J. Appl. Math. & Informatics Vol. 29(2011), No. 1 - 2, pp. 173 - 190
Website: http://www.kcam.biz

# AN AFFINE SCALING INTERIOR ALGORITHM VIA CONJUGATE GRADIENT AND LANCZOS METHODS FOR BOUND-CONSTRAINED NONLINEAR OPTIMIZATION†

CHUNXIA JIA* AND DETONG ZHU

Abstract. In this paper, we construct a new approach of affine scaling interior algorithm using the affine scaling conjugate gradient and Lanczos methods for bound constrained nonlinear optimization. We get the iterative direction by solving quadratic model via affine scaling conjugate gradient and Lanczos methods. By using the line search backtracking technique, we will find an acceptable trial step length along this direction which makes the iterate point strictly feasible and the objective function nonmonotonically decreasing. Global convergence and local superlinear convergence rate of the proposed algorithm are established under some reasonable conditions. Finally, we present some numerical results to illustrate the effectiveness of the proposed algorithm..

AMS Mathematics Subject Classification : 90C30, 65K05.
*Key words and phrases* : Lanczos method, Conjugate gradient, Interior points, Affine scaling.

## 1. Introduction

In this paper we construct an affine scaling interior algorithm combining the conjugate gradient with Lanczos methods to analyze the solution of optimization subjective to the bound constraints on variable:

$$\min f(x),\ x \in \Omega = \{\ x \mid l \leq x \leq u\ \}, \tag{1}$$

where $f : \Omega \subset \Re^n \rightarrow \Re$ is a given continuously differentiable mapping. The vector $l \in (\Re \cup \{-\infty\})^n$ and $u \in (\Re \cup \{+\infty\})^n$ are specified lower and upper bounds on the variables such that $l < u$.

Received March 20, 2010. Revised July 9, 2010. Accepted July 19, 2010. *Corresponding author. †This work was supported by the Shanghai Normal University Project(SK200802), the Ph.D. Foundation Grant(0527003) of Chinese Education Ministry, and the Scientific Computing Key Laboratory of Shanghai Universities.

There are quite a few literatures proposing affine-scaling algorithm for solving problems appeared during the last few years. Sun in [13] gave a convergence proof for an affine-scaling algorithm for convex quadratic programming without nondegeneracy assumptions, and Ye [14] introduced affine scaling algorithm for nonconvex quadratic programming. Classical methods also can be used to solve (1), for example, conjugate gradient method, which can be easily programmed and computed, is one of the most popular and useful method for solving large-scale optimization problems. The idea of conjugate gradient path in unconstrained optimization is given in [2]. The path is defined as linear combination of a sequence of conjugate directions which are obtained by applying standard conjugate direction method to approximate quadratic function of unconstrained optimization. The Lanczos method for solving the quadratic-model trust region subproblem in a weighted $l_2$-norm are proposed by Gould et al. in [6]. Combining Lanczos method with conjugate gradient path, we can construct a new path [8, 11], which has both properties of Lanczos vectors and properties of conjugate gradient path.

Stimulated by the progress in these aspects, in this paper, we propose an affine scaling interior algorithm via the conjugate gradient and Lanczos methods to solve (1). The organization of the article is as follows: In Section 2, we state the affine scaling interior algorithm combining the conjugate gradient and Lanczos method for solving (1). In Section 3, we prove the global convergence of the proposed algorithm. Further, we establish that the proposed algorithm has strongly global convergence and local convergence rate 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 conjugate gradient and Lanczos strategy in association with nonmonotonic interior point backtracking technique for solving the bound-constrained nonlinear minimum problem (1) and present an interior point backtracking technique which enforces the variable generating strictly feasible interior point approximations to the solution.

Coleman and Li in [3] observed that the first order optimality condition of (1) are equivalent to the nonlinear system of equations

$$D(x)^{-2}\nabla f(x) = 0, \ x \in \Omega. \tag{2}$$

with a suitable scaling matrix

$$D(x) = \mathrm{diag}\{|\phi^1(x)|^{-\frac{1}{2}}, |\phi^2(x)|^{-\frac{1}{2}}, \cdots, |\phi^n(x)|^{-\frac{1}{2}}\}.$$

However, it was noted by Heinkenschloss et al. [10] that the equivalence between (2) and the optimality condition of (1) holds for a rather general class of scaling matrices satisfying the conditions

$$\phi^i(x) \begin{cases} = 0, & \text{if } x^i = l^i \text{ and } g^i > 0, \\ = 0, & \text{if } x^i = u^i \text{ and } g^i < 0, \\ \geq 0, & \text{if } x^i \in \{l^i, u^i\} \text{ and } g^i = 0, \\ > 0, & \text{else}, \end{cases} \tag{3}$$

for all $i = 1, \cdots, n$ and all $x \in \Omega$, where $g^i$ is the $i$th component of the gradient of $f$ at $x$, $l^i, x^i, u^i$ are the $i$th components of $l, x, u$, respectively. In this work, we allow the scaling matrix satisfying (3) to be from a rather general class, see Assumption 5 below. The basic successive modified Newton step is

$$D_k^{-2}(\nabla_{xx}^2 f_k + C_k)d_k = -D_k^{-2} g_k.$$

where $D_k = D(x_k), x_k$ is a sequence of iterates that would be generated by the to-be-proposed algorithm, $\nabla_{xx}^2 f_k$ is the Hessian of $f$,

$$C_k = D_k diag\{g_k^1, g_k^2, \cdots g_k^n\} J_k^{\phi}\ D_k,$$

$J_k^{\phi} = J^{\phi}(x_k)$ where $J^{\phi}(x) \in \Re^{n \times n}$ is the Jacobian matrix of $\phi(x)$ whenever

$$|\phi(x)| = (|\phi^1(x)|, |\phi^2(x)|, \cdots, |\phi^n(x)|)$$

is differentiable. Considering the transformation $\overline{p}_k = D_k p_k$, the basic idea in the proposed algorithm is based on the minimum value of affine scaling quadratic programming subproblem is

$$\min \quad \overline{\psi}_k(\overline{p}) = f(x_k) + \overline{g}_k^T \overline{p} + \frac{1}{2} \overline{p}^T \overline{H}_k \overline{p}, \tag{4}$$

where $\overline{g}_k = D_k^{-1} \nabla f(x_k)$, $\overline{H}_k = D_k^{-1}(B_k + C_k) D_k^{-1}$, $B_k$ is the Hessian of $f$ or its approximation. Define $H_k = B_k + C_k, g_k = \nabla f(x_k)$, we can get the quadratic programming model of $f$ in original space:

$$\min \quad \psi_k(p) = f(x_k) + g_k^T p + \frac{1}{2} p^T H_k p. \tag{5}$$

We are now in a position to give a precise statement of the overall method.

**Nonmonotonic affine scaling interior algorithm combining conjugate gradient and Lanczos methods**
**Initialization step**
Choose parameters $\beta \in (0, \frac{1}{2})$, $\omega \in (0, 1)$, $\varepsilon > 0$ and positive integer $M$ as nonmonotonic parameter. Let $m(0) = 0$ and $\xi \in (0, 1)$, give a starting strict feasibility interior point $x_0 \in \text{int}(\Omega) \subseteq \Re^n$. Set $k = 0$, go to the main step.
**Main step**

(1) Evaluate $f_k = f(x_k)$, $g_k = \nabla f(x_k)$, $D_k, C_k$ and $M_k = D_k^T D_k$.
(2) If $\|D_k^{-1} g_k\| \leq \varepsilon$, stop with the approximate solution $x_k$.
(3) $\omega_0 = 0$, $d_0 = 0$, $v_1 = 0$, $r_1 = \nabla \psi_k(v_1) = g_k, y_1 = M_k^{-1} r_1, d_1 = -M_k^{-1} g_k, \theta_1 = 1, \gamma_1 = \sqrt{\langle r_1, y_1 \rangle}, \omega_1 = \frac{r_1}{\gamma_1}, q_1 = \frac{y_1}{\gamma_1}$. Let $i = 1$.

(4) If

$$d_i^T H_k d_i > 0 \tag{6}$$

$$D_k^{-1} r_i \neq 0 \tag{7}$$

go to the step 5, otherwise go to the step 6.

(5) Calculate

$$\begin{aligned}
\lambda_i &= \frac{\theta_i^2 r_i^T y_i}{d_i^T H_k d_i},\\
v_{i+1} &= v_i + \lambda_i d_i,\\
\delta_i &= q_i^T H_k q_i,\\
r_{i+1} &= H_k q_i - \delta_i \omega_i - \gamma_i \omega_{i-1},\\
y_{i+1} &= M_k^{-1} r_{i+1},\\
\gamma_{i+1} &= \sqrt{\langle r_{i+1}, y_{i+1}\rangle},\\
\omega_{i+1} &= \frac{r_{i+1}}{\gamma_{i+1}},\\
q_{i+1} &= \frac{y_{i+1}}{\gamma_{i+1}},\\
\theta_{i+1} &= -\lambda_i \theta_i \gamma_i,\\
\beta_i &= \frac{\theta_{i+1} y_{i+1}^T H_k d_i}{d_i^T H_k d_i},\\
d_{i+1} &= -\theta_{i+1} y_{i+1} + \beta_i d_i.
\end{aligned}$$

Calculate

$$f(x_k) - f(x_k + v_{i+1}) \geq \xi \Big[ f(x_k) - \psi_k(v_{i+1}) \Big]. \tag{8}$$

If (8) is satisfied, set $i \Leftarrow i + 1$, go to 4.

(6) If $i = 1, p_k = d_1$, otherwise, $p_k = v_i$.

(7) Choose $\alpha_k = 1,\ \omega,\ \omega^2,\ \cdots$, until the following inequality is satisfied:

$$f(x_k + \alpha_k p_k) \quad \leq \quad f(x_{l(k)}) + \alpha_k \beta g_k^T p_k, \tag{9}$$

$$\text{with} \quad x_k + \alpha_k p_k \in \operatorname{int}(\Omega) \tag{10}$$

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

(8) Set

$$x_{k+1} = x_k + \alpha_k p_k. \tag{11}$$

(9) 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.** The scalar $\alpha_k$ given in (10) of step 7, denotes the step size along the direction $p_k$ to the boundary on the variables $l \le x_k + \alpha_k p_k \le u$, that is,

$$\alpha_k^* = \min\Big\{\max\{\frac{l^i - x_k^i}{p_k^i}, \frac{u^i - x_k^i}{p_k^i}\}, i = 1, 2, \cdots, n\Big\}, \tag{12}$$

where $\alpha_k = \theta_k \alpha_k^*$, $\theta_k \in (\theta, 1]$, for some $0 < \theta < 1$ and $\theta_k - 1 = O(\|p_k\|)$, $x_k^i$ and $p_k^i$ are the $i$th components of $x_k$ and $p_k$, respectively. If $p_k^i = 0$, then set $\frac{l^i - x_k^i}{p_k^i} = \frac{u_i - x_k^i}{p_k^i} = +\infty$. The stepsize $\alpha_k$ will ensure the step $\alpha_k d_k$ within the boundary.

**Properties of the conjugate gradient and Lanczos methods**

Now, we give some properties of the conjugate gradient and Lanczos methods.

**Lemma 1.** *Suppose that the directions $q_i$ and $d_i$ are generated by the step 5 of the Algorithm, $1 \le i \le l \le n_k$, the following properties hold:*

$$q_i^T M_k q_j = 0, \quad 1 \le j < i \le l \le n_k \tag{13}$$

$$Q_i^T H_k Q_i = T_i, \quad i = 1, 2, \cdots, n_k \tag{14}$$

$$r_i^T d_j = 0, \quad 1 \le j < i \le l \le n_k \tag{15}$$

$$d_i^T H_k d_j = 0, \quad i \ne j \tag{16}$$

$$d_i^T M_k d_j \ge 0, \quad 1 \le i, j \le n_k \tag{17}$$

*where $Q_i = [q_1,\ q_2,\ \cdots,\ q_i]$ and the tridiagonal matrix $T_i$ is*

$$T_i = \begin{bmatrix} \delta_1 & \gamma_2 & & & \\ \gamma_2 & \delta_2 & \gamma_3 & & \\ & \ddots & \ddots & \ddots & \\ & & & \delta_{i-1} & \gamma_i \\ & & & \gamma_i & \delta_i \end{bmatrix}.$$

.

*Proof.* The proof is by induction. Noting

$$\begin{aligned} q_2^T M_k q_1 &= \frac{r_2^T q_1}{\gamma_2} = \frac{1}{\gamma_2}(H_k q_1 - \delta_1 \omega_1)^T q_1 = \frac{1}{\gamma_2}(\delta_1 - \delta_1 \frac{r_1^T y_1}{\gamma_1^2}) = 0, \\ q_2^T H_k q_1 &= q_2^T (r_2 + \delta_1 \omega_1) = \gamma_2 + \frac{\delta_1}{\gamma_1} q_2^T M_k y_1 = \gamma_2, \\ r_2^T d_1 &= -(H_k q_1 - \delta_1 \omega_1)^T y_1 = -\gamma_1 q_1^T H_k q_1 + \delta_1 \frac{r_1^T y_1}{\gamma_1} = 0, \\ d_2^T H_k d_1 &= (-\theta_2 y_2 + \beta_1 d_1)^T H_k d_1 = -\theta_2 y_2^T H_k d_1 + \beta_1 d_1^T H_k d_1 = 0 \end{aligned}$$

and

$$d_2^T M_k d_1 = -\theta_2 y_2^T M_k d_1 + \beta_1 d_1^T M_k d_1 = \beta_1 d_1^T D_k^T D_k d_1 \ge 0,$$

we get that (13)-(17) hold for $l = 2$.

Assuming now that these five expressions are true for some $l$ (the induction hypothesis), we now show that they continue to hold for $l + 1$.

We prove first (13) and (14). Because of the induction hypothesis of (13), we have that

$$q_l^T\omega_{l-1} = q_l^T\frac{r_{l-1}}{\gamma_{l-1}} = \frac{1}{\gamma_{l-1}}q_l^T M_k y_{l-1} = \frac{1}{\gamma_{l-1}}q_l^T M_k \gamma_{l-1}q_{l-1} = 0.$$

By combining this equation with $q_l^T\omega_l = \frac{1}{\gamma_l^2}y_l^T r_l = 1$, we can deduce

$$q_{l+1}^T M_k q_l = \frac{1}{\gamma_{l+1}}r_{l+1}^T q_l = \frac{1}{\gamma_{l+1}}(q_l^T H_k q_l - \delta_l\omega_l^T q_l - \gamma_l\omega_{l-1}^T q_l) = 0$$

and

$$q_{l+1}^T H_k q_l = q_{l+1}^T(r_{l+1} + \delta_l\omega_l - \gamma_l\omega_{l-1}) = \frac{1}{\gamma_{l+1}}y_{l+1}^T r_{l+1} = \gamma_{l+1}.$$

When $i \le l-1$, we obtain

$$\begin{aligned} q_{l+1}^T M_k q_i &= \frac{1}{\gamma_{l+1}}(H_k q_l - \delta_l\omega_l - \gamma_l\omega_{l-1})^T q_i \\ &= \frac{1}{\gamma_{l+1}}[q_l^T(r_{i+1} + \delta_i\omega_i + \gamma_i\omega_{i-1}) - \gamma_l\omega_{l-1}^T q_i] \\ &= \frac{1}{\gamma_{l+1}}[q_l^T r_{i+1} - \gamma_l\omega_{l-1}^T q_i] = 0 \end{aligned}$$

and

$$q_{l+1}^T H_k q_i = q_{l+1}^T(r_{i+1} + \delta_i\omega_i + \gamma_i\omega_{i-1}) = 0,$$

therefore, the relations (13) and (14) continue to hold when $l$ is replaced by $l+1$, as claimed.

Next, we prove (15) and (16) with $l$ replaced by $l+1$. By the definition of $\beta_l$, we deduce that

$$d_{l+1}^T H_k d_l = (-\theta_{l+1}y_{l+1} + \beta_l d_l)^T H_k d_l = 0.$$

By the induction hypothesis for (15), we have

$$q_l^T H_k d_{l-1} = q_l^T H_k(-\theta_{l-1}y_{l-1} + \beta_{l-1}d_{l-2}) = -\theta_{l-1}q_l^T H_k\gamma_{l-1}q_{l-1} = -\theta_{l-1}\gamma_{l-1}\gamma_l$$

and

$$\omega_{l-1}^T d_{l-1} = \frac{1}{\gamma_{l-1}}r_{l-1}^T(-\theta_{l-1}y_{l-1} + \beta_{l-1}d_{l-2}) = -\theta_{l-1}\gamma_{l-1}.$$

From these two formulae, we find that the following inclusion holds:

$$\begin{aligned} r_{l+1}^T d_l &= (H_k q_l - \delta_l\omega_l - \gamma_l\omega_{l-1})^T(-\theta_l y_l + \beta_{l-1}d_{l-1}) \\ &= -\theta_l\gamma_l\delta_l - \beta_{l-1}\theta_{l-1}\gamma_{l-1}\gamma_l + \delta_l\theta_l\gamma_l + \gamma_l\beta_{l-1}\theta_{l-1}\gamma_{l-1} = 0. \end{aligned}$$

When $i \le l-1$, we also deduce that

$$r_{l+1}^T d_i = (H_k q_l - \delta_l\omega_l - \gamma_l\omega_{l-1})^T d_i = q_l^T H_k d_i - \gamma_l\omega_{l-1}^T d_i = 0$$

and

$$d_{l+1}^T H_k d_i = -\theta_{l+1}y_{l+1}^T H_k d_i + \beta_l d_l^T H_k d_i = 0,$$

where

$$\begin{aligned} y_{l+1}^T H_k d_i &= y_{l+1}^T H_k(-\theta_i y_i + \beta_{i-1} d_{i-1}) \\ &= \beta_{i-1} y_{l+1}^T H_k(-\theta_{i-1} y_{i-1} + \beta_{i-2} d_{i-2}) \\ &= \cdots \\ &= \beta_{i-1}\beta_{i-2}\cdots\beta_1 y_{l+1}^T H_k d_1 \\ &= -\beta_{i-1}\beta_{i-2}\cdots\beta_1\gamma_{l+1}\gamma_1 q_{l+1}^T H_k q_1 = 0. \end{aligned}$$

Hence, the induction arguments hold for (15) and (16) also.

Finally, we prove (17). Noting

$$\begin{aligned} q_{l+1}^T M_k d_l &= q_{l+1}^T M_k(-\theta_l y_l + \beta_{l-1} d_{l-1}) \\ &= -\theta_l \gamma_l q_{l+1}^T M_k q_l + \beta_{l-1} q_{l+1}^T M_k d_{l-1} = \beta_{l-1} q_{l+1}^T M_k d_{l-1} \\ &= \cdots = \beta_{l-1}\cdots\beta_1 q_{l+1}^T M_k d_1 = 0, \end{aligned}$$

we have that

$$d_{l+1}^T M_k d_l = (-\theta_{l+1} y_{l+1} + \beta_l d_l)^T M_k d_l = \beta_l d_l^T M_k d_l \geq 0.$$

By the induction hypothesis for (17), we conclude that

$$d_{l+1}^T M_k d_i = (-\theta_{l+1} y_{l+1} + \beta_l d_l)^T M_k d_i = \beta_l d_l^T M_k d_i \geq 0.$$

So (17) holds for all $1 \leq i,\ j \leq n_k$, as claimed. □

**Lemma 2.** *Suppose that* $\nabla\psi_k(v_{i+1}) = H_k v_{i+1} + g_k = \theta_{i+1} r_{i+1}$ *(see [6]), where* $\theta_{i+1} = \langle e_{i+1}, h_{i+1}\rangle$ *and* $h_{i+1}$ *satisfy* $T_{i+1} h_{i+1} + \gamma_1 e_1 = 0$*, then we have*

$$\theta_{i+1} = -\lambda_i \theta_i \gamma_i \ (\theta_1 = 1).$$

*Proof.* From the definition of $\gamma_{i+1}$, we can get

$$r_{i+1}^T y_{i+1} = \gamma_{i+1}^2. \tag{18}$$

Noting $v_{i+1} = v_1 + \sum_{j=1}^{i} \lambda_j d_j = \sum_{j=1}^{i} \lambda_j d_j$ $(v_1 = 0)$ and

$$g_k^T y_{i+1} = \gamma_{i+1} r_1^T q_{i+1} = \gamma_{i+1} y_1^T M_k q_{i+1} = \gamma_1 \gamma_{i+1} q_1^T M_k q_{i+1} = 0,$$

we deduce that

$$\begin{aligned} r_{i+1}^T y_{i+1} &= \frac{1}{\theta_{i+1}}(H_k v_{i+1} + g_k)^T y_{i+1} = \frac{1}{\theta_{i+1}}(\sum_{j=1}^{i} \lambda_j d_j^T H_k y_{i+1} + g_k^T y_{i+1}) \\ &= \frac{1}{\theta_{i+1}} \lambda_i d_i^T H_k y_{i+1} = -\frac{\lambda_i \gamma_{i+1}^2 \theta_i \gamma_i}{\theta_{i+1}}. \end{aligned} \tag{19}$$

Compare (18) with (19), we have

$$\theta_{i+1} = -\lambda_i \theta_i \gamma_i.$$

□

## 3. Global convergence analysis

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

$$\mathcal{L}(x_0) = \{\ x \in \Re^n \mid f(x) \leq f(x_0),\ l \leq x \leq u\ \}.$$

In order to discuss the properties of the proposed method in detail, we will summarize as follows.

**Lemma 3.** *Let the step $v_j$ be obtained from the algorithm, then*
*(1) $\|v_j\|_{M_k} \leq \|v_{j+1}\|_{M_k}$, where $\|x\|_{M_k} = (x^T M_k x)^{\frac{1}{2}}$, $\forall x \in \Re^n$.*
*(2) The quadratic function $\psi_k(v_j)$ is monotonically decreasing for $1 \leq j \leq n_k$, that is,*

$$\psi_k(v_j) \geq \psi_k(v_{j+1}).$$

*Proof.* (1) Noting $v_1 = 0$, $\lambda_i > 0$ and $v_j^T M_k d_j = \sum_{i=1}^{j-1} \lambda_i d_i M_k d_j \geq 0$, we have

$$\begin{aligned} \|v_{j+1}\|_{M_k}^2 &= v_{j+1}^T M_k v_{j+1} = (v_j + \lambda_j d_j)^T M_k (v_j + \lambda_j d_j) \\ &= v_j^T M_k v_j + 2\lambda_j v_j^T M_k d_j + \lambda_j^2 d_j^T M_k d_j \geq \|v_j\|_{M_k}^2, \end{aligned}$$

so the conclusion (1) holds.

(2) Using the expression of $\psi_k$ and $v_j$, we obtain that

$$\begin{aligned} & \psi_k(v_{j+1}) - \psi_k(v_j) \\ = \ & g_k^T (v_{j+1} - v_j) + \frac{1}{2} v_{j+1}^T H_k v_{j+1} - \frac{1}{2} v_j^T H_k v_j \\ = \ & \lambda_j g_k^T d_j + \frac{1}{2} (\sum_{i=0}^{j} \lambda_i d_i)^T H_k (\sum_{i=0}^{j} \lambda_i d_i) - \frac{1}{2} (\sum_{i=0}^{j-1} \lambda_i d_i)^T H_k (\sum_{i=0}^{j-1} \lambda_i d_i) \\ = \ & \lambda_j g_k^T d_j + \frac{1}{2} \lambda_j^2 d_j^T H_k d_j \\ = \ & \frac{1}{2} \lambda_j [2 g_k^T d_j + \theta_j^2 r_j^T y_j]. \end{aligned}$$

Noting

$$\begin{aligned} g_k^T d_j + \theta_j^2 r_j^T y_j &= d_j^T r_1 + \theta_j r_j^T (-\theta_j y_j + \beta_{j-1} d_{j-1}) = d_i^T (r_1 - H_k v_i - g_k) \\ &= d_i^T (r_1 - r_1 - \sum_{j=1}^{i-1} \lambda_j H_k d_j) = 0 \end{aligned}$$

and $\theta_j^2 r_j^T y_j \geq 0$, we get $g_k^T d_j \leq 0$, so $2 g_k^T d_j + \theta_j^2 r_j^T y_j < 0$, that is, $\psi_k(v_{j+1}) - \psi_k(v_j) \leq 0$. This completes the proof of this lemma. □

The following lemma show the relation between the gradient $g_k$ of the objective function and the step $p_k$ generated by the proposed algorithm. We can see from the following lemma that the direction of the trial step is a sufficiently descent direction.

**Lemma 4.** *Let the step $p_k = v_j$ be obtained from the algorithm, then*
*(1) $\{g_k^T v_j\}$ is monotonically decreasing, that is, $g_k^T v_{j+1} \leq g_k^T v_j,\ 1 \leq j \leq n_k$.*
*(2) $g_k^T p_k$ satisfies the following sufficient descent condition*

$$g_k^T p_k \leq -\|D_k^{-1} g_k\|^2 \min\{1, \lambda_1\}. \tag{20}$$

*Proof.* (1) From (17), we deduce

$$g_k^T v_{j+1} - g_k^T v_j = g_k^T (v_{j+1} - v_j) = \lambda_j g_k^T d_j = -\lambda_j d_1^T M_k d_j \leq 0.$$

that is, the conclusion (1) holds.

(2) If $d_1^T H_k d_1 \leq 0$, then $p_k = v_1 = d_1$ and

$$g_k^T p_k = g_k^T d_1 = -g_k^T M_k^{-1} g_k = -\|D_k^{-1} g_k\|^2.$$

If $d_1^T H_k d_1 > 0$, then there exists $j_0 \geq 2$ such that $p_k = v_{j_0}$. Noting $\{g_k^T v_j\}$ is monotonically decreasing and $g_k^T v_2 = g_k^T (v_1 + \lambda_1 d_1) = \lambda_1 g_k^T d_1 = -\lambda_1 \|D_k^{-1} g_k\|^2$, we have

$$g_k^T p_k \leq g_k^T v_2 = -\lambda_1 \|D_k^{-1} g_k\|^2 \leq -\|D_k^{-1} g_k\|^2 \min\{1, \lambda_1\}.$$

So the conclusion holds. □

If there exist $\chi_D > 0, \chi_H > 0$ such that $\|D_k^{-1}\| \leq \chi_D, \|H_k\| \leq \chi_H$, then we have

$$\lambda_1 = \frac{\theta_1^2 r_1^T y_1}{d_1^T H_k d_1} \geq \frac{\|D_k^{-1} g_k\|^2}{\|D_k^{-1} g_k\|^2 \cdot \|D_k^{-1}\|^2 \cdot \|H_k\|} \geq \frac{1}{\chi_D^2 \chi_H},$$

furthermore,

$$g_k^T p_k \leq -\|D_k^{-1} g_k\|^2 \min\{1, \frac{1}{\chi_D^2 \chi_H}\} = -C_1 \|D_k^{-1} g_k\|^2, \tag{21}$$

where $C_1 = \min\{1, \frac{1}{\chi_D^2 \chi_H}\}$. □

**Lemma 5.** *Let the step $p_k$ be obtained from the algorithm, then the predicted reduction satisfy the estimate:*

$$f(x_k) - \psi_k(p_k) \geq \|D_k^{-1} g_k\|^2 \min\{1, \frac{\lambda_1}{2}\}. \tag{22}$$

*Proof.* We consider first the case of $d_1^T H_k d_1 \leq 0$. Here, we have $p_k = v_1 = d_1$ and

$$f(x_k) - \psi_k(p_k) = -g_k^T d_1 - \frac{1}{2} d_1^T H_k d_1 \geq -g_k^T d_1 = \|D_k^{-1} g_k\|^2.$$

For the next case, consider $d_1^T H_k d_1 > 0$. Noting $\{\psi_k(v_j)\}$ is monotonically decreasing, we obtain that

$$\begin{aligned} f(x_k) - \psi_k(p_k) &\geq f(x_k) - \psi_k(\lambda_1 d_1) \\ &= -\lambda_1 g_k^T d_1 - \frac{1}{2}\lambda_1^2 d_1^T H_k d_1 = \lambda_1 \|D_k^{-1} g_k\|^2 - \frac{\lambda_1}{2}\|D_k^{-1} g_k\|^2 \\ &= \frac{\lambda_1}{2}\|D_k^{-1} g_k\|^2 \geq \|D_k^{-1} g_k\|^2 \min\{1, \frac{\lambda_1}{2}\}. \end{aligned}$$

So (22) holds. □

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

**Assumption 1:** $f : \Re^n \to \Re$ is a continuously differentiable mapping, sequence $\{x_k\}$ generated by the algorithm is contained in the compact set $\mathcal{L}(x_0)$.

**Assumption 2:** $\|p_k\|$, $D_k^{-1}$ and $H_k$ are uniformly bounded, that is, there exist constants $\chi_p$, $\chi_D$ and $\chi_H$ satisfy $\|p_k\| \leq \chi_p$, $\|D_k^{-1}\| \leq \chi_D$ and $\|H_k\| \leq \chi_H$ for all $k$.

**Assumption 3:** $g(x) = \nabla f(x)$ is Lipschitz continuous, that is, there exists a constant $\gamma$ such that

$$\|g(x) - g(y)\| \leq \gamma \|x - y\| \ \forall x, y \in \mathcal{L}(x_0).$$

The following nondegenerate property is essential for convergence of the affine scaling double trust-region approach given in [3].

**Definition 1.1** (see [3]). *A point $x \in \Omega$ is nondegenerate if, for each index $i$,*

$$g^i(x) = 0 \implies l^i < x^i < u^i. \tag{23}$$

*Problem (1) is nondegenerate if (23) holds for every $x \in \Omega$.*

**Assumption 4:** The first order optimality system associated to problem (1) has no nonisolated solutions and the nondegenerate property of the system (1) holds at any solutions of systems (1).

**Assumption 5:** $\varpi$ is a constant and there exist $\varpi_k \geq \varpi$ such that

$$\varpi_k \frac{|p_k^j|}{|\phi_k^j|} \leq \begin{cases} x_k^j - l^j, & \text{if } g_k^j > 0, \text{ and } l^j > -\infty, \\ u^j - x_k^j, & \text{if } g_k^j < 0, \text{ and } u^j < +\infty. \end{cases}$$

We now ready to state one of our main results.

**Theorem 1.** *Assume that Assumptions 1-5 hold. Let $\{x_k\} \subset int(\Omega)$ be a sequence generated by the algorithm. Then*

$$\liminf_{k \to \infty} \|D_k^{-1} g_k\| = 0. \tag{24}$$

*Proof.* According to the acceptance rule of $\alpha_k$ in step 7, we have

$$f(x_{l(k)}) - f(x_k + \alpha_k p_k) \geq -\alpha_k \beta g_k^T p_k.$$

Taking into account that $m(k+1) \leq m(k) + 1$ and $f(x_{k+1}) \leq f(x_l(k))$, we get

$$f(x_{l(k+1)}) = \max_{0 \leq j \leq m(k+1)} f(x_{k+1-j}) \leq \max_{0 \leq j \leq m(k)+1} f(x_{k+1-j}) = f(x_{l(k)}).$$

This means $\{f(x_{l(k)})\}$ is nonincreasing for all $k$ and hence $\{f(x_{l(k)})\}$ is convergent.

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

$$\|D_k^{-1} g_k\| \geq \varepsilon.$$

We can deduce from (9) and (21) that

$$\begin{aligned} f(x_{l(k)}) &= f(x_{l(k)-1} + \alpha_{l(k)-1} p_{l(k)-1}) \\ &\leq f(x_{l(l(k)-1)}) + \beta \alpha_{l(k)-1} g_{l(k)-1}^T p_{l(k)-1} \\ &\leq f(x_{l(l(k)-1)}) - \alpha_{l(k)-1} \beta \epsilon^2 C_1. \end{aligned} \tag{25}$$

By the convergence of $\{f(x_{l(k)})\}$, we can conclude

$$\lim_{k\to\infty} \alpha_{l(k)-1} = 0, \tag{26}$$

it follows from Assumption 2 that

$$\lim_{k\to\infty} \alpha_{l(k)-1} \|p_{l(k)-1}\| = 0. \tag{27}$$

Similar to the proof of theorem in [7], we have

$$\lim_{k\to\infty} f(x_k) = \lim_{k\to\infty} f(x_{l(k)}), \tag{28}$$

and therefore

$$\lim_{k\to\infty} \alpha_k = 0. \tag{29}$$

If $\alpha_k$ is determined by (9), we have

$$f(x_k + \frac{\alpha_k}{\omega} p_k) > f(x_{l(k)}) + \frac{\alpha_k}{\omega} \beta g_k^T p_k \geq f(x_k) + \frac{\alpha_k}{\omega} \beta g_k^T p_k,$$

that is,

$$f(x_k + \frac{\alpha_k}{\omega} p_k) - f(x_k) > \frac{\alpha_k}{\omega} \beta g_k^T p_k, \tag{30}$$

On the other hand,

$$\begin{aligned} f(x_k + \frac{\alpha_k}{\omega} p_k) - f(x_k) &= \frac{\alpha_k}{\omega} g_k^T p_k + \frac{\alpha_k}{\omega} \int_0^1 \left[ g(x_k + t\frac{\alpha_k}{\omega} p_k) - g(x_k) \right]^T p_k dt \\ &\leq \frac{\alpha_k}{\omega} g_k^T p_k + \frac{1}{2} \gamma (\frac{\alpha_k}{\omega})^2 \|p_k\|^2, \end{aligned} \tag{31}$$

where $\gamma$ is Lipschitz constant for $g(x)$. From (30) and (31), we deduce

$$\frac{\alpha_k}{\omega} g_k^T p_k + \frac{1}{2} \gamma (\frac{\alpha_k}{\omega})^2 \|p_k\|^2 > \beta \frac{\alpha_k}{\omega} g_k^T p_k,$$

so

$$\alpha_k \geq \frac{2\omega(\beta - 1)}{\gamma \|p_k\|^2} g_k^T p_k \geq \frac{2\omega(1-\beta)}{\gamma \chi_p^2} C_1 \epsilon^2 > 0. \tag{32}$$

From the above formula, we can conclude that $\lim_{k\to\infty} \alpha_k \geq \frac{2\omega(1-\beta)}{\gamma \chi_p^2} C_1 \epsilon^2 > 0$, which contradicts (29).

If $\alpha_k$ is determined by (10), let $x_*$ be a limit point of $\{x_k\}$, then there exists a subset $\mathcal{K}_1 \subset \{k\}$ satisfies:

$$\lim_{k\to\infty, k\in\mathcal{K}_1} \alpha_k^* = 0, \quad \lim_{k\to\infty, k\in\mathcal{K}_1} x_k = x_*.$$

From the expression of $\alpha_k^*$, we know there exists an index $j$ such that $\max\{\frac{l^j - x_*^j}{p_*^j}, \frac{u^j - x_*^j}{p_*^j}\} = 0$, so we can get a subset $\mathcal{K}_2 \subset \mathcal{K}_1$ such that:

$$\lim_{k\to\infty, k\in\mathcal{K}_2} \max\{\frac{l^j - x_k^j}{p_k^j}, \frac{u^j - x_k^j}{p_k^j}\} = 0.$$

Without loss of generality, we assume $x_*^j - l^j = 0$.

If $p_k^j > 0$, by $p_k^j \leq \|p_k\| \leq \chi_p$, we get that for sufficiently large $k$,

$$\max\{\frac{l^j - x_k^j}{p_k^j}, \frac{u^j - x_k^j}{p_k^j}\} = \frac{u^j - x_k^j}{p_k^j} > \frac{u^j - x_*^j}{2\chi_p} > 0.$$

If $p_k^j < 0$, by nondegeneration and the optimization condition, we get $g_*^j > 0$, so when $k$ is large enough, $g_k^j > 0$. By Assumption 5, we get $\lim_{k\to\infty} \phi_k^j = \phi_*^j = 0$, furthermore,

$$\max\{\frac{l^j - x_k^j}{p_k^j}, \frac{u^j - x_k^j}{p_k^j}\} = \frac{l^j - x_k^j}{p_k^j} = \frac{x_k^j - l^j}{|p_k^j|} \geq \frac{\varpi}{|\phi_k^j|} \to +\infty \ (k \to \infty) \qquad (33)$$

which contradicts $\lim_{k\to\infty, k\in\mathcal{K}_2} \max\{\frac{l^j - x_k^j}{p_k^j}, \frac{u^j - x_k^j}{p_k^j}\} = 0$, so $\lim_{k\to\infty} \alpha_k \neq 0$.

Similarly, when $x_*^j - u^j = 0$, we get $\lim_{k\to\infty} \alpha_k \neq 0$, which contradicts (29), hence the conclusion of the theorem is true. □

## 4. Properties of the local convergence

Theorem 1 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 2.** *Assume that the Assumptions 1-5 hold. Let $\{x_k\}$ be a sequence generated by the proposed algorithm. Then*

$$\lim_{k\to+\infty} \|D_k^{-1} g_k\| = 0. \qquad (34)$$

*Proof.* Assume that the conclusion is not true, then there is an $\epsilon_1 \in (0, 1)$ and a subsequence $\{D_{m_i}^{-1} g_{m_i}\}$ such that for all $m_i, i = 1, 2, \cdots$

$$\|D_{m_i}^{-1} g_{m_i}\| \geq \epsilon_1.$$

Consider any index $m_i$ such that $\|D_{m_i}^{-1} g_{m_i}\| \geq \epsilon_1$. Using Assumption 3, we have

$$\|g(x) - g(x_{m_i})\| \leq \gamma \|x - x_{m_i}\|.$$

Noting $D(x)$ is continuous for $x$, we have there exists some $\delta > 0$ such that

$$\|D^{-1}(x) - D^{-1}(x_{m_i})\| \le \frac{n-1}{2n\chi_g}\epsilon_1$$

for all $x$ which satisfy $\|x - x_{m_i}\| \le \delta$, where $n$ can be some very large integer. Hence, by defining the scalar

$$R = \min\{\frac{n-1}{2n\gamma\chi_D}\epsilon_1, \delta\}$$

and the ball

$$\mathcal{B}(x_{m_i}, R) = \{x | \|x - x_{m_i}\| \le R\},$$

we can get if $x \in \mathcal{B}(x_{m_i}, R)$, then

$$\begin{aligned}
\|D^{-1}(x)g(x)\| &\ge \|D_{m_i}^{-1}g_{m_i}\| - \|D^{-1}(x)g(x) - D_{m_i}^{-1}g_{m_i}\| \\
&\ge \epsilon_1 - \|D^{-1}g - D^{-1}g_{m_i}\| - \|D^{-1} - D_{m_i}^{-1}\| \cdot \|g_{m_i}\| \\
&\ge \epsilon_1 - \chi_D\|g - g_{m_i}\| - \chi_g\|D^{-1} - D_{m_i}^{-1}\| \\
&\ge \epsilon_1 - \frac{n-1}{2n}\epsilon_1 - \frac{n-1}{2n}\epsilon_1 = \frac{1}{n}\epsilon_1 = \epsilon_2,
\end{aligned}$$

where $\epsilon_2 = \frac{1}{n}\epsilon_1$. If the entire sequence $\{x_k\}_{k \ge m_i}$ stays the ball $\mathcal{B}(x_{m_i}, R)$, we would have $\|D_k^{-1}g_k\| \ge \epsilon_2 > 0$ for all $k \ge m_i$. The reasoning in the proof of Theorem 1 can be used to show that this scenario does not occur. Therefore, the sequence $\{x_k\}_{k \ge m_i}$ eventually leaves $\mathcal{B}(x_{m_i}, R)$, and there exist another subsequence $\{D_{n_i}^{-1}g_{n_i}\}$ such that

$$\|D_k^{-1}g_k\| \ge \epsilon_2, \text{ for } m_i \le k < n_i$$

and

$$\|D_{n_i}^{-1}g_{n_i}\| \le \epsilon_2,$$

for an $\epsilon_2 \in (0, \epsilon_1)$.

The reasoning in the proof of Theorem 1 can be used to show that

$$\lim_{k\to\infty, m_i \le k < n_i} f(x_{l(k)}) = \lim_{k\to\infty, m_i \le k < n_i} f(x_k). \tag{35}$$

According to the acceptance rule in step 7, we have

$$f(x_{l(k)}) - f(x_k + \alpha_k p_k) \ge -\alpha_k\beta g_k^T p_k \ge \alpha_k\beta\tau\epsilon_2 C_1 \ge 0.$$

Similarly, we also get

$$\lim_{k\to\infty, m_i \le k < n_i} \alpha_k = 0. \tag{36}$$

If $\alpha_k$ is determined by (10), similar to the proof in Theorem 1, we have $\lim\limits_{k\to\infty, k\in\mathcal{K}} \alpha_k > 0$, so $\alpha_k$ is determined by (9). From the acceptance rule of $\alpha_k$ in (9) and (32), we have

$$0 = \lim_{k\to\infty, k\in\mathcal{K}} \alpha_k \ge \lim_{k\to\infty, k\in\mathcal{K}} \frac{2\omega(1-\beta)}{\gamma\chi_p^2} C_1\varepsilon_2^2 > 0,$$

which contradicts (36), so (34) holds. □

We now discuss the convergence rate for the proposed algorithm. For this purpose, it is show that for large enough $k$, the step size $\alpha_k \equiv 1$, $\lim_{k\to\infty} \theta_k = 1$, but it requires more assumptions.

**Assumption 6:** The strong second-order sufficient condition holds, that is, there exists $\zeta > 0$ such that

$$(D_* q)^T (D_*^{-1} H_* D_*^{-1})(D_* q) \geq \zeta \|D_* q\|^2, \ \forall q \in \Re^n. \tag{37}$$

**Assumption 7:**

$$\lim_{k\to\infty} \frac{\|[B_k - \nabla^2 f(x_k)]v_j\|}{\|v_j\|} = 0, \ \forall j = 1, 2, \cdots, n. \tag{38}$$

Because $C_k \to C_* = 0$, by Assumption 7, we have

$$\lim_{k\to\infty} \frac{\|[H_k - \nabla^2 f(x_k)]v_j\|}{\|v_j\|} \leq \lim_{k\to\infty} \frac{\|[B_k - \nabla^2 f(x_k)]v_j\| + \|C_k v_j\|}{\|v_j\|} = 0, \tag{39}$$

which means

$$v_j^T [\nabla^2 f(x_k) - H_k] v_j = o(\|v_j\|^2). \tag{40}$$

**Theorem 3.** *Assume that $H_k$ is positive definite, Assumptions 1-7 hold and $\{x_k\}$ is a sequence produced by algorithm which convergence to $x_*$, then the convergence is superlinear, i.e.,*

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} = 0. \tag{41}$$

*Proof.* We prove first $p_k = -D_k^{-1}(D_k^{-1} H_k D_k^{-1})^{-1} D_k^{-1} g_k$ for sufficiently large $k$.

From Lemma 1, we have

$$0 = \theta_j r_j^T (\sum_{i=1}^{j-1} \lambda_i d_i) = \theta_j r_j^T v_j = (g_k + H_k v_j)^T v_j = g_k^T v_j + v_j^T H_k v_j. \tag{42}$$

Using Assumption 6, we can get that for sufficiently large $k$,

$$v_j^T H_k v_j \geq \frac{\zeta}{2} \|D_k v_j\|^2. \tag{43}$$

So, for sufficiently large $k$,

$$\frac{\zeta}{2} \|D_k v_j\|^2 \leq v_j^T H_k v_j = -g_k^T v_j = -(D_k^{-1} g_k)^T D_k v_j \leq \|D_k^{-1} g_k\| \cdot \|D_k v_j\|, \tag{44}$$

that is, $\|D_k v_j\| \leq \frac{2}{\zeta} \|D_k^{-1} g_k\|$, combining this formula with Theorem 2, we can get

$$\|v_j\| \leq \|D_k^{-1}\| \cdot \|D_k v_j\| \leq \frac{2}{\zeta} \chi_D \|D_k^{-1} g_k\| \to 0. \tag{45}$$

Therefore,

$$
\begin{aligned}
& |\psi_k(v_j) - f(x_k + v_j)| \\
= \ & |g_k^T v_j + \frac{1}{2} v_j^T H_k v_j - (g_k^T v_j + \frac{1}{2} v_j^T \nabla^2 f(x_k) v_j + o(\|v_j\|^2)| \\
= \ & |\frac{1}{2} v_j^T (H_k - \nabla^2 f(x_k)) v_j - o(\|v_j\|^2)| = o(\|v_j\|^2).
\end{aligned}
$$

By Assumption 6, we can get that $D_k^{-1} H_k D_k^{-1}$ is positive definite uniformly for sufficiently large $k$, so

$$
\begin{aligned}
& f(x_k) - \psi_k(v_j) = -g_k^T v_j - \frac{1}{2} v_j^T H_k v_j \\
= \ & (-\theta_j r_j + H_k v_j)^T v_j - \frac{1}{2} v_j^T H_k v_j \qquad (46) \\
= \ & -\theta_j r_j^T v_j + v_j^T H_k v_j - \frac{1}{2} v_j^T H_k v_j = -\theta_j r_j^T (\sum_{i=0}^{j-1} \lambda_i d_i) + \frac{1}{2} v_j^T H_k v_j \\
= \ & \frac{1}{2} v_j^T H_k v_j \geq \frac{\zeta}{4} \|D_k v_j\|^2. \qquad (47)
\end{aligned}
$$

Therefore,

$$
\frac{f(x_k) - f(x_k + v_j)}{f(x_k) - \psi_k(v_j)} \geq 1 - \frac{o(\|v_j\|^2)}{f(x_k) - \psi_k(v_j)} \geq 1 - \frac{o(\|v_j\|^2)}{\frac{\zeta}{4}\|D_k v_j\|^2}. \qquad (48)
$$

Since $\|v_j\| = \|D_k^{-1} D_k v_j\| \leq \|D_k^{-1}\| \|D_k v_j\| \leq \chi_D \|D_k v_j\|$, we have $\frac{\|v_j\|}{\|D_k v_j\|} \leq \chi_D$ and hence

$$
\lim_{k \to \infty} \frac{o(\|v_j\|^2)}{\|D_k v_j\|^2} = \lim_{k \to \infty} \frac{o(\|v_j\|^2)}{\|v_j\|^2} \cdot \frac{\|v_j\|^2}{\|D_k v_j\|^2} = 0. \qquad (49)
$$

Combining (48) with (49), we deduce that $p_k = -D_k^{-1}(D_k^{-1} H_k D_k^{-1})^{-1} D_k^{-1} g_k$ for sufficiently large $k$.

Next, we prove that $p_k = -D_k^{-1}(D_k^{-1} H_k D_k^{-1})^{-1} D_k^{-1} g_k$ satisfies (9). Using (45), we have

$$
\lim_{k \to \infty} \|p_k\| = 0. \qquad (50)
$$

Because $f(x_k)$ is twice continuously differentiable, $g_k^T p_k = -p_k^T H_k p_k$, by (37) and (40), we have that

$$\begin{aligned}
f(x_k+p_k) &= f(x_k) + g_k^T p_k + \frac{1}{2} p_k^T \nabla^2 f(x_k) p_k + o(\|p_k\|^2) \\
&= f(x_k) + \beta g_k^T p_k + (\frac{1}{2} - \beta) g_k^T p_k + \frac{1}{2}(g_k^T p_k + p_k^T H_k p_k) \\
&\quad + \frac{1}{2} p_k^T [\nabla^2 f(x_k) - H_k] p_k + o(\|p_k\|^2) \\
&\leq f(x_k) + \beta g_k^T p_k - (\frac{1}{2} - \beta) p_k^T H_k p_k + o(\|p_k\|^2) \\
&\leq f(x_k) + \beta g_k^T p_k - (\frac{1}{2} - \beta) \frac{\zeta}{2} \|D_k p_k\|^2 + o(\|p_k\|^2).
\end{aligned}$$

By (49), we deduce that $f(x_k+p_k) \leq f(x_k) + \beta g_k^T p_k$ for large enough $k$, namely, $p_k = -D_k^{-1}(D_k^{-1} H_k D_k^{-1})^{-1} D_k^{-1} g_k$ satisfies (9).

Finally, we prove that (10) holds. Similar to the proof of Theorem 1, we know there exists an index $j$ such that

$$\min\{\max\{\frac{l^i - x_*^i}{p_*^i}, \frac{u^i - x_*^i}{p_*^i}\}\ i = 1,\ 2,\ \cdots,\ n\} = \max\{\frac{l^j - x_*^j}{p_*^j}, \frac{u^j - x_*^j}{p_*^j}\}.$$

In the case of $l^j < x_*^j < u^j$, noting (50), we have $\lim_{k\to\infty} \alpha_k^* = +\infty$. Otherwise, without loss of generality, we assume $x_*^j = l^j$. Consider the following two cases: If $p_k^j > 0$, combing $\max\{\frac{l^j - x_k^j}{p_k^j}, \frac{u^j - x_k^j}{p_k^j}\} = \frac{u^j - x_k^j}{p_k^j}$ with (50), we also have $\lim_{k\to\infty} \alpha_k^* = +\infty$.

If $p_k^j < 0$, similar to the proof of (33), we get

$$\lim_{k\to\infty} \max\{\frac{l^j - x_k^j}{p_k^j}, \frac{u^j - x_k^j}{p_k^j}\} \geq 1. \tag{51}$$

From $\theta_k - 1 = O(\|p_k\|)$ and (50), we have $\lim_{k\to\infty} \theta_k = 1$, so $\alpha_k = 1$ when $k$ is large enough and $p_k = -D_k^{-1}(D_k^{-1} H_k D_k^{-1})^{-1} D_k^{-1} g_k$ satisfies (10).

From above discussions, we obtain that if $D_k^{-1} H_k D_k^{-1}$ is positive definite, the new iterate step is $x_{k+1} = x_k + p_k$, $p_k$ is Newton or quasi-Newton step, so (41) holds. □

Theorem 3 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 $x_*$ quadratically.

## 5. Numerical experiments

In this section we present the numerical results. In order to check effectiveness of the method, we take the elements of $D(x)$(see [3]) as

$$v^i(x) = \begin{cases} x^i - u^i, & \text{if } g^i < 0, \text{ and } u^i < +\infty, \\ x^i - l^i, & \text{if } g^i \geq 0, \text{ and } l^i > -\infty, \\ -1, & \text{if } g^i < 0, \text{ and } u^i = +\infty, \\ 1, & \text{if } g^i \geq 0, \text{ and } l^i = -\infty, \end{cases} \tag{52}$$

and select the parameters as following: $\epsilon = 10^{-8}$, $\xi = 0.02$, $\beta = 0.4$, $\omega = 0.5$. The experiments are carried out on 6 test problems which are quoted from [5] and [12]. NF and NG stand for the numbers of function evaluations and gradient evaluations, respectively, $M$ denotes the nonmonotonic parameter. The results of numerical experiments are reported to show the effectiveness of the proposed algorithm.

Experimental results

| Problem | the optimal solution and the optimal value | | $M$=0 | | $M$=3 | |
|---|---|---|---|---|---|---|
| name | reference results | results of algorithm | NG | NF | NG | NF |
| SC229 | $x^* = (1,1)^T$ $f^* = 0$ | $x^* = (1,1)^T$ $f^* = 5.5122 \times 10^{-16}$ | 156 | 158 | 159 | 160 |
| SC208 | $x^* = (1,1)^T$ $f^* = 0$ | $x^* = (1,1)^T$ $f^* = 3.5821 \times 10^{-18}$ | 53 | 64 | 54 | 60 |
| SC206 | $x^* = (1,1)^T$ $f^* = 0$ | $x^* = (1,1)^T$ $f^* = 1.9771 \times 10^{-29}$ | 5 | 5 | 5 | 5 |
| SC201 | $x^* = (5,6)^T$ $f^* = 0$ | $x^* = (5,6)^T$ $f^* = 0$ | 2 | 2 | 2 | 3 |
| Ferraris Tronconi | $x^* = (0.5, 3.14159)^T$ $f^* = 0$ | $x^* = (0.5, 3.1416)^T$ $f^* = 5.5164 \times 10^{-20}$ | 13 | 13 | 13 | 13 |
| Reklaitis Ragsdell | $x^* = (3,2)^T$ $f^* = 0$ | $x^* = (3,2)^T$ $f^* = 0$ | 15 | 15 | 15 | 15 |

## References


1. S.Bellavia, M.Macconi and B.Morini, *An affine scaling trust-region approach to bound-constrained nonlinear systems*, Appl. Nume. Mathematics **44** (2003), 257-280.
2. J.P.Bulteau and J.Ph.Vial, *Curvilinear path and trust region in unconstrained optimization, a convergence analysis*, Math. Prog. Study, **30**(1987), 82-101.
3. T.F.Coleman and Y.Li, *An interior trust region approach for minimization subject to bounds*, SIAM J. Optimization **6(3)**(1996), 418-445.
4. J.E.Jr.Dennis and R.B.Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall Series in Computation; Mathematics, Englewood Cliffs. NJ, 1983.
5. Floudas,C.A.,et al., *Handbook of Test Problems In Local and Global Optimization,* Kluwer Academic, Dordrecht, **33**(1999).
6. N.I.M.Gould, S.Lucidi, M.Roma and P.L.Toint, *Solving the trust-region subproblem using the Lanczos method* , SIAM J. Optimization **9(2)**(1999), 504-525.

7. L.Gripp, F.Lampariello and S.Lucidi, *A nonmonotone line search technique for Newton's methods*, SIAM J. Nume. Analysis, **23**(1986), 707-716.
8. Y.Guan, and D.Zhu, *Solving the unconstrained nonlinear optimization using the Lanczos path method* , Nume. Math. J. Chin. Universities, **S1**(2005), 46-51.
9. P.Guo and D.Zhu, *A nonmonotonic interior point algorithn via optimal path for nonlinear optimization with bounds to variables* , J. Shan. Norm. University **33(3)**(2004), 23-29.
10. 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. Programming **86**(1999), 615-635.
11. C.Jia and D.Zhu, *An Affine Scaling Interior Algorithm Via Lanczos Path for Solving Bound-constrained Nonlinear Systems*, Appl. Math. and Computation, **195**(2008), 558-575.
12. K.Schittkowski, *More Test Examples for Nonlinear Programming Codes*, Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin **282**(1987).
13. J.Sun, *A convergence proof for an affine-scaling algorithm for convex quadratic programming without nondegeneracy assumptions*, Mathe. Programming, **60**(1993), 69-79.
14. Y.Ye, *On affine scaling algorithms for nonconvex quadratic programming*, Math. Programming, **56** (1992) 285-300.
15. D.Zhu, *Curvilinear paths and trust region methods with nonmonotonic back tracking techniaue for unconstrained optimization*, J. Comp. Mathematics, **19**(2001) 241-258.
16. D.Zhu, *An interior point algorithm with nonmonotonic backtracking technique for linear constrained optimization* , J. Comp. and Appl. Mathematics, **155**(2003) 285-305.

**Chunxia Jia** received her M.Sc. from Fudan University and Ph.D from Shanghai Normal University. Her research interest is nonlinear optimization.

Mathematics and Science College, Shanghai Normal University, Shanghai, 200234, China.
e-mail: jiachx@shnu.edu.cn

**Detong Zhu** received his Ph.D from City University of Hong Kong. His research interest is nonlinear optimization.

Business College, Shanghai Normal University, Shanghai, 200234, China.
e-mail: dtzhu@shnu.edu.cn