Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Journal of Computational and Applied Mathematics 184 (2005) 343–361

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

www.elsevier.com/locate/cam

An affine scaling trust-region algorithm with interior backtracking technique for solving bound-constrained nonlinear systems

Detong Zhu

Department of Mathematics, Shanghai Normal University, Shanghai 200234, PR China

Received 15 January 2004; received in revised form 5 December 2004

Abstract

In this paper, we propose a new affine scaling trust-region algorithm in association with nonmonotonic interior backtracking line search technique for solving nonlinear equality systems subject to bounds on variables. The trust-region subproblem is defined by minimizing a squared Euclidean norm of linear model adding the augmented quadratic affine scaling term subject only to an ellipsoidal constraint. By using both trust-region strategy and interior backtracking line search technique, each iterate switches to backtracking step generated by the general trust-region subproblem and satisfies strict interior point feasibility by line search backtracking technique. The global convergence and fast local convergence rate of the proposed algorithm are established under some reasonable conditions. A nonmonotonic criterion should bring about speeding up the convergence progress in some ill-conditioned cases. The results of numerical experiments are reported to show the effectiveness of the proposed algorithm.

MSC: 90 C 30; 65 K 05

Keywords: Nonmonotone technique; Nonlinear equations; Interior point; Trust region

1. Introduction

In this paper, we analyze the solution of nonlinear systems subjective to the bound constraints on variable

$$F(x) = 0, \quad x \in \Omega = \{x \mid \leqslant x \leqslant u\}, \tag{1.1}$$

E-mail address: dtzhu@shtu.edu.cn.

0377-0427/$ - see front matter
doi:10.1016/j.cam.2005.01.013

where $F : \mathcal{X} \to \mathbb{R}^n$ is a given continuously differentiable mapping and $\mathcal{X} \subseteq \mathbb{R}^n$ is an open set containing the n-dimensional box constraint Ω. The vector $l \in (\mathbb{R} \cup \{-\infty\})^n$ and $u \in (\mathbb{R} \cup \{+\infty\})^n$ are specified lower and upper bounds on the variables such that $\text{int}(\Omega) \stackrel{\text{def}}{=} \{x \mid l < x < u\}$ is nonempty. The problem (1.1) arises naturally in systems of equations modeling real-life problems when not all the solutions of the model have physical meaning. For example, cross-sectional properties of structural elements, dimensions of mechanical linkages, concentrations of chemical species, etc., are modeled by nonlinear equations where Ω is the positive orthant of $\mathbb{R}^n$ or a closed box constraint. Various sources of nonlinear equations with the box constraint Ω drawn from complimentarily, optimization and several related problems have been described. In the classic methods for solving the unconstrained nonlinear equations (1.1) when the function $F(x)$ is a continuously differentiable function, the Newton methods or quasi-Newton methods can be used. Much analysis of many Newton algorithms have been done on smooth nonlinear equations based on convergent analysis. These methods by using the Jacobian or version of Newton's methods often solve the unconstrained problem (1.1), which is known to have locally very rapid convergence (see [5,6]). However, the Newton methods used for smooth systems (1.1) did not ensure global convergence, that is, the convergence is only local. Therefore, the methods are available only when the initial start point is good enough. In the use of these methods, difficulties arise when the step lies outside the region where the linear model $F(x) + F'(x)s$ is a good approximation $F(x+s)$ where$F'(x)$ is the Jacobian of $F(x)$. One effective remedy when this occurs is to restrict the step s to a region where the linear model can be trusted. Globally, convergent methods for the unconstrained systems $F(x) = 0$ may be unsuited for the purpose of solving the bound-constrained systems (1.1), since a vector x^* satisfy $F(x) = 0$, but does not belong to Ω. Generally, two basis approaches, namely the line search and trust-region, have been used in order to ensure global convergence towards local minima. At each iterations, most modern global fit within determining an initial trial step and testing the trial step to determine whether it gives adequate progress toward a solution. Recently, Eisenstat and Walker in [7] introduced arbitrary norms as the merit function, and Brown and Saad [2] used the Euclidean norm, i.e., l_2 norm as the merit function to combine the line search to solving the unconstrained nonlinear systems (1.1) and proved the global convergence of the proposed algorithms. For most versions for solving smooth equation, these approaches only rather restrictive guarantees of global convergence have only been based on the line search procedure such as Armijo rule, damped Newton methods. Trial steps are determined in a variety ways to enforce a monotone decrease of the merit function at each step.

Classical trust-region Newton method for solving the nonlinear systems (1.1) and the affine scaling double trust-region approach for solving the bounded constrained optimization problems given in [3]. Recently, Bellavia et al. in [1] further extended the ideas and presented an affine scaling trust-region approach for solving the bound-constrained smooth nonlinear systems (1.1). The trust-region method is a well-accepted technique in nonlinear optimization to assure global convergence. However, the search direction generated in trust-region subproblem must satisfy strict interior feasibility which results in computational difficulties. It is possible that the trust-region subproblem with the strict feasibility constraints needs to be resolved many times before obtaining an acceptable step, and hence the total computational effort for completing one iteration might be expensive and difficulties. The idea of combining the trust-region and line search backtracking technique suggested by Nocedal and Yuan [11] motivates to switch to the line search technique by employing a trial step which may be unacceptable in the trust-region method, since the trial step should provide a direction of sufficient descent. Another nonmonotone technique is developed to combine with, respectively, line search technique and trust-region strategy for solving unconstrained optimization in [4,9]. In this paper, we introduce affine scaling interior point projective to generate the affine scaling trust-region Newton methods which switches to strict interior feasibility by line search backtracking technique. The trust-region subproblem is defined by minimizing a squared Euclidean norm of linear model adding the augmented quadratic affine scaling term subject only to an ellipsoidal constraint. The nonmonotone idea also motivates the study of trust-region Newton methods in association with nonmonotone interior backtracking line search technique for approximating zeros of the smooth equations (1.1).

In this research, nonmonotone global convergence of the affine scaling trust-region Newton method in association with two criterions of nonmonotone backtracking line search and strict interior feasibility accepting step for solving the smooth equations (1.1) is presented and analyzed. In order to describe and design the algorithms for solving the bound-constrained smooth equations (1.1), we first introduce the squared Euclidean norm of linear model of the unconstrained systems (1.1) and the augmented quadratic affine scaling term, and state the nonmonotone affine scaling trust-region algorithm with backtracking interior point technique for the nonlinear equations in Section 2. In Section 3, we prove the global convergence of the proposed algorithm. We discuss some further convergence properties such as strong global convergence and characterize the order of local convergence of the Newton methods in terms of the rates of the relative residuals in Section 4. Finally, the results of numerical experiments of the proposed algorithm are reported in Section 5.

2. Algorithm

In this section, we describe and design the affine scaling trust-region strategy in association with nonmonotonic interior point backtracking technique for solving the bound-constrained nonlinear minimization transformed by the bound-constrained systems (1.1) and present an interior point backtracking technique which enforces the variable generating strictly feasible interior point approximations to solution of the bound-constrained nonlinear minimization.

A classical algorithm for solving the unconstrained problem (1.1) is the Newton method. In the context of unconstrained nonlinear systems (1.1) if x_k is a very good approximation of a solution, the Newton process is that find the step s_k which satisfies

$$F_k' s_k = -F_k. \tag{2.1}$$

However, Newton method can be incorporated into a globally convergent trust-region scheme. Bellavia et al. in [1] presented the affine scaling trust-region approach scheme. The basic idea is based on the trust-region subproblem at the kth iteration

$$\begin{aligned} \min \quad & q_k(d) \stackrel{\text{def}}{=} \tfrac{1}{2}\|F_k' d + F_k\|^2 = \tfrac{1}{2}\|F_k\|^2 + F_k^{\mathrm{T}} F_k' d + \tfrac{1}{2} d^{\mathrm{T}} (F_k'^{\mathrm{T}} F_k') d, \\ \text{s.t.} \quad & \|D_k d\| \leqslant \Delta_k, \end{aligned} \tag{2.2}$$

where Δ_k is the trust-region radius and $q_k(d)$ is trusted to be an adequate representation of the merit function

$$f(x) \stackrel{\text{def}}{=} \tfrac{1}{2}\|F(x)\|^2. \tag{2.3}$$

The scaling matrix $D_k = D(x_k)$ arise naturally from examining the first-order necessary conditions for the bound-constrained nonlinear minimization transformed by the bound-constrained problem (1.1), where $D(x)$ is the diagonal scaling matrix such that

$$D(x) \stackrel{\text{def}}{=} \text{diag}\{|v_1(x)|^{-\frac{1}{2}}, \ldots, |v_n(x)|^{-\frac{1}{2}}\} \tag{2.4}$$

and the ith component of vector $v(x)$ defined componentwise as follows:

$$v_i(x) \stackrel{\text{def}}{=} \begin{cases} x_i - u_i & \text{if } g_i < 0, \text{ and } u_i < +\infty, \\ x_i - l_i & \text{if } g_i \geqslant 0, \text{ and } l_i > -\infty, \\ -1 & \text{if } g_i < 0, \text{ and } u_i = +\infty, \\ 1 & \text{if } g_i \geqslant 0, \text{ and } l_i = -\infty \end{cases} \tag{2.5}$$

here $g(x) \stackrel{\text{def}}{=} F'(x)^{\mathrm{T}} F(x)$ and g_i is the ith component of vector $g(x)$. We remark that, even though $D(x)$ may be undefined on the boundary of Ω, $D(x)^{-1}$ can be extended continuously to it. We will denote this extension as a convention by $D(x)^{-1}$ for all $x \in \Omega$.

The following nondegenerate property is essential for convergence of the affine scaling double trust-region approach for solving the bounded constrained optimization problems transformed by the bound-constrained systems (1.1).

Definition 2.1 (*see Coleman and Li [3]*). A point $x \in \Omega$ is nondegenerate if, for each index i,

$$g_i(x) = 0 \Longrightarrow l_i < x_i < u_i. \tag{2.6}$$

A transformed problem (1.1) is nondegenerate if (2.6) holds for every $x \in \Omega$.

Moreover, regarding the solution d_k of the subproblem (2.2), from [1] we know that the stepsize along d_k to the boundary need to be bounded away from zero in the global convergence and eventually satisfy to tend to 1 in the local quadratical convergence. The relevance of the used scaling matrix depends on the fact that the scaled trust-region trial step d_k is angled away from the approaching bound. Consequently, the bounds will not prevent a relatively large stepsize along d_k from being taken. In order to maintain the strict interior feasibility, a step-back tracking along the solution d_k of the following augmented quadratic affine scaling subproblem (S_k) in this algorithm, rather than the solution of the subproblem (2.2), could be required to satisfy the strict interior feasibility by nonmonotomic interior point backtracking line research technique. Following the suggestion in [3], we can make some modifications on the trust-region subproblem (2.2) for solving the nonlinear problem (1.1). The basic idea in the proposed algorithm can be summarized as follows: assume that $x_k \in \text{int}(\Omega)$, we define the diagonal matrix suggested in [3]

$$C_k \stackrel{\text{def}}{=} \text{diag}\{g_k\} J_k^v, \tag{2.7}$$

where $J^v(x) \in \mathbb{R}^{n \times n}$ is the Jacobian matrix of $|v(x)|$ whenever $|v(x)|$ is differentiable and $\text{diag}\{g_k\} \stackrel{\text{def}}{=} \text{diag}\{(g_k)_1, \ldots, (g_k)_n\}$, here $(g_k)_i$ is the ith component of g_k. Each diagonal component of the diagonal matrix J^v equals zero or ± 1. The augmented affine scaling trust-region subproblem at the kth iteration is defined as follows

$$(S_k)\quad \begin{array}{ll} \min & \psi_k(d) \stackrel{\text{def}}{=} \frac{1}{2}\|F_k' d + F_k\|^2 + \frac{1}{2} d^{\mathrm{T}} D_k C_k D_k d \\ & = \frac{1}{2}\|F_k\|^2 + F_k^{\mathrm{T}} F_k' d + \frac{1}{2} d^{\mathrm{T}} (F_k'^{\mathrm{T}} F_k') d_k + \frac{1}{2} d^{\mathrm{T}} D_k C_k D_k d \\ \text{s.t.} & \|D_k d\| \leqslant \Delta_k, \end{array}$$

where Δ_k is the trust-region radius.

Now, we describe an affine scaling trust-region algorithm with nonmonotonic strict interior feasible backtracking line search technique for solving the bound-constrained systems (1.1).

Nonmonotonic affine scaling interior trust-region (ASITR) algorithm

Initialization step

Choose parameters $\beta \in (0, \frac{1}{2})$, $\omega \in (0,1)$, $0<\eta_1<\eta_2<1$, $0<\gamma_1<\gamma_2<1<\gamma_3$, $\varepsilon>0$ and positive integer M as nonmonotonic parameter. Let $m(0)=0$. Select an initial trust-region radius $\Delta_0>0$ and a maximal trust-region radius $\Delta_{\max} \geqslant \Delta_0$, give a starting strict feasibility interior point $x_0 \in \operatorname{int}(\Omega) \subseteq \mathbb{R}^n$. Set $k=0$, go to the main step.

Main step

1. Evaluate $f_k = f(x_k) \stackrel{\text{def}}{=} \frac{1}{2}\|F(x_k)\|^2$, C_k, $g_k = \nabla f(x_k) \stackrel{\text{def}}{=} (F_k')^{\mathrm{T}} F_k$ and D_k given in (2.4).
2. If $\|D_k^{-1} g_k\| = \|D_k^{-1}(F_k')^{\mathrm{T}} F_k\| \leqslant \varepsilon$, stop with the approximate solution x_k.
3. Solve a step d_k, based on the augmented trust-region subproblem

$$(S_k)\quad \begin{array}{ll} \min & \psi_k(d) \stackrel{\text{def}}{=} \frac{1}{2}\|F_k' d + F_k\|^2 + \frac{1}{2} d^{\mathrm{T}} D_k C_k D_k d \\ \text{s.t.} & \|D_k d\| \leqslant \Delta_k. \end{array}$$

4. Choose $\alpha_k = 1, \omega, \omega^2, \ldots$, until the following inequality is satisfied:

$$f(x_k + \alpha_k d_k) \leqslant f(x_{l(k)}) + \alpha_k \beta g_k^{\mathrm{T}} d_k \tag{2.8}$$
$$\text{with } x_k + \alpha_k d_k \in \Omega, \tag{2.9}$$

   where $f(x_{l(k)}) = \max_{0 \leqslant j \leqslant m(k)} \{f(x_{k-j})\}$.
5. Set

$$h_k \stackrel{\text{def}}{=} \begin{cases} \alpha_k d_k & \text{if } x_k + \alpha_k d_k \in \operatorname{int}(\Omega), \\ \theta_k \alpha_k d_k & \text{otherwise}, \end{cases}$$

   where $\theta_k \in (\theta_l, 1]$, for some $0<\theta_l<1$ and $\theta_k - 1 = \mathrm{O}(\|d_k\|)$ and then set

$$x_{k+1} = x_k + h_k. \tag{2.10}$$

   Calculate

$$\operatorname{Pred}(h_k) = \psi_k(0) - \psi_k(h_k), \tag{2.11}$$
$$\widehat{\operatorname{Ared}}(h_k) = f(x_{l(k)}) - f(x_k + h_k), \tag{2.12}$$
$$\hat{\rho}_k = \frac{\widehat{\operatorname{Ared}}(h_k)}{\operatorname{Pred}(h_k)}. \tag{2.13}$$

6. Updating trust-region size Δ_{k+1} from Δ_k

$$\Delta_{k+1} \stackrel{\text{def}}{=} \begin{cases} [\gamma_1\Delta_k, \gamma_2\Delta_k] & \text{if } \hat{\rho}_k \leqslant \eta_1, \\ (\gamma_2\Delta_k, \Delta_k] & \text{if } \eta_1 < \hat{\rho}_k < \eta_2, \\ (\Delta_k, \min\{\gamma_3\Delta_k, \Delta_{\max}\}] & \text{if } \hat{\rho}_k \geqslant \eta_2. \end{cases} \tag{2.14}$$

7. Take the nonmonotone control parameter $m(k+1) = \min\{m(k)+1, M\}$. Then set $k \leftarrow k+1$ and go to step 1.

Remark 1. In the subproblem (S_k), $\psi_k(d)$ is a local squared Euclidean norm of linear model of the vector function F around x_k adding the augmented quadratic affine scaling term. A candidate iterative direction d_k is generated by minimizing $\psi_k(d)$ within the ellipsoidal ball centered at x_k with radius Δ_k. A key property of the line search transformation in trust-region subproblem (S_k) is that the candidate iterative step h_k must be a strict interior feasibility. Note that in each iteration the algorithm solves only one general trust-region subproblem. If the solution d_k fails to meet the acceptance criterions (2.8)–(2.9) (take $\alpha_k = 1$), then we turn to line search, i.e., retreat from $x_k + h_k$ until the criterion is satisfied. Comparing usual monotone technique with nonmonotonic technique, when $M > 1$, the accepted step h_k only guarantees that $f(x_k + h_k)$ is smaller than $f(x_{l(k)})$. It is easy to see that the usual monotone algorithm can be viewed as a special case of the proposed algorithm when $M = 0$.

Remark 2. The scalar α_k given in (2.9) of step 4, denotes the step size along the direction d_k to the boundary on the variables $l \leqslant x_k + \alpha_k d_k \leqslant u$, that is,

$$\alpha_k \stackrel{\text{def}}{=} \min\left\{\max\left\{\frac{l_i - x_{k,i}}{d_{k,i}}, \frac{u_i - x_{k,i}}{d_{k,i}}\right\}, i = 1, 2, \ldots, n\right\}, \tag{2.15}$$

where l_i, u_i, $x_{k,i}$ and $d_{k,i}$ are the ith components of l, u, x_k and d_k, respectively.

3. Global convergence analysis

Throughout this section, we assume that $F : \mathscr{X} \subset \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable and bounded from below. Given $x_0 \in \text{int}(\Omega) \subset \mathbb{R}^n$, the algorithm generates a sequence $\{x_k\} \subset \Omega \subseteq \mathbb{R}^n$. In our analysis, we denote the level set of f by

$$\mathscr{L}(x_0) = \{x \in \mathbb{R}^n \,|\, f(x) \leqslant f(x_0), l \leqslant x \leqslant u\}.$$

The following assumption is commonly used in convergence analysis of most methods for the box constrained systems.

Assumption 1. Sequence $\{x_k\}$ generated by the algorithm is contained in a compact set $\mathscr{L}(x_0)$ on $\mathbb{R}^n$.

Assumption 2. There exist some positive constants χ_g and χ_D such that $\|F'^{\mathrm{T}}(x)F(x)\| \leqslant \chi_g$, $\|D(x)^{-1}\| \leqslant \chi_D$, for all $x \in \mathscr{L}(x_0)$.

Based on solving the augmented trust-region subproblem (S_k), similar to use the proof of Lemma 3.4 in [12] (also see [10]) which is due to Sorensen's paper. The following lemma establishes the necessary and sufficient conditions concerning the pair λ_k, d_k when d_k solves the subproblem (S_k).

Lemma 3.1. *d_k is a solution to the subproblem (S_k) if and only if d_k is a solution to the following equations of the forms*

$$[D_k^{-1}(F_k'^{\mathrm{T}}F_k')D_k^{-1}+C_k+\lambda_k I]D_k d_k=-D_k^{-1}(F_k')^{\mathrm{T}}F_k, \tag{3.1}$$

$$\lambda_k(\|D_k d_k\|-\Delta_k)=0, \quad \lambda_k \geqslant 0 \tag{3.2}$$

and $[D_k^{-1}(F_k'^{\mathrm{T}}F_k')D_k^{-1}+C_k+\lambda_k I]$ is positive semidefinite.

It is well known from solving the trust-region algorithms in order to assure the global convergence of the proposed algorithm, it is a sufficient condition to show that at kth iteration the predicted reduction defined by $\mathrm{Pred}(d_k)=\psi_k(0)-\psi_k(d_k)$ which is obtained by the step d_k from trust-region subproblem (S_k), satisfies a sufficient descent condition. The following lemma is due to Lemma 3.1 in [13].

Lemma 3.2. *Let the step d_k be the solution of the trust-region subproblem (S_k), assume that Assumptions* 1–2 *hold, then there exists $\tau \in (0, \frac{1}{2}]$ such that the step d_k satisfies the following sufficient descent condition.*

$$\mathrm{Pred}(d_k) \geqslant \tau \|D_k^{-1}(F_k')^{\mathrm{T}}F_k\| \min\left\{\Delta_k, \frac{\|D_k^{-1}(F_k')^{\mathrm{T}}F_k\|}{\|D_k^{-1}(F_k')^{\mathrm{T}}F_k'D_k^{-1}+C_k\|}\right\} \tag{3.3}$$

for all F_k', F_k, C_k, D_k and Δ_k.

The following lemma show the relation between the gradient $g_k=(F_k')^{\mathrm{T}}F_k$ of the objective function and the step d_k generated by the proposed algorithm. We can see from the lemma that the direction of the trial step is a sufficiently descent direction.

Lemma 3.3. *At the iteration, let d_k be generated in trust-region subproblem (S_k), then*

$$\nabla f(x_k)^{\mathrm{T}}d_k \leqslant -\tau \|D_k^{-1}(F_k')^{\mathrm{T}}F_k\| \min\left\{\Delta_k, \frac{\|D_k^{-1}(F_k')^{\mathrm{T}}F_k\|}{\|D_k^{-1}(F_k')^{\mathrm{T}}F_k'D_k^{-1}+C_k\|}\right\}, \tag{3.4}$$

where the constant τ given in (3.3).

Proof. Since d_k is generated in trust-region subproblem (S_k), Lemma 3.2 ensures that (3.3) holds. Since $(F_k')^{\mathrm{T}}F_k'$ and $D_kC_kD_k$ are semidefinite, we have

$$\begin{aligned}(g_k)^{\mathrm{T}}d_k &= [(F_k')^{\mathrm{T}}F_k]^{\mathrm{T}}d_k \\ &= \psi_k(d_k)-\psi_k(0)-\tfrac{1}{2}[d_k^{\mathrm{T}}(F_k')^{\mathrm{T}}F_k'd_k+d_k^{\mathrm{T}}D_kC_kD_kd_k] \\ &\leqslant \psi_k(d_k)-\psi_k(0)=-\mathrm{Pred}(d_k)\end{aligned} \tag{3.5}$$

so, (3.4) holds. □

Lemma 3.4. *Let f be differentiable and assume that its gradient is such that*

$$\|\nabla f(x)-\nabla f(y)\|_2 \leqslant \gamma \|x-y\|_2 \quad \forall x, y \in \mathbb{R}^n, \tag{3.6}$$

where γ *is the Lipschitz constant. Let* $\beta \in (0,1)$ *and* d_k *be proposed by the subproblem* (S_k)*. If* $\|D_k^{-1} g_k\| \neq 0$*, then ASITR Algorithm will produce an iterate* $x_{k+1} = x_k + \alpha_k d_k$ *in a finite number of backtracking steps in* (2.8).

Proof. Using the mean value theorem, we have the equality

$$f(x_k + \alpha_k d_k) = f(x_k) + \alpha_k \nabla f(x_k + \theta_k \alpha_k d_k)^{\mathrm{T}} d_k,$$

where $0 \leqslant \theta_k \leqslant 1$. We rewrite the above equation as

$$\begin{aligned} f(x_k + \alpha_k d_k) &= f(x_k) + \alpha_k \nabla f(x_k)^{\mathrm{T}} d_k + \alpha_k [\nabla f(x_k + \theta_k \alpha_k d_k)^{\mathrm{T}} d_k - \nabla f(x_k)^{\mathrm{T}} d_k] \\ &= f(x_k) + \beta \alpha_k \nabla f(x_k)^{\mathrm{T}} d_k + (1-\beta) \alpha_k \nabla f(x_k)^{\mathrm{T}} d_k \\ &\quad + \alpha_k [\nabla f(x_k + \theta_k \alpha_k d_k)^{\mathrm{T}} d_k - \nabla f(x_k)^{\mathrm{T}} d_k] \\ &= f(x_k) + \beta \alpha_k \nabla f(x_k)^{\mathrm{T}} d_k + \alpha_k [(1-\beta) \nabla f(x_k)^{\mathrm{T}} d_k + \alpha_k \|d_k\| \xi_k], \end{aligned} \tag{3.7}$$

where for convenience we have set

$$\xi_k \stackrel{\text{def}}{=} \frac{[\nabla f(x_k + \theta_k \alpha_k d_k) - \nabla f(x_k)]^{\mathrm{T}} d_k}{\alpha_k \|d_k\|}. \tag{3.8}$$

Note that from the assumptions we have

$$|\xi_k| = \left| \frac{[\nabla f(x_k + \theta_k \alpha_k d_k) - \nabla f(x_k)]^{\mathrm{T}} d_k}{\alpha_k \|d_k\|} \right| \leqslant \gamma \theta_k \|d_k\| \leqslant \gamma \|d_k\| \leqslant \gamma \|D_k^{-1}\| \Delta_k \leqslant \gamma \chi_D \Delta_k,$$

where χ_D is given in Assumption 2. By Lemma 3.3 and the condition $\|D_k^{-1} g(x_k)\| \neq 0$. After a finite number of reductions, the last term in brackets in the right-handed side of (3.7) will become negative and the corresponding α_k will be acceptable, that is, we have that in a finite number of backtracking steps, α_k must satisfy

$$f(x_k + \alpha_k d_k) \leqslant f(x_k) + \beta \alpha_k \nabla f(x_k)^{\mathrm{T}} d_k. \tag{3.9}$$

Since $f(x_k) \leqslant f(x_{l(k)})$, the conclusion of the lemma holds. □

In this section, we are now ready to state one of our main results of the proposed algorithm, but it requires the following assumptions.

Assumption 3. $D_k^{-1} (F_k')^{\mathrm{T}} F_k' D_k^{-1}$ and C_k are bounded, i.e., there exist some constants $\chi_F > 0$ and $\chi_C > 0$ such that

$$b_k \stackrel{\text{def}}{=} \|D_k^{-1} (F_k')^{\mathrm{T}} F_k' D_k^{-1}\| \leqslant \chi_F \quad \text{and} \quad c_k \stackrel{\text{def}}{=} \|C_k\| \leqslant \chi_C \quad \forall k,$$

where without loss of generality, assume that $\chi_F + \chi_C \leqslant \gamma \chi_D$.

Assumption 4. The first-order optimality system associated to problem (1.1) has no nonisolated solutions and the nondegenerate property of the system (1.1) holds at any solutions of systems (1.1).

Theorem 3.5. *Assume that Assumptions* 1–4 *hold. Let* $\{x_k\} \subset \Omega \subset \mathbb{R}^n$ *be a sequence generated by the algorithm. If nondegenerate property of the system* (1.1) *holds at any limit point, then*

$$\liminf_{k\to\infty} \|D_k^{-1}(F_k')^{\mathrm{T}} F_k\| = 0. \tag{3.10}$$

Proof. According to the acceptance rule in step 4, we have

$$f(x_{l(k)}) - f(x_k + \alpha_k d_k) \geqslant -\alpha_k \beta g_k^{\mathrm{T}} d_k = -\alpha_k \beta [D_k^{-1}(F_k')^{\mathrm{T}} F_k]^{\mathrm{T}} (D_k d_k). \tag{3.11}$$

Taking into account that $m(k+1) \leqslant m(k)+1$, and $f(x_{k+1}) \leqslant f(x_{l(k)})$, we have $f(x_{l(k+1)}) \leqslant \max_{0\leqslant j\leqslant m(k)+1}\{f(x_{k+1-j})\} = f(x_{l(k)})$. This means that the sequence $\{f(x_{l(k)})\}$ is nonincreasing for all k and hence $\{f(x_{l(k)})\}$ is convergent.

By (2.8) and (3.4), for all $k > M$, we get

$$\begin{aligned} &f(x_{l(k)}) \\ &\quad\leqslant \max_{0\leqslant j\leqslant m(l(k)-1)} \{f(x_{l(k)-j-1})\} + \alpha_{l(k)-1}\beta \nabla f_{l(k)-1}^{\mathrm{T}} d_{l(k)-1} \\ &\quad\leqslant \max_{0\leqslant j\leqslant m(l(k)-1)} \{f(x_{l(k)-j-1})\} \\ &\qquad - \alpha_{l(k)-1}\beta\tau \|D_{l(k)-1}^{-1}(F_{l(k)-1}')^{\mathrm{T}} F_{l(k)-1}\| \min\left\{\Delta_{l(k)-1}, \frac{\|D_{l(k)-1}^{-1}(F_{l(k)-1}')^{\mathrm{T}} F_{l(k)-1}\|}{b_{l(k)-1}+c_{l(k)-1}}\right\}. \end{aligned} \tag{3.12}$$

If the conclusion of the theorem is not true, there exists some $\varepsilon > 0$ such that

$$\|D_k^{-1}(F_k')^{\mathrm{T}} F_k\| \geqslant \varepsilon, \quad k = 1, 2, \ldots. \tag{3.13}$$

As $\{f(x_{l(k)})\}$ is convergent, we obtain that from (3.13)

$$\lim_{k\to\infty} \alpha_{l(k)-1}\Delta_{l(k)-1} = 0$$

which implies that either

$$\lim_{k\to\infty} \alpha_{l(k)-1} = 0 \tag{3.14}$$

or

$$\liminf_{k\to\infty} \Delta_{l(k)-1} = 0. \tag{3.15}$$

By the updating formula of Δ_k, we have $\gamma_1^{M+1}\Delta_{l(k)-1} \leqslant \Delta_k \leqslant \gamma_2^{M+1}\Delta_{l(k)-1}$ which means that from (3.15)

$$\liminf_{k\to\infty} \Delta_k = 0. \tag{3.16}$$

Assume that α_k given in step 4 is the stepsize to the boundary of box constraints along d_k. From (2.15), we have

$$\alpha_k \stackrel{\text{def}}{=} \min\left\{\max\left\{\frac{l_i - x_{k,i}}{d_{k,i}}, \frac{u_i - x_{k,i}}{d_{k,i}}\right\}, i = 1, 2, \ldots, n\right\}.$$

(3.16) means that

$$d_{k,i} \to 0 \quad \text{for all } i.$$

Assume that α_k given in (2.9) of step 4 is the step size to the boundary of box constraints along d_k. From (2.15), we have that there exists a subset $\kappa \subset \mathscr{K}$ such that

$$\lim_{k\to\infty, k\in\mathscr{K}} \alpha_k = 0$$

and hence, without loss of generality, assume $x_{*,i} = l_i$ for some i. Recall (3.1) and left multiplying D_k^{-1} at the side of (3.1),

$$[\mathrm{diag}\{g_k\} J_k^{\gamma} + \lambda_k I] d_k = -D_k^{-2}[F_k'^{\mathrm{T}} F_k + (F_k'^{\mathrm{T}} F_k') d_k]. \tag{3.17}$$

Since $\{(F_k'^{\mathrm{T}} F_k') d_k\}$ converges to zero, $[\mathrm{diag}\{g_k\} J_k^{\gamma} + \lambda_k I]$ is a positive semidefinite diagonal matrix, and x_* is nondegenerate with $D_*^{-1} g_* = 0$, for any i with $(v_*)_i = 0$, without loss of generality, assume $x_{*,i} = l_i$ for some i, from $x_{*,i} = l_i < u_i$, we have that $(d_k)_i$ and $(g_k)_i$ have the same sign for k sufficiently large. Hence, if α_k is defined by some $(v_*)_j = 0$ and hence $(g_*)_j \neq 0$, then $\alpha_k = \frac{|(v_k)_j|}{|(d_k)_j|}$ for k sufficiently large. Using (3.1), again,

$$\alpha_k = \frac{|(g_k)_j| + \lambda_k}{|(g_k)_j + ((F_k'^{\mathrm{T}} F_k') d_k)_j|} \geqslant \frac{|(g_k)_j| + \lambda_k}{\|g_k + (F_k'^{\mathrm{T}} F_k') d_k\|_\infty}. \tag{3.18}$$

Taking norm in Eq. (3.1), we can obtain

$$\lambda_k \Delta_k = \lambda_k \|D_k d_k\| \geqslant \|D_k^{-1} g_k\| - (\|D_k^{-1}(F_k'^{\mathrm{T}} F_k') D_k^{-1}\| + \|C_k\|)\|D_k d_k\|. \tag{3.19}$$

Dividing (3.18) by Δ_k and note $\|D_k d_k\| \leqslant \Delta_k$,

$$\lambda_k \geqslant \frac{\|D_k^{-1} g_k\|}{\Delta_k} - (\|D_k^{-1}(F_k'^{\mathrm{T}} F_k') D_k^{-1}\| + \|C_k\|). \tag{3.20}$$

It is clear that from (3.20) and $\|D_k^{-1} g(x_k)\| \neq 0$, $\lim_{k\to\infty} \lambda_k = +\infty$, as $\Delta_k \to 0$. (3.18) means that we conclude that

$$\lim_{k\to\infty} \alpha_k = +\infty, \tag{3.21}$$

where α_k given in the step size to the boundary of box constraints along d_k. Furthermore, by the condition on the strictly feasible stepsize $\theta_k \in (\theta_0, 1]$, for some $0 < \theta_0 < 1$ and $\theta_k - 1 = \mathrm{O}(\|d_k\|^2)$, $\lim_{k\to\infty} \theta_k = 1$, comes from $\lim_{k\to\infty} d_k = 0$.

We now prove that if

$$\Delta_k \leqslant \frac{\tau\varepsilon(1-\beta)}{\gamma \chi_D} \tag{3.22}$$

then $\alpha_k = 1$ must satisfy the accepted condition (2.8) in step 4, that is,

$$f(x_k + d_k) \leqslant f(x_{l(k)}) + \beta g_k^{\mathrm{T}} d_k.$$

If the above formula is not true, we have

$$f(x_k + d_k) > f(x_{l(k)}) + \beta g_k^{\mathrm{T}} d_k \geqslant f(x_k) + \beta g_k^{\mathrm{T}} d_k. \tag{3.23}$$

Because $f(x)$ is Lipschitz continuously differentiable with constant γ, we have that from (3.23)

$$\begin{aligned} 0 &< (1-\beta) g_k^{\mathrm{T}} d_k + [\nabla f(x_k) - \nabla f(x_k + \xi_k d_k)]^{\mathrm{T}} d_k \\ &\leqslant (1-\beta) g_k^{\mathrm{T}} d_k + \gamma \chi_D \|D_k d_k\|^2 \\ &\leqslant (1-\beta) g_k^{\mathrm{T}} d_k + \gamma \chi_D \Delta_k^2, \end{aligned}$$

where $\xi_k \in [0, 1]$, which means that

$$0 < (1-\beta) g_k^{\mathrm{T}} d_k + \gamma \chi_D \Delta_k^2.$$

By (3.11) and (3.22), we can get

$$-\tau\varepsilon(1-\beta) \min\left\{\Delta_k, \frac{\varepsilon}{\chi_F + \chi_C}\right\} + \gamma\chi_D \Delta_k^2 > 0.$$

Since we assume $\chi_F + \chi_C \leqslant \gamma\chi_D$, and hence $\Delta_k \leqslant \frac{\tau\varepsilon(1-\beta)}{\gamma\chi_D} \leqslant \frac{\varepsilon}{\chi_F+\chi_C}$, we have

$$[-\tau\varepsilon(1-\beta) + \gamma\chi_D \Delta_k]\Delta_k > 0.$$

This means that, by $\Delta_k > 0$, $\tau\varepsilon(1-\beta) < \gamma\chi_D\Delta_k$, which contradicts (3.22).

From above, we see that if (3.21) holds, then the step size will be determined by (2.8). So, for large k, $\alpha_k = 1$, and $\theta_k = 1$, comes from $\lim_{k\to\infty} d_k = 0$, i.e., $h_k = d_k$ and hence $x_{k+1} = x_k + d_k$. We know that

$$\begin{aligned} &|f(x_k + d_k) - f_k - [\psi_k(0) - \psi_k(d_k)]| \\ &\quad = \tfrac{1}{2} |\|F(x_k) + F'(x_k) d_k\|^2 + d_k^{\mathrm{T}} D_k C_k D_k d_k - \|F(x_k + d_k)\|^2| \\ &\quad \leqslant \|F(x_k) + F'(x_k) d_k\| \|w(x_k, d_k)\| + \tfrac{1}{2}\|w(x_k, d_k)\|^2 + \tfrac{1}{2} c_k \|D_k d_k\|^2 \\ &\quad \leqslant [\|F(x_k)\| + \tfrac{1}{2}\|w(x_k, d_k)\|]\|w(x_k, d_k)\| + \tfrac{1}{2} c_k \|D_k d_k\|^2, \end{aligned} \tag{3.24}$$

where $w(x_k, d_k) = \int_0^1 [F'(x_k + \xi d_k) - F'(x_k)] d_k \, \mathrm{d}\xi$. From the Lipschitz continuality of F' with the Lipschitz constant γ_F, we obtain $\|w(x_k, d_k)\| \leqslant \gamma_F \|d_k\|^2$. (3.24) implies that

$$|f(x_k + d_k) - f_k - [\psi_k(0) - \psi_k(d_k)]| \leqslant (\|F_k\| + \tfrac{1}{2}\gamma_F)\|d_k\|^2 + \tfrac{1}{2}\chi_C \Delta_k^2.$$

Since Lemma 3.2 holds, from $h_k = d_k$ we readily obtain that for large k, $\mathrm{Pred}(d_k) \geqslant \tau\varepsilon\Delta_k$, and $\|d_k\| \leqslant \chi_D \Delta_k$ if setting

$$\rho_k = \frac{f_k - f(x_k + h_k)}{\mathrm{Pred}_k(h_k)} \tag{3.25}$$

then $\{\rho_k - 1\}$ converges to zero, as $\Delta_k \to 0$. $\hat{\rho}_k \geqslant \rho_k \geqslant \hat{\eta}_k$ implies that $\{\Delta_k\}$ is not decreased for sufficiently large k and hence bounded away from zero. Thus $\{\Delta_k\}$ cannot converge to zero, contradicting (3.16).

From above, we have obtained that (3.15) is not true. So, (3.14) must hold. Similar to the proof of theorem in [9], we have that if (3.14) holds, then

$$\lim_{k\to\infty} \alpha_k = 0.$$

Now, assume that α_k given in step 4 is the step size to the boundary of box constraints along d_k. From (2.15)

$$\alpha_k \stackrel{\text{def}}{=} \min\left\{\max\left\{\frac{l_i - x_{k,i}}{d_{k,i}}, \frac{u_i - x_{k,i}}{d_{k,i}}\right\}, i = 1, 2, \ldots, n\right\}.$$

Similar to prove (3.18), if for sufficiently large k, $\min\{|l_i - x_{k,i}|, |u_i - x_{k,i}|\} > 0$, $i = 1, \ldots, m$, and α_k is defined by some $(v_*)_j = 0$, by nondegenerate of the problem (1.1) at the limit point then $(g_*)_j \neq 0$ which implies $\alpha_k = \frac{|(v_k)_j|}{|(d_k)_j|}$. Similar to prove (3.18), we also have that

$$\alpha_k \geqslant \frac{|(g_k)_j| + \lambda_k}{\|g_k + (F_k'^{\mathrm{T}} F_k')d_k\|_\infty}.$$

It is clearly to see that from $|(g_k)_j| > \frac{1}{2}|(g_*)_j| > 0$ for sufficiently large k and some j,

$$\alpha_k \nrightarrow 0 \tag{3.26}$$

when α_k given in the step size to the boundary of box constraints along d_k.

Furthermore, the acceptance rule (2.8) means that, for large k

$$f\left(x_k + \frac{\alpha_k}{\omega} d_k\right) - f_k \geqslant f\left(x_k + \frac{\alpha_k}{\omega} d_k\right) - f(x_{l(k)}) \geqslant \beta \frac{\alpha_k}{\omega} g_k^{\mathrm{T}} d_k.$$

Since

$$f\left(x_k + \frac{\alpha_k}{\omega} d_k\right) - f_k = \frac{\alpha_k}{\omega} g_k^{\mathrm{T}} d_k + o\left(\frac{\alpha_k}{\omega}\|d_k\|\right),$$

we have

$$(1-\beta)\frac{\alpha_k}{\omega} g_k^{\mathrm{T}} d_k + o\left(\frac{\alpha_k}{\omega}\|d_k\|\right) \geqslant 0. \tag{3.27}$$

Dividing (3.27) by $\frac{\alpha_k}{\omega}\|d_k\|$ and noting $g_k^{\mathrm{T}} d_k \leqslant 0$, we have

$$\lim_{k\to+\infty} \frac{g_k^{\mathrm{T}} d_k}{\|d_k\|} = 0. \tag{3.28}$$

From (3.4) and (3.27), we have that (3.28) means

$$\begin{aligned} 0 = \lim_{k\to+\infty} \frac{g_k^{\mathrm{T}} d_k}{\|d_k\|} &\leqslant \lim_{k\to+\infty} -\tau\varepsilon \frac{1}{\|d_k\|} \min\left\{\Delta_k, \frac{\varepsilon}{b_k + c_k}\right\} \\ &\leqslant -\tau\varepsilon \min\left\{\lim_{k\to+\infty} \frac{\Delta_k}{\|d_k\|}, \lim_{k\to+\infty} \frac{\varepsilon}{(\chi_F + \chi_C)\|d_k\|}\right\} \leqslant 0, \end{aligned} \tag{3.29}$$

which contradicts $\frac{\Delta_k}{\|d_k\|} \geqslant \frac{\Delta_k}{\|\chi_D\|\|D_k d_k\|} \geqslant \frac{1}{\chi_D} > 0$ and hence the conclusion of the theorem is true. □

4. Properties of the local convergence

Theorem 3.5 indicts that at least one limit point of $\{x_k\}$ is a stationary point. In this section, we shall first extend this theorem to a stronger result and the local convergent rate.

Theorem 4.1. *Assume that the Assumptions* 1–4 *hold. Let* $\{x_k\}$ *be a sequence generated by the proposed algorithm. If nondegenerate property of the system* (1.1) *holds at any limit point, then*

$$\lim_{k\to+\infty} \|D_k^{-1}(F_k')^{\mathrm{T}}F_k\| = 0. \tag{4.1}$$

Proof. Assume that there are an $\varepsilon_1 \in (0, 1)$ and a subsequence $\{D_{m_i}^{-1}(F_{m_i}')^{\mathrm{T}}F_{m_i}\}$ of $\{D_k^{-1}(F_k')^{\mathrm{T}}F_k\}$ such that for all $m_i,\ i = 1, 2, \ldots$

$$\|D_{m_i}^{-1}(F_{m_i}')^{\mathrm{T}}F_{m_i}\| \geqslant \varepsilon_1. \tag{4.2}$$

Theorem 3.5 guarantees the existence of another subsequence $\{D_{l_i}^{-1}(F_{l_i}')^{\mathrm{T}}F_{l_i}\}$ such that

$$\|D_k^{-1}(F_k')^{\mathrm{T}}F_k\| \geqslant \varepsilon_2 \quad \text{for } m_i \leqslant k < l_i \tag{4.3}$$

and

$$\|D_{l_i}^{-1}(F_{l_i}')^{\mathrm{T}}F_{l_i}\| \leqslant \varepsilon_2 \tag{4.4}$$

for an $\varepsilon_2 \in (0, \varepsilon_1)$.

Similar to the proof of Theorem 3.5, we have that the sequence $\{f(x_{l(k)})\}$ is nonincreasing for $m_i \leqslant k < l_i$, and hence $\{f(x_{l(k)})\}$ is convergent. (4.3) and (3.12) mean that

$$f(x_{l(k)}) \leqslant f(x_{l(l(k)-1)}) - \beta\tau\alpha_{l(k)-1}\varepsilon_2 \min\left\{\Delta_{l(k)-1}, \frac{\varepsilon_2}{b_{l(k)-1} + c_{l(k)-1}}\right\}. \tag{4.5}$$

That $\{f(x_{l(k)})\}$ is convergent means

$$\lim_{k\to\infty} \alpha_{l(k)-1}\Delta_{l(k)-1} = 0.$$

Similar to the proof of theorem in [9], we have

$$\lim_{k\to\infty} f(x_{l(k)}) = \lim_{k\to\infty} f(x_k). \tag{4.6}$$

According to the acceptance rule in step 5, we have

$$f(x_{l(k)}) - f(x_k + \alpha_k d_k) \geqslant -\alpha_k\beta g_k^{\mathrm{T}}d_k \geqslant -\beta\tau\alpha_k\varepsilon_2 \min\left\{\Delta_k, \frac{\varepsilon_2}{b_k + c_k}\right\}.$$

Similarly, we also get

$$\lim_{k\to\infty} \alpha_k\Delta_k = 0. \tag{4.7}$$

Therefore, similar to prove (3.21) and (3.26), we can also get that there exists a subset $\mathscr{K} \subset \{k\}$

$$\alpha_k \nrightarrow_{k\in\mathscr{K}} 0,$$

where α_k give in the step size to the boundary of box constraints along d_k, that is, the step size $\{\alpha_k\}$ cannot converge to zero.

By accepting the rule of the step $\alpha_k d_k$, for large enough i such that $m_i \leqslant k < l_i$

$$f\left(x_k + \frac{\alpha_k}{\omega} d_k\right) - f(x_k) \geqslant f\left(x_k + \frac{\alpha_k}{\omega} d_k\right) - f(x_{l(k)}) > \beta \frac{\alpha_k}{\omega} \nabla f(x_k)^{\mathrm{T}} d_k. \tag{4.8}$$

Using the mean value theorem we have the following equality:

$$\begin{aligned} f\left(x_k + \frac{\alpha_k}{\omega} d_k\right) - f(x_k) &= \frac{\alpha_k}{\omega} \nabla f\left(x_k + \xi_k \frac{\alpha_k}{\omega} d_k\right)^{\mathrm{T}} d_k \\ &= \frac{\alpha_k}{\omega}\left[\nabla f\left(x_k + \xi_k \frac{\alpha_k}{\omega} d_k\right) - \nabla f(x_k)\right]^{\mathrm{T}} d_k + \frac{\alpha_k}{\omega} \nabla f(x_k)^{\mathrm{T}} d_k \end{aligned} \tag{4.9}$$

with $\xi_k \in (0, 1)$. Since $f(x)$ is Lipschitz continuously differentiable with constant γ, we have

$$\left|\left[\nabla f\left(x_k + \xi_k \frac{\alpha_k}{\omega} d_k\right) - \nabla f(x_k)\right]^{\mathrm{T}} d_k\right| \leqslant \gamma \xi_k \frac{\alpha_k}{\omega} \|d_k\|^2 \leqslant \gamma \frac{\alpha_k}{\omega} \|d_k\|^2. \tag{4.10}$$

(4.8)–(4.10) mean that

$$-(1-\beta) \frac{\alpha_k}{\omega} \nabla f(x_k)^{\mathrm{T}} d_k \leqslant \gamma \left(\frac{\alpha_k}{\omega}\right)^2 \|d_k\|^2. \tag{4.11}$$

Dividing $\frac{\alpha_k}{\omega}\|d_k\|$ in the two side of (4.11) and noting $(1-\beta) > 0$, $\nabla f(x_k)^{\mathrm{T}} d_k \leqslant 0$, and (3.9), we obtain that

$$\begin{aligned} \alpha_k \|d_k\|^2 &\geqslant -\omega(1-\beta) \nabla f(x_k)^{\mathrm{T}} d_k \\ &\geqslant \omega(1-\beta)\tau \|D_k^{-1}(F_k')^{\mathrm{T}} F_k\| \min\left\{\Delta_k, \frac{\|D_k^{-1}(F_k')^{\mathrm{T}} F_k\|}{\|D_k^{-1}(F_k')^{\mathrm{T}} F_k' D_k^{-1} + C_k\|}\right\} \end{aligned} \tag{4.12}$$

which (4.3) and (4.7) imply

$$\lim_{k \to \infty,\, l_i \leqslant k \leqslant m_i} \Delta_k = 0 \tag{4.13}$$

and hence

$$\lim_{k \to \infty,\, l_i \leqslant k \leqslant m_i} \|d_k\| = 0.$$

Because $f(x)$ is Lipschitz continuously differentiable, we have

$$\begin{aligned} f(x_k + d_k) &= f(x_k) + \beta \nabla f(x_k)^{\mathrm{T}} d_k + (1-\beta) \nabla f(x_k)^{\mathrm{T}} d_k + [\nabla f(x_k + \xi_k d_k) - \nabla f(x_k)]^{\mathrm{T}} d_k \\ &\leqslant f(x_{l(k)}) + \beta \nabla f(x_k)^{\mathrm{T}} d_k + \{(1-\beta) \nabla f(x_k)^{\mathrm{T}} d_k + \gamma \|d_k\|^2\}, \end{aligned}$$

where $\xi_k \in [0, 1]$ and γ given in (3.6). By Lemma 3.3 and (4.3), when (4.13) holds, that is, when $\|d_k\|$ is small enough, the term bracket in the right-hand side of the above inequality will become negative, and hence the corresponding $\theta_k \to 1$, as $\|d_k\| \to 0$, $\alpha_k \to 1$ will be accepted, that is,

$$f(x_k + d_k) \leqslant f(x_{l(k)}) + \beta \nabla f(x_k)^{\mathrm{T}} d_k, \quad x_k + d_k \in \Omega. \tag{4.14}$$

Similar to Theorem 3.4, we also have

$$|f(x_k+d_k)-f(x_k)-[\psi_k(0)-\psi_k(d_k)]|\leqslant(\|F_k\|+\tfrac{1}{2}\gamma_F)\|d_k\|^2+\tfrac{1}{2}c_k\Delta_k^2.$$

From (3.7) and (4.3), for large enough i, $m_i\leqslant k<l_i$,

$$\mathrm{Pred}_k(d_k)\geqslant\tau\|D_k^{-1}(F_k')^{\mathrm{T}}F_k\|\min\left\{\Delta_k,\frac{\|D_k^{-1}(F_k')^{\mathrm{T}}F_k\|}{b_k+c_k}\right\}\geqslant\tau\varepsilon\min\left\{\Delta_k,\frac{\varepsilon}{\chi_F+\chi_C}\right\}.$$

As $d_k=h_k$, for large i, $m_i\leqslant k<l_i$, we obtain that

$$\rho_k=1+\frac{f_k-f(x_k+d_k)+\psi_k(d_k)}{\mathrm{Pred}_k(d_k)}\geqslant1-\frac{(\|F_k\|+\frac{1}{2}\gamma_2)\|d_k\|^2+\frac{1}{2}\chi_C\Delta_k^2}{\tau_1\varepsilon_2\min\{\Delta_k,\frac{\varepsilon_2}{\chi_F+\chi_C}\}}\geqslant\eta_2. \tag{4.15}$$

This means that for large i, $m_i\leqslant k<l_i$, when Δ_k is sufficiently small,

$$f_k-f(x_k+d_k)\geqslant\eta_2\,\mathrm{Pred}_k(h_k)\geqslant\eta_2\tau\varepsilon_2\min\left\{\Delta_k,\frac{\varepsilon_2}{\chi_F+\chi_C}\right\}. \tag{4.16}$$

It follows that for sufficiently large i, when $\Delta_k\leqslant\frac{\varepsilon_2}{\chi_F+\chi_C}$,

$$f_k-f(x_k+d_k)\geqslant\sigma\Delta_k, \tag{4.17}$$

where $\sigma\overset{\mathrm{def}}{=}\eta_2\tau\varepsilon_2$. From $\|x_{k+1}-x_k\|\leqslant\|D_k^{-1}\|\|D_kd_k\|\leqslant\chi_D\Delta_k$, we then deduce from this bound that for i sufficiently large,

$$\|x_{m_i}-x_{l_i}\|\leqslant\sum_{k=m_i}^{l_i-1}\chi_D\Delta_k\leqslant\frac{\chi_D}{\sigma}\sum_{k=m_i}^{l_i-1}[f_k-f(x_k+d_k)]=\frac{\chi_D}{\sigma}(f_{m_i}-f_{l_i}). \tag{4.18}$$

Therefore, (4.6) implies that $f_{m_i}-f_{l_i}$ tends to zero as i tends to infinity. Finally, from (4.3) (4.4) and triangle inequality, we get that from $\|F_{m_i}'^{\mathrm{T}}F_{m_i}-F_{l_i}'^{\mathrm{T}}F_{l_i}\|\leqslant\gamma_F\|x_{m_i}-x_{l_i}\|$ and (2.5) implies that $|(v_{m_i})_j-(v_{l_i})_j|\leqslant|(x_{m_i})_j-(x_{l_i})_j|$ for sufficiently large. Consequently, $\|(D_{m_i}^{-1}-D_{l_i}^{-1})(F_{l_i}'^{\mathrm{T}}F_{l_i})\|\to0$ as i tends to infinity and therefore, assuming $\|x_{m_i}-x_{l_i}\|\leqslant\varepsilon_2$,

$$\begin{aligned}\varepsilon_1&\leqslant\|D_{m_i}^{-1}(F_{m_i}')^{\mathrm{T}}F_{m_i}\|\\&\leqslant\|D_{m_i}^{-1}\|\|F_{m_i}'F_{m_i}-F_{l_i}'F_{l_i}\|+\|(D_{m_i}^{-1}-D_{l_i}^{-1})(F_{l_i}'^{\mathrm{T}}F_{l_i})\|+\|D_{l_i}^{-1}(F_{l_i}')^{\mathrm{T}}F_{l_i}\|\\&\leqslant(\chi_D\varepsilon_2+\chi_F\varepsilon_2+\varepsilon_2)=(\chi_D+\chi_F+1)\varepsilon_2\end{aligned}$$

which contradicts $\varepsilon_2\in(0,\varepsilon_1)$, for arbitrarily small. □

We now discuss the convergence rate for the proposed algorithm. For this purpose, it is show that for large enough k, the step size $\alpha_k\equiv1$, $\lim_{k\to\infty}\theta_k=1$, and there exists $\hat{\Delta}>0$ such that $\Delta_k\geqslant\hat{\Delta}$.

Theorem 4.2. *Assume that Assumptions* 1–5 *hold. If nondegenerate property of the system* (1.1) *holds at every limit point* x^* *of* $\{x_k\}$, *and* $F(x^*)=0$ *and* $F'(x^*)$ *is nonsingular. For sufficiently large* k, *then the* *step* $\alpha_k \equiv 1, \lim_{k\to\infty} \theta_k = 1$ *and the trust-region constraints is inactive, that is, there exists* $\hat{\Delta} > 0$ *such that* $\Delta_k \geqslant \Delta_{K'},\ \forall k \geqslant K'$ *where* K' *is a large enough index.*

Proof. For sufficiently large k, from (3.1), then we have that there exists $\lambda_k \geqslant 0$ such that

$$\begin{aligned}\nabla f(x_k)^{\mathrm{T}} d_k &= [D_k^{-1}(F_k')^{\mathrm{T}} F_k']^{\mathrm{T}} D_k d_k \\ &= -(D_k d_k)^{\mathrm{T}}[D_k^{-1}(F_k')^{\mathrm{T}} F_k' D_k^{-1} + C_k + \lambda_k I](D_k d_k) \\ &\geqslant -\tfrac{1}{2}\|F'(x^*)\|^2 \|d_k\|^2, \end{aligned} \tag{4.19}$$

where the last inequality is deduced by $\lambda_k \geqslant 0,\ C_k$ being positive semidefinite and the continuous of $F'(x)$. According to the acceptance rule in step 4, we have

$$f(x_{l(k)}) - f(x_k + \alpha_k d_k) \geqslant -\alpha_k \beta g_k^{\mathrm{T}} d_k \geqslant \tfrac{1}{2}\alpha_k \beta \|F'(x^*)\|^2 \|d_k\|^2. \tag{4.20}$$

Similar to prove (4.13), we also have

$$\lim_{k\to\infty} \alpha_k \|d_k\| = 0. \tag{4.21}$$

Let the step size scalar α_k be given in (2.15) along the direction d_k to the boundary (2.9) of the box constraints. Since nondegenerate property of the systems (1.1) holds at every limit point x^* of $\{x_k\}$, similar to proof (3.18) in Theorem 3.4, we can also obtain that $\alpha_k = \frac{|(g_k)_j| + \lambda_k}{|(g_k)_j + ((F_k'^{\mathrm{T}} F_k') d_k)_j|}$. Hence, $\lim_{k\to+\infty} \alpha_k \neq 0$ when d_k is given in (2.9) along d_k to the boundary of the box constraints. By (4.14) and (4.19), we also obtain that at the kth iteration for large k,

$$f(x_{l(k)}) - f(x_k + d_k) \geqslant -\beta g_k^{\mathrm{T}} d_k \geqslant \tfrac{1}{4}\beta \|F'(x^*)\|^2 \|d_k\|^2. \tag{4.22}$$

Similar to the proof of Theorem 4.1, we can also prove that (4.6) holds, that is, $\lim_{k\to\infty} f(x_{l(k)}) = \lim_{k\to\infty} f(x_k)$ and hence $d_k \to 0$ by (4.22). Therefore, again using (3.18), we have that

$$\lim_{k\to+\infty} \min\{1, \alpha_k\} = 1, \quad \text{as } d_k \to 0$$

where α_k given in (2.9) along d_k to the boundary of the box constraints. This means that the step size $\alpha_k = 1$, for large enough k if α_k is determined by (2.8) and (2.9). Therefore, by the condition on the strictly feasible step size $\theta_k - 1 = \mathrm{O}(\|d_k\|)$, we get $\lim_{k\to+\infty} \theta_k = 1$, which means that the step size $\alpha_k \equiv 1$, i.e., $h_k = d_k$ for large enough k.

By the above inequality, we know that

$$x_{k+1} = x_k + d_k.$$

By (3.24), we have

$$\begin{aligned}&|f(x_k + d_k) - f(x_k) - [\psi_k(0) - \psi_k(d_k)]| \\ &\quad\leqslant \left[\|F(x_k)\| + \frac{\|\omega(x_k, d_k)\|}{2}\right]\|\omega(x_k, d_k)\| + \frac{\|D_k C_k D_k\|}{2}\|d_k\|^2.\end{aligned}$$

From the Lipschitz continuity of F', we get there exist γ_F and γ_L such that

$$\|\omega(x_k, d_k)\| \leqslant \int_0^1 \|F'(x_k + \xi d_k) - F'(x_k)\| \cdot \|d_k\| \mathrm{d}\xi \leqslant \gamma_F \|d_k\|^2$$

and from the $F(x^*)=0$ and $F'(x^*)$ is nonsingular, we also get $\|F(x_k)\| \leqslant \gamma_L \|d_k\|$. Further,

$$c_k = \|C_k\| = \|\mathrm{diag}\{g_k\} J_k^\nu\| \leqslant \|F_k' F_k\| \leqslant \gamma_F \gamma_L \|d_k\|.$$

By the nondegenerate of the problem (1.1) at the limit point x^* then $g_* = 0$ implies $|(v_k)_j| \geqslant \frac{1}{2}|(v_*)_j| > 0$, $\forall j$, for large enough k. Therefore, D_k is bounded, i.e., there exists a constant $\chi > 0$ such that $\|D_k^2\| \leqslant \chi$.

By Assumptions 1–4 and the continuous of $F'(x)$, for large enough k, there exists $\lambda_k \geqslant 0$

$$\begin{aligned} \mathrm{Pred}(d_k) &= \tfrac{1}{2}\|F_k\|^2 - \tfrac{1}{2}\|F_k + F_k' d_k\|^2 - \tfrac{1}{2} d_k^{\mathrm{T}} D_k C_k D_k d_k \\ &= \tfrac{1}{2}(D_k d_k)^{\mathrm{T}} [D_k^{-1} (F_k')^{\mathrm{T}} F_k' D_k^{-1} + C_k + \lambda_k I](D_k^{-1} d_k) \\ &\geqslant \tfrac{1}{4}\|F'(x^*)\|^2 \|d_k\|^2, \end{aligned} \tag{4.23}$$

we can obtain that

$$\begin{aligned} \rho_k &= 1 + \frac{\mathrm{Ared}(h_k) - \mathrm{Pred}(h_k)}{\mathrm{Pred}(h_k)} \\ &\geqslant 1 - \frac{[\|F(x_k)\| + \frac{1}{2}\|\omega(x_k, h_k)\|]\|\omega(x_k, h_k) + \frac{1}{2}\|C_k\|\|D_k^2\|\|h_k\|^2\|}{|\mathrm{Pred}(h_k)|} \\ &\geqslant 1 - \frac{(\gamma_L + \frac{1}{2}\gamma_F\gamma_L\chi + \frac{1}{2}\gamma_F\|h_k\|)\|h_k\|^3}{|\mathrm{Pred}(h_k)|} \\ &\geqslant 1 - \frac{4(\gamma_L + \frac{1}{2}\gamma_F\gamma_L\chi + \frac{1}{2}\gamma_F\|h_k\|)\|h_k\|}{\|F'(x^*)\|^2}. \end{aligned} \tag{4.24}$$

Hence, (4.23) and (4.24) mean that $\rho_k \to 1$ as $\|h_k\| \to 0$. Hence there exists $\hat{\Delta} > 0$ such that when $\|D_k d_k\| \leqslant \hat{\Delta}$, $\hat{\rho}_k \geqslant \rho_k \geqslant \eta_2$, and therefore, $\Delta_{k+1} \geqslant \Delta_k$. As $h_k \to 0$, there exists index K' such that $\|D_k d_k\| \leqslant \hat{\Delta}$ whenever $k \geqslant K'$. Thus $\Delta_k \geqslant \Delta_{K'}$, $\forall k \geqslant K'$ which implies that the conclusion of the theorem holds. $\square$

Theorem 4.2 means that the local convergence rate for the proposed algorithm depends on the Hessian of objective function at x^* and the local convergence rate of the step. If d_k becomes the Newton step, then the sequence $\{x_k\}$ generated by the algorithm converges x^* quadratical.

5. Numerical experiments

Numerical experiments on the new affine scaling trust-region algorithm in association with nonmonotonic interior backtracking line search technique given in this paper have been performed on an IBM 586 personal computer. In this section we present the numerical results by the ASITR algorithm. The ASITR algorithm was implemented as a MATLAB code and run under MATLAB version 6.5. In our implementation the constant θ_l in step 5 was set equal to 0.5×10^{-4}. For the sake of comparison to check effectiveness of the backtracking technique, we select the same stopping criteria parameter as used in [1]. The computation terminates when one of the following stopping criterions is satisfied which is either $\|D_k^{-1} g_k\| = \|D_k^{-1}(F_k')^{\mathrm{T}} F_k\| \leqslant 10^{-6}$ or $\|F_{k+1} - F_k\| \leqslant 10^{-6}$. The selected parameter values are: $\eta_1 = 0.001$, $\eta_2 = 0.75$, $\gamma_1 = 0.2$, $\gamma_2 = 0.5$, $\gamma_3 = 2$, $\omega = 0.5$, $\Delta_{\max} = 10$, $\beta = 0.2$, and initially $\Delta_0 = 5$. We compare with different nonmonotonic parameters $M = 0$, $M = 4$ and $M = 8$, respectively, for the proposed algorithms. The monotonic algorithms are realized by taking $M = 0$.

Table 1
Experimental results of nonmonotonic affine scaling interior trust region algorithm

Problem	Name and source	Initial point	ASITR Alg	
		x_{0i}	NF	IT
1	Himmelblau ([8, Pb. 14.1.1])	$\varpi = 1$	8	7
		$\varpi = 2$	9	8
		$\varpi = 3$	12	11
2	Equilibrium combustion ([8, Pb.14.1.2])	$\varpi = 1$	7	6
		$\varpi = 2$	6	5
		$\varpi = 3$	10	9
3	Ferraris–Tronconi system ([8, Pb. 14.1.4])	$\varpi = 1$	8	7
		$\varpi = 2$	10	9
		$\varpi = 3$	13	12
4	Brown's almost linear system ([8, Pb. 14.1.5])	$\varpi = 1$	34	28
		$\varpi = 2$	31	25
		$\varpi = 2.5$	25	21
5	Robot design problem ([8, Pb. 14.1.6])	$\varpi = 1$	14	13
		$\varpi = 2.5$	10	9
		$\varpi = 3$	12	11
6	Series of CSTRs, $R = 0.950$ ([8, Pb. 14.1.8])	$\varpi = 1$	17	12
		$\varpi = 2$	12	10
		$\varpi = 3$	11	10
7	Series of CSTRs, $R = 0.960$ ([8, Pb. 14.1.8])	$\varpi = 1$	12	9
		$\varpi = 2$	10	8
		$\varpi = 3$	13	12
8	Series of CSTRs, $R = 0.965$ ([8, Pb. 14.1.8])	$\varpi = 1$	11	9
		$\varpi = 2$	13	11
		$\varpi = 3$	13	12
9	Series of CSTRs, $R = 0.970$ ([8, Pb. 14.1.8])	$\varpi = 1$	9	7
		$\varpi = 2$	11	9
		$\varpi = 3$	15	14
10	Series of CSTRs, $R = 0.975$ ([8, Pb. 14.1.8])	$\varpi = 1$	8	6
		$\varpi = 2$	10	9
		$\varpi = 3$	14	13

The experiments are carried out on 10 standard test problems which are quoted from [8]. We also test the method with the recommended starting points in [8], $x_{0\varpi} = l + 0.25\varpi(u - l)$, for the problems have finite lower and upper bounds. However, since the choice $\varpi = 3$ corresponds to an initial point x_{03} that is solution of Problem 4 and the Jacobian matrices of Problem 5 is singular at the starting guess obtained with $\varpi = 2$. The computational results for updating the real Hessian $H_k = (F_k')^{\mathrm{T}} F_k'$ are presented at the following table, where ASITR denote the nonmonotonic affine scaling interior trust-region algorithm proposed in this paper with nonmonotonic technique. NF and IT stand for the numbers of function evaluations and performed iterations respectively. The number of gradient evaluations is not presented in the following table because it always equals the numbers of performed iterations IT.

However, the nonmonotonic technique does almost not bring in noticeable improvement in most test problems, the number of iterations in which nonmonotonic decreasing situation occurs, that is, the number of times $\|F_k\|^2 < \|F_{k+1}\|^2$ is not presented in the following table (Table 1).

Acknowledgements

The author gratefully acknowledges the partial support of the National Science Foundation Grant of China (10471094).

References

[1] S. Bellavia, M. Macconi, B. Morini, An affine scaling trust-region approach to bound-constrained nonlinear systems, Appl. Numer. Math. 44 (2003) 257–280.
[2] P.N. Brown, Y. Saad, Convergence theory of nonlinear Newton–Kryloy algorithms, SIAM J. Optim. 4 (1994) 297–330.
[3] T.F. Coleman, Y. Li, An interior trust-region approach for minimization subject to bounds, SIAM J. Optim. 6 (3) (1996) 418–445.
[4] N.Y. Deng, Y. Xiao, F.J. Zhou, A nonmonotonic trust-region algorithm, J. Optim. Theory Appl. 76 (1993) 259–285.
[5] J.E. Dennis Jr., J.J. Moré, Quasi-Newton methods, motivation and theory, SIAM Rev. 19 (1977) 46–89.
[6] J.E. Dennis Jr., R.B. Schnabel, Numerical Methods for Unconstrained Optimization and Nonlinear Equations, Prentice-Hall Series in Computation, Mathematics, Englewood Cliffs, NJ, 1983.
[7] S.C. Eisenstat, H.F. Walker, Globally convergence inexact Newton methods, SIAM J. Optim. 4 (1994) 393–422.
[8] C.A. Floudas, et al., Handbook of Test Problems in Global Optimization, vol. 33, Kluwer Academic, Dordrecht, 1999.
[9] L. Gripp, F. Lampariello, S. Lucidi, A nonmonotone line search technique for Newton's methods, SIAM J. Numer. Anal. 23 (1986) 707–716.
[10] J.J. Moré, D.C. Sorensen, Computing a trust-region step, SIAM J. Sci. Statist. Comput. 4 (1983) 553–572.
[11] J. Nocedal, Y. Yuan, Combining trust-region and line search techniques, in: Y. Yuan et al. (eds.), Advances in Nonlinear Programming, Kluwer, Dordrecht, 1998, pp. 153–175.
[12] D.C. Sorensen, Newton's method with a model trust-region modification, SIAM J. Numer. Anal. 19 (1982) 409–426.
[13] D. Zhu, Curvilinear paths and trust-region methods with nonmonotonic back tracking technique for unconstrained optimization, J. Comput. Math. 19 (2001) 241–258.