1+\delta}})$ *and* $J_f = O(\epsilon^{-\frac{1}{1+\delta}})$ *become better as we take a larger* δ*. However, we need* $\delta \le 1/2$ *for Lemma 5.1 and Theorem 5.1. Thus, the upper bounds of* J_g *and* J_f *are* $O(\epsilon^{-\frac{5}{3}})$ *and* $O(\epsilon^{-\frac{2}{3}})$*, respectively.*

5.3 Strongly convex case

In this subsection, we show that the global complexity bound of the proposed algorithm is $J_g = O(\epsilon^{-\frac{2}{1+\delta}})$ when f is strongly convex. Moreover, we show that a sequence $\{f(x_k) - f(x^*)\}$ globally linearly converges to 0 as well as the steepest descent method [11] and the cubic regularization of Newton's method [13].

From Remarks 4.2 and 5.3, we expect that the proposed algorithm behaves well as we take a larger δ. Therefore, it is worth considering the case where $\delta > 1/2$. When $\delta > 1/2$, Lemma 5.1 and Theorem 5.1 do not always hold. However, when f is strongly convex, we can relax the assumption $\delta \le 1/2$ to $\delta \le 1$, and prove properties similar to Lemma 5.1 and Theorem 5.1.

Now, we formally state assumptions used in this subsection.

Assumption 7.

(a) $\delta \le 1$.

(b) $\nabla^2 f$ *is Lipschitz continuous on* $\Omega + B(0, b_7)$ *with modulus* L_H.

(c) f *is strongly convex with modulus* $\sigma > 0$.

Under Assumption 7 (c), $\lambda_{\min}(\nabla^2 f(x)) \ge \sigma$ for all $x \in \mathbb{R}^n$ and $\Lambda_k = 0$ for all $k \ge 0$.

First, we give an upper bound of $\|d_k(\nu)\|$.

Lemma 5.9. *Suppose that* $\|g_k\| \ne 0$*. Suppose also that Assumption 7 holds. Then,*

$$\|d_k(\nu)\| \le \frac{1}{\sigma}\|g_k\|, \quad \forall \nu \in [\nu_{\min}, \infty).$$

Proof. It directly follows from the inequality (3.1) of Lemma 3.1 and $\lambda_{\min}(H_k + c\Lambda_k I + \nu\|g_k\|^\delta I) \ge \sigma$. □

From the above lemma, we show that the regularized parameter ν_k^* is bounded above by some positive constant independent of k.

Lemma 5.10. *Suppose that Assumptions 1 and 7 hold. Then,*

$$\nu \le \hat{\nu}_{\max},$$

where

$$\hat{\nu}_{\max} := \max\left(\nu_0, \frac{\gamma_2 L_H U_g^{1-\delta}}{\sigma}\right).$$

Proof. We have from (5.4) of Lemma 5.1 that

$$\begin{aligned} f(x_k + d_k(\nu)) - m_k(d_k(\nu), \nu) &\le \frac{1}{2}(L_H\|d_k(\nu)\| - \nu\|g_k\|^\delta)\|d_k(\nu)\|^2 \\ &\le \frac{1}{2}\left(\frac{L_H\|g_k\|}{\sigma} - \nu\|g_k\|^\delta\right)\|d_k(\nu)\|^2 \\ &\le \frac{1}{2}\left(\frac{L_H U_g^{1-\delta}}{\sigma} - \nu\right)\|g_k\|^\delta\|d_k(\nu)\|^2, \end{aligned}$$

where the second inequality follows from Lemma 5.9, and the third inequality follows from (3.2). Now we suppose that $\nu \ge L_H U_g^{1-\delta}/\sigma$. Then, we have

$$f(x_k + d_k(\nu)) \le m_k(d_k(\nu), \nu),$$

and hence

$$\rho_k(d_k(\nu), \nu) = \frac{f(x_k) - f(x_k + d_k(\nu))}{f(x_k) - m_k(d_k(\nu), \nu)} \geq 1.$$

Therefore, from the updating rule of $\bar{\nu}_{l_k}$, ν_k^* must satisfy

$$\nu_k^* \leq \max\left(\nu_{k-1}^*, \left(\frac{L_H U_g^{1-\delta}}{\sigma}\right)\gamma_2\right) \leq \cdots \leq \max\left(\nu_0, \left(\frac{L_H U_g^{1-\delta}}{\sigma}\right)\gamma_2\right).$$

This completes the proof. □

From the above lemma, we show that the number of inner iteration l_k^* at k-th iteration is bounded above by some positive constant independent of k.

Theorem 5.6. *Suppose that Assumptions 1 and 7 hold. Then,*

$$l_k \leq \hat{l}_{\max},$$

where

$$\hat{l}_{\max} := \left\lceil \log_{\gamma_1}\left(\frac{\hat{\nu}_{\max}}{\nu_{\min}}\right) + 1 \right\rceil.$$

Proof. In a way similar to the proof of Theorem 5.1, we obtain the desired inequality. □

By using Lemmas 5.4 and 5.5, we give a lower bound of the reduction of the model function.

Lemma 5.11. *Suppose that Assumptions 1 and 7 hold. Then,*

$$f(x_k) - m_k^* \geq p_3 \|g_{k+1}\|^{\frac{2}{1+\delta}},$$

where

$$p_3 := \min\left(\frac{\sigma}{2((1+c)U_H + \hat{\nu}_{\max}U_g^\delta)^2}\left(\frac{\sigma}{\kappa_4}\right)^{\frac{2}{1+\delta}}, \frac{\sigma}{2\kappa_4 U_g^{\frac{1-\delta}{1+\delta}}}\right).$$

Proof. We have from the equality (5.7) of Lemma 5.6 and $\lambda_{\min}(H_k) \geq \sigma$ that

$$f(x_k) - m_k^* \geq \frac{1}{2}\sigma\|d_k^*\|^2. \tag{5.21}$$

From Lemma 5.5, $\|g_{k+1}\| \leq \kappa_4 \max(\|g_k\|^\delta \|d_k^*\|, \|d_k^*\|^2)$ holds. Now we consider two cases: (i) $\|d_k^*\|^2 \leq \|g_k\|^\delta \|d_k^*\|$ and (ii) $\|d_k^*\|^2 \geq \|g_k\|^\delta \|d_k^*\|$.

Case (i): In this case, we have from Lemma 5.5 that

$$\|g_{k+1}\| \leq \kappa_4 \|g_k\|^\delta \|d_k^*\| \leq \frac{\kappa_4}{\sigma}\|g_k\|^{1+\delta},$$

where the second inequality follows from Lemma 5.9, and the last inequality follows from Lemma 5.10. Thus we have

$$\|g_k\| \geq \left(\frac{\sigma}{\kappa_4}\|g_{k+1}\|\right)^{\frac{1}{1+\delta}}.$$

From Lemma 5.4 and Lemma 5.10, we have

$$\|d_k^*\| \geq \frac{1}{(1+c)U_H + \hat{\nu}_{\max}U_g^\delta}\|g_k\|.$$

It then follows from (5.21) that

$$f(x_k) - m_k^* \geq \frac{\sigma}{2((1+c)U_H + \hat{\nu}_{\max} U_g^\delta)^2} \|g_k\|^2$$
$$\geq \frac{\sigma}{2((1+c)U_H + \hat{\nu}_{\max} U_g^\delta)^2} \left(\frac{\sigma}{\kappa_4}\right)^{\frac{2}{1+\delta}} \|g_{k+1}\|^{\frac{2}{1+\delta}}. \tag{5.22}$$

Case (ii): In this case, we have from Lemma 5.5 that

$$\|g_{k+1}\| \leq \kappa_4 \|d_k^*\|^2.$$

It then follows from (5.21) that

$$f(x_k) - m_k^* \geq \frac{\sigma}{2\kappa_4}\|g_{k+1}\| \geq \frac{\sigma}{2\kappa_4}\|g_{k+1}\|^{\frac{2}{1+\delta} - \frac{1-\delta}{1+\delta}} \geq \frac{\sigma}{2\kappa_4 U_g^{\frac{1-\delta}{1+\delta}}} \|g_{k+1}\|^{\frac{2}{1+\delta}}, \tag{5.23}$$

where the last inequality follows from (3.2).

Therefore, we obtain from (5.22) and (5.23) that

$$f(x_k) - m_k^* \geq \min\left(\frac{\sigma}{2((1+c)U_H + \hat{\nu}_{\max} U_g^\delta)^2}\left(\frac{\sigma}{\kappa_4}\right)^{\frac{2}{1+\delta}}, \frac{\sigma}{2\kappa_4 U_g^{\frac{1-\delta}{1+\delta}}}\right) \|g_{k+1}\|^2.$$

This completes the proof. □

By using the above lemma, we give a lower bound of the reduction $f(x_k) - f(x_{k+1})$.

Lemma 5.12. *Suppose that Assumptions 1 and 7 hold. Then,*

$$f(x_k) - f(x_{k+1}) \geq \eta_1 p_3 \|g_{k+1}\|^{\frac{2}{1+\delta}}$$

Proof. In a way similar to the proof of Lemma 3.5, we obtain the desired inequality. □

Now, by using Lemma 5.12, we obtain the global complexity bound J_g in the case where f is strongly convex.

Theorem 5.7. *Suppose that Assumptions 1 and 7 hold. Let $\{x_k\}$ be a sequence generated by the proposed algorithm. Let J_g be the first iteration such that $\|g_{J_g}\| \leq \epsilon$. Then,*

$$J_g \leq \frac{f(x_0) - f_{\min}}{\eta_1 p_3} \epsilon^{-\frac{2}{1+\delta}} + 1.$$

In particular, if $\delta = 1$, then

$$J_g \leq \frac{f(x_0) - f_{\min}}{\eta_1 p_3} \epsilon^{-1} + 1.$$

Proof. It directly follows from the proof of Theorem 5.2. □

By using a technique similar to [13, Theorem 7], we can show that $\{f(x_k) - f(x^*)\}$ converges to 0 linearly.

Theorem 5.8. *Suppose that Assumptions 1 and 7 hold. Let $\{x_k\}$ be a sequence generated by the proposed algorithm. Then, $\{f(x_k) - f(x^*)\}$ globally linearly converges to 0. Thus, the first iteration J_f such that $f(x_{J_f}) - f(x^*) \leq \epsilon$ satisfies*

$$J_f = O\left(\log \epsilon^{-1}\right).$$

Proof. Since f is strongly convex, we have

$$f(x_{k+1}) - f(x^*) \le g_{k+1}^T(x_{k+1} - x^*) \le \|g_{k+1}\| \cdot \|x_{k+1} - x^*\| \le \frac{1}{\sigma}\|g_{k+1}\|^2.$$

It then follows from Lemma 5.12 that

$$f(x_k) - f(x_{k+1}) \ge \eta_1 p_3 \sigma^{\frac{1}{1+\delta}} \left(f(x_{k+1}) - f(x^*)\right)^{\frac{1}{1+\delta}}.$$

Denoting $\alpha_k := f(x_k) - f(x^*)$ and $\gamma := \eta_1 p_3 \sigma^{\frac{1}{1+\delta}}$, we obtain

$$\alpha_k \ge \alpha_{k+1} + \gamma \alpha_{k+1}^{\frac{1}{1+\delta}}.$$

Then, we have from $\alpha_{k+1} \le \alpha_0$ that

$$\alpha_{k+1} \le \frac{1}{1 + \gamma \alpha_k^{-\frac{\delta}{1+\delta}}} \alpha_k \le \frac{1}{1 + \gamma \alpha_0^{-\frac{\delta}{1+\delta}}} \alpha_k. \tag{5.24}$$

Therefore, $f(x_k) - f(x^*)$ globally linearly converges to 0.

Next, we show the second part of the theorem. From (5.24), we have

$$\alpha_k \le \left(\frac{1}{1 + \gamma \alpha_0^{-\frac{\delta}{1+\delta}}}\right)^k \alpha_0,$$

and hence if

$$k \ge \frac{1}{1 + \gamma \alpha_0^{-\frac{\delta}{1+\delta}}} \log \frac{\alpha_0}{\epsilon},$$

then $\alpha_k \le \epsilon$. This completes the proof. □

6 Numerical results

In this section, we report some results on the following numerical experiments for the proposed algorithm.

1. Examination of the effects of the updating rules of the regularized parameter;
2. Comparison of the proposed algorithm and the existing Newton-type methods.

In each experiment, benchmark problems were chosen from CUTEr [7]. All algorithms were coded in MATLAB 7.4, and run on a machine with 3.2GHz Pentium 4 CPU and 3.2GB memory. We used an initial point x_0 given by CUTEr, and set the termination criterion as $\|g_k\| \le 10^{-5}$. If the number of inner iterations at the k-th iteration or the number of outer iterations exceeds 10^4, then we terminated all methods as failing.

We consider the following two updating rules of the regularized parameter μ_k.

(A) $\mu_k = c\Lambda_k + \nu_k \|g_k\|^\delta$;

(B) $\mu_k = c\Lambda_k + \nu_k \min(1, \|g_k\|^\delta)$.

The updating rule (B) prevents $\|d_k(\bar{\nu}_{l_k})\|$ from becoming too small when $\|g_k\|^\delta$ is large. Note that the convergence properties given in Sections 3 – 5 still hold even if we replace the above updating rule (A) with (B). We updated ν_k in Steps 2 and 3 as follows.

$$\begin{aligned} \rho_k(d_k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) < \eta_1 &\Rightarrow \bar{\nu}_{l_k+1} = \gamma_b \bar{\nu}_{l_k}, \\ \eta_2 > \rho_k(d_k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) \ge \eta_1 &\Rightarrow \nu_{k+1} = \bar{\nu}_{l_k}, \\ \rho_k(d_k(\bar{\nu}_{l_k}), \bar{\nu}_{l_k}) \ge \eta_2 &\Rightarrow \nu_{k+1} = \max\left(\nu_{\min}, \gamma_a \bar{\nu}_{l_k}\right), \end{aligned}$$

where γ_a and γ_b are positive parameters such that $\gamma_a < 1$ and $\gamma_b > 1$. In all numerical experiments, except for γ_a, γ_b and δ, the parameters of the proposed algorithm are chosen as follows.

$$\nu_0 = 1,\ \nu_{\min} = 10^{-5},\ c = 2,\ \eta_1 = 0.01,\ \eta_2 = 0.8.$$

In Subsections 6.1 and 6.2, we will compare algorithms by using the distribution function proposed in [6]. We denote a set of solvers as $\mathcal{S}$, and a set of problems that can be solved by all methods in $\mathcal{S}$ as $\mathcal{P}_{\mathcal{S}}$. We also denote a measure for evaluation required to solve a problem p by a solver s as $t_{p,s}$, and the best $t_{p,s}$ for each p as t_p^*, i.e., $t_p^* := \min\{t_{p,s} \mid a \in \mathcal{S}\}$. The distribution function $F_s^{\mathcal{S}}(\tau)$ for a method s is defined by

$$F_s^{\mathcal{S}}(\tau) = \frac{|\{p \in \mathcal{P}_{\mathcal{S}} \mid t_{p,s} \le \tau t_p^*\}|}{|\mathcal{P}_{\mathcal{S}}|}, \quad \tau \ge 1.$$

The algorithm whose $F_s^{\mathcal{S}}(\tau)$ is close to 1 is considered to be superior to the other algorithms in $\mathcal{S}$.

6.1 Influences of the updating rule of the regularized parameter

First, we investigate influences of the parameter δ and the updating rules (A) and (B). We set γ_a and γ_b as $\gamma_a = 0.5$ and $\gamma_b = 2$, respectively.

Table 1 shows the number of the function evaluations for $\delta = 1/2, 1, 2$ and the updating rules (A) and (B). The symbol "−" in the table means that the number of inner or outer iterations of the proposed algorithm exceeds 10^4.

Figure 2 shows the distribution functions for the proposed algorithm with various δ and the updating rules (A) and (B) in terms of the number of the function evaluations. Figure 2 shows that for $\delta = 0.5$, the updating rule (A) is almost same as the updating rule (B). On the other hand, for $\delta = 1$ and 2, the updating rule (B) is better than the updating rule (A). The reason is that when $\|g_k\|^\delta$ is large, $\|d_k(\bar{\nu}_{l_k})\|$ becomes too small, and a sequence of the proposed algorithm changes only slightly. Moreover, from the same reason, the number of the function evaluations tends to become large as δ become large for the updating rule (A). Finally, for the updating rule (B), the proposed algorithm does not have much difference among $\delta = 0.5, 1, 2$. From the above fact, the proposed algorithm has good numerical performance when we use the updating rule (B).

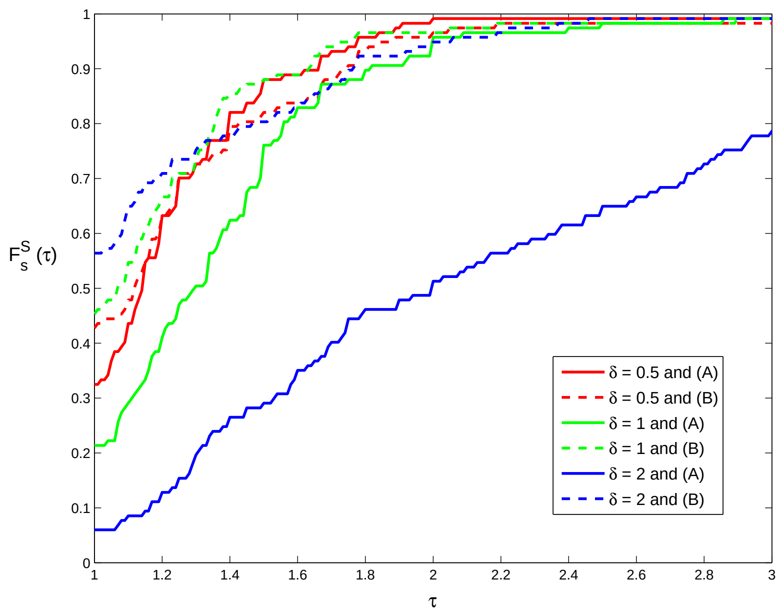

Figure 2: Comparison of δ and the updating rules (A) and (B)

Next, we examine the influences of (γ_a, γ_b). We set $\delta = 1$ and used the updating rule (B), and tested the proposed algorithm for each (γ_a, γ_b) in $\{\frac{1}{2}, \frac{1}{5}, \frac{1}{10}\} \times \{2, 5, 10\}$.

Table 2 shows the number of the function evaluations for each (γ_a, γ_b). Figure 3 shows the comparisons of (γ_a, γ_b) in terms of the number of the function evaluations. From Figure 3, we see that $\gamma_b = 5$ and 10 have good performances as compared to $\gamma_b = 2$.

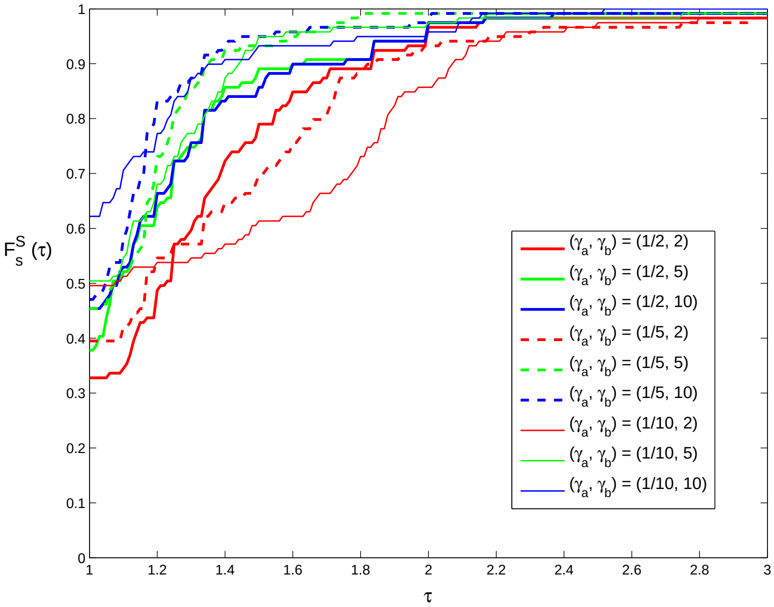

Figure 3: Comparison of (γ_a, γ_b)

6.2 Comparison with the existing Newton-type methods

We compare the proposed adaptive regularized Newton method (ARNM) with the regularized Newton method with Armijo's step size rule (RNM) and the TR-Newton method. We denote the TR-Newton method solving subproblems exactly as "TR-NM", and the TR-Newton method solving subproblems approximately by using the conjugate gradient method as "TRCG-NM".

The regularized Newton method with Armijo's step size rule is described as follows.

The Regularized Newton Method with Armijo's Step Size Rule

Step 0 : Choose a starting point x_0. Set $k := 0$.

Step 1 : If the stopping criterion is satisfied, then terminate. Otherwise, go to Step 2.

Step 2 : Compute

$$d_k = -(H_k + 2\Lambda_k I + \min(1, \|g_k\|)I)^{-1} g_k.$$

Step 3 : Find the smallest nonnegative integer l_k such that

$$f(x_k) - f(x_k + (0.5)^{l_k} d_k) \geq -0.01 \times (0.5)^{l_k} g_k^T d_k.$$

Step 4 : Update $x_{k+1} = x_k + (0.5)^{l_k} d_k$. Set $k := k + 1$, and go to Step 1.

The TR-Newton method is described as follows.

The TR-Newton Method

Step 0 : Choose a starting point x_0. Set $\Delta_0 := 1$ and $k := 0$.

Step 1 : If the stopping criterion is satisfied, then terminate. Otherwise, go to Step 2.

Step 2 : **Step 2.0 :** Set $l_k := 1$ and $\bar{\Delta}_{l_k} = \Delta_k$.

Step 2.1 : Compute an approximate solution $d_k(\bar{\Delta}_{l_k})$ of the trust-region subproblem

$$\begin{aligned} &\underset{d\in\mathbb{R}^n}{\text{minimize}}\ f(x_k) + g_k^T d + \frac{1}{2} d^T H_k d, \\ &\text{subject to } \|d\| \le \bar{\Delta}_{l_k}. \end{aligned}$$

Step 2.2 : Compute

$$\rho_k(d_k(\bar{\Delta}_{l_k}), \bar{\Delta}_{l_k}) = \frac{f(x_k) - f(x_k + d_k(\bar{\Delta}_{l_k}))}{f(x_k) - (f(x_k) + g_k^T d_k(\bar{\Delta}_{l_k}) + \frac{1}{2} d_k(\bar{\Delta}_{l_k})^T H_k d_k(\bar{\Delta}_{l_k}))}.$$

If $\rho_k(d_k(\bar{\Delta}_{l_k}), \bar{\Delta}_{l_k}) < 0.05$, then update $\bar{\Delta}_{l_k+1} = 0.25\bar{\Delta}_{l_k}$. Set $l_k := l_k + 1$, and go to Step 2.1. Otherwise, go to Step 3.

Step 3 : If $0.9 > \rho_k(d_k(\bar{\Delta}_{l_k}), \bar{\Delta}_{l_k}) \ge 0.05$, then update $\Delta_{k+1} = \bar{\Delta}_{l_k}$.
If $\rho_k(d_k(\bar{\Delta}_{l_k}), \bar{\Delta}_{l_k}) \ge 0.9$, then update $\Delta_{k+1} = \max(10^5, 2.5\bar{\Delta}_{l_k})$.
Update $x_{k+1} = x_k + d_k(\bar{\Delta}_{l_k})$. Set $k := k + 1$, and go to Step 1.

In solving subproblems of the TR-NM, we used Algorithm 7.3.4 in [4], and employed the terminate condition (7.3.20) in [4], where we set a parameter κ_{easy} as $\kappa_{\text{easy}} = 10^{-4}$. On the other hand, in solving subproblems of the TRCG-NM, we used Algorithm 7.5.1 in [4]. We set the upper bound of the number of iterations in the trust-region subproblems as 5×10^4. In the proposed algorithm, we adopted the updating rule (B) of μ_k, and set $\delta = 1$, $\gamma_a = 1/10$ and $\gamma_b = 10$.

Table 3 shows the number of the function evaluations (N_f) and the number of solving linear equations (N_L) for each method. Note that the computational complexity of calculating the minimum eigenvalue of H_k is not contained in N_L. Note also that since the TRCG-NM does not solve a linear equation exactly, we do not consider N_L for the TRCG-NM.

The ARNM cannot solve 'MARATOSB', and the TR-NM cannot solve 'BROWNAL', 'FREUROTH' and 'SBRYBND', and the TRCG-NM cannot solve 'CURLY10', 'CURLY20', 'MOREBV', 'NONDIA', 'QUARTC', 'SBRYBND', 'TESTQUAD' and 'TOINTGSS'.

Figures 4 and 5 show the comparisons of the ARNM and the RNM for N_f and N_L, Figures 6 and 7 show the comparisons of the ARNM and the TR-NM for N_f and N_L, and Figure 8 shows the comparison of the ARNM and the TRCG-NM for N_f.

Figures 4 and 5 show that both N_f and N_L of the ARNM are much less than those of the RNM, that is, the proposed algorithm is much superior to the traditional regularized Newton method. Figure 6 shows that N_f of the ARNM is almost same as that of the TR-NM. On the other hand, from Figure 7, we see that N_L of the ARNM is much less than that of the TR-NM. These results show that the ARNM can solve subproblems more easily as compared to the TR-NM. Finally, Figure 8 shows that N_f of the proposed algorithm is slightly than that of the TRCG-NM. Note that since the TRCG-NM solves subproblems approximately, it is faster than the ARNM for some problems.

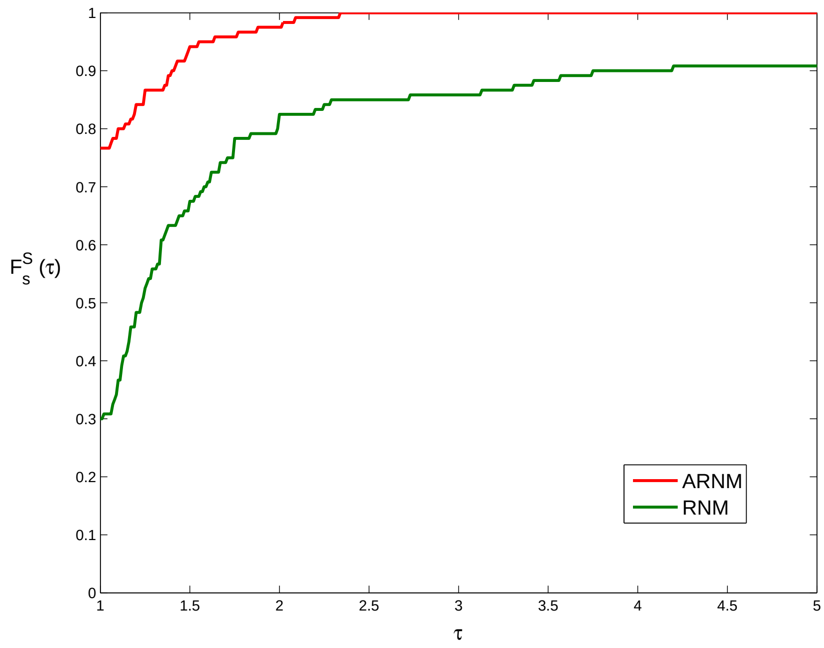

Figure 4: Comparison of ARNM and RNM for N_f

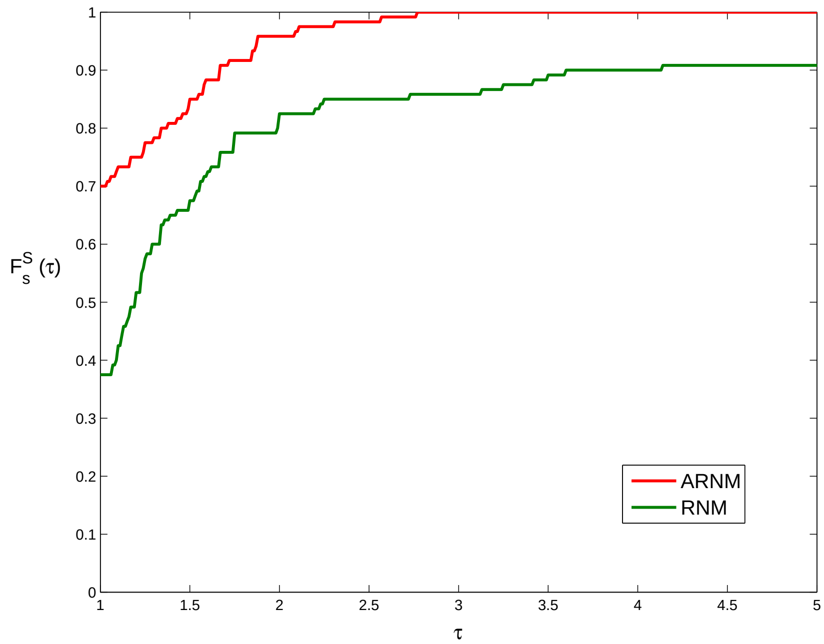

Figure 5: Comparison of ARNM and RNM for N_L

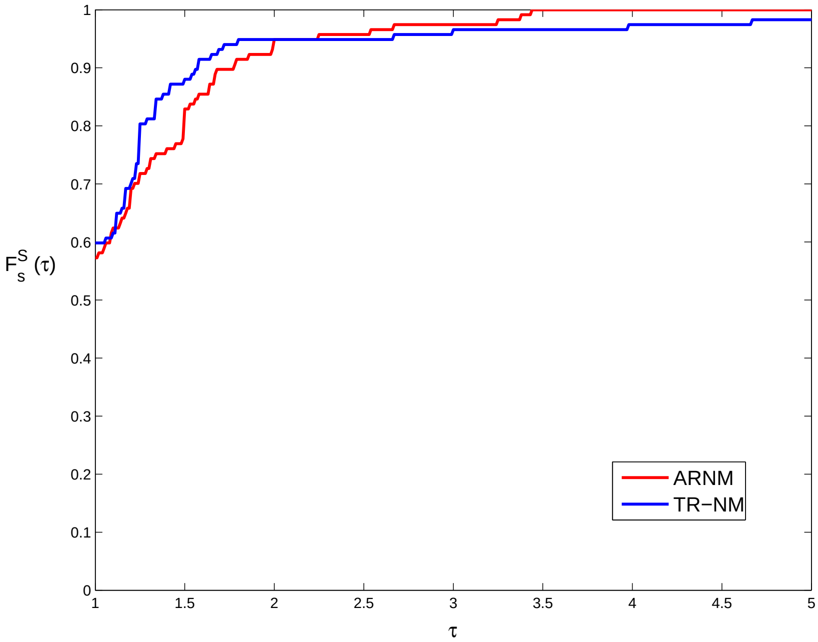

Figure 6: Comparison of ARNM and TR-NM for N_f

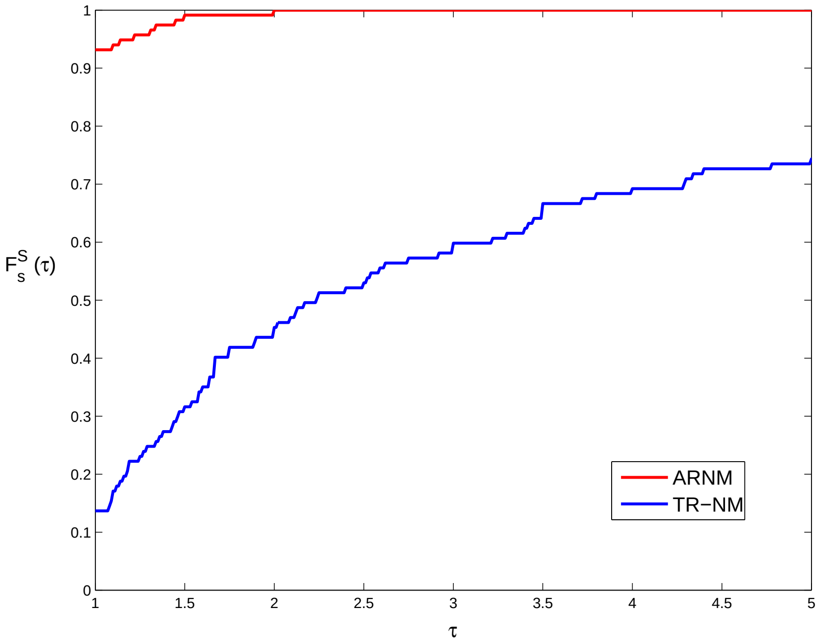

Figure 7: Comparison of ARNM and TR-NM for N_L

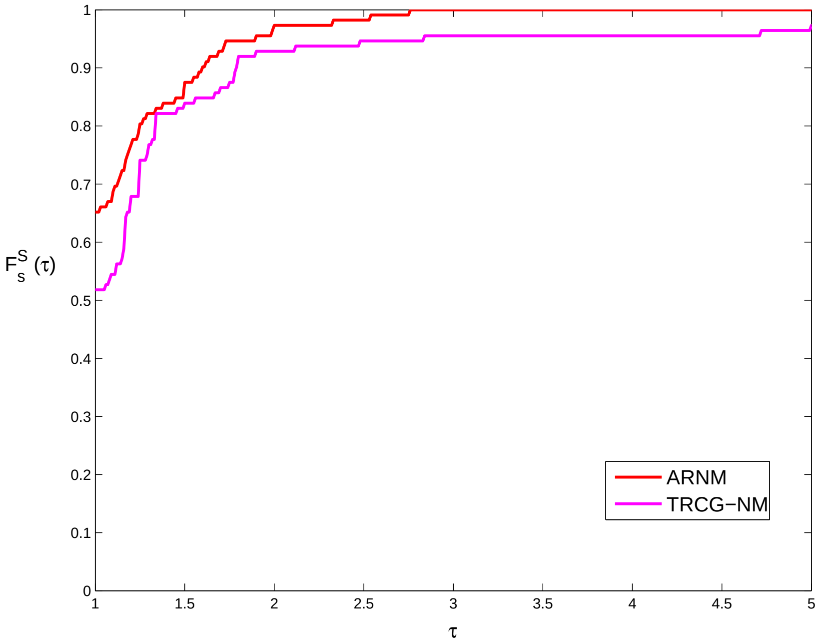

Figure 8: Comparison of ARNM and TRCG-NM for N_f

7 Concluding remarks

In this paper, we have proposed a regularized Newton method without line search. We have shown the global and superlinear convergence of the proposed algorithm, and given its global complexity bounds. In particular, we have given the conditions under which the global complexity bound J_g of the proposed algorithm is better than that of the steepest descent method $J_g = O(\epsilon^{-2})$ when f is not convex. Moreover, we have presented some numerical results, which shows that the proposed algorithm is competitive with the existing Newton-type methods.

The most time-consuming tasks of the proposed algorithm are to solve linear equations for a search direction and to calculate the minimum eigenvalue of $\nabla^2 f(x_k)$. Therefore, it is important to calculate them efficiently for large-scale problems. For the unconstrained convex optimization, Li and Li [10] proposed the regularized Newton method using an inexact solution of a regularized Newton equation as a search direction. We expect that the proposed algorithm is accelerated by exploiting their idea. On the other hand, we may use the approximating value $\bar{\Lambda}_k$ of Λ_k such that

$$0.5\Lambda_k \leq \bar{\Lambda}_k \leq 2\Lambda_k$$

instead of Λ_k in (2.1), that is, we adopt μ_k such that

$$\mu_k = c\bar{\Lambda}_k + \nu_k \|g_k\|^\delta, \quad c > 2. \tag{7.1}$$

The proposed algorithm with this modification has the same convergence properties given in Sections 3 – 5. In fact, by denoting $c_k := c\bar{\Lambda}_k/\Lambda_k$, we have

$$\mu_k = c\bar{\Lambda}_k + \nu_k \|g_k\|^\delta = c_k \Lambda_k + \nu_k \|g_k\|^\delta. \tag{7.2}$$

Moreover, we obtain that $c_k \geq c/2 > 1, \forall k \geq 0$ and $\{c_k\}$ is bounded. Then, in a way similar to the proofs in Sections 3 – 5, we can show that the proposed algorithm with (7.2) (that is (7.1)) has the same convergence properties.

Acknowledgment

The author wishes to express his sincerest thanks and appreciation to Associate Professor Nobuo Yamashita for his kind guidance and invaluable discussions in this study, and constructive criticisms in the writing of the manuscripts. The author wishes to tender his acknowledgments to Professor Masao Fukushima for his constructive comments and kind guidance. The author also wishes to express his thanks to Assistant Professor Shunsuke Hayashi for his invaluable discussions and appropriate comments. Finally, the author greatly appreciates the help of all members of Fukushima's Laboratory.

References

[1] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, New York, 1995.

[2] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.

[3] C. Cartis, N. I. M. Gould, and P. L. Toint, *Adaptive cubic overestimation methods for unconstrained optimization*, Technical Report 07/05, Department of Mathematics, FUNDP - University of Namur, 2007.

[4] A. R. Conn, N. I. M. Gould, and P. L. Toint, *Trust-Region Methods*, SIAM, Philadelphia, 2000.

[5] J. P. Crouzeix, *On second order conditions for quasiconvexity*, Mathematical Programming, 18 (1980), pp. 349–352.

[6] E. D. Dolan and J. J. Moré, *Benchmarking optimization software with performance profiles*, Mathematical Programming, 91 (2002), pp. 201–213.

[7] N. I. M. Gould, D. Orban, and P. L. Toint, *CUTEr (and SifDec), a Constrained and Unconstrained Testing Environment, revisited*, ACM Transactions on Mathematical Software, 29 (2003), pp. 373–394.

[8] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Canbridge University Press, 1985.

[9] D. H. Li, M. Fukushima, L. Qi, and N.Yamashita, *Regularized Newton methods for convex minimization problems with singular solutions*, Computational Optimization and Applications, 28 (2004), pp. 131–147.

[10] Y. J. Li and D. H. Li, *Truncated regularized Newton method for convex minimizations*, Computational Optimization and Applications, (2007). DOI: 10.1007/s10589-007-9128-7.

[11] Yu. Nesterov, *Introductory Lectures on Convex Optimization*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 2004.

[12] ——, *Accelerating the cubic regularization of Newton's method on convex problems*, Mathematical Programming, 112 (2008), pp. 159–181.

[13] Yu. Nesterov and B. T. Polyak, *Cubic regularization of Newton method and its global performance*, Mathematical Programming, 108 (2006), pp. 177–205.

[14] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, 1999.

[15] J. S. Pang, *Error bounds in mathematical programming*, Mathematical Programming, 79 (1997), pp. 299–332.

[16] R. A. Polyak, *Regularized Newton method for unconstrained convex optimization*, Mathematical Programming, (2007). DOI: 10.1007/s10107-007-0143-3.

[17] K. Ueda and N. Yamashita, *Convergence properties of the regularized Newton method for the unconstrained nonconvex optimization*, Technical Report 2008-015, Department of Applied Mathematics and Physics, Graduate School of Informatics, Kyoto University, 2008. http://www.amp.i.kyoto-u.ac.jp/tecrep/ps_file/2008/2008-015.pdf.

A Tables of numerical results

Table 1: Influence of δ

		$\delta = 1/2$		$\delta = 1$		$\delta = 2$	
Name	n	(A)	(B)	(A)	(B)	(A)	(B)
3PK	30	15	15	14	12	20	9
AKIVA	2	6	6	7	6	10	6
ALLINITU	4	13	9	15	9	11	9
ARGLINA	200	7	6	9	5	13	5
ARWHEAD	100	6	6	6	6	13	6
BARD	3	8	8	9	8	14	9
BDQRTIC	100	10	9	12	9	25	9
BEALE	2	10	10	8	9	11	9
BIGGS6	6	115	111	111	111	144	136
BOX3	3	8	8	8	8	9	7
BRKMCC	2	5	4	5	4	7	4
BROWNAL	200	4	4	8	4	42	4
BROWNBS	2	25	18	46	17	–	17
BROWNDEN	4	8	8	12	8	105	8
BROYDN7D	100	37	25	39	25	30	25
BRYBND	100	9	16	10	16	18	16
CHNROSNB	50	72	68	82	68	107	68
CLIFF	2	27	27	27	27	4898	27
COSINE	100	13	14	10	13	12	13
CRAGGLVY	100	13	13	15	13	28	13
CUBE	2	43	40	47	43	64	46
CURLY10	100	34	43	26	43	21	42
CURLY20	100	28	38	21	37	23	37
DECONVU	61	10	34	13	29	51	36
DENSCHNA	2	6	5	6	5	8	5
DENSCHNB	2	6	6	6	6	8	6
DENSCHNC	2	10	10	11	10	19	10
DENSCHND	3	28	43	29	43	884	43
DENSCHNE	3	20	31	28	18	23	32
DENSCHNF	2	6	6	7	6	15	6
DIXMAANA	300	8	7	10	7	17	6
DIXMAANB	300	8	8	10	8	18	7
DIXMAANC	300	9	8	11	8	20	8
DIXMAAND	300	10	9	12	9	22	9
DIXMAANE	300	10	10	11	9	18	8
DIXMAANF	300	13	13	23	12	37	13
DIXMAANG	300	22	13	25	13	39	13
DIXMAANH	300	21	14	28	14	41	14
DIXMAANI	300	12	12	12	10	20	10
DIXMAANJ	300	24	25	20	26	28	33
DIXMAANK	15	11	13	17	12	29	12
DIXMAANL	300	19	27	22	31	32	37
DIXON3DQ	100	11	11	9	8	8	6
DQDRTIC	100	6	5	10	5	22	4
EDENSCH	36	11	12	16	12	30	12
ENGVAL1	100	8	7	11	7	20	7
ENGVAL2	3	17	17	23	17	26	17
ERRINROS	50	71	79	89	96	112	124
EXPFIT	2	10	14	10	13	13	13
FLETCBV2	100	8	8	4	4	1	1
FREUROTH	100	12	24	11	24	23	24
GENROSE	100	168	176	161	176	178	175
GROWTHLS	3	148	121	200	121	332	121
GULF	3	38	42	36	49	61	57
HAIRY	2	63	108	68	108	68	107
HATFLDD	3	19	17	17	17	28	16
HATFLDE	3	21	18	16	17	27	17
HEART6LS	6	1620	1410	1869	1445	3442	2142
HEART8LS	8	166	146	155	160	182	159
HELIX	3	15	9	13	9	34	9
HIELOW	3	6	10	10	10	16	10
HILBERTA	2	7	7	6	6	7	5

Table 1: Influence of δ

Name	n	$\delta = 1/2$ (A)	$\delta = 1/2$ (B)	$\delta = 1$ (A)	$\delta = 1$ (B)	$\delta = 2$ (A)	$\delta = 2$ (B)
HILBERTB	10	5	4	7	4	13	4
HIMMELBB	2	13	11	20	11	40	11
HIMMELBF	4	136	159	164	153	159	167
HIMMELBG	2	8	7	7	7	7	7
HIMMELBH	2	6	7	5	7	5	6
HUMPS	2	385	497	1098	633	502	509
KOWOSB	4	9	9	7	7	8	8
LIARWHD	100	11	11	14	11	26	11
LOGHAIRY	2	4498	4496	3028	3030	5234	5236
MARATOSB	2	1494	1212	2169	1209	3543	1235
MEXHAT	2	46	47	42	47	69	47
MOREBV	100	4	4	3	3	2	2
NONCVXU2	100	49	50	41	48	83	48
NONCVXUN	100	28	47	36	46	91	45
NONDIA	100	9	11	8	11	22	11
OSBORNEA	5	58	66	64	69	84	101
OSBORNEB	11	15	15	20	16	29	22
PALMER1C	8	14	14	14	12	3485	10
PALMER1D	7	13	12	13	11	316	9
PALMER2C	8	14	14	13	11	296	9
PALMER3C	8	14	14	12	11	107	8
PALMER4C	8	15	15	12	11	107	9
PALMER5C	6	7	5	10	5	18	5
PALMER6C	8	16	16	12	12	31	9
PALMER7C	8	17	17	13	13	57	9
PALMER8C	8	16	16	13	12	32	9
PFIT1LS	3	569	547	748	714	1135	1049
PFIT2LS	3	244	189	395	256	550	388
PFIT3LS	3	232	259	282	389	439	548
PFIT4LS	3	423	417	620	567	953	922
POWELLSG	4	15	15	16	15	23	15
QUARTC	100	24	24	31	24	–	24
ROSENBR	2	32	32	32	33	43	35
S308	2	11	10	15	10	15	10
SBRYBND	100	–	96	48	96	43	95
SCHMVETT	100	5	5	6	4	8	4
SINEVAL	2	82	72	107	72	125	76
SINQUAD	100	16	21	14	20	18	20
SISSER	2	12	12	13	12	14	12
SNAIL	2	112	115	130	130	147	158
SPARSINE	100	6	6	8	6	19	6
SPARSQUR	100	16	16	17	16	26	16
SPMSRTLS	100	12	10	14	10	18	10
SROSENBR	100	8	9	8	9	17	8
STRATEC	10	27	33	29	33	64	33
TESTQUAD	1000	7	6	12	5	2529	5
TOINTGOR	50	8	7	10	6	19	6
TOINTGSS	100	8	6	9	6	14	6
TOINTPSP	50	23	31	26	30	41	30
TOINTQOR	50	7	6	8	5	14	5
TQUARTIC	100	15	15	13	15	13	14
TRIDIA	100	6	5	8	4	16	4
VARDIM	200	29	29	31	29	–	29
VAREIGVL	50	12	11	15	15	36	27
VIBRBEAM	8	103	88	86	88	54	88
WATSON	12	9	9	9	9	13	9
WOODS	4	68	65	81	65	103	65
YFITU	3	70	59	90	59	166	59
ZANGWIL2	2	5	5	5	4	5	4

Table 2: Influence of γ_a and γ_b

		$\gamma_a = 1/2$			$\gamma_a = 1/5$			$\gamma_a = 1/10$		
Name	n	$\gamma_b = 2$	$\gamma_b = 5$	$\gamma_b = 10$	$\gamma_b = 2$	$\gamma_b = 5$	$\gamma_b = 10$	$\gamma_b = 2$	$\gamma_b = 5$	$\gamma_b = 10$
3PK	30	12	12	12	7	7	7	6	6	6
AKIVA	2	6	6	6	6	6	6	6	6	6
ALLINITU	4	9	9	9	8	8	8	8	8	8
ARGLINA	200	5	5	5	4	4	4	4	4	4
ARWHEAD	100	6	6	6	5	5	5	5	5	5
BARD	3	8	8	8	7	7	7	7	7	7
BDQRTIC	100	9	9	9	9	9	9	9	9	9
BEALE	2	9	9	9	10	8	9	11	9	8
BIGGS6	6	111	99	93	124	104	104	154	114	100
BOX3	3	8	8	8	7	7	7	7	7	7
BRKMCC	2	4	4	4	4	4	4	3	3	3
BROWNAL	200	4	4	4	4	4	4	4	4	4
BROWNBS	2	17	15	16	16	13	12	16	13	12
BROWNDEN	4	8	8	8	8	8	8	8	8	8
BROYDN7D	100	25	25	25	25	25	25	42	32	33
BRYBND	100	16	14	11	19	14	12	21	13	12
CHNROSNB	50	68	58	57	103	71	72	118	85	119
CLIFF	2	27	27	27	27	27	27	27	27	27
COSINE	100	13	11	12	9	9	9	9	9	9
CRAGGLVY	100	13	13	13	13	13	13	13	13	13
CUBE	2	43	40	38	66	47	42	77	48	46
CURLY10	100	43	33	32	45	31	27	43	30	26
CURLY20	100	37	29	27	41	28	25	40	27	24
DECONVU	61	29	17	22	50	30	33	36	23	25
DENSCHNA	2	5	5	5	5	5	5	5	5	5
DENSCHNB	2	6	6	6	8	5	7	9	7	5
DENSCHNC	2	10	10	10	10	10	10	10	10	10
DENSCHND	3	43	35	33	52	37	36	54	36	36
DENSCHNE	3	18	18	18	16	16	16	13	13	13
DENSCHNF	2	6	6	6	6	6	6	6	6	6
DIXMAANA	300	7	7	7	7	7	7	13	11	9
DIXMAANB	300	8	8	8	8	8	8	30	22	17
DIXMAANC	300	8	8	8	9	9	9	9	9	9
DIXMAAND	300	9	9	9	9	9	9	9	9	9
DIXMAANE	300	9	9	9	8	8	8	8	8	8
DIXMAANF	300	12	12	12	12	12	12	12	12	12
DIXMAANG	300	13	13	13	13	13	13	13	13	13
DIXMAANH	300	14	14	14	14	14	14	14	14	14
DIXMAANI	300	10	10	10	9	9	9	17	13	12
DIXMAANJ	300	26	20	19	29	23	19	28	23	21
DIXMAANK	15	12	12	13	28	19	14	22	24	18
DIXMAANL	300	31	23	25	33	26	22	32	26	22
DIXON3DQ	100	8	8	8	6	6	6	5	5	5
DQDRTIC	100	5	5	5	4	4	4	4	4	4
EDENSCH	36	12	12	12	12	12	12	12	12	12
ENGVAL1	100	7	7	7	7	7	7	7	7	7
ENGVAL2	3	17	17	17	17	17	17	17	17	17
ERRINROS	50	96	76	80	109	66	63	106	63	57
EXPFIT	2	13	11	10	14	10	10	14	12	12
FLETCBV2	100	4	4	4	3	3	3	3	3	3
FREUROTH	100	24	19	19	18	13	–	17	13	13
GENROSE	100	176	121	107	222	155	136	265	154	146
GROWTHLS	3	121	96	92	183	155	122	221	131	184
GULF	3	49	41	38	61	33	37	60	49	36
HAIRY	2	108	79	106	192	101	115	196	120	70
HATFLDD	3	17	17	17	19	18	18	22	19	21
HATFLDE	3	17	17	18	23	17	19	25	19	17
HEART6LS	6	1445	1507	1654	2333	1732	1515	2975	1794	1875
HEART8LS	8	160	158	151	224	159	153	260	181	175
HELIX	3	9	9	9	9	9	9	10	10	10
HIELOW	3	10	9	8	13	10	9	15	11	10
HILBERTA	2	6	6	6	5	5	5	4	4	4
HILBERTB	10	4	4	4	4	4	4	3	3	3
HIMMELBB	2	11	11	11	12	12	12	12	12	12

Table 2: Influence of γ_a and γ_b

Name	n	$\gamma_a = 1/2$ $\gamma_b = 2$	$\gamma_b = 5$	$\gamma_b = 10$	$\gamma_a = 1/5$ $\gamma_b = 2$	$\gamma_b = 5$	$\gamma_b = 10$	$\gamma_a = 1/10$ $\gamma_b = 2$	$\gamma_b = 5$	$\gamma_b = 10$
HIMMELBF	4	153	153	153	167	161	162	165	163	158
HIMMELBG	2	7	5	6	8	7	6	9	7	7
HIMMELBH	2	7	7	8	6	5	6	6	6	6
HUMPS	2	633	434	422	1049	327	652	676	476	340
KOWOSB	4	7	7	7	10	9	8	12	9	12
LIARWHD	100	11	11	11	10	10	10	10	10	10
LOGHAIRY	2	3030	1411	2504	2385	4391	2507	3649	1649	51
MARATOSB	2	1209	–	907	1848	1424	1171	2346	1502	–
MEXHAT	2	47	36	35	60	42	37	65	46	44
MOREBV	100	3	3	3	2	2	2	2	2	2
NONCVXU2	100	48	41	40	30	30	30	47	38	36
NONCVXUN	100	46	39	37	40	32	30	27	27	27
NONDIA	100	11	10	8	13	10	8	15	11	10
OSBORNEA	5	69	60	46	73	71	44	93	62	59
OSBORNEB	11	16	17	17	18	17	22	23	21	17
PALMER1C	8	12	12	12	7	7	7	6	6	6
PALMER1D	7	11	11	11	7	7	7	6	6	6
PALMER2C	8	11	11	11	7	7	7	6	6	6
PALMER3C	8	11	11	11	7	7	7	6	6	6
PALMER4C	8	11	11	11	7	7	7	6	6	6
PALMER5C	6	5	5	5	4	4	4	4	4	4
PALMER6C	8	12	12	12	7	7	7	6	6	6
PALMER7C	8	13	13	13	8	8	8	6	6	6
PALMER8C	8	12	12	12	7	7	7	6	6	6
PFIT1LS	3	714	695	751	893	771	861	1236	751	613
PFIT2LS	3	256	237	263	398	397	272	475	231	238
PFIT3LS	3	389	364	373	416	439	491	515	328	245
PFIT4LS	3	567	445	735	714	649	649	808	458	419
POWELLSG	4	15	15	15	15	15	15	15	15	15
QUARTC	100	24	24	24	24	24	24	24	24	24
ROSENBR	2	33	27	27	41	33	31	49	30	40
S308	2	10	9	9	14	11	10	16	12	11
SBRYBND	100	96	65	71	61	40	38	62	35	30
SCHMVETT	100	4	4	4	4	4	4	4	4	4
SINEVAL	2	72	57	57	102	73	66	127	83	123
SINQUAD	100	20	21	16	25	21	17	30	23	16
SISSER	2	12	12	12	12	12	12	12	12	12
SNAIL	2	130	94	100	171	110	122	202	126	236
SPARSINE	100	6	6	6	6	6	6	6	6	6
SPARSQUR	100	16	16	16	16	16	16	16	16	16
SPMSRTLS	100	10	10	10	11	11	11	10	10	10
SROSENBR	100	9	9	9	8	8	8	7	7	7
STRATEC	10	33	23	20	46	28	20	35	28	26
TESTQUAD	1000	5	5	5	4	4	4	4	4	4
TOINTGOR	50	6	6	6	5	5	5	5	5	5
TOINTGSS	100	6	6	6	6	6	6	5	5	5
TOINTPSP	50	30	23	25	38	26	33	41	29	41
TOINTQOR	50	5	5	5	4	4	4	4	4	4
TQUARTIC	100	15	14	13	14	12	10	15	15	13
TRIDIA	100	4	4	4	4	4	4	3	3	3
VARDIM	200	29	29	29	29	29	29	29	29	29
VAREIGVL	50	15	16	13	12	12	12	30	25	25
VIBRBEAM	8	88	67	62	89	52	41	78	51	44
WATSON	12	9	9	9	9	9	9	9	9	9
WOODS	4	65	54	56	93	64	64	106	66	67
YFITU	3	59	55	52	78	54	57	92	62	62
ZANGWIL2	2	4	4	4	4	4	4	4	4	4

Table 3: Comparison with other methods

Name	n	ARNM N_f, N_L	ARNM f	RNM N_f	RNM N_L	RNM f	TR-NM N_f	TR-NM N_L	TR-NM f	TRCG-NM N_f	TRCG-NM f
3PK	30	6	1.72E + 00	333	333	1.72E + 00	8	33	1.72E + 00	8	1.72E + 00
AKIVA	2	6	6.17E + 00	7	7	6.17E + 00	6	6	6.17E + 00	6	6.17E + 00
ALLINITU	4	8	5.74E + 00	11	9	5.74E + 00	8	18	5.74E + 00	5	5.74E + 00
ARGLINA	200	4	2.00E + 02	8	8	2.00E + 02	5	5	2.00E + 02	5	2.00E + 02
ARWHEAD	100	5	8.79E − 14	6	6	0.00E + 00	5	5	6.59E − 14	5	6.59E − 14
BARD	3	7	8.21E − 03	10	10	8.21E − 03	12	30	8.21E − 03	8	1.00E + 00
BDQRTIC	100	9	3.79E + 02	10	10	3.79E + 02	10	15	3.79E + 02	10	3.79E + 02
BEALE	2	8	8.23E − 13	8	8	2.50E − 11	7	16	7.52E − 14	7	1.95E − 16
BIGGS6	6	100	8.17E − 08	102	97	1.11E − 06	398	2576	2.43E − 01	472	2.43E − 01
BOX3	3	7	2.28E − 12	9	9	4.05E − 11	7	10	1.52E − 11	7	1.49E − 11
BRKMCC	2	3	1.69E − 01	4	4	1.69E − 01	2	2	1.69E − 01	2	1.69E − 01
BROWNAL	200	4	3.80E − 13	4	4	3.21E − 09	–	–	–	20	2.24E − 14
BROWNBS	2	12	1.00E − 13	22	20	0.00E + 00	32	74	1.97E − 31	34	5.08E − 26
BROWNDEN	4	8	8.58E + 04	8	8	8.58E + 04	10	17	8.58E + 04	10	8.58E + 04
BROYDN7D	100	33	3.28E + 01	30	30	3.28E + 01	22	99	4.01E + 01	26	4.12E + 01
BRYBND	100	12	3.56E − 15	11	9	3.61E − 13	17	112	4.69E − 18	13	1.29E − 22
CHNROSNB	50	119	3.55E − 19	51	43	1.79E − 17	71	161	1.86E − 23	69	4.22E − 14
CLIFF	2	27	2.00E − 01	34	34	2.00E − 01	27	29	2.00E − 01	27	2.00E − 01
COSINE	100	9	−9.90E + 01	12	11	−9.90E + 01	6	15	−9.90E + 01	19	−9.90E + 01
CRAGGLVY	100	13	3.23E + 01	15	15	3.23E + 01	13	20	3.23E + 01	14	3.23E + 01
CUBE	2	46	6.77E − 15	34	29	4.85E − 13	38	54	6.80E − 18	39	1.97E − 18
CURLY10	100	26	−1.00E + 04	22	20	−1.00E + 04	8	30	−1.00E + 04	–	–
CURLY20	100	24	−1.00E + 04	20	18	−1.00E + 04	9	33	−1.00E + 04	–	–
DECONVU	61	25	7.31E − 08	17	15	1.46E − 07	21	146	8.19E − 09	62	3.40E − 11
DENSCHNA	2	5	1.35E − 12	6	6	1.02E − 15	5	5	2.21E − 12	5	2.21E − 12
DENSCHNB	2	5	1.85E − 20	4	4	4.77E − 18	3	8	1.88E − 15	5	1.31E − 15
DENSCHNC	2	10	4.87E − 20	11	11	2.27E − 17	10	10	2.18E − 20	10	2.18E − 20
DENSCHND	3	36	1.82E − 08	29	29	1.35E − 08	40	123	8.19E − 09	35	6.99E − 09
DENSCHNE	3	13	7.78E − 13	26	26	1.01E − 18	9	19	1.07E − 18	11	6.87E − 19
DENSCHNF	2	6	6.86E − 22	6	6	8.44E − 20	6	6	6.51E − 22	6	6.51E − 22
DIXMAANA	300	9	1.00E + 00	10	10	1.00E + 00	7	17	1.00E + 00	7	1.00E + 00
DIXMAANB	300	17	1.00E + 00	11	11	1.00E + 00	18	73	1.00E + 00	11	1.00E + 00
DIXMAANC	300	9	1.00E + 00	12	12	1.00E + 00	14	60	1.00E + 00	12	1.00E + 00
DIXMAAND	300	9	1.00E + 00	12	12	1.00E + 00	11	43	1.00E + 00	14	1.00E + 00
DIXMAANE	300	8	1.00E + 00	25	25	1.00E + 00	11	44	1.00E + 00	10	1.00E + 00
DIXMAANF	300	12	1.00E + 00	21	21	1.00E + 00	17	87	1.00E + 00	21	1.00E + 00
DIXMAANG	300	13	1.00E + 00	21	21	1.00E + 00	20	108	1.00E + 00	22	1.00E + 00
DIXMAANH	300	14	1.00E + 00	22	22	1.00E + 00	22	106	1.00E + 00	25	1.00E + 00
DIXMAANI	300	12	1.00E + 00	41	41	1.00E + 00	15	78	1.00E + 00	14	1.00E + 00
DIXMAANJ	300	21	1.00E + 00	27	27	1.00E + 00	24	141	1.00E + 00	28	1.00E + 00
DIXMAANK	15	18	1.00E + 00	17	17	1.00E + 00	15	62	1.00E + 00	16	1.00E + 00
DIXMAANL	300	22	1.00E + 00	27	27	1.00E + 00	27	170	1.00E + 00	32	1.00E + 00
DIXON3DQ	100	5	4.96E − 14	35	35	2.39E − 08	4	19	1.09E − 29	5	0.00E + 00

Table 3: Comparison with other methods

Name	n	ARNM N_f, N_L	ARNM f	RNM N_f	RNM N_L	RNM f	TR-NM N_f	TR-NM N_L	TR-NM f	TRCG-NM N_f	TRCG-NM f
DQDRTIC	100	4	2.67E − 25	6	6	2.99E − 18	5	8	6.28E − 29	5	4.93E − 30
EDENSCH	36	12	2.19E + 02	12	12	2.19E + 02	15	48	2.19E + 02	20	2.19E + 02
ENGVAL1	100	7	1.09E + 02	8	8	1.09E + 02	9	11	1.09E + 02	9	1.09E + 02
ENGVAL2	3	17	1.28E − 14	21	21	8.39E − 22	13	17	9.71E − 17	19	4.08E − 17
ERRINROS	50	57	3.99E + 01	130	127	3.99E + 01	54	115	3.99E + 01	52	3.99E + 01
EXPFIT	2	12	2.41E − 01	10	8	2.41E − 01	8	21	2.41E − 01	10	2.41E − 01
FLETCBV2	100	3	−5.14E − 01	4	4	−5.14E − 01	2	4	−5.14E − 01	2	−5.14E − 01
FREUROTH	100	13	1.20E + 04	15	12	1.20E + 04	–	–	–	9	1.20E + 04
GENROSE	100	144	1.00E + 00	105	77	1.00E + 00	88	362	1.00E + 00	130	1.00E + 00
GROWTHLS	3	184	1.00E + 00	366	366	1.00E + 00	99	199	1.00E + 00	97	1.00E + 00
GULF	3	36	2.84E − 11	119	117	2.41E − 06	30	93	4.36E − 20	36	2.21E − 15
HAIRY	2	70	2.00E + 01	78	60	2.00E + 01	69	225	2.00E + 01	84	2.00E + 01
HATFLDD	3	21	6.62E − 08	21	21	6.76E − 08	20	47	6.62E − 08	13	6.62E − 08
HATFLDE	3	17	5.12E − 07	23	23	5.12E − 07	19	36	5.12E − 07	18	5.12E − 07
HEART6LS	6	1875	7.93E − 23	3193	2923	1.75E − 10	555	4064	1.78E − 26	680	4.41E − 29
HEART8LS	8	175	4.00E − 23	107	83	1.90E − 19	78	444	6.80E − 21	111	3.58E − 23
HELIX	3	10	3.74E − 23	11	11	1.91E − 13	10	34	4.22E − 15	18	6.02E − 23
HIELOW	3	10	8.74E + 02	7	6	8.74E + 02	8	30	8.74E + 02	8	8.74E + 02
HILBERTA	2	4	3.39E − 15	9	9	1.92E − 13	3	7	2.05E − 33	3	2.05E − 33
HILBERTB	10	3	1.23E − 12	5	5	2.44E − 19	3	5	1.28E − 29	3	6.36E − 30
HIMMELBB	2	12	1.99E − 18	12	12	6.13E − 26	15	66	5.53E − 21	15	3.45E − 24
HIMMELBF	4	158	3.19E + 02	993	993	3.19E + 02	46	259	3.19E + 02	68	3.19E + 02
HIMMELBG	2	7	1.05E − 14	6	6	1.17E − 12	5	9	8.86E − 12	7	8.40E − 18
HIMMELBH	2	6	−1.00E + 00	7	6	−1.00E + 00	4	6	−1.00E + 00	4	−1.00E + 00
HUMPS	2	340	3.39E − 12	1275	1221	4.20E − 13	2712	10595	1.42E − 10	2607	1.07E − 15
KOWOSB	4	12	3.08E − 04	8	8	3.08E − 04	10	35	3.08E − 04	7	3.08E − 04
LIARWHD	100	10	1.52E − 13	12	12	1.19E − 12	12	19	6.12E − 14	13	1.47E − 26
LOGHAIRY	2	51	6.53E + 00	214	211	6.48E + 00	2757	10486	1.82E − 01	4576	1.82E − 01
MARATOSB	2	–	–	948	672	−1.00E + 00	1036	1419	−1.00E + 00	1036	−1.00E + 00
MEXHAT	2	44	−4.00E − 02	31	28	−4.00E − 02	44	52	−4.00E − 02	44	−4.00E − 02
MOREBV	100	2	7.69E − 07	3	3	5.44E − 07	1	1	7.89E − 10	–	–
NONCVXU2	100	36	2.32E + 02	98	98	2.34E + 02	44	206	2.32E + 02	43	2.34E + 02
NONCVXUN	100	27	2.34E + 02	42	42	2.37E + 02	36	154	2.32E + 02	35	2.32E + 02
NONDIA	100	10	4.93E − 16	8	7	6.78E − 21	6	24	1.50E − 18	–	–
OSBORNEA	5	59	5.51E − 05	52	32	5.51E − 05	38	98	5.46E − 05	35	5.46E − 05
OSBORNEB	11	17	4.01E − 02	26	26	4.01E − 02	28	88	8.76E − 02	20	4.01E − 02
PALMER1C	8	6	9.76E − 02	282	282	9.76E − 02	7	37	9.76E − 02	7	9.76E − 02
PALMER1D	7	6	6.53E − 01	189	189	6.53E − 01	7	35	6.53E − 01	7	6.53E − 01
PALMER2C	8	6	1.44E − 02	165	165	1.44E − 02	6	32	1.44E − 02	7	1.44E − 02
PALMER3C	8	6	1.95E − 02	170	170	1.95E − 02	6	30	1.95E − 02	6	1.95E − 02
PALMER4C	8	6	5.03E − 02	227	227	5.03E − 02	7	33	5.03E − 02	7	5.03E − 02
PALMER5C	6	4	2.13E + 00	7	7	2.13E + 00	5	10	2.13E + 00	5	2.13E + 00
PALMER6C	8	6	1.64E − 02	365	365	1.64E − 02	7	36	1.64E − 02	7	1.64E − 02

Table 3: Comparison with other methods

		ARNM		RNM			TR-NM			TRCG-NM	
Name	n	N_f, N_L	f	N_f	N_L	f	N_f	N_L	f	N_f	f
PALMER7C	8	6	6.02E − 01	1139	1139	6.02E − 01	9	35	6.02E − 01	9	6.02E − 01
PALMER8C	8	6	1.60E − 01	495	495	1.60E − 01	8	43	1.60E − 01	8	1.60E − 01
PFIT1LS	3	613	1.14E − 10	808	357	1.85E − 09	374	561	6.99E − 15	376	5.50E − 15
PFIT2LS	3	238	2.01E − 08	255	129	1.00E − 11	133	210	9.91E − 13	274	1.05E − 19
PFIT3LS	3	245	2.41E − 24	311	156	1.18E − 14	161	274	1.14E − 15	211	5.96E − 18
PFIT4LS	3	419	2.56E − 13	473	283	2.52E − 11	322	460	1.94E − 16	348	5.77E − 21
POWELLSG	4	15	4.43E − 09	16	16	5.64E − 09	15	19	4.64E − 09	15	4.62E − 09
QUARTC	100	24	2.31E − 08	27	27	1.11E − 07	29	84	2.78E − 08	–	–
ROSENBR	2	40	6.25E − 16	27	24	5.93E − 15	27	33	7.30E − 26	29	4.14E − 22
S308	2	11	7.73E − 01	8	8	7.73E − 01	10	12	7.73E − 01	10	7.73E − 01
SBRYBND	100	30	1.24E − 13	17	13	3.89E − 14	–	–	–	–	–
SCHMVETT	100	4	−2.94E + 02	5	5	−2.94E + 02	4	6	−2.94E + 02	5	−2.94E + 02
SINEVAL	2	123	3.41E − 15	61	48	2.56E − 12	62	94	1.46E − 14	62	7.36E − 15
SINQUAD	100	16	−4.01E + 03	15	13	−4.01E + 03	9	23	−4.01E + 03	13	−4.01E + 03
SISSER	2	12	1.14E − 08	13	13	5.90E − 09	12	12	1.07E − 08	12	1.07E − 08
SNAIL	2	236	4.84E − 13	126	126	5.58E − 18	93	163	2.63E − 17	93	3.88E − 14
SPARSINE	100	6	9.52E − 22	7	7	8.59E − 21	28	216	7.96E − 16	41	1.93E − 14
SPARSQUR	100	16	7.66E − 09	17	17	5.50E − 09	16	19	1.48E − 08	16	1.51E − 08
SPMSRTLS	100	10	4.05E − 11	11	11	1.30E − 16	11	33	6.97E − 13	19	2.70E − 16
SROSENBR	100	7	7.74E − 15	14	14	2.16E − 17	6	11	1.14E − 27	6	2.47E − 30
STRATEC	10	26	2.21E + 03	38	36	2.21E + 03	41	91	2.21E + 03	31	2.21E + 03
TESTQUAD	1000	4	2.53E − 20	7	7	4.31E − 14	5	14	2.14E − 26	–	–
TOINTGOR	50	5	1.37E + 03	11	11	1.37E + 03	9	22	1.37E + 03	9	1.37E + 03
TOINTGSS	100	5	1.01E + 01	8	8	1.01E + 01	15	53	1.01E + 01	–	–
TOINTPSP	50	41	2.26E + 02	56	22	2.26E + 02	26	67	2.26E + 02	54	2.26E + 02
TOINTQOR	50	4	1.18E + 03	7	7	1.18E + 03	4	11	1.18E + 03	4	1.18E + 03
TQUARTIC	100	13	1.12E − 24	21	20	1.73E − 15	12	34	1.27E − 14	15	5.92E − 24
TRIDIA	100	3	1.60E − 11	5	5	1.59E − 14	4	13	7.65E − 31	4	4.19E − 30
VARDIM	200	29	2.53E − 24	29	29	2.33E − 25	29	33	2.33E − 25	29	9.34E − 25
VAREIGVL	50	25	2.01E − 11	12	12	2.16E − 09	22	93	1.02E − 10	22	1.44E − 10
VIBRBEAM	8	44	1.56E − 01	63	54	1.56E − 01	74	341	1.56E − 01	78	1.56E − 01
WATSON	12	9	6.60E − 12	11	11	7.77E − 09	10	77	2.62E − 07	16	3.96E − 09
WOODS	4	67	2.38E − 14	48	46	4.09E − 16	57	140	5.57E − 16	54	4.15E − 26
YFITU	3	62	6.67E − 13	221	217	6.29E − 09	57	90	6.67E − 13	58	6.67E − 13
ZANGWIL2	2	4	−1.82E + 01	5	5	−1.82E + 01	2	3	−1.82E + 01	2	−1.82E + 01