Journal of Computational Mathematics, Vol.26, No.4, 2008, 578–597.

AN AFFINE SCALING INTERIOR ALGORITHM VIA CONJUGATE GRADIENT PATH FOR SOLVING BOUND-CONSTRAINED NONLINEAR SYSTEMS*

Chunxia Jia
Mathematics and Sciences College, Shanghai Normal University, Shanghai 200234, China
Email: jiachx@shnu.edu.cn
Detong Zhu
Business College, Shanghai Normal University, Shanghai 200234, China
Email: dtzhu@shnu.edu.cn

Abstract

In this paper we propose an affine scaling interior algorithm via conjugate gradient path for solving nonlinear equality systems subject to bounds on variables. By employing the affine scaling conjugate gradient path search strategy, we obtain an iterative direction by solving the linearize model. By using the line search technique, we will find an acceptable trial step length along this direction which is strictly feasible and makes the objective function nonmonotonically decreasing. The global convergence and fast local convergence rate of the proposed algorithm are established under some reasonable conditions. Furthermore, the numerical results of the proposed algorithm indicate to be effective.

Mathematics subject classification: 90C30, 65K05.
Key words: Conjugate gradient path, Interior points, Affine scaling.

1. Introduction

In this paper we use an affine scaling interior conjugate gradient path method to analyze the solution of nonlinear systems subject to the bound constraints on variable:

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

where $F : \mathcal{X} \to \Re^n$ is a given continuously differentiable mapping and $\mathcal{X} \subseteq \Re^n$ is an open set containing the n-dimensional box constraint Ω. 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

$$\mathrm{int}(\Omega) \overset{\mathrm{def}}{=} \{\ x \mid l < x < u\ \}$$

is nonempty, where $l < u$. 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 $\Re^n$ or a closed box constraint. In the classical methods for solving the unconstrained nonlinear equations (1.1) when the function $F(x)$ is a continuously differentiable function, the Newton method or quasi-Newton method can be used. These methods by using the Jacobian or version of Newton's method often solve the unconstrained problem (1.1), which is known to

* Received July 3, 2006 / Revised version received October 15, 2007 / Accepted December 4, 2007 /

have locally very rapid convergence (see, e.g., [3, 4]). However, the Newton methods used for smooth systems (1.1) does not ensure global convergence, that is, the convergence is only local. Other methods for solving (1.1) can be found in, e.g., [11, 14].

Many papers about affine-scaling algorithm for solving problems appeared during the last few years. Sun in [9] gave a convergence proof for an affine-scaling algorithm for convex quadratic programming without nondegeneracy assumptions, and Ye [12] introduced affine scaling algorithm for nonconvex quadratic programming. Classical trust-region Newton method for solving the nonlinear system (1.1) and the affine scaling double trust-region approach for solving the bounded constrained optimization problems are given in [2]. Recently, Bellavia et al. in [1] further extended the idea and presented an affine scaling trust-region approach for solving the bounded-constrained smooth nonlinear systems (1.1). However, the search direction generated in trust-region subproblem must satisfy strict interior feasibility which results in computational difficulties. In this paper, we introduce an affine scaling interior algorithm via conjugate gradient path to solve the bound-constrained nonlinear systems (1.1).

In order to describe and design the affine scaling interior conjugate gradient path algorithm for solving the bound-constrained smooth equations (1.1), we first introduce the squared Euclidean norm of linearize model of the unconstrained systems (1.1) and the augmented quadratic affine scaling model, and state the affine scaling conjugate gradient path with backtracking interior point technique for the bound-constrained 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 method 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 conjugate gradient 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.

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

$$\begin{aligned} \min \quad & q_k(d) \stackrel{\text{def}}{=} \frac{1}{2}\|F_k^{'}d + F_k\|^2 = \frac{1}{2}\|F_k\|^2 + F_k^T F_k^{'} d + \frac{1}{2} d^T (F_k^{'T} F_k^{'}) d \\ \text{s.t.} \quad & \|D_k d\| \le \Delta_k, \end{aligned} \tag{2.1}$$

where F' is the Jacobi matrix of F, Δ_k is the trust region radius and $q_k(d)$ is trusted to be an adequate representation of the merit function

$$f(x) \stackrel{\text{def}}{=} \frac{1}{2}\|F(x)\|^2. \tag{2.2}$$

The scaling matrix $D_k = D(x_k)$ arises 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

$$D(x) \stackrel{\text{def}}{=} \text{diag}\{|v^1(x)|^{-\frac{1}{2}}, \cdots, |v^n(x)|^{-\frac{1}{2}}\} \tag{2.3}$$

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

$$v^i(x)\overset{\text{def}}{=}\begin{cases} x^i - u^i, & \text{if } g^i < 0, \text{ and } u^i < +\infty, \\ x^i - l^i, & \text{if } g^i \geq 0, \text{ and } l^i > -\infty, \\ -1, & \text{if } g^i < 0, \text{ and } u^i = +\infty, \\ 1, & \text{if } g^i \geq 0, \text{ and } l^i = -\infty, \end{cases} \tag{2.4}$$

here $g(x)\overset{\text{def}}{=}F'(x)^T F(x)$ 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 given in [2].

Definition 2.1 (see [2]). *A point $x \in \Omega$ is nondegenerate, if for each index i,*

$$g^i(x) = 0 \implies l^i < x^i < u^i. \tag{2.5}$$

A transformed problem (1.1) *is nondegenerate if* (2.5) *holds for every $x \in \Omega$.*

In order to maintain the strict interior feasibility, a step-back tracking along the solution p_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. Following the suggestion in [2], we can make some modifications on the trust region subproblem (2.1) 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 by Coleman and Li in [2],

$$C_k\overset{\text{def}}{=}\text{diag}\{g_k\}J_k^\nu, \tag{2.6}$$

where $J^\nu(x) \in \Re^{n\times n}$ is the Jacobian matrix of $|\nu(x)|$ whenever $|\nu(x)|$ is differentiable and $\text{diag}\{g_k\}\overset{\text{def}}{=}\text{diag}\{g_k^1, \cdots, g_k^n\}$, here g_k^i is the ith component of g_k. Each diagonal component of the diagonal matrix J^ν equals zero or ± 1. The augmented affine scaling trust region subproblem at the kth iteration is defined as follows

$$(S_k)\begin{cases} \min & \psi_k(p)\overset{\text{def}}{=}\frac{1}{2}\|F_k'p + F_k\|^2 + \frac{1}{2}p^T D_k C_k D_k p \\ & \qquad = f(x_k) + g_k^T p + \frac{1}{2}p^T H_k p \\ \text{s.t.} & \|D_k p\| \leq \Delta_k, \end{cases}$$

where

$$g_k\overset{\text{def}}{=}F_k'^T F_k, \ \ H_k\overset{\text{def}}{=}F_k'^T F_k' + D_k C_k D_k,$$

Δ_k is the trust region radius. For solving subproblem (S_k), we first introduce the affine conjugate gradient path $\Gamma_k(\tau)$.

2.1. Affine scaling conjugate gradient path

Starting from $v_1 = 0$, $r_1 = \nabla\psi_k(v_1) = g_k$, $s_1 = M_k^{-1}g_k$, $d_1 = -s_1$, then we generate a sequence of points $v_1, v_2, \cdots, v_{q+1}$, and a sequence of conjugate directions $d_1, d_2, \cdots, d_{q+1}$, which satisfy

$$M_k s_{i+1} = r_{i+1}, \qquad i = 1, 2, \cdots, q, \tag{2.7}$$

$$d_{i+1} = -s_{i+1} + \beta_i d_i, \quad i = 1, 2, \cdots, q, \tag{2.8}$$

$$v_{i+1} = v_i + \lambda_i d_i, \qquad i = 1, 2, \cdots, q, \tag{2.9}$$

$$d_i^T H_k d_i > 0, \qquad i = 1, 2, \cdots, q. \tag{2.10}$$

where

$$M_k \overset{\text{def}}{=} D_k^T D_k, \quad \lambda_i = \frac{r_i^T s_i}{d_i^T H_k d_i} > 0, \quad \beta_i = \frac{s_{i+1}^T H_k d_i}{d_i^T H_k d_i} > 0,$$
$$r_{i+1} = \nabla\psi_k(v_{i+1}) = H_k v_{i+1} + g_k = r_i + \lambda_i H_k d_i. \tag{2.11}$$

The procedure stops either because $r_{i+1} = 0$ or $r_{i+1} \neq 0$ ($d_{i+1} \neq 0$), but $d_{i+1}^T H_k d_{i+1} \leq 0$. In the former case v_{i+1} is a critical point of ψ_k; in the latter d_{i+1} is a (descent) direction of negative curvature. We define the conjugate gradient path by

$$\Gamma_k(\tau) = \sum_{i=1}^{q} t_i(\tau) d_i - t_{q+1}(\tau) d_{q+1}, \tag{2.12}$$

$$t_i(\tau) = \min\Big\{\lambda_i, \max\{0, \tau - \sum_{j=1}^{i-1} \lambda_j\}\Big\}, \tag{2.13}$$

where the conjugate direction d_i and λ_i are defined by (2.8) and (2.11), respectively. In this formula we take $\sum_{j=1}^{i-1} \lambda_j = 0$ for $i = 1$.

2.2. Properties of the conjugate gradient method

In this section, we give some properties of the conjugate gradient method.

Lemma 2.1. *Suppose that the directions d_i are generated by* (2.7)-(2.9), $1 \leq i, j \leq l \leq q+1$, *the following properties hold:*

$$r_i^T d_j = 0, \qquad 1 \leq j < i \leq l \leq q+1, \tag{2.14}$$
$$d_i^T H_k d_j = 0, \qquad i \neq j, \tag{2.15}$$
$$r_i^T M_k^{-1} r_j = 0, \qquad i \neq j, \tag{2.16}$$
$$d_i^T r_i = -r_i^T M_k^{-1} r_i, \qquad i = 1, 2, \cdots, q+1, \tag{2.17}$$
$$d_i^T M_k d_j > 0, \qquad i \neq j. \tag{2.18}$$

Proof. The proof is done by induction. Noting $d_1 = -M_k^{-1} g_k = -M_k^{-1} r_1$, we have $d_1^T r_1 = -r_1^T M_k^{-1} r_1$, that is, (2.17) holds for $i = 1$. By (2.7), (2.8) and (2.11), we can get

$$\begin{aligned}
&d_2^T H_k d_1 = (-s_2 + \beta_1 d_1)^T H_k d_1 = -s_2^T H_k d_1 + \beta_1 d_1^T H_k d_1 = 0,\\
&r_2^T M_k^{-1} r_1 = (r_1 + \lambda_1 H_k d_1)^T (-d_1) = -r_1^T d_1 - \lambda_1 d_1^T H_k d_1 = 0,\\
&r_2^T d_1 = r_2^T (-M_k^{-1} r_1) = -r_2 M_k^{-1} r_1 = 0,\\
&d_2^T M_k d_1 = (-s_2 + \beta_1 d_1)^T M_k d_1 = -(M_k^{-1} r_2)^T M_k d_1 + \beta_1 d_1^T M_k d_1 = \beta_1 d_1^T M_k d_1 > 0,
\end{aligned}$$

that is, (2.14)-(2.18) hold for $l = 2$.

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

$$r_{l+1} = H_k v_{l+1} + g_k = H_k(v_l + \lambda_l d_l) + g_k = r_l + \lambda_l H_k d_l, \tag{2.19}$$

we have that

$$r_{l+1}^T d_j = (r_l + \lambda_l H_k d_l)^T d_j = r_l^T d_j + \lambda_l d_l^T H_k d_j. \tag{2.20}$$

Because of the induction hypothesis, we have from (2.14) and (2.15) that $r_{l+1}^T d_j = 0$ for $1 \leq j \leq l-1$, by applying (2.7), (2.11) and (2.19), we find that

$$r_{l+1}^T d_l = r_l^T d_l - \frac{r_l^T s_l}{d_l^T H_k d_l} d_l^T H_k d_l = r_l^T d_l - r_l^T M_k^{-1} r_l = 0, \tag{2.21}$$

therefore, the relation (2.14) continues to hold when l is replaced by $l+1$, as claimed.

By applying (2.7), (2.8) and (2.19), we find that

$$\begin{aligned} r_{l+1}^T M_k^{-1} r_i &= (r_l + \lambda_l H_k d_l)^T M_k^{-1} r_i \\ &= r_l^T M_k^{-1} r_i + \lambda_l d_l^T H_k(-d_i + \beta_{i-1} d_{i-1}) \\ &= r_l^T M_k^{-1} r_i - \lambda_l d_l^T H_k d_i + \lambda_l \beta_{i-1} d_l^T H_k d_{i-1}. \end{aligned} \tag{2.22}$$

By combining (2.22) with the induction hypothesis for (2.15) and (2.16), we conclude that $r_{l+1}^T M_k^{-1} r_i = 0$ for $1 \leq i \leq l-1$, noting

$$\begin{aligned} r_{l+1}^T M_k^{-1} r_l &= r_l^T M_k^{-1} r_l - \lambda_l d_l^T H_k d_l + \lambda_l \beta_{l-1} d_l^T H_k d_{l-1} \\ &= r_l^T M_k^{-1} r_l - \frac{r_l^T s_l}{d_l^T H_k d_l} d_l^T H_k d_l \\ &= r_l^T M_k^{-1} r_l - r_l^T M_k^{-1} r_l = 0, \end{aligned}$$

we have that (2.16) holds when l is replaced by $l+1$.

To prove (2.15) holds as well, we use (2.7), (2.8), (2.19) and the induction hypothesis for (2.15) and (2.16) to deduce that

$$\begin{aligned} d_{l+1}^T H_k d_i &= (-s_{l+1} + \beta_l d_l)^T H_k d_i \\ &= -s_{l+1}^T \frac{r_{i+1} - r_i}{\lambda_i} + \beta_l d_l^T H_k d_i \\ &= -r_{l+1}^T M_k^{-1} \frac{r_{i+1} - r_i}{\lambda_i} + \beta_l d_l^T H_k d_i \\ &= -\left(r_{l+1}^T M_k^{-1} r_{i+1} - r_{l+1}^T M_k^{-1} r_i \right) / \lambda_i + \beta_l d_l^T H_k d_i = 0. \end{aligned} \tag{2.23}$$

From (2.8) and (2.11), we can get

$$\begin{aligned} d_{l+1}^T H_k d_l &= (-s_{l+1} + \beta_l d_l)^T H_k d_l \\ &= -s_{l+1}^T H_k d_l + \frac{s_{l+1}^T H_k d_l}{d_l^T H_k d_l} d_l^T H_k d_l = 0, \end{aligned}$$

so the induction argument holds for (2.15) also.

Noting (2.8) and (2.14), we show that (2.17) continues to hold when k is replaced by $k+1$ by the following argument:

$$d_{l+1}^T r_{l+1} = (-s_{l+1} + \beta_l d_l)^T r_{l+1} = -s_{l+1}^T r_{l+1} = -r_{l+1}^T M_k^{-1} r_{l+1}.$$

Finally, by applying (2.7), (2.8), (2.14) and the induction hypothesis for (2.18), we obtain that

$$\begin{aligned} d_{l+1}^T M_k d_i &= (-s_{l+1} + \beta_l d_l)^T M_k d_i \\ &= -r_{l+1}^T M_k^{-1} M_k d_i + \beta_l d_l^T M_k d_i \\ &= -r_{l+1}^T d_i + \beta_l d_l^T M_k d_i \\ &= \beta_l d_l^T M_k d_i > 0, \end{aligned}$$

so that $d_i^T M_k d_j > 0$ when $i \neq j$, as claimed. This completes the proof of the lemma. □

2.3. Nonmonotonic affine scaling interior conjugate gradient path algorithm

Now, we describe an affine scaling conjugate gradient path algorithm with nonmonotonic strick interior feasible backtracking line search technique for solving the bound-constrained systems (1.1).

Algorithm 1.

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 $x_0 \in \mathrm{int}(\Omega) \subseteq \Re^n$. Set $k = 0$, go to the main step.

Main step

1. Evaluate
$$f_k = f(x_k) \stackrel{\mathrm{def}}{=} \frac{1}{2}\|F(x_k)\|^2, \quad C_k, \quad g_k = \nabla f(x_k) \stackrel{\mathrm{def}}{=} (F_k^{'})^T F_k$$
and D_k given in (2.3).
2. If $\|D_k^{-1} g_k\| = \|D_k^{-1}(F_k^{'})^T F_k\| \le \varepsilon$, stop with the approximate solution x_k.
3. Form the affine scaling conjugate gradient path $\Gamma_k(\tau)$, set $p_k(\tau) = \Gamma_k(\tau)$, choose $\tau = \infty, \omega^{-n}, \omega^{-(n-1)}, \cdots$, until the following inequality is satisfied
$$f(x_k) - f(x_k + p_k(\tau)) \ge \xi \Big[f(x_k) - \psi_k(p_k(\tau)) \Big]. \tag{2.24}$$
4. Choose $\alpha_k = 1, \omega, \omega^2, \cdots$, until the following inequality is satisfied:
$$\begin{aligned} f(x_k + \alpha_k p_k(\tau_k)) \quad &\le \quad f(x_{l(k)}) + \alpha_k \beta g_k^T p_k(\tau_k), \\ &\text{with } \ x_k + \alpha_k p_k(\tau_k) \in \mathrm{int}(\Omega), \end{aligned} \tag{2.25}$$
where $f(x_{l(k)}) = \max_{0 \le j \le m(k)} \{f(x_{k-j})\}$.
5. Set
$$x_{k+1} = x_k + \alpha_k p_k. \tag{2.26}$$
6. 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 2.1. The scalar α_k given in (2.25) of Step 4 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,
$$\alpha_k^* \stackrel{\mathrm{def}}{=} \min \Big\{ \max\{ \frac{l^i - x_k^i}{p_k^i}, \frac{u^i - x_k^i}{p_k^i} \}, i = 1, \cdots, n \Big\}, \tag{2.27}$$
where $\alpha_k = \theta_k \alpha_k^*$, $\theta_k \in (\theta_l, 1]$, for some $0 < \theta_l < 1$ and $\theta_k - 1 = \mathcal{O}(\|p_k(\tau_k)\|)$, l^i, u^i, x_k^i and p_k^i are the ith components of l, u, x_k and p_k, respectively.

Remark 2.2. In fact, we will sequentially compute the conjugate directions $d_1, d_2, \cdots, d_q, d_{q+1}$ in Step 3, which satisfies (2.24). In practical, we can sequentially compute the conjugate gradient path $\Gamma_k(\tau) = \sum_{i=1}^q \lambda_i(\tau) d_i$ until (2.24) is satisfied, if $\Gamma_k(\tau) = \sum_{i=1}^{q-1} \lambda_i(\tau) d_i$ does not satisfy (2.24). So we compute τ given in (2.13) such that (2.24) holds.

3. Global Convergence Analysis

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

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

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 $\mathcal{L}(x_0)$ on $\Re^n$.

Assumption 2. There exist some positive constants χ_g and χ_D such that

$$\|F^{'T}(x)F(x)\| \le \chi_g, \ \|D(x)^{-1}\| \le \chi_D, \quad \text{for all } x \in \mathcal{L}(x_0).$$

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

Lemma 3.1. *Let the step $p_k(\tau) = \Gamma_k(\tau)$ be obtained from the affine scaling conjugate gradient path. We have*
(1) *The norm function of the path $\|p_k(\tau)\|_{M_k}$ is monotonically increasing for $\tau \in (0, +\infty)$, where $\|x\|_{M_k} = (x^T M_k x)^{\frac{1}{2}}$, $\forall x \in \Re^n$;*
(2) *The quadratic function $\psi_k(p_k(\tau))$ is monotonically decreasing for $\tau \in (0, +\infty)$;*
(3) *If H_k is positive definite, then*

$$\lim_{\tau \to \infty} p_k(\tau) = -D_k^{-1}(D_k^{-1} H_k D_k^{-1})^{-1} D_k^{-1} g_k. \tag{3.1}$$

Proof. (1) By the definition of $\| \cdot \|_{M_k}$, we have

$$\|\Gamma_k(\tau)\|_{M_k} = (\Gamma_k(\tau)^T M_k \Gamma_k(\tau))^{\frac{1}{2}}$$

and

$$\begin{aligned}\frac{d\|\Gamma_k(\tau)\|}{d\tau} &= \frac{1}{2\Gamma_k(\tau)^T M_k \Gamma_k(\tau)} \frac{d(\Gamma_k(\tau)^T M_k \Gamma_k(\tau))}{d\tau} \\ &= \frac{1}{2\Gamma_k(\tau)^T M_k \Gamma_k(\tau)} \Big[(\frac{d\Gamma_k(\tau)}{d\tau})^T M_k \Gamma_k(\tau) + \Gamma_k(\tau)^T M_k \frac{d\Gamma_k(\tau)}{d\tau} \Big] \\ &= \frac{1}{\Gamma_k(\tau)^T M_k \Gamma_k(\tau)} \Gamma_k(\tau)^T M_k \frac{d\Gamma_k(\tau)}{d\tau}.\end{aligned}$$

Noting $dt_i(\tau)/d\tau = 1$ or 0, $t_i(\tau) \ge 0$, and the definition of the conjugate gradient, we have $d\|\Gamma_k(\tau)\|/d\tau \ge 0$, the conclusion (1) holds.

(2) From (2.13), we have that

$$t_l(\tau) = \begin{cases} \tau - \sum_{j=1}^{i-1} \lambda_j & \text{if } \ l = i, \\ \lambda_l & \text{if } \ l < i, \\ 0 & \text{if } \ i < l \le q, \end{cases} \tag{3.2}$$

for $\sum_{j=1}^{i-1}\lambda_j < \tau \le \sum_{j=1}^{i}\lambda_j$ $(i \le q)$ and $t_l(\tau) = \lambda_l$ for $\tau > \sum_{j=1}^{q}\lambda_j$. Therefore,

$$\Gamma_k(\tau) = \begin{cases} \sum_{j=1}^{i-1}\lambda_j d_j + \Big(\tau - \sum_{j=1}^{i-1}\lambda_j\Big) d_i & \text{if } \sum_{j=1}^{i-1}\lambda_j \le \tau < \sum_{j=1}^{i}\lambda_j\ ; \\ \sum_{j=1}^{q}\lambda_j d_j & \text{if } \tau \ge \sum_{j=1}^{q}\lambda_j\ . \end{cases} \tag{3.3}$$

Noting that $\psi_k(\Gamma_k(\tau)) = f(x_k) + g_k^T\Gamma_k(\tau) + \frac{1}{2}\Gamma_k(\tau)^T H_k \Gamma_k(\tau)$, we have

$$\begin{aligned} \frac{d\psi_k(\Gamma_k(\tau))}{d\tau} &= \Big(\frac{d\psi_k(\Gamma_k(\tau))}{d\Gamma_k}\Big)^T d_i \\ &= g_k^T d_i + \Big[\sum_{j=1}^{i-1}\lambda_j d_j + (\tau - \sum_{j=1}^{i-1}\lambda_j) d_i\Big]^T H_k d_i \\ &= g_k^T d_i + \Big(\tau - \sum_{j=1}^{i-1}\lambda_j\Big) d_i^T H_k d_i \le r_1^T d_i + \lambda_i d_i^T H_k d_i \\ &= r_1^T d_i + r_i^T s_i = r_1^T d_i + r_i^T M_k^{-1} r_i = r_1^T d_i - r_i^T d_i \\ &= r_1^T d_i - \Big(r_1 + \sum_{j=1}^{i-1}\lambda_j H_k d_j\Big)^T d_i = -\sum_{j=1}^{i-1}\lambda_j d_j^T H_k d_i = 0, \end{aligned} \tag{3.4}$$

for $\sum_{j=1}^{i-1}\lambda_j < \tau \le \sum_{j=1}^{i}\lambda_j$ $(i \le q)$ and $d\psi_k(\Gamma_k(\tau))/\mathrm{d}\tau = 0$ for $\tau \ge \sum_{j=1}^{q}\lambda_j$. So the quadratic function $\psi_k(\Gamma_k(\tau))$ is monotonically decreasing for $\tau \in (0, +\infty)$.

(3) If H_k is positive definite, by the termination condition, we have $r_{q+1} = 0$. Noting $H_k v_{q+1} + g_k = r_{q+1} = 0$, then $v_{q+1} = -H_k^{-1} g_k$. By $\lambda_{q+1} = r_{q+1}^T s_{q+1}/(d_{q+1}^T H_k d_{q+1})$, we get

$$\lambda_{q+1} = 0, \quad \lim_{\tau\to\infty} t_{q+1}(\tau) = 0.$$

By (2.9) and (3.3), we have

$$\lim_{\tau\to\infty}\Gamma_k(\tau) = \sum_{j=1}^{q}\lambda_j d_j = v_{q+1} = -H_k^{-1} g_k = -D_k^{-1}(D_k^{-1} H_k D_k^{-1})^{-1} D_k^{-1} g_k.$$

So the conclusion (3) holds. This completes the proof of this lemma. □

The following lemma shows the relation between the gradient $g_k = F_k^{'T} F_k$ of the objective function and the step p_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 3.2. *Let the step $p_k(\tau) = \Gamma_k(\tau)$ be obtained from the affine scaling conjugate gradient path. Then:*
(1) *The function $\Phi_k(\tau) = g_k^T p_k(\tau)$ is monotonically decreasing for $\tau \in (0, +\infty)$.*
(2) *For $\tau \in (0, +\infty)$, the function $\Phi_k(\tau)$ satisfies the following sufficient descent condition:*

$$g_k^T p_k(\tau) \le -\min\Big\{\tau, \frac{1}{\|D_k^{-1} H_k D_k^{-1}\|}\Big\}\|D_k^{-1} g_k\|^2. \tag{3.5}$$

Proof. (1) By Lemma 3.1, if $\sum_{j=1}^{i-1}\lambda_j \le \tau < \sum_{j=1}^{i}\lambda_j$ $(i \le q)$, then

$$\begin{aligned}\Phi_k(\tau) = g_k^T\Gamma_k(\tau) &= g_k^T\Big(\sum_{j=1}^{i-1}\lambda_j d_j + (\tau - \sum_{j=1}^{i-1}\lambda_j)d_i\Big)\\ &= \sum_{j=1}^{i-1}\lambda_j g_k^T d_j + \Big(\tau - \sum_{j=1}^{i-1}\lambda_j)g_k^T d_i\Big).\end{aligned}$$

Noting $d_1 = -M_k^{-1}g_k$ and $d_i^T M_k d_j > 0$, we can get

$$\Phi_k'(\tau) = g_k^T d_i = -d_1^T M_k d_i < 0. \tag{3.6}$$

(2) If $0 < \tau \le \lambda_1$, where

$$\lambda_1 = \frac{r_1^T s_1}{d_1^T H_k d_1} = \frac{\|D_k^{-1}g_k\|^2}{(D_k^{-1}g_k)^T(D_k^{-1}H_kD_k^{-1})(D_k^{-1}g_k)},$$

then

$$g_k^T\Gamma_k(\tau) = g_k^T(\tau d_1) = -\tau g_k^T M_k^{-1}g_k = -\tau g_k^T D_k^{-1}D_k^{-1}g_k = -\tau\|D_k^{-1}g_k\|^2.$$

Noting $\Phi_k(\tau)$ is monotonically decreasing for $\tau \in (0,+\infty)$, we have $\Phi_k(\tau) \le \Phi_k(\lambda_1)$ for $\tau > \lambda_1$, that is,

$$\begin{aligned}g_k^T\Gamma_k(\tau) &\le g_k^T\Gamma_k(\lambda_1) = -\lambda_1\|D_k^{-1}g_k\|^2\\ &= -\frac{1}{(D_k^{-1}g_k)^T(D_k^{-1}H_kD_k^{-1})(D_k^{-1}g_k)}\|D_k^{-1}g_k\|^4\\ &\le -\frac{1}{\|D_k^{-1}H_kD_k^{-1}\|}\|D_k^{-1}g_k\|^2,\end{aligned}$$

which gives (3.5). □

Lemma 3.3. *Let the step $p_k(\tau) = \Gamma_k(\tau)$ be obtained from the affine scaling conjugate gradient path. Then the predicted reduction satisfy the estimate:*

$$f(x_k) - \psi_k(p_k(\tau)) \ge \frac{1}{2}\min\Big\{\tau, \frac{1}{\|D_k^{-1}H_kD_k^{-1}\|}\Big\}\|D_k^{-1}g_k\|^2 \tag{3.7}$$

for all F_k', F_k, C_k, D_k and τ.

Proof. We consider first the case of $0 < \tau \le \lambda_1$. Here, we have

$$\begin{aligned}&f(x_k) - \psi_k(\Gamma_k(\tau)) = f(x_k) - \psi_k(\tau d_1)\\ =& -\tau g_k^T d_1 - \frac{1}{2}\tau^2 d_1^T H_k d_1 \quad (d_1 = -M_k^{-1}g_k = -D_k^{-1}D_k^{-1}g_k)\\ =& -\tau g_k^T d_1 - \frac{1}{2}\tau^2(D_k^{-1}g_k)^T D_k^{-1}H_kD_k^{-1}(D_k^{-1}g_k) \quad (\tau \le \lambda_1)\\ \ge& -\tau g_k^T d_1 - \frac{\tau}{2}\lambda_1\frac{\|D_k^{-1}g_k\|^2}{\lambda_1} = \frac{\tau}{2}\|D_k^{-1}g_k\|^2 \ge \frac{1}{2}\min\{\tau, \frac{1}{\|D_k^{-1}H_kD_k^{-1}\|}\}\|D_k^{-1}g_k\|^2\end{aligned}$$

and so (3.7) certainly holds.

For the next case, consider $\tau > \lambda_1$. Noting the quadratic function $\psi_k(\Gamma_k(\tau))$ is monotonically decreasing for $\tau \in (0, +\infty)$, we obtain that

$$\begin{aligned}
&f(x_k) - \psi_k(\Gamma_k(\tau)) \geq f(x_k) - \psi_k(\Gamma_k(\lambda_1))\\
=& - \lambda_1 g_k^T d_1 - \frac{1}{2}\lambda_1^2 d_1^T H_k d_1 = -\frac{\|D_k^{-1} g_k\|^2}{d_1^T H_k d_1} g_k^T d_1 - \frac{1}{2}\frac{\|D_k^{-1} g_k\|^4}{d_1^T H_k d_1}\\
=&\frac{1}{2}\frac{\|D_k^{-1} g_k\|^4}{(D_k^{-1} g_k)^T (D_k^{-1} H_k D_k)(D_k^{-1} g_k)} \geq \frac{1}{2}\min\{\tau, \frac{1}{\|D_k^{-1} H_k D_k^{-1}\|}\}\|D_k^{-1} g_k\|^2
\end{aligned}$$

for $\tau > \lambda_1$. This proves (3.7). □

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

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

Lemma 3.4. *Assume that Assumptions 1-3 hold. If $\|D_k^{-1} g_k\| \neq 0$, then Algorithm 1 produces an τ_k in a finite number steps which satisfies*

$$f(x_k) - f(x_k + p_k(\tau)) \geq \xi[f(x_k) - \psi_k(p_k(\tau))].$$

Proof. Using the triagonal inequality, we have

$$\begin{aligned}
&f(x_k) - f(x_k + p_k(\tau))\\
=& f(x_k) - \psi_k(p_k(\tau)) + \psi_k(p_k(\tau)) - f(x_k + p_k(\tau))\\
\geq& f(x_k) - \psi_k(p_k(\tau)) - |\psi_k(p_k(\tau)) - f(x_k + p_k(\tau))|.
\end{aligned}$$

Using Taylor's Theorem, we can get

$$\begin{aligned}
&|\psi_k(p_k(\tau)) - f(x_k + p_k(\tau))|\\
=& \Big|\frac{1}{2}\|F_k\|^2 + g_k^T p_k(\tau) + \frac{1}{2} p_k(\tau)^T H_k p_k(\tau) - \frac{1}{2}\|F(x_k + p_k(\tau))\|^2\Big|\\
=& \Big|\frac{1}{2}\|F_k\|^2 + g_k^T p_k(\tau) + \frac{1}{2} p_k(\tau)^T H_k p_k(\tau) - \frac{1}{2}\|F_k + \nabla F(x_k + tp_k) p_k(\tau)\|^2\Big|\\
\leq& \Big[\|\nabla F_k^T F_k - \nabla F(x_k + tp_k(\tau))^T F_k\| + \frac{1}{2}\|\nabla F_k^T \nabla F_k\\
&- \nabla F(x_k + tp_k(\tau))^T \nabla F(x_k + tp_k(\tau)) + D_k C_k D_k\| \cdot \|p_k(\tau)\|\Big] \cdot \|p_k(\tau)\|\\
\overset{\text{def}}{=}& \epsilon(x_k, p_k)\|p_k(\tau)\|,
\end{aligned}$$

where

$$\begin{aligned}
\epsilon(x_k, p_k) = \|\nabla F_k^T F_k - \nabla F^T (x_k + tp_k(\tau)) F_k\| + \frac{1}{2}\|\nabla F_k^T \nabla F_k\\
- \nabla F(x_k + tp_k(\tau))^T \nabla F(x_k + tp_k(\tau)) + D_k C_k D_k\| \cdot \|p_k(\tau)\|.
\end{aligned}$$

We assume $0 < \tau < \lambda_1$. Then

$$0 \leq \|p_k(\tau)\| = \|\tau d_1\| = \tau\|M_k^{-1} g_k\| = \tau\|D_k^{-2} g_k\| \leq \tau\chi_D^2\chi_g. \tag{3.8}$$

If $\|D_k^{-1}g_k\| \neq 0$, then there exists $\varepsilon > 0$ such that $\|D_k^{-1}g_k\| \geq \varepsilon$. Noting (3.8), we have that for sufficiently small τ,

$$\begin{aligned} f(x_k) - \psi_k(p_k(\tau)) &\geq \frac{1}{2}\min\{\tau, \frac{1}{D_k^{-1}H_kD_k^{-1}}\}\|D_k^{-1}g_k\|^2 \\ &\geq \frac{1}{2}\varepsilon^2\tau \geq \frac{1}{2}\frac{\varepsilon^2}{\chi_D^2\chi_g}\|p_k(\tau)\| = c_1\|p_k(\tau)\|, \end{aligned} \tag{3.9}$$

where $c_1 = \frac{1}{2}\varepsilon^2/(\chi_D^2\chi_g)$. Furthermore,

$$\begin{aligned} \frac{f(x_k) - f(x_k + p_k(\tau))}{f(x_k - \psi_k(p_k(\tau)))} &\geq 1 - \frac{|\psi_k(p_k(\tau)) - f(x_k + p_k(\tau))|}{f(x_k) - \psi_k(p_k(\tau))} \\ &\geq 1 - \frac{\epsilon(x_k, p_k(\tau))\|p_k(\tau)\|}{c_1\|p_k(\tau)\|} = 1 - \frac{\epsilon(x_k, p_k(\tau))}{c_1}. \end{aligned}$$

Noting that $\nabla F(x)$ is continuous and (3.8), we can get $\epsilon(x_k, p_k(\tau)) < c_1/3$ for τ is small enough. This completes the proof. □

Lemma 3.5. *Assume that Assumptions* 1-3 *hold and the gradient of* f *satisfies*

$$\|\nabla f(x) - \nabla f(y)\|_2 \leq \gamma\|x - y\|_2, \ \forall \ x, y \in \Re^n,$$

where γ *is the Lipschitz constant. Let* $\beta \in (0,1)$ *and the step* $p_k(\tau) = \Gamma_k(\tau)$ *be obtained from the affine scaling conjugate gradient path. If* $\|D_k^{-1}g_k\| \neq 0$*, then Algorithm 1 produces an iterate* $x_{k+1} = x_k + \alpha_k p_k$ *in a finite number of backtracking steps in* (**??**).

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

$$f(x_k + \alpha_k p_k(\tau_k)) = f(x_k) + \alpha_k \nabla f(x_k + \eta_k\alpha_k p_k(\tau_k))^T p_k(\tau_k),$$

where $0 \leq \eta_k \leq 1$. We rewrite the above equation as:

$$\begin{aligned} f(x_k + \alpha_k p_k(\tau_k)) &= f(x_k) + \alpha_k g_k^T p_k(\tau_k) + \alpha_k\big[\nabla f(x_k + \eta_k\alpha_k p_k(\tau_k)) - \nabla f(x_k)\big]^T p_k(\tau_k) \\ &= f(x_k) + \alpha_k\beta g_k^T p_k(\tau_k) + \alpha_k\big[(\nabla f(x_k + \eta_k\alpha_k p_k(\tau_k)) \\ &\quad - \nabla f(x_k))^T p_k(\tau_k) + (1-\beta)\nabla f(x_k)^T p_k(\tau_k)\big]. \end{aligned}$$

Note that

$$[\nabla f(x_k + \eta_k\alpha_k p_k(\tau_k)) - \nabla f(x_k)]^T p_k(\tau_k) \leq \gamma\eta_k\alpha_k\|p_k(\tau_k)\|^2,$$

from Lemma 3.1 and the condition $\|D_k^{-1}g(x_k)\| \neq 0$. After a finite number of reductions, the above formula will become negative and the corresponding α_k will be acceptable. That is, in a finite number of backtracking steps, α_k must satisfy

$$f(x_k + \alpha_k p_k(\tau_k)) \leq f(x_k) + \beta\alpha_k\nabla f(x_k)^T p_k(\tau_k),$$

equivalently,

$$f(x_k) \leq f(x_{l(k)}) + \beta\alpha_k\nabla f(x_k)^T p_k(\tau_k).$$

Consequently, the conclusion of the lemma holds. □

We are now ready to state one of our main results of the proposed algorithm. Before doing this, we need the following assumptions.

Assumption 4. $\|p_k(\tau_k)\|$ and H_k are uniformly bounded, that is, there exist constants χ_p, χ_H satisfy $\|p_k(\tau_k)\| \le \chi_p$ and $\|H_k\| \le \chi_H$ for all k.

Assumption 5. Assume

$$\frac{|v_k^j|}{|p_k^j|} \ge \frac{\varpi_k}{\sigma |g_k^j|},$$

where $v_i(x)$ are defined by (2.4), $\varpi_k \in (\varpi_l, 1]$, $0 < \varpi_l < 1$ and $\varpi_k - 1 = \mathcal{O}(\|p_k\|)$, $\sigma > 0$ is a constant.

Assumption 6. 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 3.6. *Assume that Assumptions* 1-6 *hold. Let* $\{x_k\} \subset \Re^n$ *be a sequence generated by the algorithm. If the nondegenerate property of the system* (1.1) *holds at any limit point, then*

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

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

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

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

$$\begin{aligned} f(x_{l(k+1)}) &= \max_{0 \le j \le m(k+1)} f(x_{k+1-j}) \\ &\le \max_{0 \le j \le m(k)+1} f(x_{k+1-j}) = \max_{0 \le j \le m(k)} f(x_{k-j}) = f(x_{l(k)}). \end{aligned}$$

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

$$\|D_k^{-1} g_k\| = \|D_k^{-1} F_k^{'T} F_k\| \ge \varepsilon.$$

From (**??**) and Lemma 3.2, we obtain

$$\begin{aligned} f(x_{l(k)}) &= f(x_{l(k)-1} + \alpha_{l(k)-1} p_{l(k)-1}(\tau_{l(k)-1})) \\ &\le \max_{0 \le j \le m(l(k)-1)} f(x_{l(l(k)-1-j)}) + \beta \alpha_{l(k)-1} g_{l(k)-1}^T p_{l(k)-1} \\ &\le f(x_{l(l(k)-1)}) - \beta \alpha_{l(k)-1} \min\left\{\tau_{l(k)-1}, \frac{1}{\chi_F + \chi_c}\right\} \epsilon^2. \end{aligned} \tag{3.11}$$

Since $\lim_{k\to+\infty} f(x_{l(k)})$ exists, we can conclude that

$$\lim_{k\to\infty} \alpha_{l(k)-1} \tau_{l(k)-1} = 0; \tag{3.12}$$

moreover, from (3.12) we can deduce

$$\lim_{k\to\infty} \alpha_{l(k)-1} \|p_{l(k)-1}\| = 0. \tag{3.13}$$

Similar to the proof of a theorem in [6], we have

$$\lim_{k\to\infty} f(x_k) = \lim_{k\to\infty} f(x_{l(k)}). \tag{3.14}$$

According to the acceptance rule (**??**), we have

$$\begin{aligned}&f(x_{l(k)})-f(x_k+\alpha_k p_k(\tau_k))\geq-\alpha_k\beta g_k^T p_k(\tau_k)\\ \geq\ &\alpha_k\beta\min\{\tau_k,\frac{1}{\|D_k^{-1}H_kD_k^{-1}\|}\}\|D_k^{-1}g_k\|^2\geq\alpha_k\beta\varepsilon^2\min\{\tau_k,\frac{1}{\chi_F+\chi_c}\}\geq 0.\end{aligned}$$

Noting (3.14) and the above formula, we obtain $\lim_{k\to\infty}\alpha_k\tau_k=0$, which implies that either

$$\lim_{k\to\infty}\inf\tau_k=0, \tag{3.15}$$

or

$$\lim_{k\to\infty}\alpha_k=0. \tag{3.16}$$

If (3.15) holds, from the acceptance rule of τ_k, we have

$$f(x_k)-f(x_k+p_k(\frac{\tau_k}{\omega}))<\xi\Big[f(x_k)-\psi_k(p_k(\frac{\tau_k}{\omega}))\Big],$$

which gives

$$-\Big[g_k^Tp_k(\frac{\tau_k}{\omega})+o(\|p_k(\frac{\tau_k}{\omega})\|)\Big]<\xi\Big[-g_k^Tp_k(\frac{\tau_k}{\omega})-\frac{1}{2}(p_k(\frac{\tau_k}{\omega}))^TH_kp_k(\frac{\tau_k}{\omega})\Big], \tag{3.17}$$

where

$$\|p_k(\tau)\|=\|\tau d_1\|=\tau\|M_k^{-1}g_k\|=\tau\|D_k^{-2}g_k\|\leq\tau\chi_D^2\chi_g\quad\text{for }0<\tau\leq\lambda_1.$$

From (3.17), we obtain

$$\lim_{k\to\infty}\frac{g_k^Tp_k(\tau_k/\omega)}{\|p_k(\tau_k/\omega)\|}=0,$$

which contradicts

$$\begin{aligned}\lim_{k\to\infty}\frac{g_k^Tp_k(\tau_k/\omega)}{\|p_k(\tau_k/\omega)\|}&\leq\lim_{k\to\infty}\frac{-\min\{\tau_k/\omega,\frac{1}{\|D_k^{-1}H_kD_k^{-1}\|}\}\|D_k^{-1}g_k\|^2}{\frac{\tau_k}{\omega}\|d_1\|}\\&=\lim_{k\to\infty}\frac{-\|D_k^{-1}g_k\|^2}{\|D_k^{-1}D_k^{-1}g_k\|}\leq-\frac{1}{\chi_D}\|D_k^{-1}g_k\|\leq-\frac{\epsilon}{\chi_D}.\end{aligned}$$

As a result (3.16) holds. If α_k is determined by (**??**), we have

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

which gives

$$f(x_k+\frac{\alpha_k}{\omega}p_k(\tau_k))-f(x_k)>\frac{\alpha_k}{\omega}\beta g_k^Tp_k(\tau_k). \tag{3.18}$$

On the other hand,

$$\begin{aligned}&f(x_k+\frac{\alpha_k}{\omega}p_k(\tau_k))-f(x_k)\\=\ &\frac{\alpha_k}{\omega}g_k^Tp_k(\tau_k)+\frac{\alpha_k}{\omega}\int_0^1 g(x_k+t\frac{\alpha_k}{\omega}p_k(\tau_k)-g(x_k))^Tp_k(\tau_k)dt\\ \leq\ &\frac{\alpha_k}{\omega}g_k^Tp_k(\tau_k)+\frac{1}{2}\gamma(\frac{\alpha_k}{\omega})^2\|p_k(\tau_k)\|^2,\end{aligned} \tag{3.19}$$

where γ is the Lipschitz constant. From (3.18) and (3.19), we have

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

Therefore,

$$\begin{aligned}\alpha_k &\geq \frac{2\omega(\beta-1)}{\gamma\|p_k(\tau_k)\|^2}g_k^T p_k(\tau_k) \geq \frac{2\omega(\beta-1)}{\gamma\chi_p^2}g_k^T p_k(\tau_k)\\ &\geq \frac{2\omega(1-\beta)}{\gamma\chi_p^2}\min\Big\{\tau_k, \frac{1}{\|D_k^{-1}H_kD_k^{-1}\|}\Big\}\|D_k^{-1}g_k\|^2\\ &\geq \frac{2\omega(1-\beta)}{\gamma\chi_p^2}\min\Big\{\tau_k, \frac{1}{\chi_F+\chi_c}\Big\}\epsilon^2 \geq 0. \qquad (3.20)\end{aligned}$$

From (3.20), we can conclude that $\lim_{k\to\infty}\tau_k = 0$, which contradicts $\lim_{k\to\infty}\inf\tau_k \neq 0$.

If α_k is determined by (2.25), let x_* be a limit point of $\{x_k\}$. Then there exists a subset $K_1 \subset \{k\}$ satisfies:

$$\lim_{k\to\infty, k\in K_1}\alpha_k^* = 0, \qquad \lim_{k\to\infty, k\in K_1}x_k = x_*.$$

From the expression of α_k^*, we know there exists an index j such that

$$\max\Big\{\frac{l^j - x_*^j}{p_*^j}, \frac{u^j - x_*^j}{p_*^j}\Big\} = 0,$$

so we can get a subset $K_2 \subset K_1$ such that:

$$\lim_{k\to\infty, k\in K_2}\max\Big\{\frac{l^j - x_k^j}{p_k^j}, \frac{u^j - x_k^j}{p_k^j}\Big\} = 0.$$

Without loss of generality, assume $x_*^j - l^j = 0$. If $p_k^j > 0$, by $p_k^j \leq \|p_k\| < \chi_p$, we get that for sufficiently large k,

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

If $p_k^j < 0$, by nondegeneration, we get $g_*^j > 0$; by the optimization condition, we get

$$\mu_*^j > 0, \quad v_*^j = 0, \quad g_*^j = \mu_*^j - v_*^j > 0,$$

so when k is large enough, $g_k^j > 0$. By the definition of $D(x)$, we know

$$v^j(x) = x^j - l^j, \quad \max\Big\{\frac{l^j - x_k^j}{p_k^j}, \frac{u^j - x_k^j}{p_k^j}\Big\} = \frac{l^j - x_k^j}{p_k^j} = \frac{|v_k^j|}{|p_k^j|}.$$

By Assumption 5, we get

$$\frac{|v_k^i|}{|p_k^i|} \geq \frac{\varpi_k}{\sigma|g_k|} \geq \frac{\varpi_l}{\sigma\chi_g} > 0,$$

which contradicts

$$\lim_{k\to\infty, k\in K_2}\max\Big\{\frac{l^j - x_k^j}{p_k^j}, \frac{u^j - x_k^j}{p_k^j}\Big\} = 0,$$

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 (3.17). Hence the conclusion of the theorem is true. □

4. Properties of the Local Convergence

Theorem 3.6 indicates 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 then prove a local convergence rate.

Theorem 4.1. *Assume that the Assumptions* 1-6 *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*

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

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

$$\|D_{m_i}(F_{m_i}^{'})^T F_{m_i}\| \geq \epsilon_1.$$

Theorem 3.6 guarantees the existence of another subsequence $\{D_{l_i}^{-1}(F_{l_i}^{'})^T F_{l_i}\}$ such that

$$\|D_k^{-1}(F_k^{'})^T F_k\| \geq \epsilon_2, \quad \text{for } m_i \leq k < l_i; \quad \|D_{l_i}^{-1}(F_{l_i}^{'})^T F_{l_i}\| \leq \epsilon_2,$$

for an $\epsilon_2 \in (0, \epsilon_1)$. Similar to the proof of Theorem 3.6, we have

$$\lim_{k\to\infty, m_i\leq k<l_i} f(x_{l(k)}) = \lim_{k\to\infty, m_i\leq k<l_i} f(x_k). \tag{4.2}$$

According to the acceptance rule in step 4, we have

$$f(x_{l(k)}) - f(x_k + \alpha_k p_k(\tau_k)) \geq -\alpha_k\beta g_k^T p_k(\tau_k) \geq \beta\tau\alpha_k\epsilon_2 \min\Big\{\tau_k, \frac{1}{\chi_D^2\chi_H}\Big\} \geq 0.$$

Similarly, we also get

$$\lim_{k\to\infty, m_i\leq k<l_i} \alpha_k\tau_k = 0.$$

For simplicity, we rewrite the above formula as

$$\lim_{k\to\infty} \alpha_k\tau_k = 0. \tag{4.3}$$

Assume there exists $\mathcal{K} \subset \{k\}$, such that $\lim_{k\to\infty, k\in\mathcal{K}}\tau_k > 0$. Then $\lim_{k\to\infty, k\in\mathcal{K}}\alpha_k = 0$. If α_k is determined by (2.25), similar to the proof of Theorem 3.6, we have $\lim_{k\to\infty, k\in\mathcal{K}}\alpha_k > 0$, so α_k is determined by (**??**). From the acceptance rule of α_k in (**??**) and (3.20), we have

$$0 = \lim_{k\to\infty, k\in\mathcal{K}} \alpha_k \geq \lim_{k\to\infty, k\in\mathcal{K}} \frac{2\omega(1-\beta)}{\gamma\chi_p^2} \min\Big\{\tau_k, \frac{1}{\chi_F + \chi_c}\Big\}\varepsilon_2^2 \geq 0. \tag{4.4}$$

Consequently,

$$\lim_{k\to\infty, k\in\mathcal{K}} \tau_k = 0, \tag{4.5}$$

which contradicts the assumption $\lim_{k\to\infty, k\in\mathcal{K}} \tau_k > 0$. Therefore,

$$\lim_{k\to\infty} \tau_k = 0. \tag{4.6}$$

By the definition of $p_k(\tau_k)$, we have

$$\lim_{k\to\infty}\|p_k(\tau_k)\| = 0. \tag{4.7}$$

Noting that

$$\begin{aligned} f(x_k + p_k(\tau_k)) &= f(x_k) + g_k^T p_k + \mathcal{O}(\|p_k(\tau_k)\|^2) \\ &\le f(x_{l(k)}) + \beta g_k^T p_k + (1-\beta) g_k^T p_k + \mathcal{O}(\|p_k(\tau_k)\|^2) \end{aligned}$$

for $m_i \le k < l_i$ and

$$\begin{aligned} &(1-\beta) g_k^T p_k + \mathcal{O}(\|p_k(\tau_k)\|^2) \\ &\le -\frac{1}{2}(1-\beta)\min\Big\{\tau_k, \frac{1}{\|D_k^{-1}H_kD_k^{-1}\|}\Big\}\|D_k^{-1}g_k\|^2 + \mathcal{O}(\|p_k(\tau_k)\|^2) \\ &\le -\frac{1}{2}(1-\beta)\min\Big\{\tau_k, \frac{1}{\chi_F+\chi_c}\Big\}\epsilon_2^2 + c(\chi_D^2\chi_g\tau_k)^2 < 0 \end{aligned}$$

for sufficiently large k, we have that $x_{k+1} = x_k + p_k(\tau_k)$. By (3.9), we know

$$f(x_k) - \psi_k(p_k(\tau_k)) \ge c_1\|p_k(\tau_k)\|;$$

hence, by (2.24), we can get

$$f(x_k) - f(x_k + p_k(\tau_k)) \ge \xi[f(x_k) - \psi_k(p_k(\tau_k))] \ge \xi c_1\|p_k(\tau_k)\| = c_2\|p_k(\tau_k)\|,$$

where $c_1 = \frac{1}{2}\epsilon_2^2/(\chi_D^2\chi_g), c_2 = \xi c_1$. From the above formula, we obtain

$$\begin{aligned} \|x_{m_i} - x_{n_i}\| &\le \sum_{k=m_i}^{n_i-1}\|x_{k+1} - x_k\| = \sum_{k=m_i}^{n_i-1}\|p_k(\tau_k)\| \\ &\le \frac{1}{c_2}\sum_{k=m_i}^{l_i-1}[f(x_k) - f(x_k + p_k(\tau_k))] = \frac{1}{c_2}[f(x_{m_i}) - f(x_{n_i})]. \end{aligned} \tag{4.8}$$

It follows from (4.8) and (3.14) that

$$\lim_{i\to\infty}\|x_{m_i} - x_{l_i}\| = 0. \tag{4.9}$$

From (4.9) and

$$\|F_{m_i}^{'}F_{m_i} - F_{l_i}^{'}F_{l_i}\| = \|\nabla f(x_{m_i}) - \nabla f(x_{l_i})\| \le \gamma\|x_{m_i} - x_{l_i}\|,$$

we have

$$\|x_{m_i} - x_{l_i}\| \le \epsilon_2, \quad \|F_{m_i}^{'}F_{m_i} - F_{l_i}^{'}F_{l_i}\| \le \gamma\epsilon_2$$

for sufficiently large i.

On the other hand, the level set $\mathcal{L}$ is compact, so $g(x)$ is uniformly continuous. Furthermore, we can get

$$\|g_{m_i} - g_{l_i}\| \le \varepsilon_2.$$

If there exists subsequence $\{l_i\}$ satisfy $\lim\limits_{i\to\infty}\{x_{l_i}\} = x_*$, then $\lim\limits_{i\to\infty}\{x_{m_i}\} = x_*$. By the definition of $v(x)$, we have

$$\lim_{i\to\infty}\{\mathrm{diag}(|v_{m_i}|^{\frac{1}{2}} - |v_{l_i}|^{\frac{1}{2}})g_{l_i}\} = 0, \tag{4.10}$$

and hence

$$\|(D_{m_i}^{-1} - D_{n_i}^{-1})g_{l_i}\| = \|\text{diag}(|v_{m_i}|^{\frac{1}{2}} - |v_{l_i}|^{\frac{1}{2}})g_{l_i}\| \le \varepsilon_2 \tag{4.11}$$

for sufficiently large i. Consequently,

$$\begin{aligned}\epsilon_1 &\le \|D_{m_i}^{-1}(F_{m_i}^{'})^T F_{m_i}\| \\ &\le \|D_{m_i}^{-1}\|\|F_{m_i}^{'T}F_{m_i} - F_{l_i}^{'T}F_{l_i}\| + \|D_{m_i}^{-1} - D_{l_i}^{-1}\|\|F_{l_i}^{'T}F_{l_i}\| + \|D_{l_i}^{-1}(F_{l_i}^{'})^T F_{l_i}\| \\ &\le (\chi_D\gamma + \chi_g + 1)\epsilon_2,\end{aligned}$$

which contradicts the assumption that $\epsilon_2 \in (0, \epsilon_1)$ can be arbitrarily small. □

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

Assumption 7. $D_*^{-1}H_*D_*^{-1}$ satisfies the strong second-order sufficient condition, i.e., there exists $\zeta > 0$ such that:

$$(D_*q)^T(D_*^{-1}H_*D_*^{-1})(D_*q) \ge \zeta\|D_*q\|^2, \quad \forall q. \tag{4.12}$$

Assumption 8.

$$\lim_{k\to\infty}\frac{\|[B_k - \nabla^2 f(x_k)]p_k(\tau_k)\|}{\|p_k(\tau_k)\|} = 0. \tag{4.13}$$

Because $C_k \to C_* = 0$, by Assumption 8, we have

$$\begin{aligned}&\lim_{k\to\infty}\frac{\|[H_k - \nabla^2 f(x_k)]p_k(\tau_k)\|}{\|p_k(\tau_k)\|} \\ \le &\lim_{k\to\infty}\frac{\|[B_k - \nabla^2 f(x_k)]p_k(\tau_k)\| + \|C_k p_k(\tau_k)\|}{\|p_k(\tau_k)\|} = 0,\end{aligned} \tag{4.14}$$

which implies that

$$p_k(\tau_k)^T(\nabla^2 f(x_k) - H_k)p_k(\tau_k) = o(\|p_k(\tau_k)\|^2). \tag{4.15}$$

Theorem 4.2. *Assume that Assumptions* 1-8 *hold and* $\{x_k\}$ *is a sequence produced by Algorithm 1 which converges to* x_*. *Then the convergence is superlinear. i.e.,*

$$\lim_{k\to\infty}\frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} = 0. \tag{4.16}$$

Proof. Since x^* is a point at which the second-order sufficient optimality conditions are satisfied, $D_*^{-1}H_*D_*^{-1}$ is positive definite. It is not difficult to verify that $D_k^{-1}H_kD_k^{-1}$ is also positive definite for large enough k. By Lemma 3.1 (3), we know $p_k(+\infty) = -H_k^{-1}g_k$. Now we prove that for large enough k, $p_k(+\infty) = -H_k^{-1}g_k$ satisfies (2.24)-(2.25) in the algorithm.

We first prove $p_k(+\infty)$ satisfies (2.24) for sufficiently large k.

By Assumption 7 we get that

$$(D_kp)^T(D_k^{-1}H_kD_k^{-1})D_kp \ge \zeta\|D_kp\|^2,$$

so we can deduce that $\|(D_k^{-1}H_kD_k^{-1})^{-1}\|$ is bounded, by combining this result with

$$\|D_k^{-1}(D_k^{-1}H_kD_k^{-1})^{-1}D_k^{-1}g_k\| \le \|D_k^{-1}\|\|(D_k^{-1}H_kD_k^{-1})^{-1}\|\|D_k^{-1}g_k\|$$

and (3.10), we have that

$$\lim_{k\to\infty}\|p_k(+\infty)\| = \lim_{k\to\infty}\|D_k^{-1}(D_k^{-1}H_kD_k^{-1})^{-1}D_k^{-1}g_k\| = 0. \tag{4.17}$$

By(4.15), we can get

$$\begin{aligned}&|\psi_k(p_k(+\infty)) - f(x_k + p_k(+\infty))|\\ =\ &|g_k^T p_k(+\infty) + \frac{1}{2}p_k(+\infty)^T H_k p_k(+\infty) - (g_k^T p_k(+\infty)\\ &+ \frac{1}{2}p_k(+\infty)^T \nabla^2 f(x_k)p_k(+\infty) + o(\|p_k(+\infty)\|^2))|\\ =\ &|\frac{1}{2}p_k(+\infty)^T(H_k - \nabla^2 f(x_k))p_k(+\infty) - o(\|p_k(+\infty)\|^2)| = o(\|p_k(+\infty)\|^2).\end{aligned} \tag{4.18}$$

It follows from Assumption 7 that $D_k^{-1}H_kD_k^{-1}$ is positive definite uniformly for sufficiently large k. Hence,

$$\begin{aligned}&f(x_k) - \psi_k(p_k(+\infty)) = -g_k^T p_k(+\infty) - \frac{1}{2}p_k(+\infty)^T H_k p_k(+\infty)\\ =\ &\frac{1}{2}p_k(+\infty)^T H_k p_k(+\infty) \geq \frac{\zeta}{2}\|D_k p_k(+\infty)\|^2.\end{aligned} \tag{4.19}$$

Therefore,

$$\begin{aligned}&\frac{f(x_k) - f(x_k + p_k(+\infty))}{f(x_k) - q_k(p_k(+\infty))}\\ \geq\ &1 - \frac{o(\|p_k(+\infty)\|^2)}{f(x_k) - q_k(p_k(+\infty))} \geq 1 - \frac{o(\|p_k(+\infty)\|^2)}{\frac{\zeta}{2}\|D_k p_k(+\infty)\|^2}.\end{aligned} \tag{4.20}$$

Since

$$\|p_k\| = \|D_k^{-1}D_kp_k\| \leq \|D_k^{-1}\|\|D_kp_k\| \leq \chi_D\|D_kp_k\|,$$

we have $\|p_k\|/\|D_kp_k\| \leq \chi_D$ and hence

$$\lim_{k\to\infty}\frac{o(\|p_k\|^2)}{\|D_kp_k\|^2} = \lim_{k\to\infty}\frac{o(\|p_k\|^2)}{\|p_k\|^2}\cdot\frac{\|p_k\|^2}{\|D_kp_k\|^2} = 0. \tag{4.21}$$

Combining (4.20) and (4.21), we deduce that $p_k(+\infty) = -H_k^{-1}g_k$ satisfies (2.24).

Next, we prove that $p_k(+\infty)$ satisfies (**??**). Let $p_k \overset{\text{def}}{=} p_k(+\infty) = -H_k^{-1}g_k$. Because $f(x_k)$ is twice continuously differentiable, $\beta \in (0, \frac{1}{2})$, $g_k^Tp_k = -p_k^TH_kp_k$. By (4.12) and (4.15), we have that

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

By (4.21), we deduce that $f(x_k+p_k) \le f(x_k)+\beta g_k^T p_k$ for large enough k, i.e., p_k satisfies (**??**).

Finally, we prove that (2.25) holds. By (4.17), $\lim_{k\to\infty} p_k^i(+\infty)=0$. Noting

$$\lim_{k\to\infty} \|D_k^{-1}g_k\|=0, \quad \lim_{k\to\infty} \|D_k^{-1}g_k\|=\|D_*^{-1}g_*\|,$$

we can get $\|D_*^{-1}g_*\|=0$ and hence $g_*=0$. By the definition of nodegeneration, we obtain $l<x_*<u$ and $l^i<x_*^i<u^i$ $(\forall\ 1\le i\le n)$, so there exists some $\epsilon_0>0$ satisfying $\min\{x_*^i-l^i, u^i-x_*^i\}>2\epsilon_0$, which implies $\min\{x_k^i-l^i, u^i-x_k^i\}>\epsilon_0$ for sufficiently large k. Therefore,

$$\lim_{k\to\infty} \max\left\{\frac{l^i-x_k^i}{p_k^i}, \frac{u^i-x_k^i}{p_k^i}\right\}=0, \tag{4.22}$$

which implies that $\alpha_k^*=+\infty$. Therefore, $p_k(+\infty)=-H_k^{-1}g_k$ satisfies (2.25).

From the above discussion, we obtain if H_k is positive definite, the new iterate step is $x_{k+1}=x_k+p_k(+\infty)$. $p_k(+\infty)$ is a Newton or quasi-Newton step, so (4.16) holds. □

Theorem 4.2 means that the local convergence rate for the proposed algorithm depends on the Hessian of the 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 quadratically to x_*.

5. Numerical Experiments

In this section we present some numerical results. In order to check the effectiveness of the method, we select the parameters as follows: $\epsilon=10^{-6}$, $\xi=0.02$, $\beta=0.4$, $\omega=0.5$.

Table 5.1: Numerical comparison.

Problem name	the optimal solution and the optimal value		M=0		M=5	
	reference results	results of algorithm	NG	NF	NG	NF
SC229	$x^*=(1,1)^T$ $f^*=0$	$x^*=(1,1)^T$ $f^*=1.6934\times10^{-12}$	21	25	20	20
SC208	$x^*=(1,1)^T$ $f^*=0$	$x^*=(1,1)^T$ $f^*=5.5417\times10^{-29}$	17	24	11	13
SC209	$x^*=(1,1)^T$ $f^*=0$	$x^*=(1,1)^T$ $f^*=6.8002\times10^{-24}$	55	74	35	37
SC201	$x^*=(5,6)^T$ $f^*=0$	$x^*=(5,6)^T$ $f^*=7.8886\times10^{-31}$	2	3	2	3
Ferraris Tronconi	$x^*=(0.5,3.14159)^T$ $f^*=0$	$x^*=(0.5,3.1416)^T$ $f^*=1.6668\times10^{-15}$	12	13	12	13
Reklaitis Ragsdell	$x^*=(3,2)^T$ $f^*=0$	$x^*=(3,2)^T$ $f^*=5.2577\times10^{-18}$	10	11	10	11

The experiments are carried out on 6 test problems which are quoted from [5] and [10]. NF and NG stand for the numbers of function evaluations and gradient evaluations, respectively, M denotes the nonmonotonic parameter. The results of the numerical experiments are reported in Table 5.1 to show the effectiveness of the proposed algorithm.

Acknowledgments. The authors gratefully acknowledge the partial supports of the National Science Foundation of China Grant (10471094), the Ph.D. Foundation Grant (0527003) of

Chinese Education Ministry, the Shanghai Leading Academic Discipline Project (T0401), the Scientific Computing Key Laboratory of Shanghai Universities, and the Science Foundation Grant (05DZ11) of Shanghai Education Committee. The authors would like to thank the anonymous referee for his/her detailed comments and helpful suggestions.

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] T.F. Coleman, Y. Li, An interior trust region approach for minimization subject to bounds, *SIAM J. Optimiz.*, **6**:3 (1996), 418-445.

[3] Dennis, J.E.Jr., Moré, J.J., Quasi-Newton methods, motivation and theory, *SIAM Rev.*, **19** (1977), 46-89.

[4] Dennis, J.E.Jr., Schnabel, R.B., Numerical Methods for Unconstrained Optimization and Nonlinear Equations Prentice-Hall Series in Computation; Mathematics, Englewood Cliffs, NJ, 1983.

[5] C.A. Floudas et al., Handbook of Test Problems in Local and Global Optimization, Vol. 33, Kluwer Academic, Dordrecht, 1999.

[6] L. Gripp, F. Lampariello, S. Lucidi, A nonmonotone line search technique for Newton's methods, *SIAM J. Numer. Anal.*, **23** (1986), 707-716.

[7] I. Dikin, Iterative solution of problems of linear and quadratic programming, *Soviet Math. Dokl.*, **8** (1967), 674-675.

[8] J. Nocedal, Y. Yuan, Combining trust region and line search techniques, in Y. Yuan Advances in Nonlinear Programming (Kluwer, Dordrecht, 1998), 153-175.

[9] J. Sun, A convergence proof for an affine-scaling algorithm for convex quadratic programming without nondegeneracy assumptions, *Math. Program.*, **60** (1993), 69-79.

[10] K. Schittkowski, More Test Examples for Nonlinear Programming Codes, Lecture Notes in Economics and Mathematical Systems, 282, Springer-Verlag, Berlin (1987).

[11] X. Xu, C. Li, Convergence of Newton's method for systems of equations with constant rank derivatives, *J. Comput. Math.*, **25** (2007), 705-718.

[12] Y. Ye, On affine scaling algorithms for nonconvex quadratic programming, *Math. Program.*, **56** (1992) 285-300.

[13] Y. Yuan, On the truncated conjugate gradient method, *Math. Program.*, **87** (2000), 561-571.

[14] W. Zhou, D. Li, Limited memory BFGS method for nonlinear monotone equations, *J. Comput. Math.*, **25** (2007), 89-96.

[15] D. Zhu, Curvilinear paths and trust region methods with nonmonotonic back tracking techniaue for unconstrained optimization, *J. Comput. Math.*, **19** (2001), 241-258.

[16] D. Zhu, An interior point algorithm with nonmonotonic backtracking technique for linear constrained optimization, *J. Comput. Appl. Math.*, **155** (2003), 285-305.