RAIRO-Oper. Res. 53 (2019) 787–805
https://doi.org/10.1051/ro/2017054

RAIRO Operations Research
www.rairo-ro.org

NONMONOTONE CONIC TRUST REGION METHOD WITH LINE SEARCH TECHNIQUE FOR BOUND CONSTRAINED OPTIMIZATION*

Lijuan Zhao[1,*]

Abstract. In this paper, we propose a nonmonotone trust region method for bound constrained optimization problems, where the bounds are dealt with by affine scaling technique. Differing from the traditional trust region methods, the subproblem in our algorithm is based on a conic model. Moreover, when the trial point isn't acceptable by the usual trust region criterion, a line search technique is used to find an acceptable point. This procedure avoids resolving the trust region subproblem, which may reduce the total computational cost. The global convergence and Q-superlinear convergence of the algorithm are established under some mild conditions. Numerical results on a series of standard test problems are reported to show the effectiveness of the new method.

Mathematics Subject Classification. 65K05, 90C30.

Received December 20, 2016. Accepted June 24, 2017.

1. Introduction

In this paper, we consider the following bound constrained optimization problem

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

$$\text{s.t. } x \in \Omega = \{x \mid l \le x \le u\}, \tag{1.2}$$

where $f:\mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable, l and u are the upper and lower bounds of the variables (whose components may be infinite), respectively.

Minimization problems with upper and lower bounds on some of the variables form an important and common class of minimization problems. There are many methods for solving problems of this kind, such as active set type method, projected Newton method, pattern search method and affine scaling trust region method. In this paper, we propose an affine scaling trust region method for solving problem (1.1)–(1.2). The main advantages of this type of method are the global convergence and the robustness of the algorithm.

Traditional trust region methods [19, 23] compute the ratio between the actual reduction and the predicted reduction. If the ratio is satisfactory, we accept the trial point and enlarge the trust region radius. Otherwise,

Keywords. Nonmonotone technique, conic model, line search, trust region, bound constrained optimization.

* *This work was supported by the project of Nanjing Vocational Institute of Railway Technology (grant No. Y160027), the National Natural Science Foundation of China (grant No. 11571178).*

[1] Department of Social Science Teaching, Nanjing Vocational Institute of Railway Technology, Nanjing 210031, China.

*Corresponding author: zzlljj210@163.com

the trial point is rejected, the trust region radius is shrunk and the trust region subproblem is resolved until an acceptable trial point is found. Thus, the trust region subproblem may be resolved several times at each iteration, which is likely to increase the total cost of computation for large scale optimization problems. Nocedal and Yuan [20] presented a method which combines the trust region method with the line search technique. Their method performs line search technique to find a successful iteration when the trial point isn't accepted by trust region, where the trust region subproblem needs to be solved only once at each iteration, so the total cost of computation may be reduced to some extent.

In traditional trust–region methods, if the ratio between the actual reduction and the predicted reduction is satisfactory at each iteration, the objective function decreases. Otherwise, the objective function doesn't change. It may cause that the iterations are trapped at the bottom of the curved narrow valley such that it has slow convergence rate. To overcome this drawback, Sun [24], Fu and Sun [9], Toint [28] pointed out that the nonmonotone technique may be helpful. Nonmonotone technique was first proposed by Grippo, Lampariello and Lucidi [11] in the following form

$$f(x_k+\alpha_k d_k) \le \max_{0\le j\le m(k)} f(x_{k-j}) + \delta\alpha_k g_k^T d_k,$$

where α_k is a step-size, $m(k+1)=\min\{m(k)+1, M\}$, $M>0$ and $\delta\in(0,1)$ are two constants. However, this kind of nonmonotone technique has some drawbacks. First, it is dependent on the choices of M. Second, due to the max-value choice, many good values may be discarded at each iteration. To overcome these drawbacks, Zhang and Hager [31] presented another nonmonotone line search technique as follows,

$$\begin{aligned} f(x_k+\alpha_k d_k) &\le C_k + \delta\alpha_k g_k^T d_k,\\ g(x_k+\alpha_k d_k)^T d_k &\ge \sigma g_k^T d_k, \end{aligned}$$

where $\delta\in(0,1)$ and $\sigma\in(0,1)$ are two constants,

$$C_k = \begin{cases} f(x_k), & \text{if} \quad k=0,\\ (\tau_{k-1}Q_{k-1}C_{k-1}+f(x_k))/Q_k, & \text{if} \quad k\ge 1, \end{cases} \tag{1.3}$$

$$Q_k = \begin{cases} 1, & \text{if} \quad k=0,\\ \tau_{k-1}Q_{k-1}+1, & \text{if} \quad k\ge 1, \end{cases}$$

here, $\tau_k\in[\tau_{\min},\tau_{\max}]$, $\tau_{\min}\in[0,1)$ and $\tau_{\max}\in[\tau_{\min},1)$ are two given constants.

From (1.3), we can see that C_k is a combination of C_{k-1} and $f(x_k)$. In theory, τ_k should be close to 1 when the iterates are far from the optimal solution and τ_k should be close to 0 when the iterates are near the optimal solution. Gu and Mo [12] introduced another nonmonotone technique as follows,

$$f(x_k+\alpha_k d_k) \le E_k + \delta\alpha_k g_k^T d_k, \tag{1.4}$$

$$g(x_k+\alpha_k d_k)^T d_k \ge \sigma g_k^T d_k, \tag{1.5}$$

where E_k is a simple convex combination of E_{k-1} and $f(x_k)$, *i.e.*,

$$E_k = \begin{cases} f(x_k), & \text{if} \quad k=0,\\ \mu_{k-1}E_{k-1}+(1-\mu_{k-1})f(x_k), & \text{if} \quad k\ge 1, \end{cases} \tag{1.6}$$

for $\mu_{k-1}\in(0,1)$. When $\mu_{k-1}=\tau_{k-1}Q_{k-1}/Q_k$, E_k is C_k; when $\mu_k=0$ for all k, E_k becomes $f(x_k)$. So (1.4)–(1.5) can be regarded as an extension of the traditional line search and Zhang–Hager's nonmonotone line search.

The conic model was first proposed by Davidon and Sorensen [7,25] for unconstrained optimization problem in the following form

$$\min_{d\in\mathbb{R}^n} \varphi_k(d) = \frac{g_k^T d}{1+b_k^T d} + \frac{1}{2}\frac{d^T B_k d}{(1+b_k^T d)^2},$$

where $g_k = \nabla f(x_k)$ is the gradient of the objective function f at x_k, $B_k \in \mathbb{R}^{n\times n}$ is the Hessian of the objective function or its approximation, $b_k \in \mathbb{R}^n$ is the horizontal vector. Due to the global convergence of the trust region method, Di and Sun [8] proposed the following conic trust region subproblem for unconstrained optimization problem

$$\begin{aligned} \min_{d\in\mathbb{R}^n} \ & \varphi_k(d) = \frac{g_k^T d}{1+b_k^T d} + \frac{1}{2}\frac{d^T B_k d}{(1+b_k^T d)^2}, \\ \text{s.t.} \ & \|d\| \le \Delta_k, \end{aligned}$$

where $\Delta_k > 0$ is a trust region radius, $\|\cdot\|$ refers to Euclidean norm. Ni [21] proposed a new conic trust region subproblem for unconstrained optimization problem as follows,

$$\min_{d\in\mathbb{R}^n} \ \varphi_k(d) = \frac{g_k^T d}{1+b_k^T d} + \frac{1}{2}\frac{d^T B_k d}{(1+b_k^T d)^2}, \tag{1.7}$$

$$\text{s.t.} \ \|d\| \le \Delta_k, \tag{1.8}$$

$$|1+b_k^T d| \ge \epsilon_0, \tag{1.9}$$

where $\epsilon_0 \in (0,1)$ is a constant, to overcome the case in which the conic function $\varphi_k(d)$ may be unbounded. In [21] he divided the problem (1.7)–(1.9) into three subproblems and analyzed the optimality condition for each subproblem. In [26], Salahi gave an SOCP/SDP formulation for the conic trust region subproblem. However, this method is not applicable for large scale problems. In [27], they presented a generalized Newton algorithm to overcome this limition.

In [34], Zhu proposed an combined affine scaling trust region and interior backtracking line search method for bound constrained nonlinear systems, where he used Grippo et al's nonmonotone line search, which has some drawbacks that we have mentioned before. In [6], Cui, Wu and Qu proposed a method which combines nonmonotone conic trust region method and line search technique for unconstrained optimization problem, where they used the nonmonotone line search which is proposed by Zhang and Hager [31]. The numerical results show that the method is quite effective. In [32], we proposed a conic affine scaling dogleg method for bound constrained optimization problem. The numerical results show that the method has some drawbacks. If the trial point isn't accepted by trust region, we shrink the trust region radius and solve the trust region subproblem, which may increase the total cost of computation to some extent. In this paper, we propose a nonmonotone conic trust region method with line search technique for bound constrained optimization problem, where the bounds are dealt with by affine scaling technique, which was first proposed by Coleman and Li [4] and later studied by many authors [1–3, 13, 15, 16, 18, 29, 33–36]. Our method produces a prediction of the objective function better than that obtained from the quadratic model. When the trial point isn't accepted by the trust region, we perform line search until an acceptable trial point is found, which reduces the total cost of computation.

The paper is organized as follows. In Section 2, we describe our nonmonotone conic trust region method with line search technique for bound constrained optimization problem. The global convergence and Q-superlinear convergence are established in Section 3. Numerical results are reported in Section 4. We give some conclusions in Section 5.

2. Nonmonotone conic trust region method with line search technique for bound constrained optimization problem

Ignoring primal and dual feasibility of the problem (1.1)–(1.2), we give the first order necessary conditions for x^* to be a local minimizer as follows,

$$\begin{cases} (g_*)_i = 0, & \text{if} \quad l_i < (x^*)_i < u_i; \\ (g_*)_i \ge 0, & \text{if} \quad (x^*)_i = l_i; \\ (g_*)_i \le 0, & \text{if} \quad (x^*)_i = u_i. \end{cases} \tag{2.10}$$

where $g_* \stackrel{\text{def}}{=} g(x^*)$. We can rewrite (2.10) in the form

$$D(x^*)g(x^*) = 0,$$

where $D(x)$ is Coleman-Li scaling matrix [4]. The diagonal entries of $D(x)$ are

$$d_i(x) = \begin{cases} u_i - x_i, & \text{if} \quad g_i < 0 \quad \text{and} \quad u_i < +\infty; \\ x_i - l_i, & \text{if} \quad g_i > 0 \quad \text{and} \quad l_i > -\infty; \\ \min\{x_i - l_i, u_i - x_i\}, & \text{if} \quad g_i = 0 \quad \text{and} \quad l_i > -\infty \quad \text{or} \quad u_i < \infty; \\ 1, & \text{otherwise.} \end{cases}$$

The diagonal scaling matrix $D(x)$ handles the bounds implicitly.

2.1. Algorithm

In this paper, we consider the following conic trust region subproblem

$$\min_{d\in\mathbb{R}^n} c(x_k + d) = f(x_k) + \frac{g_k^T d}{1 + b_k^T d} + \frac{1}{2}\frac{d^T B_k d}{(1 + b_k^T d)^2}, \tag{2.11}$$

$$\text{s.t.} \quad \|S_k d\| \le \Delta_k, \tag{2.12}$$

$$x_k + d \in \text{int}(\Omega), \tag{2.13}$$

where S_k is a scaling matrix and $D_k = D(x_k)$.

Next, we are going to state our Nonmonotone Conic Trust Region method with Line Search technique (abbreviated as NCTRLS) for bound constrained optimization problem.

Algorithm 2.1. $\{NCTRLS\}$

Step 0. Choose parameters $0 < \eta_1 < \eta_2 < 1$, $0 < \gamma_1 < \gamma_2 < 1 < \gamma_3$, $\delta \in (0, 1/2)$, $b_0 \in \mathbb{R}^n$, $\lambda \in (0, 1)$, $\mu_0 \in (0, 1)$, $\epsilon > 0$, $\sigma \in (0, 1)$. Given an initial trust region radius Δ_0, a maximal trust region radius $\Delta_{\max} \ge \Delta_0$, a positive definite matrix $B_0 \in \mathbb{R}^{n\times n}$, and a strictly feasible iterate point $x_0 \in \text{int}(\Omega)$. Compute g_0, D_0, and $f(x_0)$. Set $E_0 = f(x_0)$ and $k = 0$.

Step 1. If $\|D_k^{\frac{1}{2}} g_k\| < \epsilon$, stop and x_k is an approximate solution to (1.1)–(1.2). Otherwise, go to Step 2.

Step 2. Compute a trial step d_k such that

$$c(x_k) - c(x_k + d_k) \ge \kappa_c[c(x_k) - c(x_k + d_k^C)], \tag{2.14}$$

where d_k^C is the generalized Cauchy point, which is pointed out by Zhao and Sun [32], and $\kappa_c \in (0, 1)$ is a constant.

Step 3. Compute

$$\widehat{Ared}(d_k) = E_k - f(x_k + d_k),$$

$$Pred(d_k) = c(x_k) - c(x_k + d_k),$$

$$\widehat{\rho}_f(d_k) = \frac{\widehat{Ared}(d_k)}{Pred(d_k)}.$$

Step 4. If $\widehat{\rho}_f(d_k) \ge \eta_1$, set $x_{k+1} = x_k + d_k$, and go to Step 6. Otherwise, go to Step 5.

Step 5. Find the smallest $i_k \ge 0$ such that $\alpha_k = \lambda^{i_k}$ ensures the following inequalities,

$$f(x_k + \alpha_k d_k) \le E_k + \delta\alpha_k g_k^T d_k, \tag{2.15}$$

$$g(x_k + \alpha_k d_k)^T d_k \ge \sigma g_k^T d_k, \tag{2.16}$$

$$x_k + \alpha_k d_k \in \text{int}(\Omega). \tag{2.17}$$

Set $x_{k+1} = x_k + \alpha_k d_k$.
Step 6. Update the trust region radius Δ_{k+1} from Δ_k:

$$\Delta_{k+1} = \begin{cases} [\gamma_1\Delta_k, \gamma_2\Delta_k), & \text{if} \quad \widehat{\rho}_f(d_k) \le \eta_1; \\ [\gamma_2\Delta_k, \Delta_k), & \text{if} \quad \eta_1 < \widehat{\rho}_f(d_k) \le \eta_2; \\ [\Delta_k, \min\{\gamma_3\Delta_k, \Delta_{\max}\}), & \text{if} \quad \widehat{\rho}_f(d_k) > \eta_2. \end{cases}$$

Step 7. Compute $b_{k+1} \in \mathbb{R}^n$, $B_{k+1} \in \mathbb{R}^{n\times n}$, $\mu_{k+1} \in (0,1)$ and E_{k+1} by (1.6). Set $k := k+1$ and go to Step 1.

Remarks.

(1) In order to guarantee the global convergence, we choose a sufficiently small constant γ such that

$$\Delta_k\|b_k\| \le \gamma \quad \text{and} \quad 1 - \gamma M_S > 0, \tag{2.18}$$

where $M_S = 1 + \max_k \|S_k^{-1}\|$. In our numerical results, we set $S_k = I$ and $S_k = D_k^{-\frac{1}{2}}$, respectively.

(2) If $\mu_k = 0$ for all k, (2.15)–(2.16) is the traditional line search. If $\mu_k = \eta_k Q_k/Q_{k+1}$, where $\eta_k \in (0,1)$ is a constant, (2.15)–(2.16) becomes Zhang–Hager's nonmonotone line search.

(3) If $b_k = 0$, $c(x_k + d)$ is reduced to a quadratic model. Furthermore, one can see that $c(x_k + d)$ is quadratic along any direction $d \in \mathbb{R}^n$ satisfying $b_k^T d = 0$. In Step 2 of Algorithm 2.1, the conic trust region subproblem (2.11)–(2.13) may be solved by the dogleg method described in [32].

(4) In Step 7, B_{k+1} may be computed by using the damped BFGS formula in order to preserve the positive definiteness of these matrices.

(5) In Algorithm 2.1, we set $\mu_k = \tau_k Q_k/Q_{k+1}$ and μ_k a constant in $(0,1)$, respectively.

3. Convergence analysis

In this section, we establish the global convergence of Algorithm 2.1. Denote the level set as

$$L(x_0) = \{x | f(x) \le f(x_0),\, l \le x \le u\}.$$

First, we give some assumptions as follows.

(A1) $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable and bounded below.

(A2) There exists a constant $K_1 > 0$ such that for all $d \in \mathbb{R}^n$,

$$d^T B_k d \ge K_1\|d\|^2,\ k = 0, 1, 2, \ldots.$$

(A3) There exists a constant $\chi_B > 0$ such that

$$\|D_k^{\frac{1}{2}} B_k D_k^{\frac{1}{2}}\| \le \chi_B.$$

(A4) The level set $L(x_0)$ is bounded.

Assumptions A1 and A4 mean that there exist positive constants M_1 and M_2 such that

$$\|\nabla^2 f(x)\| \le M_1, \quad \|g(x)\| \le M_2, \quad \forall x \in L(x_0).$$

From [1] we know that $D(x)^{-1}$ is bounded in $\Omega \cap B_\rho(x)$ for any $x \in \text{int}(\Omega)$ and $\rho > 0$, where $B_\rho(x) = \{y : \|y - x\| < \rho\}$, so there exists a positive constant χ_D such that

$$\|D_k^{-\frac{1}{2}}\| \le \chi_D, \quad \forall x \in L(x_0). \tag{3.19}$$

Suppose that $\|b_k\| \le M_b$ and $\|B_k\| \le M_B$, where M_b and M_B are positive constants, respectively.

For the convenience of proof, we denote

$$I = \{k|\widehat{\rho}_f(d_k) \geq \eta_1\}, \quad J = \{k|\widehat{\rho}_f(d_k) < \eta_1\}.$$

The following lemma shows that the model function $c(x_k + d_k)$ descends sufficiently. Its proof is similar to the proof of Lemma 4.1 in [32].

Lemma 3.1. *Suppose that Assumptions* A1 *and* A3 *hold, then the following inequality*

$$Pred(d_k) \geq \frac{1}{2}\beta\delta_1\kappa_c\|D_k^{\frac{1}{2}}g_k\| \min\left\{\Delta_k, \frac{\|D_k^{\frac{1}{2}}g_k\|}{\|D_k^{\frac{1}{2}}B_kD_k^{\frac{1}{2}}\|}, \frac{\theta_1\|D_k^{\frac{1}{2}}g_k\|}{\|g_k\|_\infty}\right\}$$

holds for all k, *where* β, δ_1, $\theta_1 \in (0,1)$ *are three constants,* $\kappa_c > 0$ *is a constant independent of* k, d_k *is a trial step that satisfies* (2.14).

The next lemma tells us that the direction of the trial step d_k decreases sufficiently.

Lemma 3.2. *Suppose that Assumptions* A1 *and* A3 *hold, and that* d_k *is generated by the conic trust region subproblem* (2.11)–(2.13)*, then*

$$g_k^Td_k \leq -\frac{1}{2}\beta\delta_1\kappa_c(1-\gamma M_S)\|D_k^{\frac{1}{2}}g_k\| \min\left\{\Delta_k, \frac{\|D_k^{\frac{1}{2}}g_k\|}{\|D_k^{\frac{1}{2}}B_kD_k^{\frac{1}{2}}\|}, \frac{\theta_1\|D_k^{\frac{1}{2}}g_k\|}{\|g_k\|_\infty}\right\}.$$

Proof. Because B_k is updated by the damped BFGS formula, it is positive definite. Then

$$\begin{aligned} Pred(d_k) &= -\frac{g_k^Td_k}{1+b_k^Td_k} - \frac{1}{2}\frac{d_k^TB_kd_k}{(1+b_k^Td_k)^2} \\ &\leq -\frac{g_k^Td_k}{1+b_k^Td_k}. \end{aligned} \tag{3.20}$$

Also, from (2.18), we get

$$|b_k^Td_k| \leq \|b_k\|\|d_k\| = \|b_k\|\|S_k^{-1}S_kd_k\| < M_S\|b_k\|\Delta_k \leq \gamma M_S,$$

and then

$$0 < 1-\gamma M_S < 1+b_k^Td_k < 1+\gamma M_S. \tag{3.21}$$

From (3.20), (3.21) and Lemma 3.1, we obtain

$$\begin{aligned} g_k^Td_k &\leq -(1+b_k^Td_k)Pred(d_k) \\ &\leq -\frac{1}{2}\beta\delta_1\kappa_c(1-\gamma M_S)\|D_k^{\frac{1}{2}}g_k\| \min\left\{\Delta_k, \frac{\|D_k^{\frac{1}{2}}g_k\|}{\|D_k^{\frac{1}{2}}B_kD_k^{\frac{1}{2}}\|}, \frac{\theta_1\|D_k^{\frac{1}{2}}g_k\|}{\|g_k\|_\infty}\right\}, \end{aligned}$$

and the proof is complete. □

Lemma 3.3. *Suppose that Assumptions* A1 *and* A3 *hold, then the following inequality*

$$E_k \geq E_{k+1} \geq f(x_{k+1})$$

holds for all k.

Proof. From (1.6), we have $E_{k+1} - f(x_{k+1}) = \mu_k(E_k - f(x_{k+1}))$. Consider the following two cases.
Case a. If $k \in I$, $x_{k+1} = x_k + d_k$. Then we have

$$\begin{aligned} E_k - f(x_{k+1}) &\geq \frac{\eta_1 \beta \delta_1 \kappa_c}{2} \|D_k^{\frac{1}{2}} g_k\| \min \left\{ \Delta_k, \frac{\|D_k^{\frac{1}{2}} g_k\|}{\|D_k^{\frac{1}{2}} B_k D_k^{\frac{1}{2}}\|}, \frac{\theta_1 \|D_k^{\frac{1}{2}} g_k\|}{\|g_k\|_\infty} \right\} \\ &\geq 0, \end{aligned}$$

which induces $E_k \geq f(x_{k+1})$.
Case b. If $k \in J$, $x_{k+1} = x_k + \alpha_k d_k$. From Lemma 3.2, we get

$$\begin{aligned} g_k^T d_k &\leq -\frac{(1-\gamma M_S)\beta \delta_1 \kappa_c}{2} \|D_k^{\frac{1}{2}} g_k\| \min \left\{ \Delta_k, \frac{\|D_k^{\frac{1}{2}} g_k\|}{\|D_k^{\frac{1}{2}} B_k D_k^{\frac{1}{2}}\|}, \frac{\theta_1 \|D_k^{\frac{1}{2}} g_k\|}{\|g_k\|_\infty} \right\} \\ &\leq 0. \end{aligned}$$

Combining it with (2.15), we also obtain $E_k \geq f(x_{k+1})$.

In both cases, as $E_{k+1} - f(x_{k+1}) = \mu_k(E_k - f(x_{k+1}))$, we have $E_{k+1} \geq f(x_{k+1})$. Furthermore,

$$E_{k+1} - E_k = (1 - \mu_k)(f(x_{k+1}) - E_k) \leq 0,$$

then $E_{k+1} \leq E_k$. Thus, the proof is complete. □

The next lemma shows that the line search step-size α_k is bounded below.

Lemma 3.4. *Suppose that Assumptions* A1, A2, A3 *and* A4 *hold, then*

$$\alpha_k \geq \omega$$

for all $k \in J$, *where* $\omega > 0$ *is a constant.*

Proof. For all $k \in J$, from (2.15) and Lemma 3.3, we have

$$f(x_k + \lambda^{-1}\alpha_k d_k) > E_k + \delta \lambda^{-1} \alpha_k g_k^T d_k \geq f(x_k) + \delta \lambda^{-1} \alpha_k g_k^T d_k. \tag{3.22}$$

Taylor's expansion yields that

$$f(x_k + \lambda^{-1}\alpha_k d_k) = f(x_k) + \lambda^{-1}\alpha_k g_k^T d_k + \frac{\lambda^{-2}}{2} \alpha_k^2 d_k^T \nabla^2 f(\overline{x}_k) d_k, \tag{3.23}$$

where $\overline{x}_k = x_k + \varsigma_k \lambda^{-1} \alpha_k d_k$, $\varsigma_k \in (0,1)$ is a constant. Next, we prove $\overline{x}_k = x_k + \varsigma_k \lambda^{-1}\alpha_k d_k \in L(x_0)$. First, by first order Taylor's expansion, we have

$$f(x_k + \varsigma_k \lambda^{-1}\alpha_k d_k) = f(x_k) + \varsigma_k \lambda^{-1} \alpha_k g_k^T d_k + o(|\varsigma_k \lambda^{-1} \alpha_k g_k^T d_k|), \tag{3.24}$$

where $\varsigma_k \in (0,1)$, $o(|\varsigma_k \lambda^{-1}\alpha_k g_k^T d_k|)$ is high-order infinite smallness of $|\varsigma_k \lambda^{-1} \alpha_k g_k^T d_k|$. Combining it with Lemma 3.2, we know

$$f(x_k + \varsigma_k \lambda^{-1}\alpha_k d_k) \leq f(x_k) \leq f(x_0). \tag{3.25}$$

Second, we will prove $l \leq \overline{x}_k = x_k + \varsigma_k \lambda^{-1}\alpha_k d_k \leq u$. Because $l \leq x_k \leq u$, $l \leq x_k + \lambda^{i_k - 1} d_k \leq u$ (which means that the line search iteration is proceeded in the box constraint for every $k \in J$), we consider the following two cases.

Case a. If $(d_k)_i \geq 0$, by using $\varsigma_k \in (0,1)$, we have

$$l_i \leq (x_k)_i \leq (x_k)_i + \varsigma_k \lambda^{-1} \alpha_k (d_k)_i \leq (x_k)_i + \lambda^{-1}\alpha_k (d_k)_i = (x_k)_i + \lambda^{i_k - 1}(d_k)_i \leq u_i.$$

Case b. If $(d_k)_i < 0$, by using $\varsigma_k \in (0,1)$, we obtain

$$l_i \le (x_k)_i + \lambda^{i_k - 1}(d_k)_i = (x_k)_i + \lambda^{-1}\alpha_k(d_k)_i \le (x_k)_i + \varsigma_k\lambda^{-1}\alpha_k(d_k)_i \le (x_k)_i \le u_i.$$

Here, l_i, $(x_k)_i$, $(d_k)_i$, u_i are the ith components of l, x_k, d_k, u, respectively. Thus, we have $\overline{x}_k = x_k + \varsigma_k\lambda^{-1}\alpha_k d_k \in L(x_0)$. Combining (3.22), (3.23), Assumptions A1 and A4 yields that

$$\delta\lambda^{-1}\alpha_k g_k^T d_k < \lambda^{-1}\alpha_k g_k^T d_k + \frac{\lambda^{-2}}{2}\alpha_k^2 d_k^T \nabla^2 f(\xi_k) d_k < \lambda^{-1}\alpha_k g_k^T d_k + \frac{\lambda^{-2}}{2}\alpha_k^2 M_1 \|d_k\|^2,$$

then

$$-(1-\delta) g_k^T d_k < \frac{\lambda^{-1}\alpha_k}{2} M_1 \|d_k\|^2. \tag{3.26}$$

Because

$$\begin{aligned} Pred(d_k) &= -\frac{g_k^T d_k}{1 + b_k^T d_k} - \frac{1}{2}\frac{d_k^T B_k d_k}{(1 + b_k^T d_k)^2} \\ &\ge \frac{\beta\delta_1\kappa_c}{2}\|D_k^{\frac{1}{2}} g_k\| \min\left\{\Delta_k, \frac{\|D_k^{\frac{1}{2}} g_k\|}{\|D_k^{\frac{1}{2}} B_k D_k^{\frac{1}{2}}\|}, \frac{\theta_1\|D_k^{\frac{1}{2}} g_k\|}{\|g_k\|_\infty}\right\} \\ &\ge 0, \end{aligned}$$

combining (3.21) and Assumption A2 yields that

$$-g_k^T d_k \ge \frac{d_k^T B_k d_k}{2(1 + b_k^T d_k)} \ge \frac{K_1\|d_k\|^2}{2(1 + \gamma M_S)}. \tag{3.27}$$

From (3.26) and (3.27), we have

$$\alpha_k \ge \frac{\lambda(1-\delta)K_1}{M_1(1+\gamma M_S)} \stackrel{\text{def}}{=} \omega,$$

and the proof is complete. □

Lemma 3.5. *Suppose that Assumptions* A1, A2 *and* A3 *hold, and that the sequence* $\{x_k\}$ *is generated by Algorithm* 2.1, *then* $\{x_k\} \in L(x_0)$.

Proof. If $k = 0$, the result holds obviously. Assume that $x_k \in L(x_0)$, we prove first that $E_k \le f(x_0)$.

When $j = 0$, $E_0 = f(x_0)$, it is obvious. Assume that the inequality $E_j \le f(x_0)$ holds when $j = k-1$. When $j = k$, from (1.6), we have

$$\begin{aligned} E_k &= \mu_{k-1}E_{k-1} + (1-\mu_{k-1})f(x_k) \\ &\le \mu_{k-1}f(x_0) + (1-\mu_{k-1})f(x_0) \\ &= f(x_0). \end{aligned}$$

So for all $x_k \in L(x_0)$, we have $E_k \le f(x_0)$. From Lemma 3.3, we have

$$f(x_{k+1}) \le E_{k+1} \le E_k \le f(x_0).$$

Furthermore, $l \le x_{k+1} \le u$, so $x_{k+1} \in L(x_0)$. □

Lemma 3.6. *Suppose that Assumptions* A1, A2, A3 *and* A4 *hold, then*

$$E_k - E_{k+1} \ge (1-\mu_k)\delta_2\|D_k^{\frac{1}{2}} g_k\| \min\left\{\Delta_k, \frac{\|D_k^{\frac{1}{2}} g_k\|}{\|D_k^{\frac{1}{2}} B_k D_k^{\frac{1}{2}}\|}, \frac{\theta_1\|D_k^{\frac{1}{2}} g_k\|}{\|g_k\|_\infty}\right\},$$

where $\delta_2 > 0$ *is a constant.*

Proof.

(1) For $k \in I$, $x_{k+1} = x_k + d_k$, from Lemma 3.1, we have

$$\begin{aligned} E_k - f(x_{k+1}) &\geq \eta_1 Pred(d_k) \\ &\geq \frac{\eta_1 \beta \delta_1 \kappa_c}{2} \|D_k^{\frac{1}{2}} g_k\| \min\left\{ \Delta_k, \frac{\|D_k^{\frac{1}{2}} g_k\|}{\|D_k^{\frac{1}{2}} B_k D_k^{\frac{1}{2}}\|}, \frac{\theta_1 \|D_k^{\frac{1}{2}} g_k\|}{\|g_k\|_\infty} \right\}. \end{aligned}$$

(2) For $k \in J$, $x_{k+1} = x_k + \alpha_k d_k$, from Lemma 3.2, we have

$$g_k^T d_k \leq -\frac{(1-\gamma M_S)\beta \delta_1 \kappa_c}{2} \|D_k^{\frac{1}{2}} g_k\| \min\left\{ \Delta_k, \frac{\|D_k^{\frac{1}{2}} g_k\|}{\|D_k^{\frac{1}{2}} B_k D_k^{\frac{1}{2}}\|}, \frac{\theta_1 \|D_k^{\frac{1}{2}} g_k\|}{\|g_k\|_\infty} \right\}.$$

Thus, from (2.15) we obtain

$$f(x_{k+1}) \leq E_k - \frac{\delta \alpha_k (1-\gamma M_S)\beta \delta_1 \kappa_c}{2} \|D_k^{\frac{1}{2}} g_k\| \min\left\{ \Delta_k, \frac{\|D_k^{\frac{1}{2}} g_k\|}{\|D_k^{\frac{1}{2}} B_k D_k^{\frac{1}{2}}\|}, \frac{\theta_1 \|D_k^{\frac{1}{2}} g_k\|}{\|g_k\|_\infty} \right\}.$$

From Lemma 3.4, we get

$$f(x_{k+1}) \leq E_k - \frac{\delta \omega (1-\gamma M_S)\beta \delta_1 \kappa_c}{2} \|D_k^{\frac{1}{2}} g_k\| \min\left\{ \Delta_k, \frac{\|D_k^{\frac{1}{2}} g_k\|}{\|D_k^{\frac{1}{2}} B_k D_k^{\frac{1}{2}}\|}, \frac{\theta_1 \|D_k^{\frac{1}{2}} g_k\|}{\|g_k\|_\infty} \right\}.$$

Set $\delta_2 \stackrel{\text{def}}{=} \min\left\{ \frac{\eta_1 \beta \delta_1 \kappa_c}{2}, \frac{\delta \omega (1-\gamma M_S)\beta \delta_1 \kappa_c}{2} \right\}$, then we have

$$f(x_{k+1}) \leq E_k - \delta_2 \|D_k^{\frac{1}{2}} g_k\| \min\left\{ \Delta_k, \frac{\|D_k^{\frac{1}{2}} g_k\|}{\|D_k^{\frac{1}{2}} B_k D_k^{\frac{1}{2}}\|}, \frac{\theta_1 \|D_k^{\frac{1}{2}} g_k\|}{\|g_k\|_\infty} \right\}.$$

Combining it with (1.6), we obtain

$$\begin{aligned} E_{k+1} &= \mu_k E_k + (1-\mu_k) f(x_{k+1}) \\ &\leq E_k - (1-\mu_k)\delta_2 \|D_k^{\frac{1}{2}} g_k\| \min\left\{ \Delta_k, \frac{\|D_k^{\frac{1}{2}} g_k\|}{\|D_k^{\frac{1}{2}} B_k D_k^{\frac{1}{2}}\|}, \frac{\theta_1 \|D_k^{\frac{1}{2}} g_k\|}{\|g_k\|_\infty} \right\}, \end{aligned}$$

and the proof is complete. □

Next, we prove that the trust region radius Δ_k is bounded away from zero. The proof is discussed by considering two cases. The first one is the trust region iteration, *i.e.*, $x_{k+1} = x_k + d_k$; the second one is the line search iteration, *i.e.*, $x_{k+1} = x_k + \alpha_k d_k$.

For trust region iteration, we first prove some technical lemmas.

The following lemma shows the discrepancy between the objective function and the model function.

Lemma 3.7. *Suppose that Assumptions* A1, A3 *and* A4 *hold,* d_k *is the solution of the problem* (2.11)–(2.13), *then*

$$|[f(x_k) - f(x_k + d_k)] - [c(x_k) - c(x_k + d_k)]| \leq \overline{M} \Delta_k^2.$$

Proof. From (3.21), we have

$$0 < 1 - \gamma M_S < 1 + b_k^T d_k < 1 + \gamma M_S.$$

Taylor's expansion yields that

$$f(x_k + d_k) = f(x_k) + g_k^T d_k + \frac{1}{2} d_k^T \nabla^2 f(x_k + \zeta_k d_k) d_k, \tag{3.28}$$

where $\zeta_k \in (0,1)$ is a constant. Next, we prove $x_k + \zeta_k d_k \in L(x_0)$. First, by first order Taylor's expansion, we obtain

$$f(x_k + \zeta_k d_k) = f(x_k) + \zeta_k g_k^T d_k + o(\|\zeta_k d_k\|),$$

where $o(\|\zeta_k d_k\|)$ is high-order infinite smallness of $\|\zeta_k d_k\|$. Combining it with Lemma 3.2, we know

$$f(x_k + \zeta_k d_k) \le f(x_k) \le f(x_0). \tag{3.29}$$

Second, we will prove $l \le x_k + \zeta_k d_k \le u$. Because $l \le x_k \le u$ and $l \le x_k + d_k \le u$ hold for $k \in I$, we consider the following two cases.

Case a. If $(d_k)_i \ge 0$, by using $\zeta_k \in (0,1)$, we get

$$l_i \le (x_k)_i \le (x_k)_i + \zeta_k (d_k)_i \le (x_k)_i + (d_k)_i \le u_i.$$

Case b. If $(d_k)_i < 0$, by using $\zeta_k \in (0,1)$, we obtain

$$l_i \le (x_k)_i + (d_k)_i \le (x_k)_i + \zeta_k (d_k)_i \le (x_k)_i \le u_i.$$

Here, l_i, $(x_k)_i$, $(d_k)_i$, u_i are the i-th component of l, x_k, d_k, u, respectively. Thus, we have $x_k + \zeta_k d_k \in L(x_0)$. Using (3.28), Assumptions A1 and A4 yield that

$$\begin{aligned}
&|[f(x_k) - f(x_k + d_k)] - [c(x_k) - c(x_k + d_k)]| \\
&= \left| g_k^T d_k + \frac{1}{2} d_k^T \nabla^2 f(x_k + \zeta_k d_k) d_k - \frac{g_k^T d_k}{1 + b_k^T d_k} - \frac{1}{2} \frac{d_k^T B_k d_k}{(1 + b_k^T d_k)^2} \right| \\
&= \left| \frac{g_k^T d_k b_k^T d_k}{1 + b_k^T d_k} + \frac{1}{2} d_k^T \nabla^2 f(x_k + \zeta_k d_k) d_k - \frac{1}{2} \frac{d_k^T B_k d_k}{(1 + b_k^T d_k)^2} \right| \\
&\le \left(\frac{M_2 M_b}{1 - \gamma M_S} + \frac{M_1}{2} + \frac{M_B}{2(1 - \gamma M_S)^2} \right) \|d_k\|^2 \\
&\le M_S^2 \left(\frac{M_2 M_b}{1 - \gamma M_S} + \frac{M_1}{2} + \frac{M_B}{2(1 - \gamma M_S)^2} \right) \Delta_k^2 \\
&= \overline{M} \Delta_k^2,
\end{aligned}$$

where $\overline{M} \overset{\text{def}}{=} M_S^2 \left(\frac{M_2 M_b}{1 - \gamma M_S} + \frac{M_1}{2} + \frac{M_B}{2(1 - \gamma M_S)^2} \right)$. □

Lemma 3.8. *Suppose that Assumptions* A1, A2, A3 *and* A4 *hold, and that there exists a positive constant* $\epsilon \ge 0$ *such that* $\|D_k^{\frac{1}{2}} g_k\| \ge \epsilon$, *and* $\Delta_k < \min\left\{\Delta_{\max}, \frac{\epsilon}{\chi_B}, \frac{\theta_1}{\chi_D}, \frac{(1-\eta_2)\beta\delta_1\kappa_c\epsilon}{2\overline{M}}\right\}$, *then* $\widehat{\rho}_f(d_k) > \eta_2$ *and* $\Delta_{k+1} \ge \Delta_k$.

Proof. From (3.19) and Lemma 3.5, we get

$$\|g_k\|_\infty \le \|g_k\| = \left\| D_k^{-\frac{1}{2}} D_k^{\frac{1}{2}} g_k \right\| \le \chi_D \left\| D_k^{\frac{1}{2}} g_k \right\|. \tag{3.30}$$

From Lemma 3.1, Lemma 3.7 and Assumption A3, we have

$$\begin{aligned}\left|\frac{f(x_k)-f(x_k+d_k)}{Pred(d_k)}-1\right| &= \left|\frac{[f(x_k)-f(x_k+d_k)]-[c(x_k)-c(x_k+d_k)]}{Pred(d_k)}\right| \\ &\le \frac{\overline{M}\Delta_k^2}{\frac{\beta\delta_1\kappa_c}{2}\|D_k^{\frac{1}{2}}g_k\|\min\left\{\Delta_k, \frac{\|D_k^{\frac{1}{2}}g_k\|}{\|D_k^{\frac{1}{2}}B_kD_k^{\frac{1}{2}}\|}, \frac{\theta_1\|D_k^{\frac{1}{2}}g_k\|}{\|g_k\|_\infty}\right\}} \\ &\le \frac{\overline{M}\Delta_k^2}{\frac{\beta\delta_1\kappa_c\epsilon}{2}\min\left\{\Delta_k, \frac{\epsilon}{\chi_B}, \frac{\theta_1}{\chi_D}\right\}} \\ &= \frac{2\overline{M}\Delta_k}{\beta\delta_1\kappa_c\epsilon} \\ &< 1-\eta_2,\end{aligned}$$

then

$$\frac{f(x_k)-f(x_k+d_k)}{Pred(d_k)} > \eta_2. \tag{3.31}$$

From (3.31) and Lemma 3.3, we have

$$\widehat{\rho}_f(d_k) = \frac{E_k - f(x_k+d_k)}{Pred(d_k)} \ge \frac{f(x_k)-f(x_k+d_k)}{Pred(d_k)} > \eta_2.$$

So $\Delta_{k+1} \ge \Delta_k$. □

The next lemma shows that if the current iterate isn't a first order critical point, the trust region radius Δ_k is bounded away from zero for trust region iteration.

Lemma 3.9. *Suppose that Assumptions* A1, A2, A3 *and* A4 *hold, and that there exists a constant* $\epsilon > 0$ *such that* $\|D_k^{\frac{1}{2}}g_k\| \ge \epsilon$, *then there exists a constant* $\Delta_{lbd1} > 0$ *such that* $\Delta_k \ge \Delta_{lbd1}$ *for* $k \in I$.

Proof. Suppose that k is the first index satisfying

$$\Delta_{k+1} \le \gamma_1 \min\left\{\Delta_{\max}, \frac{\epsilon}{\chi_B}, \frac{\theta_1}{\chi_D}, \frac{(1-\eta_2)\beta\delta_1\kappa_c\epsilon}{2\overline{M}}\right\} \stackrel{\text{def}}{=} \gamma_1\delta_0, \tag{3.32}$$

which means that the index $k-1$ doesn't satisfy (3.32), then

$$\Delta_k > \gamma_1\delta_0 \ge \Delta_{k+1}. \tag{3.33}$$

From Step 6 of Algorithm 2.1, we have $\gamma_1\Delta_k \le \Delta_{k+1}$, so

$$\Delta_k \le \delta_0 = \min\left\{\Delta_{\max}, \frac{\epsilon}{\chi_B}, \frac{\theta_1}{\chi_D}, \frac{(1-\eta_2)\beta\delta_1\kappa_c\epsilon}{2\overline{M}}\right\}.$$

From Lemma 3.8, we know $\Delta_{k+1} \ge \Delta_k$, which contradicts (3.33). So $\Delta_k > \Delta_{lbd1}$, where $\Delta_{lbd1} \stackrel{\text{def}}{=} \gamma_1 \min\{\Delta_{\max}, \frac{\epsilon}{\chi_B}, \frac{\theta_1}{\chi_D}, \frac{(1-\eta_2)\beta\delta_1\kappa_c\epsilon}{2\overline{M}}\}$. We complete the proof. □

Next, we prove that for line search iteration, if the current iterate isn't a first order critical point, then the trust region radius Δ_k is bounded away from zero for sufficiently large k.

Lemma 3.10. *Suppose that Assumptions* A1, A2, A3 *and* A4 *hold, and that there exists a constant* $\epsilon > 0$ *such that* $\|D_k^{\frac{1}{2}} g_k\| \geq \epsilon$, *then there exists a constant* $\Delta_{lbd2} > 0$ *such that* $\Delta_k \geq \Delta_{lbd2}$ *for* $k \in J$ *sufficiently large.*

Proof. For $k \in J$, $x_{k+1} = x_k + \alpha_k d_k$, from Lemmas 3.2 and 3.3, we have

$$\begin{aligned}
&0\\
&> E_k - f(x_k + \lambda^{-1}\alpha_k d_k) + \delta\lambda^{-1}\alpha_k g_k^T d_k\\
&\geq f(x_k) - f(x_k + \lambda^{-1}\alpha_k d_k) + \delta\lambda^{-1}\alpha_k g_k^T d_k\\
&\geq -(1-\delta)\lambda^{-1}\alpha_k g_k^T d_k - \frac{\lambda^{-2}}{2}\alpha_k^2 M_1 \|d_k\|^2\\
&\geq \frac{\beta\delta_1\kappa_c\epsilon}{2}(1-\delta)\lambda^{-1}\alpha_k(1-\gamma M_S)\min\left\{\Delta_k, \frac{\epsilon}{\chi_B}, \frac{\theta_1}{\chi_D}\right\} - \frac{\lambda^{-2}}{2}\alpha_k^2 M_1\|d_k\|^2\\
&= \lambda^{-1}\alpha_k\Big[\frac{\beta\delta_1\kappa_c\epsilon}{2}(1-\delta)(1-\gamma M_S)\min\left\{\Delta_k, \frac{\epsilon}{\chi_B}, \frac{\theta_1}{\chi_D}\right\} - \frac{\lambda^{-1}}{2}M_1 M_S \Delta_k\|x_{k+1}-x_k\|\Big]\\
&= \lambda^{-1}\alpha_k\Delta_k\left[\frac{\beta\delta_1\kappa_c\epsilon}{2}(1-\delta)(1-\gamma M_S)\min\left\{1, \frac{\epsilon}{\Delta_k\chi_B}, \frac{\theta_1}{\Delta_k\chi_D}\right\} - \frac{\lambda^{-1}}{2}M_1 M_S\|x_{k+1}-x_k\|\right].
\end{aligned}$$

So

$$\|x_{k+1} - x_k\| > \frac{\lambda(1-\delta)(1-\gamma M_S)\beta\delta_1\kappa_c\epsilon\min\left\{1, \frac{\epsilon}{\Delta_k\chi_B}, \frac{\theta_1}{\Delta_k\chi_D}\right\}}{M_1 M_S}. \tag{3.34}$$

Next, we prove $\Delta_k \leq \min\{\frac{\epsilon}{\chi_B}, \frac{\theta_1}{\chi_D}\}$ holds for $k \in J$ sufficiently large by contradiction. If $\Delta_k > \min\left\{\frac{\epsilon}{\chi_B}, \frac{\theta_1}{\chi_D}\right\}$ holds for $k \in J$ sufficiently large, denoting

$$\Delta_l \overset{\text{def}}{=} \min\left\{\frac{\epsilon}{\chi_B}, \frac{\theta_1}{\chi_D}, \gamma_1\Delta_{\max}, \frac{\gamma_1(1-\eta_2)\beta\delta_1\kappa_c\epsilon}{2M_2}\right\},$$

then we have $\Delta_k > \Delta_l$. From Lemma 3.6, we have

$$E_k - E_{k+1} \geq (1-\mu_k)\delta_2\epsilon\min\left\{\Delta_l, \frac{\epsilon}{\chi_B}, \frac{\theta_1}{\chi_D}\right\} \geq 0. \tag{3.35}$$

From Assumption A1, we know $E_k \geq f(x_k) \geq f_{lbd}$, where f_{lbd} is the lower bound of $f(x_k)$. Combining it with the decreasing sequence $\{E_k\}$, we have $\{E_k\}$ converges. From (3.35), we obtain $\lim\limits_{k\to\infty}\mu_k = 1$, which contradicts the fact that $\mu_k \in (0,1)$. So $\Delta_k \leq \min\left\{\frac{\epsilon}{\chi_B}, \frac{\theta_1}{\chi_D}\right\}$ holds for $k \in J$ sufficiently large.

From (3.34), we have

$$\|x_{k+1} - x_k\| > \frac{\lambda(1-\delta)(1-\gamma M_S)\beta\delta_1\kappa_c\epsilon}{M_1 M_S}.$$

On the other hand, $\|x_{k+1} - x_k\| = \alpha_k\|d_k\| < M_S\Delta_k$, so

$$\Delta_k > \frac{\lambda(1-\delta)(1-\gamma M_S)\beta\delta_1\kappa_c\epsilon}{M_1 M_S^2} \overset{\text{def}}{=} \Delta_{lbd2}.$$

We complete the proof. □

From Lemmas 3.9 and 3.10, we know that for sufficiently large k, the trust region radius Δ_k is bounded below, the lower bound is

$$\Delta_{lbd} = \min\{\Delta_{lbd1}, \Delta_{lbd2}\}. \tag{3.36}$$

The next theorem shows that at least one of the limit points (if any) is a first order critical point.

Theorem 3.11. *Suppose that Assumptions* A1, A2, A3 *and* A4 *hold, then the sequence* $\{x_k\}$ *generated by Algorithm* 2.1 *satisfies*

$$\liminf_{k\to\infty} \|D_k^{\frac{1}{2}} g_k\| = 0.$$

Proof. Suppose that the conclusion doesn't hold, *i.e.*, there exists a constant $\epsilon > 0$ such that $\|D_k^{\frac{1}{2}} g_k\| \geq \epsilon$. From Lemma 3.6, we have

$$\begin{aligned} E_k - E_{k+1} &\geq (1-\mu_k)\delta_2 \|D_k^{\frac{1}{2}} g_k\| \min\left\{\Delta_k, \frac{\|D_k^{\frac{1}{2}} g_k\|}{\|D_k^{\frac{1}{2}} B_k D_k^{\frac{1}{2}}\|}, \frac{\theta_1 \|D_k^{\frac{1}{2}} g_k\|}{\|g_k\|_\infty}\right\} \\ &\geq (1-\mu_k)\delta_2 \epsilon \min\left\{\Delta_k, \frac{\epsilon}{\chi_B}, \frac{\theta_1}{\chi_D}\right\}. \end{aligned} \tag{3.37}$$

Case a. when $\mu_k \in (0,1)$ is a constant for all k, from (3.37), we have $\lim_{k\to\infty} \Delta_k = 0$, which contradicts (3.36).

Case b. when $\mu_k = \tau_k Q_k / Q_{k+1}$, from the definition of Q_{k+1}, we have $Q_{k+1} \leq k+2$. Combining (3.36) and (3.37) yields that

$$\begin{aligned} E_k - E_{k+1} &\geq \frac{1}{Q_{k+1}} \delta_2 \epsilon \min\left\{\Delta_{lbd}, \frac{\epsilon}{\chi_B}, \frac{\theta_1}{\chi_D}\right\} \\ &\geq \frac{1}{k+2} \delta_2 \epsilon \min\left\{\Delta_{lbd}, \frac{\epsilon}{\chi_B}, \frac{\theta_1}{\chi_D}\right\}. \end{aligned}$$

Summing up the above inequality from $k = 0$ to j, we obtain

$$E_0 - E_{j+1} \geq \sum_{k=0}^{j} \frac{1}{k+2} \delta_2 \epsilon \min\left\{\Delta_{lbd}, \frac{\epsilon}{\chi_B}, \frac{\theta_1}{\chi_D}\right\}.$$

From Lemma 3.3, we have $E_{j+1} \geq f(x_{j+1}) \geq f_{lbd}$, then

$$f(x_0) - f_{lbd} \geq f(x_0) - E_{j+1} = E_0 - E_{j+1} \geq \sum_{k=0}^{j} \frac{1}{k+2} \delta_2 \epsilon \min\left\{\Delta_{lbd}, \frac{\epsilon}{\chi_B}, \frac{\theta_1}{\chi_D}\right\}.$$

Thus

$$\sum_{k=0}^{j} \frac{1}{k+2} \leq \frac{f(x_0) - f_{lbd}}{\delta_2 \epsilon \min\left\{\Delta_{lbd}, \frac{\epsilon}{\chi_B}, \frac{\theta_1}{\chi_D}\right\}} \stackrel{\text{def}}{=} \omega_1. \tag{3.38}$$

Set j tending to $+\infty$, then (3.38) contradicts the fact that the series $\sum_{k=0}^{\infty} \frac{1}{k+2}$ diverges. We complete the proof. ☐

In order to prove the local convergence rate, we add some additional assumptions.

(A5) The sequence $\{x_k\}$ generated by Algorithm 2.1 converges to a stationary point x^*, *i.e.*,

$$\lim_{k\to\infty} x_k = x^* \quad \text{and} \quad \lim_{k\to\infty} \|g_k\| = \|g^*\| = 0.$$

(A6) If

$$\frac{\|B_k^{-1}g_k\|}{1+b_k^TB_k^{-1}g_k} \le \Delta_k,$$

then

$$d_k = -\frac{B_k^{-1}g_k}{1+b_k^TB_k^{-1}g_k}.$$

The local convergence is similar to Theorem 4.2 in [17], Theorem 5.10 in [22] and Theorem 3.2 in [12], so we omit the proof here.

Theorem 3.12. *Suppose that Assumptions* A1, A2, A3, A4, A5 *and* A6 *hold,* $\nabla^2 f(x)$ *is positive definite and Lipschitz continuous in the neighborhood of* x^*. *If*

$$\lim_{k\to\infty} \frac{\|(B_k - \nabla^2 f(x_k))d_k\|}{\|d_k\|} = 0,$$

then the sequence $\{x_k\}$ *converges to* x^* *Q-superlinearly.*

4. Numerical results

In this section, the Algorithm 2.1 is tested on a set of standard testing problems which appeared in [5,14,30]. A MATLAB program is coded to perform the experiments.

The fixed constants are $\eta_1 = 0.001$, $\eta_2 = 0.75$, $\gamma_1 = 0.2$, $\gamma_2 = 0.5$, $\gamma_3 = 2$, $\delta = 0.2$, $b_0 = 0$, $\lambda = 0.5$, $\tau_k = 0.85$, $\epsilon = 10^{-6}$, $\Delta_0 = 5$, $\Delta_{\max} = 10$, $B_0 = I$, $\sigma = 0.9$, $\zeta = 1 - \epsilon$. The stopping criterion is $\|D_k^{\frac{1}{2}} g_k\| \le \epsilon$, the maximal number of iteration is 5000. In order to compute the optimal solution by our conic model method, we set

$$d_{k-1} = x_k - x_{k-1},$$

$$\rho_k = (f(x_{k-1}) - f(x_k))^2 - (g_{k-1}^T d_{k-1})(g_k^T d_{k-1}),$$

$$\beta_k = \begin{cases} \frac{(f(x_{k-1})-f(x_k))+\sqrt{\rho_k}}{-g_{k-1}^T d_{k-1}}, & \text{if} \quad \rho_k > 0; \\ 1, & \text{otherwise.} \end{cases}$$

$$b_k = \frac{\beta_k - 1}{g_{k-1}^T d_{k-1}} g_{k-1},$$

$$y_{k-1} = \beta_k g_k - \beta_k^3 g_{k-1},$$

$$B_k = \begin{cases} B_{k-1} + \dfrac{y_{k-1}y_{k-1}^T}{d_{k-1}^T y_{k-1}} - \dfrac{B_{k-1}d_{k-1}d_{k-1}^T B_{k-1}}{d_{k-1}^T B_{k-1} d_{k-1}}, & \text{if} \quad d_{k-1}^T y_{k-1} > 0; \\ B_{k-1}, & \text{otherwise.} \end{cases}$$

In our nonmonotone conic trust region method with line search technique, we set the initial point $x_0 \in \text{int}(\Omega)$. If the initial point $x_0 \notin \text{int}(\Omega) = \{x | l < x < u\}$, we project it to Ω by $x_0 = \max(l, \min(u, x_0))$; when $x_0 = l$, we set $x_0 := l + \zeta l$; when $x_0 = u$, we set $x_0 := u - \zeta u$. So, we can ensure $x_0 \in \text{int}(\Omega) = \{x | l < x < u\}$.

Tables 1 and 2 are the numerical results when $\mu_k = \tau_k Q_k / Q_{k+1}$, Table 3 is the numerical results of conic trust region method (abbreviated as CTR) and NCTRLS for different choices of μ_k, where n is the dimension of the test problems, n_g is the number of gradient evaluations, n_f is the number of function evaluations, $f_{\min}$ is the final objective function value, and $\|D^{\frac{1}{2}} g\|$ is the norm of the final gradient value. $m(n)$ stands for $m \times 10^n$. The sign $'F'$ means that when the number of iteration reaches 5000, the algorithm fails to reach a minimum, and $'O'$ means overflow. For the numerical experiments in Tables 1–3, the performance profiles are displayed in Figures 1–4 according to Dolan and Moré [10].

TABLE 1. Computation results when $S_k = I$ and $\mu_k = \tau_k Q_k / Q_{k+1}$.

		NTRLS				NCTRLS			
Fun. Name	n	n_g	n_f	$f_{\min}$	$\|D^{\frac{1}{2}} g\|$	n_g	n_f	$f_{\min}$	$\|D^{\frac{1}{2}} g\|$
HS003	2	5	5	2.250(−17)	4.743(−09)	6	7	1.125(−21)	3.354(−11)
Gen. Ros	8	258	278	3.781(−16)	7.731(−07)	160	176	2.552(−19)	2.036(−08)
Cha. Sin	20	F	F	F	F	159	182	1.903(−11)	8.121(−07)
Cragglevy	50	F	F	F	F	448	485	1.467(+01)	8.856(−07)
Brown1	10	O	O	O	O	158	166	9.989(−01)	5.215(−08)
Brown1	50	F	F	F	F	280	319	4.995(+00)	8.105(−07)
Brown1	100	O	O	O	O	468	562	9.989(+00)	5.943(−07)
Gen. Wood	20	199	213	1.000(+00)	7.026(−07)	242	263	1.000(+00)	2.288(−07)
Cha. Wood	8	148	156	1.000(+00)	3.375(−07)	209	223	1.000(+00)	5.245(−07)
Brown3	100	19	19	3.125(−15)	1.511(−07)	20	20	3.210(−14)	4.983(−07)

Table 1 is the numerical results of NCTRLS (nonmonotone conic trust region method with line search) and NTRLS (nonmonotone trust region method with line search based on quadratic model). From Figures 1 and 2, we can see that NCTRLS performs better than NTRLS, especially for ill-conditioned problems.

Table 2 is the numerical comparison of CTR and NCTRLS. In our numerical results, the trust region scaling matrix S_k is set to $S_k = I$ and $S_k = D_k^{-\frac{1}{2}}$, respectively. When $S_k = I$, in 8 out of 11 problems, NCTRLS can solve the problems with less function and gradient evaluations than CTR. Both NCTRLS and CTR have the same results in solving HS005 problem. For Cragglevy problem and Brown1 problem, NCTRLS performs worse than CTR. When $S_k = D_k^{-\frac{1}{2}}$, NCTRLS performs almost the same as CTR. However, for ill-conditioned problems, NCTRLS performs better than CTR. These numerical tests indicate that, for conic trust-region methods, the nonmonotone strategy is also efficient.

Table 3 is the numerical results of CTR and NCTRLS for different choices of μ_k, where μ_k is a constant and $S_k = I$. Since the case $S_k = D_k^{-\frac{1}{2}}$ is similar to the case $S_k = I$ for our tests, so we omit here. In our numerical results, we choose $\mu_k = 0.15$, 0.5, 0.85, respectively. From Table 3, we can see that when $\mu_k = 0.15$, NCTRLS performs better than the other two choices. We also make comparisons between CTR and NCTRLS, and find that NCTRLS can solve more problems with less function or gradient evaluations than CTR, especially when $\mu_k = 0.15$.

From Figures 3 and 4, we can see that NCTRLS performs better than CTR in the sense that NCTRLS method needs less evaluations of function and gradient values for most of problems, especially for ill-conditioned problems.

5. CONCLUSIONS

In this paper, we propose a nonmonotone conic trust region method with line search technique for bound constrained optimization problem. When the trial point isn't accepted by the trust region, we perform the line search technique until an acceptable trial point is found instead of resolving the trust region subproblem, which reduces the total cost of computation to some extent. The global convergence and Q-superlinear convergence are established under some reasonable conditions. Numerical results show that the new method is effective for bound constrained optimization problems. The algorithm and theory in this paper can be extended to general constrained optimization problem, which is our future work.

TABLE 2. Computation results when $\mu_k = \tau_k Q_k / Q_{k+1}$.

			CTR				NCTRLS			
Scal. Matrix	Fun. Name	n	n_g	n_f	$f_{\min}$	$\|D^{\frac{1}{2}} g\|$	n_g	n_f	$f_{\min}$	$\|D^{\frac{1}{2}} g\|$
$S = I$	HS001	2	41	47	1.440(−20)	1.465(−09)	27	28	2.310(−17)	9.229(−08)
	HS003	2	19	26	2.318(−18)	1.522(−09)	6	7	1.125(−21)	3.354(−11)
	HS005	2	10	11	−1.913(+00)	1.170(−07)	10	11	−1.913(+00)	9.731(−08)
	HS038	4	93	105	1.285(−18)	8.325(−08)	87	88	4.590(−17)	3.921(−07)
	Broydenu	50	190	208	1.000(+00)	5.396(−07)	161	171	1.008(+00)	5.654(−07)
	Cha. Sin	20	178	200	1.232(−11)	9.828(−07)	159	182	1.903(−11)	8.121(−07)
	Gen. Wood	100	761	832	1.000(+00)	8.110(−07)	642	729	1.000(+00)	7.998(−07)
	Cha. Wood	50	559	606	1.000(+00)	2.590(−07)	449	510	1.000(+00)	9.879(−07)
	Cragglevy	8	87	90	8.977(−10)	4.976(−07)	153	155	2.478(−10)	7.836(−07)
	Brown1	10	79	93	9.989(−01)	4.630(−07)	158	166	9.989(−01)	5.215(−08)
	Brown1	50	348	368	4.995(+00)	3.775(−07)	280	319	4.995(+00)	8.105(−07)
$S = D^{-\frac{1}{2}}$	HS001	2	41	47	1.440(−20)	1.465(−09)	27	28	2.310(−17)	9.229(−08)
	HS003	2	19	26	2.318(−18)	1.522(−09)	6	7	1.125(−21)	3.354(−11)
	HS035	2	9	10	−1.913(+00)	1.212(−08)	13	15	−1.913(+00)	3.392(−08)
	HS038	4	60	64	4.485(−17)	4.538(−07)	141	153	9.114(−20)	2.973(−08)
	Broydenu	50	130	139	1.008(+00)	3.265(−07)	169	180	1.008(+00)	3.591(−07)
	Cha. Sin	20	178	200	1.232(−11)	9.828(−07)	159	182	1.903(−11)	8.121(−07)
	Gen. Wood	100	761	832	1.000(+00)	8.110(−07)	642	729	1.000(+00)	7.998(−07)
	Cha. Wood	50	559	606	1.000(+00)	2.590(−07)	449	510	1.000(+00)	9.879(−07)
	Cragglevy	8	87	90	8.977(−10)	4.976(−07)	153	155	2.478(−10)	7.836(−07)
	Brown1	10	18	22	9.989(−01)	1.649(−08)	111	113	9.989(−01)	5.690(−07)
	Brown1	50	F	F	F	F	340	386	4.995(+00)	3.485(−07)

TABLE 3. Numerical comparison of CTR and different choices of μ_k for NCTRLS when $S_k = I$.

Fun. Name	n	CTR	NCTRLS($\mu_k = 0.15$)	NCTRLS($\mu_k = 0.5$)	NCTRLS($\mu_k = 0.85$)
		$n_g/n_f/f_{\min}$	$n_g/n_f/f_{\min}$	$n_g/n_f/f_{\min}$	$n_g/n_f/f_{\min}$
HS001	2	41/47/1.440(−20)	30/33/5.614(−18)	27/28/2.310(−17)	27/28/2.310(−17)
HS002	2	F/F/F	O/O/O	O/O/O	F/F/F
HS003	2	19/26/2.318(−18)	6/7/1.125(−21)	6/7/1.125(−21)	6/7/1.125(−21)
HS005	2	10/11/−1.913(+00)	10/11/−1.913(+00)	10/11/−1.913(+00)	10/11/−1.913(+00)
HS038	4	93/105/1.285(−18)	68/73/2.205(−16)	117/125/8.381(−17)	139/148/6.450(−17)
Broydenu	10	50/54/1.026(+00)	45/50/1.026(+00)	94/100/1.000(+00)	74/78/1.026(+00)
	50	190/208/1.000(+00)	79/97/1.000(+00)	205/227/1.000(+00)	212/229/1.008(+00)
	100	232/246/1.000(+00)	209/232/1.000(+00)	257/281/1.006(+00)	F/F/F
Penaltu	8	F/F/F	96/105/4.839(+02)	F/F/F	87/90/4.839(+02)
Gen. Ros	8	122/142/2.632(−16)	147/165/6.848(−18)	370/432/2.923(−18)	160/176/2.552(−19)
	100	2236/2415/1.379(−17)	2146/2287/3.987(+00)	875/1008/4.706(−16)	F/F/F
	200	2101/2251/1.881(−16)	F/F/F	1675/1926/3.987(+00)	F/F/F
Cha. Sin	20	178/200/1.232(−11)	179/190/2.314(−11)	172/193/1.243(−11)	176/201/4.214(−12)
Gen. Wood	20	215/231/1.000(+00)	159/174/1.000(+00)	181/200/1.000(+00)	260/283/1.000(+00)
	100	761/832/1.000(+00)	572/635/1.000(+00)	1117/1295/1.000(+00)	663/760/1.000(+00)
Cha. Wood	8	108/117/1.000(+00)	129/145/1.000(+00)	142/158/1.000(+00)	143/156/1.000(+00)
	50	559/606/1.000(+00)	F/F/F	431/486/1.000(+00)	F/F/F
Cragglevy	8	87/90/8.977(−10)	126/128/8.629(−10)	106/109/2.699(−09)	153/155/2.478(−10)
Brown1	10	79/93/9.989(−01)	158/169/9.989(−01)	114/119/9.989(−01)	158/166/9.989(−01)
Brown3	10	17/19/1.756(−15)	18/18/2.442(−15)	18/18/2.442(−15)	18/18/2.442(−15)
	100	17/22/9.250(−14)	20/20/3.210(−14)	17/22/9.250(−14)	17/22/9.250(−14)

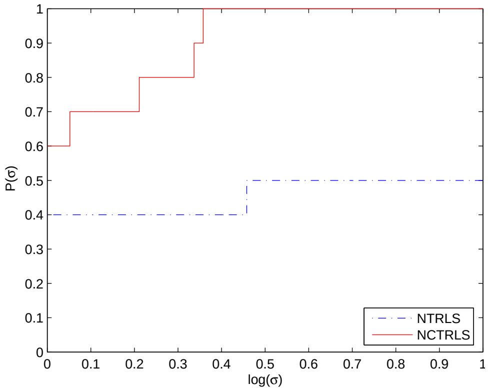

FIGURE 1. Function performance profiles of NTRLS and NCTRLS.

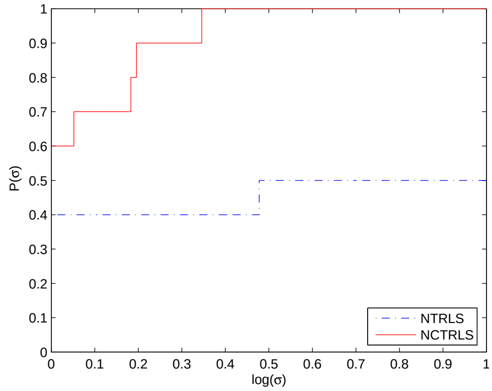

FIGURE 2. Gradient performance profiles of NTRLS and NCTRLS.

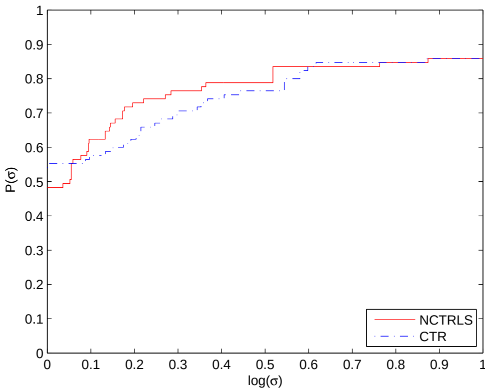

FIGURE 3. Function performance profiles of CTR and NCTRLS.

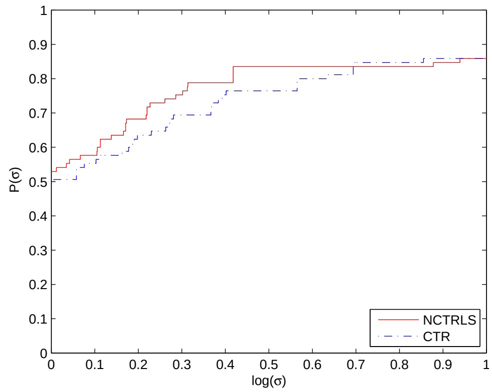

FIGURE 4. Gradient performance profiles of CTR and NCTRLS.

Acknowledgements. The author thank the referees for his/her thorough reviewing with useful suggestions and comments to the paper, which greatly improve the quality of the paper. The author also thank Professor Wenyu Sun for his instruction.

REFERENCES

[1] S. Bellavia, M. Macconi and S. Pieraccini, Constrained dogleg methods for nonlinear systems with simple bounds. *Comput. Optimiz. Appl.* **53** (2012) 771–794.

[2] S. Bellavia, M. Macconi and B. Morini, An affine scaling trust-region approach to bound-constrained nonlinear systems. *Appl. Numer. Math.* **44** (2003) 257–280.

[3] S. Bellavia and S. Pieraccini, On affine scaling inexact dogleg methods for bound constrained nonlinear systems. *Optimiz. Methods Software* **30** (2015) 276–300.

[4] T.F. Coleman and Y.Y. Li, An interior trust-region approach for minimization subject to bounds. *SIAM J. Optimiz.* **6** (1996) 418−445.

[5] A.R. Conn, N.I.M. Gould and Ph.L. Toint, Testing a class of methods for solving minimization problems with simple bounds on the variables. *Math. Comput.* **50** (1988) 399–430.

[6] Z.C. Cui, B.Y. Wu and S.J. Qu, Combining nonmonotone conic trust region and line search techniques for unconstrained optimization. *J. Comput. Appl. Math.* **235** (2011) 2432–2441.

[7] W.C. Davidon, Conic approximations and collinear scalings for optimizers. *SIAM J. Numer. Anal.* **17** (1980) 268–281.

[8] S. Di and W.Y. Sun, A trust region method for conic model to solve unconstrained optimizations. *Optimiz. Methods and Software* **6** (1996) 237–263.

[9] J.H. Fu and W.Y. Sun, Nonmonotone adaptive trust-region method for unconstrained optimization problems. *Appl. Math. Comput.* **163** (2005) 489–504.

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

[11] L. Grippo, F. Lampariello and S. Lucidi, A nonmonotone line search technique for Newton's method. *SIAM J. Numer. Anal.* **23** (1986) 707–716.

[12] N.Z. Gu and J.T. Mo, Incorporating nonmonotone strategies into the trust region method for unconstrained optimization. *Comput. Math. Appl.* **55** (2008) 2158–2172.

[13] M. Heinkenschloss, M. Ulbrich and S. Ulbrich, Superlinear and quadratic convergence of affine scaling interior point Newton methods for problems with simple bounds without strict complementarity assumption. *Math. Program.* **86** (1999) 615–635.

[14] W. Hock and K. Schittkowski, *Test examples for nonlinear programming codes*, Springer (1981).

[15] W.W. Hager, B.A. Mair and H.C. Zhang, An affine scaling interior point CBB method for box constrained optimization. *Math. Program.* **119** (2009) 1–32.

[16] W.W. Hager and H.C. Zhang, An affine scaling method for optimization with polyhedral constraints. *Comput. Optimiz. Appl.* **59** (2014) 163–183.

[17] Y. Ji, Y.J. Li, K.C. Zhang and X.L. Zhang, A new nonmonotone trust region method of conic model for solving unconstrained optimization. *J. Comput. Appl. Math.* **233** (2010) 1746–1754.

[18] C. Kanzow and A. Klug, On affine scaling interior point Newton methods for nonlinear minimization with bound constraints. *Comput. Optimiz. Appl.* **35** (2006) 177–197.

[19] J. Nocedal and S.J. Wright, Numerical Optimization, Springer, New York (1999).

[20] J. Nocedal and Y. Yuan, Combining trust region and line search techniques, edited by Y. Yuan. Advances in Nonlinear Programming, Kluwer Academic Publishers, Dordrecht (1996) 153–175.

[21] Q. Ni, Optimality conditions for trust region subproblems involving a conic model. *SIAM J. Optimiz.* **15** (2005) 826–837.

[22] S.J. Qu and S.D. Jiang, A trust region method with a conic model for unconstrained optimization. *Math. Methods Appl. Sci.* **31** (2008) 1780–1808.

[23] W.Y. Sun and Y.X. Yuan, Optimization Theory and Methods: Nonlinear Programming, Vol. 1 of *Springer Optimization and its Applications*. Springer, New York (2006).

[24] W.Y. Sun, Nonmonotone trust region method for solving optimization problems. *Appl. Math. Comput.* **156** (2004) 159–174.

[25] D.C. Sorensen, The Q-superlinear convergence of a collinear scaling algorithm for unconstrained optimization. *SIAM J. Numer. Anal.* **17** (1980) 84–114.

[26] M. Salahi, Exact two steps SOCP/SDP formulation for a modified conic trust region subproblem. *Optimiz. Lett.* **11** (2017) 1691–1697.

[27] M. Salahi, An iterative algorithm for the conic trust region subproblem. *Inter. J. Appl. Comput. Math.* **3** (2017) 2553–2558.

[28] Ph.L. Toint, A nonmonotone trust region algorithm for nonlinear optimization subject to convex constraints. *Math. Program.* **77** (1997) 69–94.

[29] X. Wang, A trust-region affine scaling method for bound constrained optimization. *Acta Math. Sci.* **29** (2013) 159–182.

[30] Z.S. Yu, Solving bound constrained optimization *via* a new nonmonotone spectral projected gradient method. *Appl. Numer. Math.* **58** (2008) 1340–1348.

[31] H. Zhang and W. Hager, A nonmonotone line search technique and its application to unconstrained optimization. *SIAM J. Optimiz.* **14** (2004) 1043–1056.

[32] L.J. Zhao and W.Y. Sun, A conic affine scaling dogleg method for nonlinear optimization with bound constraints. *Asia-Pacific J. Operat. Res.* **30** (2013) 1–30.

[33] D.T. Zhu, Affine scaling inexact generalized Newton algorithm with interior backtracking technique for solving bound-constrained semismooth equations. *J. Comput. Appl. Math.* **187** (2006) 227–252.

[34] D.T. Zhu, An affine scaling trust-region algorithm with interior backtracking technique for solving bound-constrained nonlinear systems. *J. Comput. Appl. Math.* **184** (2005) 343–361.

[35] D.T. Zhu, An interior affine scaling projective algorithm for nonlinear equality and linear inequality constrained optimization. *J. Comput. Appl. Math.* **173** (2005) 115–148.

[36] D.T. Zhu, Affine scaling interior Levenberg-Marquardt method for bound-constrained semismooth equations under local error bound conditions. *J. Comput. Appl. Math.* **219** (2008) 198–215.