Nonmonotonic back-tracking trust region interior point algorithm for linear constrained optimization

Detong Zhu

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

Received 1 February 2002; received in revised form 12 October 2002

Abstract

In this paper, we modify the trust region interior point algorithm proposed by Bonnans and Pola in (SIAM J. Optim. 7(3) (1997) 717) for linear constrained optimization. A mixed strategy using both trust region and line-search techniques is adopted which switches to back-tracking steps when a trial step produced by the trust region subproblem may be unacceptable. The global convergence and local convergence rate of the improved algorithm are established under some reasonable conditions. A nonmonotonic criterion is used to speed up the convergence progress in some ill-conditioned cases. The results of numerical experiments are reported to show the effectiveness of the proposed algorithm.

© 2003 Elsevier Science B.V. All rights reserved.

MSC: 90C.30; 65K.05; 49M.40

Keywords: Trust region methods; Back tracking; Nonmonotonic technique; Interior points

1. Introduction

In this paper, we analyze the trust region interior point algorithm for solving the linear equality constrained optimization problem:

\[ \min \ f(x) \]
\[ \text{s.t. \ } Ax = b, \]
\[ x \geq 0, \]

where \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) is smooth nonlinear function, not necessarily convex, \( A \in \mathbb{R}^{m \times n} \) is a matrix \( b \in \mathbb{R}^m \) is a vector. There are quite a few articles proposing sequential convex quadratic programming methods with trust region idea. These resulting methods generate sequences of points in the interior...
of the feasible set. Recently, Bonnans and Pola in [2] presented a trust region interior point algorithm for the linear constrained optimization. The algorithm would use two matrices at each iteration. The first is $X_k = \text{diag}\{x_1^k, \ldots, x_n^k\}$, a scaling matrix, where $x_i^k$ is the $i$th component of $x_k > 0$, the current interior feasible point. The second matrix is $B_k$, a symmetric approximation of the Hessian $\nabla^2 f(x_k)$ of the objective function in which $B_k$ assumed to be positive semidefinite in [2]. A search direction at $x_k$ by solving the trust region convex quadratic programming:

$$\begin{align*}
\min \quad & \varphi_k(d) \overset{\text{def}}{=} f_k + g_k^T d + \frac{1}{2} d^T B_k d \\
\text{s.t.} \quad & A d = 0 \\
& d^T X_k^{-2} d \leq \Delta_k^2, \quad x_k + d > 0,
\end{align*}$$

(1.2)

where $g_k = \nabla f(x_k)$, $d = x - x_k, B_k, \varphi_k(d)$ is the local quadratic approximation of $f$ and $\Delta_k$ is the trust region radius. Let $d_k$ be the solution of the subproblem. Then Bonnans and Pola in [2] used the line search to computing $\lambda_{lk}$, with $l_k$ the smallest nonnegative integer such that

$$f(x_k + \lambda_k d_k) \leq f(x_k) + \lambda_k \beta (\varphi_k(d_k) - f_k),$$

(1.3)

where $\beta \in (0, \frac{1}{2})$ and $\omega \in (0, 1)$, we obtain next step,

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

(1.4)

However, Bonnans and Pola in [2] assumed that $\varphi_k(d_k)$ is a convex function, in order to obtain global convergence of the proposed algorithm. Trust region method is a well-accepted technique in nonlinear optimization to assure global convergence. One of the advantages of the model is that it does not require the objective function to be convex. A solution that minimizes the model function within the trust region is solved as a trial step. If the actual reduction achieved on the objective function $f$ at this point is satisfactory comparing with the reduction predicted by the quadratic model, then the point is accepted as a new iterate, and hence the trust region radius is adjusted and the procedure is repeated. Otherwise, the trust region radius should be reduced and a new trial point needs to be determined. It is possible that the trust region subproblem needs to be resolved many times before obtaining an acceptable step, and hence the total computation for completing one iteration might be expensive. Nocedal and Yuan [9] suggested a combination of the trust region and line-search method for unconstrained optimization. The plausible remedy motivates to switch to the line-search technique by employing the back-tracking steps at an unsuccessful trial step. Of course, the prerequisite for being able to making this shift is that although the trial step is unacceptable as next iterative point, it should provide a direction of sufficient descent. More recently, the nonmonotonic line-search technique for solving unconstrained optimization is proposed by Grippo et al. in [6]. Furthermore, the nonmonotone technique is developed to trust region algorithm for unconstrained optimization (see [3], for instance). The nonmonotonic idea motivates to further study the back-tracking trust region interior point algorithm, because monotonicity may cause a series of very small steps if the contours of objective function $f$ are a family of curves with large curvature.

The paper is organized as follows. In Section 2, we describe the algorithm which combines the techniques of trust region interior point, back-tracking step, scaling matrix and nonmonotonic search. In Section 3, weak global convergence of the proposed algorithm is established. Some further convergence properties such as strong global convergence and local convergence rate are discussed.
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 a method which combines nonmonotonic line-search technique with a trust region interior point algorithm.

2.1. Initialization step

Choose parameters $\beta \in (0, \frac{1}{2})$, $\omega \in (0, 1)$, $0 < \eta_1 < \eta_2 < 1$, $0 < \gamma_1 < \gamma_2 < 1 < \gamma_3$, $\varepsilon > 0$ and positive integer $M$. Let $m(0) = 0$. Choose a symmetric matrix $B_0$. Select an initial trust region radius $\Delta_0 > 0$ and a maximal trust region radius $\Delta_{\text{max}} \geq \Delta_0$, give a starting strictly feasible point $x_0 \in \text{int}(\Omega) = \{x | Ax = b, x > 0\}$. Set $k = 0$, go to the main step.

2.2. Main step

1. Evaluate $f_k = f(x_k)$, $g_k = \nabla f(x_k)$.
2. If $\|P_k \bar{g}_k\| \leq \varepsilon$, stop with the approximate solution $x_k$, here the projection map $P_k = I - \tilde{A}_k^T (\tilde{A}_k \tilde{A}_k^T)^{-1} \tilde{A}_k$ of the null subspace of $\mathcal{N}(\tilde{A}_k)$ with $\tilde{A}_k \equiv A x_k$, $\bar{g}_k \equiv X g_k$.
3. Solve subproblem

$$\min \quad \varphi_k (d) \equiv f_k + g_k^T d + \frac{1}{2} d^T B_k d$$

(s.t. $A d = 0$,

$\|X_k^{-1} d\| \leq \Delta_k$, $x_k + d > 0$).

Denote by $d_k$ the solution of the subproblem $(S_k)$.
4. Choose $\alpha_k = 1, \omega, \omega^2, \ldots$, until the following inequality is satisfied:

$$f(x_k + \alpha_k d_k) \leq f(x_{l(k)}) + \alpha_k \beta g_k^T d_k,$$

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

$$h_k = \alpha_k d_k,$$

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

Calculate

$$\text{Pred}(h_k) = f_k - \varphi_k (h_k),$$

$$\text{Ared}(h_k) = f(x_{l(k)}) - f(x_k + h_k),$$

$$\hat{\rho}_k = \frac{\text{Ared}(h_k)}{\text{Pred}(h_k)}.$$
and take
\[ A_{k+1} = \begin{cases} [\gamma_1 A_k, \gamma_2 A_k] & \text{if } \hat{\rho}_k \leq \eta_1, \\ (\gamma_2 A_k, A_k) & \text{if } \eta_1 < \hat{\rho}_k < \eta_2, \\ (A_k, \min\{\gamma_3 A_k, A_{\max}\}) & \text{if } \hat{\rho}_k \geq \eta_2. \end{cases} \]

Calculate \( f(x_{k+1}) \) and \( g_{k+1} \).

6. Take \( m(k + 1) = \min\{m(k) + 1, M\} \), and update \( B_k \) to obtain \( B_{k+1} \). Then set \( k \leftarrow k + 1 \) and go to step 2.

**Remark 1.** In the subproblem \((S_k)\), \( \phi_k(d) \) is a local quadratic model of the objective function \( f \) around \( x_k \). A candidate iterative direction \( d \) is generated by minimizing \( \phi_k(d) \) along the interior points of the feasible set within the ellipsoidal ball centered at \( x_k \) with radius \( A_k \).

**Remark 2.** A key property of this transformation in trust region subproblem \((S_k)\) is that \( X_k^{-1} d_k \) is at least unit distance from all bounds in the scaled coordinates; i.e., an arbitrary step \( X_k^{-1} d_k \) to the point \( x_k + d_k \) does not violate any bound if \( d_k^T X_k^{-2} d_k < 1 \). To see this, first observe that
\[ d_k^T X_k^{-2} d_k = \sum_{i=1}^{n} \left( \frac{d_k^i}{x_k^i} \right)^2 < 1 \]
implies \( |d_k^i| < x_k^i \), for \( i = 1, \ldots, n \), where \( x_k^i \) and \( d_k^i \) are the \( i \)th components of \( x_k \) and \( d_k \), respectively. Then no matter what the sign of \( d_k^i \) is, the inequality \( x_k^i + d_k^i > 0 \) holds. One typical method for solving the subproblem \((S_k)\) was presented in [2] in detail (also see [5,8,11]).

**Remark 3.** Note that in each iteration the algorithm solves only one trust region subproblem. If the solution \( d_k \) fails to meet the acceptance criterion (2.1) (take \( x_k = 1 \)), then we turn to line search, i.e., retreat from \( x_k + h_k \) until the criterion is satisfied.

**Remark 4.** We improved the trust region interior algorithm in [2] by using back-tracking trust region technique and the trust region radius adjusted depends on the traditional trust region criterion. At the line search, we use \( g_k^T d_k \) in (2.1) instead of \( \phi_k(d_k) - f_k \) in (1.3). The line-search criterion (2.1) is satisfied easier than criterion (1.3), because if \( B_k \) is positive semidefinite, then \( g_k^T d_k \leq \phi_k(d_k) - f_k \).

**Remark 5.** Comparing the usual monotone technique with the nonmonotonic technique, when \( M \geq 1 \), the accepted step \( h_k \) only guarantees that \( f(x_k + h_k) \) is smaller than \( f(x_{\ell(k)}) \). It is easy to see that the usual monotone algorithm can be viewed as a special case of the proposed algorithm when \( M = 0 \).

### 3. Global convergence

Throughout this section we assume that \( f : \mathbb{R}^n \to \mathbb{R} \) is twice continuously differentiable and bounded from below. Given \( x_0 \in \mathbb{R}^n \), the algorithm generates a sequence \( \{x_k\} \subset \mathbb{R}^n \). In our analysis,
we denote the level set of $f$ by
$$\mathcal{L}(x_0) = \{x \in \mathbb{R}^n | f(x) \leq f(x_0), \ Ax = b, \ x \geq 0\}.$$ 

The following assumption is commonly used in convergence analysis of most methods for linear equality constrained optimization.

**Assumption 1.** Sequence $\{x_k\}$ generated by the algorithm is contained in a compact set $\mathcal{L}(x_0)$ on $\mathbb{R}^n$. Matrix $A$ has full row-rank $m$.

Based on solving the above trust region subproblem, we give the following lemma.

**Lemma 3.1.** $d_k$ is a solution of subproblem $(S_k)$ if and only if there exist $\mu_k \geq 0$, $\lambda_k \in \mathbb{R}^m$ such that

$$(B_k + \mu_k X_k^{-2})d_k = -g_k + A^T \lambda_k,$$
$$A d_k = 0, \ x_k + d_k > 0,$$
$$\mu_k (A_k^2 - d_k^T X_k^{-2} d_k) = 0$$

(3.1)

holds and $B_k + \mu_k X_k^{-2}$ is positive semidefinite in $\mathcal{N}(A)$.

**Proof.** Suppose $d_k$ is the solution of $(S_k)$, it is a straightforward conclusion of the first-order necessary conditions that there exist $\mu_k \geq 0$ and $\lambda_k \in \mathbb{R}^m$ such that (3.1) holds. It is only need to show that $B_k + \mu_k X_k^{-2}$ is positive semidefinite in $\mathcal{N}(A)$.

Suppose that $d_k \neq 0$, since $d_k$ solves $(S_k)$, it also solves

$$\min\{\psi_k(d) \defeq g_k^T d + \frac{1}{2} d^T B_k d | Ad = 0, \ x_k + d > 0 \ | |X_k^{-1} d|| = ||X_k^{-1} d_k||\}.$$ 

It follows that $\psi_k(d) \geq \psi_k(d_k)$ for all $d$ such that $||X_k^{-1} d|| = ||X_k^{-1} d_k||$ and $Ad = 0$, that is,

$$g_k^T d + \frac{1}{2} d^T B_k d \geq g_k^T d_k + \frac{1}{2} d_k^T B_k d_k.$$  

(3.2)

Since $(B_k + \mu_k X_k^{-2})d_k = -g_k + A^T \lambda_k$ and $Ad = 0$, we have

$$g_k^T d = -d_k^T (B_k + \mu_k X_k^{-2}) d \quad \text{and} \quad g_k^T d_k = -d_k^T (B_k + \mu_k X_k^{-2}) d_k.$$  

(3.3)

Replacing the above two equalities into (3.2) and rearranging terms in (3.2) gives

$$(d - d_k)^T (B_k + \mu_k X_k^{-2})(d - d_k) \geq \mu_k (d^T X_k^{-2} d - d_k^T X_k^{-2} d_k) = 0,$$

because of $d_k \neq 0$, we have the conclusion.

If $d_k = 0$, by (3.1) we have $-g_k + A^T \lambda_k = 0$, so $d_k = 0$ solves

$$\min\{\frac{1}{2} d^T B_k d \ | d^T X_k^{-2} d \leq \lambda_k^2\}.$$
and we must conclude that $B_k$ is positive semidefinite. Since $\mu_k \geq 0$ is necessary, $B_k + \mu_k X_k^{-2}$ is positive semidefinite in $N(A)$.

Let $\mu_k \geq 0, \lambda_k \in \mathbb{R}^m, d_k \in \mathbb{R}^n$ satisfy (3.1) and $B_k + \mu_k X_k^{-2}$ be positive semidefinite in $N(A)$. Then for all $d \in N(A), x_k + d > 0$, we have

$$g_k^T d + \frac{1}{2} d^T (B_k + \mu_k X_k^{-2})d \geq g_k^T d_k + \frac{1}{2} d_k^T (B_k + \mu_k X_k^{-2})d_k.$$

It means

$$\psi_k(d) = g_k^T d + \frac{1}{2} d^T B_k d$$

$$= g_k^T d + \frac{1}{2} d^T (B_k + \mu_k X_k^{-2})d - \frac{1}{2} \mu_k d^T X_k^{-2}d$$

$$\geq g_k^T d_k + \frac{1}{2} d_k^T (B_k + \mu_k X_k^{-2})d_k - \frac{1}{2} \mu_k d_k^T X_k^{-2}d_k$$

$$= \psi_k(d_k) + \frac{1}{2} (d_k^T X_k^{-2}d_k - d_k^T X_k^{-2}d_k)$$

$$\geq \psi_k(d_k),$$

which proves that $d_k$ solves $(S_k)$. □

Lemma 3.1 establishes the necessary and sufficient condition concerning $\mu_k, d_k$, and $\lambda_k$ when $d_k$ solves (2.1). It is well known from solving the trust region algorithms in order to assure the global convergence of the proposed algorithm, it is a sufficient condition to show that at $k$th iteration the predicted reduction defined by $\text{Pred}(d_k) = f_k - \varphi_k(d_k)$ which is obtained by the step $d_k$ from trust region subproblem.

**Lemma 3.2.** Let the step $d_k$ be the solution of the trust region subproblem, then there exists $\tau > 0$ such that the step $d_k$ satisfies the following sufficient descent condition:

$$\text{Pred}(d_k) \geq \tau \|P_k g_k\| \left\{ 1, A_k, \frac{\|P_k \hat{g}_k\|}{\|X_k B_k X_k\|} \right\} \tag{3.4}$$

for all $g_k, B_k$ and $A_k$, where $P_k = I - \hat{A}_k^T (\hat{A}_k \hat{A}_k^T)^{-1} \hat{A}_k$ with $\hat{A}_k = AX_k$ and $\hat{g}_k = X_k g_k$.

**Proof.** From Remark 2, we know that the inequality $x_k + d > 0$ must hold, if $A_k < 1$. Assume first that $A_k < 1$.

The affine scaling transformation is the one-to-one mapping

$$\hat{d}_k = X_k^{-1} d_k, \tag{3.5}$$

which is extended for the equality constrained trust region subproblem ($S_k$) by defining the quantities

$$\hat{g}_k = X_k g_k, \tag{3.6}$$

$$\hat{A}_k = AX_k, \tag{3.7}$$

$$\hat{B}_k = X_k B_k X_k. \tag{3.8}$$
Thus, the basic affinely scaled subproblem at the $k$th iteration is

$$\min \  \hat{\lambda}_k(d) \overset{\text{def}}{=} \hat{g}_k^T \hat{d} + \frac{1}{2} \hat{d}^T \hat{B}_k \hat{d}$$

(\hat{S}_k) \quad \text{s.t.} \quad \hat{A}_k \hat{d} = 0, \quad \|\hat{d}\| \leq \Lambda_k.

Denote by $\hat{d}_k$ the solution of the subproblem ($\hat{S}_k$). Let $w_k = -P_k \hat{g}_k$, it is clear to see that $\hat{A}_k w_k = 0$. If $P_k \hat{g}_k \neq 0$, consider the following problem:

$$\min \  \phi_k(t) \overset{\text{def}}{=} - t \|P_k \hat{g}_k\|^2 + \frac{1}{2} t^2 (P_k \hat{g}_k)^T \hat{B}_k (P_k \hat{g}_k)$$

s.t. \quad \begin{align*}
0 & \leq t \leq \frac{\Lambda_k}{\|P_k \hat{g}_k\|} \quad \text{(3.9)}
\end{align*}

Let $t_k$ be the optimal solution of the above problem and $\varphi_k$ be the optimal value of the subproblem ($S_k$). Consider two cases:

1. $w_k^T \hat{B}_k w_k > 0$, set $t_k^* = \|w_k\|^2 / w_k^T \hat{B}_k w_k$, if $\|t_k^* w_k\| \leq \Lambda_k$, then $t_k = t_k^*$ is the solution of the subproblem (3.9), we have that

$$\varphi_k \leq \phi_k(t_k) = - \frac{\|w_k\|^4}{w_k^T \hat{B}_k w_k} + \frac{1}{2} \left( \frac{\|w_k\|^2}{w_k^T \hat{B}_k w_k} \right)^2 (w_k^T \hat{B}_k w_k) \leq - \frac{1}{2} \|w_k\|^2. \quad \text{(3.10)}$$

On the other hand, if $\|t_k^* w_k\| > \Lambda_k$, i.e., $\|w_k\| > (\Lambda_k / \|w_k\|^2) (w_k^T \hat{B}_k w_k)$ then set $t_k = \Lambda_k / \|w_k\|$, we have that

$$\varphi_k \leq \phi_k(t_k) = - \left( \frac{\Lambda_k}{\|w_k\|} \right) \|w_k\|^2 + \frac{1}{2} \left( \frac{\Lambda_k}{\|w_k\|} \right)^2 (w_k^T \hat{B}_k w_k) \leq - \frac{1}{2} \|w_k\| \Lambda_k. \quad \text{(3.11)}$$

2. $w_k^T \hat{B}_k w_k \leq 0$, set $t_k = \Lambda_k / \|w_k\|$, we have that

$$\varphi_k \leq \phi_k(t_k) = - \left( \frac{\Lambda_k}{\|w_k\|} \right) \|w_k\|^2 + \frac{1}{2} \left( \frac{\Lambda_k}{\|w_k\|} \right)^2 (w_k^T \hat{B}_k w_k) \leq - \frac{1}{2} \|w_k\| \Lambda_k. \quad \text{(3.12)}$$

As above, (3.10)–(3.12) mean that $\varphi_k \leq - \frac{1}{2} \|w_k\| \min \{\Lambda_k, \|w_k\| / \|\hat{B}_k\|\}$. From (3.10)–(3.12), we have that

$$\text{Pred}(d_k) \geq \tau \|P_k \hat{g}_k\| \min \left\{ \Lambda_k, \frac{\|P_k \hat{g}_k\|}{\|X_k B_k X_k\|} \right\}, \quad \text{(3.13)}$$

where set $\tau = \frac{1}{2}$.

If $\Lambda_k \geq 1$, let $\hat{d}_k$ be the solution of the following subproblem:

$$\min \ \hat{\lambda}_k(d) \overset{\text{def}}{=} \hat{g}_k^T \hat{d} + \frac{1}{2} \hat{d}^T \hat{B}_k \hat{d}$$

(\hat{S}_k) \quad \text{s.t.} \quad \hat{A}_k \hat{d} = 0, \quad \|\hat{d}\| < 1,$$
then $\tilde{d}_k$ is a feasible solution of subproblem $(\hat{S}_k)$ and $x_k + \tilde{d}_k > 0$. So, $\tilde{d}_k$ is also a feasible solution of subproblem $(S_k)$. Hence, similar to prove (3.13), we have

$$
\varphi_k \leq \hat{\psi}_k(\tilde{d}_k) \leq \tilde{\psi}_k(\tilde{d}_k) \leq -\tau \|P_k \hat{g}_k\| \min \left\{1, \frac{\|P_k \hat{g}_k\|}{\|X_k B_k X_k\|}\right\}.
$$

(3.14)

From (3.12)–(3.14), we have that the conclusion of the lemma holds.

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

**Lemma 3.3.** At the $k$th iteration, let $d_k$ be generated in trust region subproblem $(S_k)$, Lagrange multiplier estimates $\lambda_k = (\hat{A}_k \hat{A}_k^T)^{-1} \hat{A}_k \hat{g}_k$, then

$$
g_k^T d_k \leq -\tau_1 \|P_k \hat{g}_k\| \min \left\{1, \Delta_k, \frac{\|P_k \hat{g}_k\|}{\|X_k B_k X_k\|}\right\},
$$

(3.15)

where $\tau_1 > 0$ is a constant.

**Remark.** The Lagrange multiplier estimates $\lambda_k = (\hat{A}_k \hat{A}_k^T)^{-1} \hat{A}_k \hat{g}_k$, we can use standard first-order least-squares estimates by solving

$$(\hat{A}_k \hat{A}_k^T) \lambda_k = \hat{A}_k \hat{g}_k.$$

**Proof.** The inequality $x_k + d > 0$ must hold if $\Delta_k < 1$. Without loss of generality, assume that $\Delta_k < 1$. Since $d_k$ is generated in trust region subproblem $(S_k)$, Lemma 3.1 ensures that,

$$(B_k + \mu_k X_k^{-2})d_k = -g_k + A_k^T \lambda_k.$$

Noting $X_k$ diagonal matrix and from $\lambda_k = -((\hat{A}_k \hat{A}_k^T)^{-1} \hat{A}_k \hat{g}_k$ we have

$$
\mu_k X_k^{-1} d_k = -P_k \hat{g}_k - (X_k B_k X_k) X_k^{-1} d_k.
$$

we take norm in the above equation and obtain

$$
\mu_k \Delta_k = \mu_k \|X_k^{-1} d_k\| \leq \|P_k \hat{g}_k\| + \|X_k B_k X_k\| \|X_k^{-1} d_k\|.
$$

Hence, noting $\|X_k^{-1} d_k\| \leq \Delta_k$,

$$
0 \leq \mu_k \leq \frac{\|P_k \hat{g}_k\|}{\|X_k^{-1} d_k\|} + \|X_k B_k X_k\| \leq \frac{\|P_k \hat{g}_k\|}{\Delta_k} + \|X_k B_k X_k\|.
$$

(3.17)

From (3.1), we have that from

$$
X_k^{-1} d_k = -(X_k B_k X_k + \mu_k I)^{\dagger} P_k \hat{g}_k,
$$

where $B^{\dagger}$ is the general pseudo-inverse of $B$.

(3.18)
Then we have that from $\hat{A}_k \hat{d}_k = 0$,
\[
g_k^T d_k = (X_k g_k)^T X_k^{-1} d_k \\
= (P_k \hat{g}_k)^T X_k^{-1} d_k \\
= -(P_k \hat{g}_k)^T (X_k B_k X_k + \mu_k I) P_k \hat{g}_k \\
\leq - \frac{1}{\|X_k B_k X_k\| + \mu_k} \|P_k \hat{g}_k\|^2 \\
\leq - \frac{\|P_k \hat{g}_k\|^2}{2 \|X_k B_k X_k\| + \|P_k \hat{g}_k\|/\Delta_k} \\
\leq - \frac{\|P_k \hat{g}_k\|^2}{2 \max\{2 \|X_k B_k X_k\|, \|P_k \hat{g}_k\|/\Delta_k\}} \\
\leq - \frac{1}{4} \|P_k \hat{g}_k\| \min\left\{ \frac{\|P_k \hat{g}_k\|}{\|X_k B_k X_k\|}, \Delta_k \right\}. \tag{3.19}
\]
From (3.19) and taking $\tau_1 = \frac{1}{4}$ we have that (3.15) holds. □

**Assumption 2.** $X_k B_k X_k$ and $X \nabla^2 f(x) X$ are bounded, i.e., there exist $b, \hat{b} > 0$ such that $\|X_k B_k X_k\| \leq b$, $\forall k$, and $\|X \nabla^2 f(x) X\| \leq \hat{b}$, $\forall x \in \mathcal{D}(x_0)$, here the scaling matrix $X = \text{diag}\{x^1, \ldots, x^n\}$ and $x^i$ is the $i$th component of $x > 0$. Similar to proofs of Lemma 4.1 and Theorem 4.2 in [13], we can also obtain following main results.

**Lemma 3.4.** Assume that Assumptions 1–2 hold. If there exists $\varepsilon > 0$ such that
\[
\|P_k \hat{g}_k\| \geq \varepsilon \tag{3.20}
\]
for all $k$, then there is $\alpha > 0$ such that
\[
\Delta_k \geq \alpha, \quad \forall k. \tag{3.21}
\]

**Theorem 3.5.** Assume that Assumptions 1–2 hold. Let $\{x_k\} \subset \mathbb{R}^n$ be a sequence generated by the algorithm. Then
\[
\liminf_{k \to \infty} \|P_k \hat{g}_k\| = 0. \tag{3.22}
\]

**Theorem 3.6.** Assume that Assumptions 1–2 hold. Let $\{x_k\} \subset \mathbb{R}^n$ be a sequence generated by the proposed algorithm and take $B_k = \nabla^2 f(x_k)$, and $-d_k^T (B_k + \frac{1}{2} \mu_k X_k^{-2}) d_k \geq 0$, where $\mu_k$ is the Lagrange multiplier given in (3.1) for the trust region constraint, then $X_* \nabla^2 f(x_*) X_*$ is positive semidefinite on $\mathcal{N}(A)$, where $x_*$ is a limit point of $\{x_k\}$.

**Proof.** By (3.1), we have that
\[
g_k^T d_k = -d_k^T (B_k + \mu_k X_k^{-2}) d_k - \lambda_k^T A d_k \\
= -d_k^T (B_k + \frac{1}{2} \mu_k X_k^{-2}) d_k - \frac{1}{2} \mu_k d_k^T X_k^{-2} d_k \\
\]
\[-\frac{1}{2} \mu_k^T \left( d_k^T X_k^{-2} d_k \right) \]
\[-\frac{1}{2} \mu_k \Delta_k^2. \tag{3.23} \]

First we consider the case when \( \liminf_{k \to \infty} \mu_k = 0 \) on \( \mathcal{N}(A) \). Let \( \mu_{\min}(X_k B_k X_k) \) denote the minimum eigenvalue of \( X_k B_k X_k \) on \( \mathcal{N}(A) \). By Lemma 3.1, \( B_k + \mu_k X_k^{-2} \) is positive semidefinite in \( \mathcal{N}(A) \), is equivalent to \( X_k B_k X_k + \mu_k I \) is positive semidefinite on \( \mathcal{N}(A) \). So, \( \mu_k \geq \max \{-\mu_{\min}(X_k B_k X_k), 0\} \) on \( \mathcal{N}(A) \). It is clear that when \( \liminf_{k \to \infty} \mu_k = 0 \) on \( \mathcal{N}(A) \) there must exist a limit point \( x^* \) at which \( X^* \nabla^2 f(x^*) X^* \) is positive semidefinite on \( \mathcal{N}(A) \).

Now we prove by contradiction that \( \liminf_{k \to \infty} \mu_k = 0 \). Assume that \( \mu_k \geq 2\epsilon > 0 \) for all \( k \) sufficiently large. First we show that \( \{\mu_k\} \) converges to zero. According to the acceptance rule in step 4, we have that, from (3.23)

\[ f(x_{l(k)}) - f(x_k + \alpha_k d_k) \geq -\alpha_k \beta g_k^T d_k \geq -\frac{1}{2} \alpha_k \beta \mu_k \Delta_k^2 \geq -\alpha_k \beta \epsilon \Delta_k^2. \tag{3.24} \]

Similar to the proof of Theorem in [4], we have that the sequence \( \{f(x_{l(k)})\} \) is nonincreasing for all \( k \), and hence \( \{f(x_{l(k)})\} \) is convergent.

(3.24) means that

\[ f(x_{l(k)}) \leq f(x_{l(k)-1}) - \alpha_{l(k)-1} \beta \epsilon \Delta_k^2 \tag{3.25} \]

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

\[ \lim_{k \to \infty} \alpha_{l(k)-1} \Delta_{l(k)-1}^2 = 0 \tag{3.26} \]

which implies that either

\[ \lim_{k \to \infty} \alpha_{l(k)-1} = 0 \tag{3.27} \]

or

\[ \liminf_{k \to \infty} \Delta_{l(k)-1} = 0. \tag{3.28} \]

If (3.28) holds, by the updating formula of \( \Delta_k \), for all \( j, \gamma_1'^j \Delta_k \leq \Delta_{k+j} \leq \gamma_2'^j \Delta_k \) so that \( \gamma_{M+1}' \Delta_{l(k)-1} \leq \Delta_k \leq \gamma_2^{M+1}' \Delta_k \). It means that

\[ \liminf_{k \to \infty} \Delta_k = 0. \tag{3.29} \]

If (3.27) holds, similar to the proof of Theorem in [6], we can also obtain that

\[ \lim_{k \to \infty} f(x_{l(k)}) = \lim_{k \to \infty} f(x_k). \tag{3.30} \]

(3.24) and (3.30) mean that if (3.28) is not true, then

\[ \lim_{k \to \infty} \alpha_k = 0. \tag{3.31} \]

The acceptance rule (2.1) means that, for large enough \( k \),

\[ f \left( x_k + \frac{\alpha_k}{\omega} d_k \right) - f(x_k) > \beta \frac{\alpha_k}{\omega} g_k^T d_k. \]
Since
\[
f\left(x_k + \frac{\alpha_k}{\omega} d_k\right) - f(x_k) = \frac{\alpha_k}{\omega} g_k^T d_k + o\left(\frac{\alpha_k}{\omega} \|d_k\|\right)
\]
we have that
\[
(1 - \beta) \frac{\alpha_k}{\omega} g_k^T d_k + o\left(\frac{\alpha_k}{\omega} \|d_k\|\right) \geq 0.
\]
(3.32)

Dividing (3.32) by \(\alpha_k/\omega\|d_k\|\) and noting that \(1 - \beta > 0\), we have that from (3.23) and \(\mu_k \geq \epsilon\)
\[
0 \leq \lim_{k \to \infty} \frac{g_k^T d_k}{\|d_k\|} \leq - \lim_{k \to \infty} \epsilon \frac{A_k^2}{\|d_k\|} \leq - \epsilon \lim_{k \to \infty} A_k \leq 0
\]
(3.33)
(3.29) and (3.33) imply also that
\[
\lim_{k \to \infty} A_k = 0.
\]
(3.34)

Since \(f(x)\) is twice continuously differentiable we have that
\[
f(x_k + d_k) = f_k + g_k^T d_k + \frac{1}{2} d_k^T \nabla^2 f(x_k) d_k + o(\|d_k\|^2)
\]
\[
\leq f(x_{l(k)}) + \beta g_k^T d_k + \left(\frac{1}{2} - \beta\right)g_k^T d_k + \frac{1}{2}(g_k^T d_k + d_k^T \nabla^2 f(x_k) d_k) + o(\|d_k\|^2).
\]
From (3.1), we can obtain that from \(B_k = \nabla^2 f_k\),
\[
g_k^T d_k + d^T \nabla^2 f_k d_k = -\mu_k d_k^T X_k^{-2} d_k + \hat{\lambda}_k Ad_k = -\mu_k d_k^T X_k^{-2} d_k \leq -\epsilon A_k^2.
\]
Therefore, by (3.23), we have that for large enough \(k\),
\[
f(x_k + d_k) \leq f(x_{l(k)}) + \beta_k g_k^T d_k,
\]
which means that the step size \(\alpha_k = 1\), i.e., \(h_k = d_k\) for large enough \(k\). If \(x_k\) is close enough to \(x^*\) and \(A_k\) is close to 0, we have that
\[
|f(x_k + d_k) - f_k - \psi_k(d_k)| = o(\|d_k\|^2).
\]
(3.35)

Again by (3.1), we have that from \(B_k + \mu_k X_k^{-2}\) being positive semidefinite in \(N(A)\)
\[
\operatorname{Pred}_k(d_k) = -(g_k^T d_k + \frac{1}{2} d_k^T B_k d_k)
\]
\[
= d_k^T(B_k + \mu_k X_k^{-2}) d_k - \frac{1}{2} d_k^T B_k d_k
\]
\[
= \frac{1}{2} d_k^T(B_k + \mu_k X_k^{-2}) d_k + \frac{\mu_k}{2} d_k^T X_k^{-2} d_k
\]
\[
\geq \frac{\mu_k}{2} A_k^2
\]
\[
\geq \epsilon A_k^2.
\]
(3.36)
By (3.35)–(3.36), we have that set
\[ Y_{S UBk} = A_{red}(d_k) = \text{Pred}_k(d_k) \]
with
\[ A_{red}(d_k) = f_k - f(x + d_k) \]

\[ |\rho_k - 1| = \left| \frac{|A_{red}(d_k) - \text{Pred}_k(d_k)|}{\text{Pred}_k(d_k)} \right| = o(\|d_k\|^2) \]
\[ \epsilon A_k \to 0 \] as \( k \to 0 \).

We conclude that the entire sequence \( \{\rho_k\} \) converges to unity. \( \hat{\rho}_k \to 1 \) implies that is not decreased for sufficiently large \( k \) and hence bounded away from zero. Thus, \( \{\Delta_k\} \) cannot converge to zero, contradicting (3.34). \( \square \)

4. Local convergence

Theorem 3.5 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 the local convergent rate, but it requires more assumptions.

We denote the set of active constraints by
\[ I(x) = \{i \mid x_i = 0, \ i = 1, \ldots, n\}. \]  
(4.1)

To any \( I \subset \{1, \ldots, n\} \) we associate the optimization subproblem

\[ (P)_I \quad \min f(x); \ s.t. \ Ax = b, \ x_I = 0. \]  
(4.2)

Assumption 3. For all \( I \subset \{1, \ldots, n\} \), the first-order optimality system associated to \( (P)_I \) has no nonisolated solutions.

Assumption 4. The constraints of (1.1) are qualified in the sense that \( (A^T \lambda)_i = 0, \ \forall i \not\in I(\bar{x}) \) implies that \( \lambda = 0 \).

Assuming that \( (\bar{x}, \bar{v}) \) is associated with a unique pair \( \bar{x} \) which satisfies Assumption 3. Define the set of strictly active constraints as
\[ J(\bar{x}) = \{i \mid \bar{v}_i > 0, \ i = 1, \ldots, n\} \]  
(4.3)

and the extended critical cone as
\[ T \overset{\text{def}}{=} \{d \in \mathbb{R}^n \mid Ad = 0, \ d_i = 0, \ i \in J(\bar{x})\}. \]  
(4.4)

Assumption 5. The solution \( x_* \) of problem (1.1) satisfies the strong second-order condition, that is, there exists \( \alpha > 0 \) such that

\[ p^T \nabla^2 f(x_k) p \geq \alpha \|p\|^2, \ \forall p \in T. \]  
(4.5)

This is a sufficient condition for the strong regularity (see [2] or [12]).

Given \( d \in \mathcal{N}(A) \), the null space of \( A \), we define \( d^T, d^N \) as the orthogonal projection of \( d \) onto \( T \) and \( N \), where \( N \) is the orthogonal complement of \( T \) in \( \mathcal{N}(A) \), i.e.,
\[ N = \{z \in \mathcal{N}(A) \mid z^T d = 0, \ \forall d \in T\}, \]
which means that \( d = d^T + d^N \) and \( \|d\|^2 = \|d^T\|^2 + \|d^N\|^2 \).
The following lemma is actually Lemma 3.5 in [2].

**Lemma 4.1.** Assume that Assumptions 3–5 hold. Given $\kappa > 0$, if $x_k$ is sufficiently close to $x_*$ then
\[
 d^T (B_k + \mu_k X_k^{-2}) d_k \geq \frac{\alpha}{2} \|d_k\|^2 + \kappa \|d_k^N\|^2,
\]
where $\alpha$ given in (1.1) and multiplier $\mu_k$ given in (3.1).

Similar to the proof of the above lemma, we can also obtain that given $\kappa/2 \geq \kappa_1 > 0$,
\[
 d^T (B_k + \mu_k X_k^{-2}) d_k \geq \frac{\alpha}{2} \|d_k\|^2 + \kappa_1 \|d_k^N\|^2.
\]

**Theorem 4.2.** Assume that Assumptions 3–5 hold. Let $\{x_k\}$ be a sequence generated by the algorithm. Then $d_k \to 0$. Furthermore, if $x_k$ is close enough to $x_*$, and $x_*$ is a strict local minimum of problem (1.1), then $x_k \to x_*$. 

**Proof.** By (3.1) and (4.7), we get that
\[
 g_k^T d_k = -d_k^T (B_k + \mu_k X_k^{-2}) d_k - \lambda_k^T A d_k
 = -d_k^T (B_k + \mu_k X_k^{-2}) d_k
 \leq -\frac{\alpha}{2} \|d_k\|^2 - \kappa_1 \|d_k^N\|^2.
\]
According to the acceptance rule in step 4, we have
\[
 f(x_{(l(k))}) - f(x_k + \alpha_k d_k) \geq -\alpha_k \beta g_k^T d_k \geq \frac{\alpha}{2} \beta \alpha_k \|d_k\|^2.
\]

Similar to the proof of Theorem in [6], we have that the sequence $\{f(x_{(l(k))})\}$ is nonincreasing for all $k$, and therefore $\{f(x_{(l(k))})\}$ is convergent.

(4.8) and (4.9) mean that
\[
 f(x_{(l(k))}) \leq f(x_{(l(k)-1)}) - \alpha_{l(k)-1} \beta \|d_{l(k)-1}\|^2.
\]
That $\{f(x_{(l(k))})\}$ is convergent means
\[
 \lim_{k \to \infty} \alpha_{l(k)-1} \|d_{l(k)-1}\|^2 = 0.
\]

Similar to the proof of Theorem in [6], we can also obtain that
\[
 \lim_{k \to \infty} f(x_{(l(k))}) = \lim_{k \to \infty} f(x_k).
\]
(4.9) and (4.12) imply that
\[
 \lim_{k \to \infty} \alpha_k \|d_k\|^2 = 0.
\]
Assume that there exists a subsequence $k \subseteq \{k\}$ such that
\[
 \lim_{k \to \infty, k \in \mathcal{K}} \|d_k\| > 0.
\]
Then, Assumption (4.14) means
\[ \lim_{k \to \infty, k \in \mathcal{K}} x_k = 0. \] (4.15)

Similar to prove (3.32), the acceptance rule (2.1) means that for large enough \( k \),
\[ (1 - \beta) \frac{z_k}{\omega} g_k^T d_k + o\left(\frac{z_k}{\omega} \|d_k\|\right) \geq 0. \] (4.16)

Dividing (4.16) by \( z_k/\omega \|d_k\| \) and noting that \( 1 - \beta > 0 \) and \( g_k^T d_k \leq 0 \), we obtain
\[ 0 \leq \lim_{k \to \infty} \frac{g_k^T d_k}{\|d_k\|} = \lim_{k \to \infty} - \frac{\omega}{2} \|d_k\|^2 = \lim_{k \to \infty} - \|d_k\| \leq 0. \] (4.17)

From (4.17), we have that
\[ \lim_{k \to \infty, k \in \mathcal{K}} \|d_k\| = 0, \]
which contradicts (4.14). Therefore, we have that
\[ \lim_{k \to \infty} \|d_k\| = 0. \] (4.18)

Assume that there exists a limit point \( x_* \) which is a local minimum of \( f \), let \( \{x_k\}_{\mathcal{K}} \) be a subsequence of \( \{x_k\} \) converging to \( x_* \). As \( k \geq l(k) \geq k - M \), for any \( k \),
\[ x_k = x_{l(k) + M + 1} - \alpha_k d_k - \cdots - \alpha_{l(k) + M + 1} - 1 d_{l(k) + M + 1} - 1, \]
there exists a point \( x_{l(k)} \) such that from (4.18)
\[ \lim_{k \to \infty} \|x_{l(k)} - x_k\| = 0, \] (4.19)
so that we can obtain
\[ \lim_{k \in \mathcal{K}, k \to \infty} \|x_{l(k)} - x_*\| \leq \lim_{k \in \mathcal{K}, k \to \infty} \|x_* - x_k\| + \lim_{k \in \mathcal{K}, k \to \infty} \|x_{l(k)} - x_k\| = 0. \] (4.20)

This means that also the subsequence \( \{x_{l(k)}\}_{\mathcal{K}} \) converges to \( x_* \).

As Assumption 3 necessarily holds in a neighborhood of \( x_* \), then \( x_* \) is the only limit point \( \{x_k\} \) in some neighborhood \( \mathcal{N}(x_*; \delta) \) of \( x_* \), where \( \delta > 0 \) is any constant.

On the other hand, we know that the sequence \( \{f(x_{l(k)})\} \) is nonincreasing for all large \( k \), that is,
\[ f(x_{l(k)}) \geq f(x_{l(k+1)}). \] Define
\[ \tilde{f} \overset{\text{def}}{=} \inf \{ f(x); x \in \mathcal{N}(x_*; \delta) \setminus \mathcal{N}(x_*; \delta/2M) \}. \]

Because \( x_* \) is a strict local minimum of problem (1.1), we may assume \( \tilde{f} > f_* \). Now, assuming that \( f(x_k) \leq f(x_{l(k)}) \leq \tilde{f} \) and \( x_{l(k)} \in \mathcal{N}(x_*; \delta/2M) \), it follows that \( f(x_{l(k+1)}) \leq \tilde{f} \) and \( x_{l(k+1)} \in \mathcal{N}(x_*; \delta) \); using the definition of \( \tilde{f} \), we find that \( x_{l(k+1)} \in \mathcal{N}(x_*; \delta/2M) \), and hence \( x_{l(k+2)}, \ldots, x_{l(k+M+1)} \in \mathcal{N}(x_*; \delta/2M) \) again. This implies that sequence \( \{x_{l(k)}\} \) remains in \( \mathcal{N}(x_*; \delta/2M) \). By
\[ \|x_{k+1} - x_*\| \leq \|x_* - x_{l(k+1)}\| + \|x_{l(k+1)} - x_{k+1}\| \leq \delta/2 \]
we get that \( x_k \to x_* \) which means that the conclusion of the theorem is true. \( \square \)
Theorem 4.3. Assume that Assumptions 3–5 hold. Let \( \{x_k\} \) be a sequence generated by the algorithm. Then

\[
\lim_{k \to \infty} \|P_k \hat{g}_k\| = 0. 
\] (4.21)

Proof. Assume that there are an \( \varepsilon_1 \in (0, 1) \) and a subsequence \( \{P_{m_i} \hat{g}_{m_i}\} \) of \( \{P_k \hat{g}_k\} \) such that for all \( m_i, i = 1, 2, \ldots \)

\[
\|P_{m_i} \hat{g}_{m_i}\| \geq \varepsilon_1. 
\] (4.22)

Theorem 3.5 guarantees the existence of another subsequence \( \{P_{l_i} \hat{g}_{l_i}\} \) such that

\[
\|P_k \hat{g}_k\| \geq \varepsilon_2 \quad \text{for } m_i \leq k < l_i 
\] (4.23)

and

\[
\|P_{l_i} \hat{g}_{l_i}\| \leq \varepsilon_2 
\] (4.24)

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

From Theorem 4.2, we know that

\[
\lim_{k \to \infty} \|d_k\| = 0. 
\] (4.25)

Because \( f(x) \) is twice continuously differentiable, we have that, from above,

\[
f(x_k + d_k) = f(x_k) + g_k^T d_k + \frac{1}{2} d_k^T \nabla^2 f(x_k) d_k + o(\|d_k\|^2) 
\leq f(x_{l(k)}) + \beta g_k^T d_k + \left( \frac{1}{2} - \beta \right) g_k^T d_k + \frac{1}{2} [g_k^T d_k + d_k^T B_k d_k] 
+ \frac{1}{2} d_k^T (\nabla^2 f(x_k) - B_k) d_k + o(\|d_k\|^2). 
\] (4.26)

From (3.1), we can obtain

\[
g_k^T d_k + d_k^T B_k d_k = -\mu_k d_k^T X_k^{-2} d_k - \lambda_k^T A d_k = -\mu_k d_k^T X_k^{-2} d_k \leq 0. 
\]

Similar to the proof of Theorem 3.1 in [2], we can obtain that

\[
g_k^T d_k \leq -\frac{1}{2(2 - \beta)} (d_k^T \nabla^2 f(x_k) d_k - \frac{\mu_k}{2} A_k^2, \quad (4.27)
\]

\[
\frac{1}{2} d_k^T (B_k - \nabla^2 f(x_k)) d_k \geq \frac{\beta - 1}{2(2 - \beta)} (d_k^T \nabla^2 f(x_k) d_k - \zeta \|d_k\|^2 + \frac{\varepsilon_0}{4} \|d_k\|^2, 
\] (4.28)

where \( \zeta \equiv \sup_k \{ \|\nabla^2 f(x_k) - B_k\| + \|\nabla^2 f(x_k) - B_k\|^2 / \varepsilon \} \) and \( \varepsilon_0 \equiv \sup_k \{ d_k^T (\nabla^2 f(x_k) - \nabla^2 f(x_k + \varepsilon d_k)) d_k / \|d_k\|^2 \}, \tau \in (0, 1) \). Since \(- (1/2 - \beta)/(2 - \beta) - (\beta - 1)/(2 - \beta) = \beta/2(2 - \beta) > 0 \). Hence,

\[
\left( \frac{1}{2} - \beta \right) g_k^T d_k + \frac{1}{2} d_k^T (\nabla^2 f(x_k) - B_k) d_k + o(\|d_k\|^2) < 0. 
\]
So, we have that for large enough $i$ and $m_i \leq k < l_i$,

$$f(x_k + d_k) \leq f(x_{i(k)}) + \beta g_k^T d_k,$$

(4.29)

which means that the step size $x_k = 1$, i.e., $h_k = d_k$ for large enough $i$ and $m_i \leq k < l_i$.

If $x_k$ is close enough to $x^*$, $d_k$ is then close to 0, and hence $B_k$ is close to $\nabla^2 f(x_k)$. We deduce that

$$\left| f(x_k + d_k) - f(x_k) - \psi_k(d_k) \right|
= |g_k^T d_k + \frac{1}{2} d_k^T \nabla^2 f(x_k) d_k + o(\|d_k\|^2) - (g_k^T d_k + \frac{1}{2} d_k^T B_k d_k)|
= o(\|d_k\|^2).$$

(4.30)

From (2.4) and (4.6), for large enough $i$, $m_i \leq k < l_i$,

$$\text{Pred}(d_k) = d_k^T \left( \frac{1}{2} B_k + \mu_k X_k^{-2} \right) d_k + \lambda_k^T A d_k
\geq \frac{\eta}{2} \|d_k\|^2 + \kappa \|d_k^N\|^2,$$

(4.31)

where $\kappa$ given in (4.6). As $d_k = h_k$, for large $i$, $m_i \leq k < l_i$, we obtain that

$$\hat{\rho}_k \geq \rho_k = \frac{f_k - f(x_k + h_k)}{\text{Pred}(h_k)}
= 1 + \frac{f_k - f(x_k + d_k) + \psi_k(d_k)}{\text{Pred}(h_k)}
\geq 1 - \frac{o(\|d_k\|^2)}{\frac{\eta}{2} \|d_k\|^2 + \kappa \|d_k^N\|^2}
\geq \eta_2.$$  

(4.32)

This means that for large $i$, $m_i \leq k < l_i$,

$$f_k - f(x_k + h_k) \geq \eta_2 \text{Pred}(h_k) \geq \eta_2 \frac{\eta}{2} \|d_k\|^2.$$

Therefore, we can deduce that, for large $i$,

$$\|x_{m_i} - x_{l_i}\|^2 \leq \sum_{k=m_i}^{l_i-1} \|x_{k+1} - x_k\|^2
= \sum_{k=m_i}^{l_i-1} \|h_k\|^2$$
\[ \frac{1}{\eta_2} \sum_{k=m_i}^{l_i-1} [f(x_k) - f(x_k + h_k)] \leq \frac{1}{\eta_2} (f_{m_i} - f_{l_i}). \]  

(4.33) and (4.12) mean that for large \( i \), we have
\[ \|x_{m_i} - x_{l_i}\| \leq \frac{\varepsilon_1}{L(L + 1)}, \]
where \( L \) is the Lipschitz constant of \( g(x) \) in \( \mathcal{L}(x_0) \) and \( L' \) is \( \|X\| \) and \( g(x) \) bounded in \( \mathcal{L}(x_0) \), that is, \( \|g(x)\| \leq L' \) and \( \|X\| \leq L' \). We then use the triangle inequality to show
\[ \|P_{m_i}\hat{g}_{m_i}\| \leq \|P_{m_i}\hat{g}_{m_i} - P_{m_i}\hat{g}_{l_i}\| + \|P_{m_i}\hat{g}_{l_i} - P_{l_i}\hat{g}_{l_i}\| + \|P_{l_i}\hat{g}_{l_i}\| \]
\[ \leq LL'\|x_{m_i} - x_{l_i}\| + L'\|P_{m_i} - P_{l_i}\| + \varepsilon_2 \]
\[ \leq L'(L + 1)\|x_{m_i} - x_{l_i}\| + \varepsilon_2. \]  

(4.34)

We choose \( \varepsilon_2 = \varepsilon_1/4 \), then (4.34) contradicts (4.22). This implies that (4.22) is not true, and hence the conclusion of the theorem holds. \( \square \)

**Theorem 4.4.** Let \( \{x_k\} \) be propose by the algorithm. Assume that \( \{B_k\} \) is bounded, then (a) any limit point \( x_* \) of \( \{x_k\} \) is a solution of the first-order optimality system associated to problem (P)$_{I(x_*)}$, that is,
\[ \nabla f(x_*) = A^T\lambda_* - v_* = 0, \]
\[ Ax_* = b, \]
\[ (x_*)_{I(x_*)} = 0; \quad v_*^i = 0, \quad i \not\in I(x_*), \]
where \( v_*^i \) is the \( i \)th component of \( v_* \).

(b) If Assumptions 3–4 and (4.7) hold, then \( x_* \) satisfies the first-order optimality system of (1.1); i.e., there exist \( \lambda_* \in \mathbb{R}^m, v_* \in \mathbb{R}^n \) such that
\[ \nabla f(x_*) - A^T\lambda_* - v_* = 0, \]
\[ Ax_* = b, \]
\[ x_* \geq 0, \quad v_* \geq 0, \quad x_*^T v_* = 0. \]

**Proof.** Following Lemma 2.3 in [1], we only prove that the sequence \( \{x_k\} \) satisfies the following conditions:

(i) \( \mu_k \to 0, \)
(ii) \( d_k \to 0, \)
(iii) \( X_k[\nabla f_k - A^T\lambda_k] \to 0. \)
In fact, we have proved that (a) holds, similar to prove Theorem 3.6. From Theorem 4.2, we obtain that (b) holds. Theorem 4.3 means
\[ P_k \hat{g}_k = [I - \hat{A}_k^T (\hat{A}_k \hat{A}_k^T)^{-1} \hat{A}_k] \hat{g}_k = X_k [\nabla f_k - \hat{A}_k^T \lambda_k] \]
because of \( \hat{A}_k = AX_k, \hat{g}_k = X_k g_k \) and \( \lambda_k \overset{\text{def}}{=} (\hat{A}_k^T \hat{A}_k)^{-1} \hat{A}_k^T \hat{g}_k \). By conditions (i), (ii), (iii), similar to the proofs of Theorem 2.2 in [2], we can also obtain that the conclusions of the theorem are true. \( \square \)

The conclusions of Theorem 4.4 are the same results as Theorem 2.2 in [2] under the same as assumptions in [2].

We now discuss the convergence rate for the proposed algorithm. For this purpose, it is show that for large enough \( k \), the step size \( \alpha_k \equiv 1 \), and there exists \( \hat{\lambda} > 0 \) such that
\[ \Delta_k \geq \hat{\lambda}. \]

**Theorem 4.5.** Assume that Assumptions 3–5 hold. For sufficiently large \( k \), then the step \( \alpha_k \equiv 1 \) and the trust region constraint is inactive, that is, there exists \( \hat{\lambda} > 0 \) such that
\[ \Delta_k \geq \Delta_k', \quad \forall k \geq K', \]
where \( K' \) is a large enough index.

**Proof.** Similar to prove (4.29), we can also obtain that at the sufficiently large \( k \)th iteration,
\[ f(x_k + d_k) \leq f(x_{(k)}) + \beta \hat{g}_k^T d_k, \quad (4.35) \]
which means that the step size \( \alpha_k \equiv 1 \), i.e., \( h_k = d_k \) for large enough \( k \).

By the above inequality, we know that
\[ x_{k+1} = x_k + d_k. \]

By Assumptions 3–5, we can obtain that from (4.30)
\[
\rho_k - 1 = \frac{\text{Ared}(h_k) - \text{Pred}(h_k)}{\text{Pred}(h_k)} \\
= \frac{(g_k^T h_k + \frac{1}{2} h_k^T B_k h_k) - (g_k^T h_k + \frac{1}{2} h_k^T \nabla^2 f(x_k) h_k + o(\|h_k\|^2))}{|\text{Pred}(h_k)|} \\
= \frac{o(\|h_k\|^2)}{|\text{Pred}(h_k)|}. \quad (4.36)
\]

Similar to prove (4.31), for large enough \( k \),
\[ \text{Pred}(d_k) \geq \frac{\kappa}{2} \|d_k\|^2 + \kappa \|d_k^N\|^2, \quad (4.37) \]
where \( \kappa \) given in (4.6).
Similar to the proof of Theorem 4.2, we can also obtain that \( d_k \to 0 \). Hence, (4.36) and (4.37) mean that \( d_k \to 0 \). Hence, (4.36) and (4.37) mean that \( Y_{SUB_k} \to 1 \). Hence there exists \( \hat{Y}_{SOH} > 0 \) such that when \( \|d_k\| \leq \hat{Y}_{SOH} \), \( \hat{Y}_{SUB_k} \leq Y_{DCz} \). And therefore, \( Y_{SOH_k + 1} \leq Y_{SOH_k} \). As \( h_k \to 0 \), there exists an index \( K' \) such that \( \|d_k\| \leq \hat{Y} \) whenever \( k \geq K' \). Thus

\[
\Delta_k \geq \Delta_{K'}, \quad \forall k \geq K'.
\]

The conclusion of the theorem holds. \( \square \)

Theorem 4.5 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 \( d_k \). Since the step size \( \alpha_k \equiv 1 \), and the trust region constraint is inactive, for sufficiently large \( k \).

5. Numerical experiments

Numerical experiments on the nonmonotonic back-tracking trust region interior point algorithm given in this paper have been performed on an IBM 586 personal computer. In this section, we present the numerical results. We compare with different nonmonotonic parameters \( M = 0, 4 \) and 8, respectively, for the proposed algorithm. A monotonic algorithm is realized by taking \( M = 0 \). In order to check effectiveness of the back-tracking technique, we select the same parameters as used in \([4]\). The selected parameter values are: \( \hat{\eta} = 0.01, \ \eta_1 = 0.001, \ \eta_2 = 0.75, \ \gamma_1 = 0.2, \ \gamma_2 = 0.5, \ \gamma_3 = 2, \ \omega = 0.5, \ \Delta_{max} = 10, \ \beta = 0.2, \) and initially \( \Delta_0 = 5 \). The computation terminates when one of the following stopping criterions is satisfied \( \|P_k \hat{g}_k\| \leq 10^{-6} \), or \( f_k - f_{k+1} \leq 10^{-8} \max\{1, |f_k|\} \).

The experiments are carried out on 15 standard test problems which are quoted from \([10,7]\). Besides the recommended starting points in \([10,7]\) (HS: the problems from Hock and Schittkowski [7] and SC: from Schittkowski [10]), denoted by \( x_0^{HS} \), we also test these methods with another set of starting points \( x_0^{SB} \). The computational results for \( B_k = \nabla^2 f(x_k) \), the real Hessian, are presented at the following table, where ALG denote it variation proposed in this paper with non-monotonic decreasing and back-tracking techniques. NF and NG stand for the numbers of function evaluations and gradient evaluations, respectively. NO stands for the number of iterations in which nonmonotonic decreasing situation occurs, that is, the number of times \( f_k - f_{k+1} < 0 \). The number of iterations is not presented in the following table because it always equals NG.

The results under ALG (\( M = 0 \)) represent mixture of trust region and line-search techniques via interior point of feasible set considered in this paper. Our type of approximate trust region method is very easy to resolve the subproblem \( (S_k) \) with a reduced radius via interior point. The back-tracking steps can outperform the traditional method when the trust region subproblem is solved accurately over the whole hyperball. The last three parts of the table, under the headings of \( M = 0, 4 \) and 8 respectively, show that for most test problems the nonmonotonic technique does bring in some noticeable improvement (Table 1).
<table>
<thead>
<tr>
<th>Problem name</th>
<th>Initial point</th>
<th>ALG</th>
<th>$M = 0$</th>
<th>$M = 4$</th>
<th>$M = 8$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>NF</td>
<td>NG</td>
<td>NF</td>
</tr>
<tr>
<td>Rosenbrock</td>
<td>$x_{0a}$</td>
<td>25</td>
<td>21</td>
<td>16</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>$x_{0b}$</td>
<td>14</td>
<td>12</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>(C = 100)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rosenbrock</td>
<td>$x_{0a}$</td>
<td>92</td>
<td>60</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>$x_{0b}$</td>
<td>30</td>
<td>25</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>(C = 10 000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rosenbrock</td>
<td>$x_{0a}$</td>
<td>249</td>
<td>214</td>
<td>26</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>$x_{0b}$</td>
<td>45</td>
<td>35</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>(C = 1 000 000)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freudenstein</td>
<td>$x_{0a}$</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>$x_{0b}$</td>
<td>12</td>
<td>10</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>Huang</td>
<td>$x_{0a}$</td>
<td>14</td>
<td>14</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>HS048</td>
<td>$x_{0b}$</td>
<td>18</td>
<td>18</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Huang</td>
<td>$x_{0a}$</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>HS049</td>
<td>$x_{0b}$</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>Huang</td>
<td>$x_{0a}$</td>
<td>23</td>
<td>17</td>
<td>20</td>
<td>14</td>
</tr>
<tr>
<td>HS050</td>
<td>$x_{0b}$</td>
<td>38</td>
<td>27</td>
<td>31</td>
<td>19</td>
</tr>
<tr>
<td>Hsia</td>
<td>$x_{0a}$</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>HS055</td>
<td>$x_{0b}$</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Betts(HS062)</td>
<td>$x_{0}$</td>
<td>24</td>
<td>16</td>
<td>20</td>
<td>14</td>
</tr>
<tr>
<td>Sheela(HS063)</td>
<td>$x_{0}$</td>
<td>35</td>
<td>35</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>While</td>
<td>$x_{0}$</td>
<td>231</td>
<td>201</td>
<td>209</td>
<td>183</td>
</tr>
<tr>
<td>SC265</td>
<td>$x_{0}$</td>
<td>6</td>
<td>4</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>SC347</td>
<td>$x_{0}$</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Banana</td>
<td>$x_{0a}$</td>
<td>27</td>
<td>20</td>
<td>19</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>$x_{0b}$</td>
<td>32</td>
<td>26</td>
<td>24</td>
<td>23</td>
</tr>
<tr>
<td>(n = 6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Banana</td>
<td>$x_{0a}$</td>
<td>45</td>
<td>35</td>
<td>45</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>$x_{0b}$</td>
<td>33</td>
<td>30</td>
<td>32</td>
<td>30</td>
</tr>
<tr>
<td>(n = 16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Acknowledgements

The author would like to thank an anonymous referee for his helpful comments and suggestions. The author gratefully acknowledges the partial supports of the National Science Foundation Grant (10071050) of China, the Science Foundation Grant 02ZA14070 of Shanghai Technical Sciences Committee and the Science Foundation Grant (02DK06) of Shanghai Education Committee.

References