A class of derivative-free trust-region methods with interior backtracking technique for nonlinear optimization problems subject to linear inequality constraints

Jing Gao and Jian Cao

Abstract

This paper focuses on a class of nonlinear optimization subject to linear inequality constraints with unavailable-derivative objective functions. We propose a derivative-free trust-region methods with interior backtracking technique for this optimization. The proposed algorithm has four properties. Firstly, the derivative-free strategy is applied to reduce the algorithm’s requirement for first- or second-order derivatives information. Secondly, an interior backtracking technique ensures not only to reduce the number of iterations for solving trust-region subproblem but also the global convergence to standard stationary points. Thirdly, the local convergence rate is analyzed under some reasonable assumptions. Finally, numerical experiments demonstrate that the new algorithm is effective.

MSC: 49M37; 65K05; 90C30; 90C51

Keywords: Affine scaling; Trust-region method; Inequality constraints; Derivative-free optimization; Interior backtracking technique

1 Introduction

In this paper, we analyze the solution of following nonlinear optimization problem:

\[
\min f(x)
\]

s.t. \( Ax \geq b, \) (1)

where \( f(x) \) is a nonlinear twice continuously differentiable function, but its first-order or second-order derivatives are not explicitly available, \( A \triangleq [a_1^T, a_2^T, \ldots, a_m^T]^T \in \mathbb{R}^m \times n \) with \( a_i^T \in \mathbb{R}^n \) and \( b \triangleq [b_1, b_2, \ldots, b_m]^T \in \mathbb{R}^m. \) The feasible set, in (1), is denoted by \( \Omega \triangleq \{ x \in \mathbb{R}^n | Ax \geq b \} \) and the strict interior feasible set is \( \text{int}(\Omega) \triangleq \{ x \in \mathbb{R}^n | Ax > b \}. \)
1.1 Affine-scaling matrix for inequality constraints

The KKT system of (1) is

\[
\nabla f(x) - A^T \lambda_f = 0, \\
\text{diag}(Ax - b) \lambda_f = 0, \\
Ax - b \geq 0, \\
\lambda_f \geq 0,
\]

where \( \lambda_f \in \mathbb{R}^m \). A feasibility \( x^* \) is said to be the stationary point for problem (1), if there exists a vector \( 0 \leq \lambda_f^* \in \mathbb{R}^m \) such that the KKT system (2) holds.

To solve this KKT system, some effective affine-scaling algorithms are designed. Reference [1] proposed an affine-scaling trust-region method with interior-point technique for bound-constrained semismooth equations. Reference [2] introduced affine-scaling interior-point Newton methods for bound-constrained nonlinear optimization. In particular, [3] proved the superlinear and quadratic convergence properties of affine-scaling interior-point Newton methods for bound optimization problems without strict complementarity assumption. Different affine-scaling matrix denotes different algorithm. In [4], the Dikin affine scaling was denoted by

\[
D(x) \overset{\text{def}}{=} \text{diag}(Ax - b) \quad \text{and} \quad D_k \overset{\text{def}}{=} D(x_k).
\]

Moreover, diagonal matrix \( C_{k, f} \overset{\text{def}}{=} \text{diag}(|\lambda_{f, k}|) \) was presented in [4]. Then \( \lambda_{f, k} \) could be obtained as a least-squares Lagrangian multiplier approximation computed by

\[
\begin{bmatrix} A^T \\ -D_k^{1/2} \end{bmatrix} \lambda_{f, k}^{LS} \overset{\text{def}}{=} \begin{bmatrix} \nabla f_k \\ 0 \end{bmatrix}.
\]

One efficient affine-scaling interior-point trust-region model is the one which is presented in [5] and [6], written in the form

\[
\begin{align*}
\min \quad & q_{k, f}(p) = \nabla f_k^T p + \frac{1}{2} p^T H_{f_k} p + \frac{1}{2} p^T A^T D_k^{-1} C_{k, f} A p \\
\text{subject to} \quad & \| [p; D_k^{-1/2} A p] \| \leq \Delta_k,
\end{align*}
\]

where \( \nabla f(x_k) \) is the gradient of \( f(x) \) at the current iteration, \( H_{f_k} \) is either \( \nabla^2 f(x_k) \) or its approximation. Furthermore, \( \| \nabla f_k^T H_{f_k} \| \leq \varepsilon \), where

\[
h_{f_k} = - (\nabla f(x_k) - A^T \lambda_{f, k}),
\]

and \( \varepsilon \) is a small enough constant, is usually considered as the termination criterion in this class of algorithms.

Motivation

The above discussions illustrate that the affine-scaling interior-point trust-region method is an effective way to solve the nonlinear optimization problems with inequality constraints. The trust-region frame guarantees the stable numerical performance. However, in Eqs. (4)–(6) the first- and second-order derivatives play important roles during the computational process, which maybe fail to solve the optimization problems like
If both the feasibility and the stability of the algorithm need to be guaranteed, we should consider the derivative-free trust-region methods.

### 1.2 Derivative-free technique for trust-region subproblem

Since the first- or second-order derivatives of objective functions are not explicitly available, the derivative-free optimization algorithms have been favored by researchers for a time. The application forms of the derivative-free theory are devise [7, 8] and widely applied. Reference [9] proposed a derivative-free algorithm for least-squares minimization, and proved the local convergence in [10]. Reference [11] presented a derivative-free approach to constrained multiobjective nonsmooth optimization. Reference [12] presented a higher-order contingent derivative of perturbation maps in multiobjective optimization. In [13], Conn proposed an unconstrained derivative-free trust-region method. They constructed the trust-region subproblem

$$\min_{s \in B(0; \Delta_k)} m_k = m(x_k + s) = m(x_k) + s^T g_k + \frac{1}{2} s^T H_m k s$$

by using a polynomial interpolation technique, where $\nabla m(x_k) = g_k$, and $\nabla^2 m(x_k) = H_m k$. Following this idea, we consider that $Y_k = \{y_k^0, y_k^1, \ldots, y_k^t\}$ is an interpolation sample set around the current iteration point $x_k$, and we construct the trust-region subproblem

$$\min_{q = \{p\}} q_m (p) = g_k^T p + \frac{1}{2} p^T H_m k p + \frac{1}{2} p^T A^T D_k^{-1} C_m k A p$$

subject to

$$\left\| [p; D_k^{-\frac{1}{2}} A p] \right\| \leq \Delta_k.$$  

(7)

$C_m k \overset{\text{def}}{=} \text{diag}(|\lambda_{m_k}|)$ with $\lambda_{m_k}$ obtained from

$$\begin{bmatrix}
A^T \\
-D_k^2
\end{bmatrix}
\lambda_{m_k} \overset{\text{L.S.}}{=} \begin{bmatrix} g_k \\
0
\end{bmatrix},$$

(8)

$h_{m_k} = -(g_k - A^T \lambda_{m_k}).$  

(9)

We should note that the gradient and Hessian in (5) and (7), (4) and (8), (6) and (9) are different. Meanwhile, since the algorithm in this paper adopts both the decrease direction $p$ and the stepsize $\alpha$ to update the iteration point, we give a new definition of the error bounds between the objective function $f(x_k + \alpha p)$ and the approximation function $m(x_k + \alpha p)$ to ensure the global convergence. We shall show the details after assumption (A1).

**Assumption**

(A1) Suppose that a level set $L(x_0)$ and a maximal radius $\Delta_{\text{max}}$ are given. Assume that $f$ is twice continuously differentiable with Lipschitz continuous Hessian in an appropriate open domain containing the $\Delta_{\text{max}}$ neighborhood $\bigcup_{x \in L(x_0)} B(x, \Delta_{\text{max}})$ of the set $L(x_0)$.

**Definition 1** Given a function $f$ satisfies (A1), $\mathcal{M} = \{m : \mathbb{R}^n \rightarrow \mathbb{R}, m \in C^2\}$ is a set of model functions. If there exist positive constants $\kappa_{\text{ef}}, \kappa_{\text{eg}}, \kappa_{\text{dl}},$ and $\kappa_{\text{ldl}}$, such that, for any $x \in L(x_0)$, $\Delta \in [0, \Delta_{\text{max}}]$, and $\alpha \in (0, 1)$, there is a model function $m(x + \alpha p) \in \mathcal{M}$, with Lipschitz continuous Hessian and corresponding Lipschitz constant bounded by $\kappa_{\text{ldl}},$ and such that:
the error between the Hessian of the model \( m(x + \alpha p) \) and the Hessian of the function \( f(x + \alpha p) \) satisfies
\[
\| \nabla^2 f(x + \alpha p) - \nabla^2 m(x + \alpha p) \| \leq \kappa_2 \alpha \Delta, \quad \forall p \in B(0, \Delta); \tag{10}
\]
the error between the gradient of the model \( m(x + \alpha p) \) and the gradient of the function \( f(x + \alpha p) \) satisfies
\[
\| \nabla f(x + \alpha p) - \nabla m(x + \alpha p) \| \leq \kappa_1 \alpha^2 \Delta^2, \quad \forall p \in B(0, \Delta); \tag{11}
\]
the error between the model \( m(x + \alpha p) \) and the function \( f(x + \alpha p) \) satisfies
\[
\| f(x + \alpha p) - m(x + \alpha p) \| \leq \kappa_3 \alpha^3 \Delta^3, \quad \forall p \in B(0, \Delta). \tag{12}
\]

Such a model \( m \) is called fully quadratic on \( B(x, \Delta) \).

In this paper, we aim to present a class of derivative-free trust-region method for nonlinear programming with linear inequality constraints. The main features of this paper are:

- We use the derivatives of approximation function \( m(x_k + \alpha p) \) to replace the derivatives of objective function \( f(x_k + \alpha p) \) to reduce the algorithm’s requirement for gradient and Hessian of the iteration points. We solve an affine-scaling trust-region subproblem to find a feasible search direction in each iteration.
- In the \( k \)th iteration, a feasible search direction \( p \) is obtained from an affine-scaling trust-region subproblem. Meanwhile, interior backtracking skill will be applied both for determining stepsize \( \alpha \) and for guaranteeing the feasibility of iteration point.
- We will show that the iteration points generated by the proposed algorithm could converge to the optimal points of (1).
- Local convergence will be given under some reasonable assumptions.

This paper is organized as follows: we describe a class of derivative-free trust-region method in Sect. 2. The main results including global convergence property and local convergence rate will be discussed in Sect. 3. The numerical results will be illustrated in Sect. 4. Finally, we give some conclusions.

**Notation** In this paper, \( \| \cdot \| \) is the 2-norm for a vector and the induced 2-norm for a matrix. \( B \subset \mathbb{R}^n \) is a closed ball and \( B(x, \Delta) \) is the closed ball centered at \( x \), with radius \( \Delta > 0 \). \( Y \) is a sample set and \( \mathcal{L}(\xi_0) = \{ x \in \mathbb{R}^n | f(x) \leq f(\xi_0), Ax \geq b \} \) is the level set about the objective function \( f \). We use the subscript \( f_k \) and subscript \( m_k \) to distinguish the relevant information between the original function and the approximate function. For example, \( H_{f_k} \) is the Hessian of \( f \) at \( k \)th iteration and \( H_{m_k} \) is the Hessian of \( m_k \) at \( k \)th iteration.

### 2 A derivative-free trust region method with interior backtracking technique

To solve the optimization problem (1) with not all available first- or second-order derivatives, we design a derivative-free trust-region method. An affine-scaling matrix is denoted by (3) for linear inequality constraints. We chose a stepsize \( \alpha_k \) satisfying the following inequalities:

\[
f(x_k + \alpha_k p_k) \leq f(x_k) + \alpha_k \kappa_3 g_k^T p_k, \tag{13a}
\]
with \( x_k + \alpha_k p_k \in \Omega. \tag{13b} \]
Moreover, set
\[
\theta_k = \begin{cases} 
1, & \text{if } x_k + \alpha_k p_k \in \text{int}(\Omega), \\
1 - O(\|p_k\|^2), & \text{otherwise},
\end{cases}
\tag{14}
\]
where $\theta_k \in (\theta_0, 1)$, for some $0 < \theta_0 < 1$. The $\theta_k$ is to ensure the iterative points generated by the algorithm are strictly interior. Combining with (13a), (13b) and (14), this interior backtracking technique is to guarantee the feasibility of the iterative points. The algorithm possesses the trust-region property and the derivative-free technique is reflected in the trust-region subproblem (7) since the gradient $g_k$ and Hessian $H_{mk}$ come from the approximation function, which are different from $\nabla f_k$ and $H_{ik}$ in (5), satisfying the error bounds (11) and (12). We adopt $\|g_k^T h_{mk}\|$ to be a termination criterion. Now we present the derivative-free trust-region method in detail (see Algorithm 1).

**Remark 1** We add a backtracking interior line-search technique in the algorithm. It is helpful to reducing the number of iterations. Equation (13a) is used to guarantee the descent property of $f(x)$ and (13b) ensures the feasibility of $x_k + \alpha_k p_k$.

**Remark 2** The scalar $\alpha_k$, given in step 5, denotes the stepsize along $p_k$ to the boundary (13b) of the linear inequality constraints
\[
\Gamma_k \overset{\text{def}}{=} \min \left\{ \frac{-a_i^T x_k - b_i}{a_i^T p_k} \right\} > 0, i = 1, 2, \ldots, m,
\tag{15}
\]
with $\Gamma_k \overset{\text{def}}{=} +\infty$ if $-(a_i^T x_k - b_i)/(a_i^T p_k) \leq 0$ for all $i = 1, 2, \ldots, m$. A key property of the scalar $\alpha_k$ is that an arbitrary step $\alpha_k p_k$ to the point $x_k + \alpha_k p_k$ does not violate any linear inequality constraints.

**Remark 3** Let
\[
M_{mk} = \begin{bmatrix} H_{mk} & 0 \\ 0 & C_{mk} \end{bmatrix}.
\tag{16}
\]
The first-order necessary conditions of (7) implies that there exists $v_{mk} \geq 0$ such that
\[
(M_{mk} + v_{mk} I) \begin{bmatrix} p_k \\ \hat{p}_k \end{bmatrix} = -g_k + \begin{bmatrix} A^T \\ D_k^T \end{bmatrix} \lambda_{mk+1}, \quad \text{with}
\]
\[
v_{mk} \left( \Delta_k - \begin{bmatrix} p_k \\ \hat{p}_k \end{bmatrix} \right) = 0.
\tag{17}
\]

In order to obtain a suitable approximation function, Algorithm 1 needs to update the objective function of the trust-region subproblem if necessary. The model-improvement algorithm is applied only if $\|g_k^T h_{mk}\| \leq \varepsilon$ and at least one of the following holds: The model $m(x_k + \alpha p)$ is not certifiably fully quadratic on $B(x_k, \Delta_k)$ or $\Delta_k > \varepsilon \|g_k^T h_{mk}\|$. It improves on the current approximate function $m(x_k + \alpha p)$ to meet the requirements of the error bounds so that the model function becomes fully quadratic. We display the model-improvement
Algorithm 1 Derivative-free trust-region method with interior backtracking technique

1. Initialization: Given $x_0$, $\Delta_{\text{max}} > 0$, $\Delta_0 \in (0, \Delta_{\text{max}}]$, $0 \leq \eta_0 \leq \eta_1 < 1$, $(\eta_1 \neq 0)$ and $\theta_0 \in (0, 1)$ are constants. $k := 0$.
2. Construct model function: Let $y_k = x_k$, and obtain an interpolation point set $Y_k = \{y_0^k, \ldots, y_N^k\}$. Construct a quadratic function $m(x_k + \alpha p)$, get the information such as $g_k$, $H_{mk}$. Calculate $D_k$, $h_{mk}$ and $h_{m_k}$ from (3), (8) and (9).
3. Termination criterion: If $\|g_k^T h_{mk}\| > \varepsilon$, then go to step 4. Consider the model $m_k$ on $B(x_k, \Delta_k)$, there are two cases would be happened:
   (a) If $m_k$ is not fully quadratic on $B(x_k, \Delta_k)$ or $\Delta_k > \varepsilon \|g_k^T h_{mk}\|$ holds, construct another model to modified $m_k$, $\Delta_k = \min\{\Delta_k, \beta \|g_k^T h_{mk}\|, \Delta_k\}$ and go to step 4.
   (b) Otherwise, we get the optimal point and stop.
4. Trust-region subproblem: solve trust-region subproblem (7) to find descent direction $p_k$.
5. Step size: choose $\alpha_k = 1, \alpha_k \alpha^2, \ldots$, until the inequalities (13a) and (13b) are satisfied.
6. Set $s_k = \alpha_k \theta_k p_k$ by (14).
7. Calculate $\text{Pred}(s_k) = m(x_k) - m(x_k + s_k)$, $\text{Ared}(s_k) = f(x_k) - f(x_k + s_k)$,
   \[
   \rho_k = \frac{\text{Ared}(s_k)}{\text{Pred}(s_k)}.
   \]
   (a) If $\rho_k \geq \eta_1$ or if both $\rho_k \geq \eta_0$ and $m_k$ is fully quadratic on $B(x_k, \Delta_k)$, then set $y_{k+1} = x_{k+1} = x_k + s_k$, go to step 2.
   (b) Otherwise $x_{k+1} = x_k$.
8. If $\rho_k < \eta_1$, call Algorithm 2 to guarantee $m_k$ is fully quadratic.
   (a) If $m_k$ is not fully quadratic, modify $\Delta_k$.
   (b) Otherwise, set $m_{k+1} = m_k$.
9. Trust-region radius update: set
   \[
   \Delta_{k+1} = \begin{cases} 
   \min\{\zeta \Delta_k, \Delta_{\text{max}}\} & \text{if } \rho_k \geq \eta_1 \text{ and } \Delta_k < \varepsilon \|g_k^T h_{mk}\|, \\
   [\Delta_k, \min\{\zeta \Delta_k, \Delta_{\text{max}}\}] & \text{if } \rho_k \geq \eta_1 \text{ and } \Delta_k \geq \varepsilon \|g_k^T h_{mk}\|, \\
   \zeta \Delta_k & \text{if } \rho_k < \eta_1 \text{ and } m_k \text{ is fully quadratic,} \\
   \Delta_k & \text{if } \rho_k < \eta_1 \text{ and } m_k \text{ is not fully quadratic.}
   \end{cases}
   \]
10. Set $k := k + 1$ and go to step 2.

Algorithm 2 Model-improvement mechanism

1. Initialization: set $i = 0$, $m_k^{(0)} = m_k$.
2. Repeat: $i = i + 1$; $\tilde{\Delta}_k = \omega^{i-1} \Delta_k$; Ensure $m_k^{(i-1)}$ satisfies the error bounds (10)–(12) in Definition 1 on $B(x_k, \omega^{i-1} \Delta_k)$.
   Until $\tilde{\Delta}_k \leq \varepsilon \|g_k^T h_{mk}\|^{(i)}$.

mechanism in Algorithm 2 which has the same principle as Algorithm 2 proposed in [14], with a constant $\omega \in (0, 1)$.
3 Main results and discussion

In this section, we mainly discuss some properties about the proposed algorithm, including the discussion of the error bounds, the sufficiently descent property, the global and local convergence properties. First of all, we make some necessary assumptions as follows.

Assumptions

(A2) The level set $L(x_0)$ is bounded.

(A3) There exist positive constants $\kappa_{g}$ and $\kappa_{gm}$ such that $\|\nabla f_{k}\| \leq \kappa_{g}$ and $\|g_k\| \leq \kappa_{gm}$, respectively, for all $x_k \in L(x_0)$.

(A4) There exist positive constants $\kappa_{Hf}$ and $\kappa_{Hm}$ such that $\|Hf_{k}\| \leq \kappa_{Hf}$ and $\|Hm_{k}\| \leq \kappa_{Hm}$, respectively, for all $x_k \in L(x_0)$.

(A5) $[A - D_k^T]$ is full row rank for all $x_k \in L(x_0)$.

3.1 Error bounds

Observe first that some error bounds hold immediately.

Lemma 1 Suppose that (A1)–(A5), the error bounds (10)–(12) and the fact that $\Delta_k \leq \Delta_{\max}$ hold. If $m_k$ is a fully quadratic model on $B(x_k, \Delta_k)$, then the following bound is true:

$$\|h_{f_k} - h_{m_k}\| \leq \kappa_h \alpha_k \Delta_k.$$  \hspace{1cm} (18)

Proof Using the theory of matrix perturbation analysis, Eqs. (4) and (8), we obtain

$$\|\lambda_{f_k} - \lambda_{m_k}\| \leq \left\| (AA^T + D_k)^{-1} \right\| \| A \| \| \nabla f(x_k) - g_k \|, \hspace{1cm} (19)$$

where $AA^T$ is a positive definition matrix and $D_k$ is a diagonal matrix related with $x_k \in L(x_0)$. By (A2), there exists a constant $\kappa_2 > 0$ such that $\|(AA^T + D_k)^{-1}\| \| A \| \leq \kappa_2$. Thus, from (6), (9) and the error bound (11), one has the fact that

$$\|h_{f_k} - h_{m_k}\| = \|\nabla f(x_k) - g_k - A^T(\lambda_{f_k} - \lambda_{m_k})\|$$

$$\leq \|\nabla f(x_k) - g_k\| + \kappa_2 \| A \| \| \nabla f(x_k) - g_k\|$$

$$\leq (1 + \kappa_2 \| A \|) \| \nabla f(x_k) - g_k\|$$

$$\leq (1 + \kappa_2 \| A \|) \kappa_{eg} \alpha_k^2 \Delta_k^2$$

$$\leq (1 + \kappa_2 \| A \|) \kappa_{eg} \Delta_{\max} \alpha_k \Delta_k.$$  

Clearly, the conclusion holds with $\kappa_h = (1 + \kappa_2 \| A \|) \kappa_{eg} \Delta_{\max}$. \hfill \Box

Lemma 2 Suppose that (A1)–(A5), the error bounds (10)–(12) and the fact that $\Delta_k \leq \Delta_{\max}$ hold. If $m_k$ is a fully quadratic model on $B(x_k, \Delta_k)$, for some constant $\kappa_2$, one has

$$\|\nabla f(x_k)^T h_{f_k} \| - \|g_k^T h_{m_k}\| \leq \kappa_2 \alpha_k \Delta_k.$$  \hspace{1cm} (20)

Proof Using the triangle inequality, Cauchy–Schwarz inequality, (18), (A3), the error bounds (10)–(12) and the fact that $\alpha_k \in (0, 1]$ and $\Delta_k \leq \Delta_{\max}$ successively, we obtain

$$\|\nabla f(x_k)^T h_{f_k} \| - \|g_k^T h_{m_k}\|$$
\[
\| \nabla f(x_k) - g_k \| h_j_k - h_j_k \| \leq \| \nabla f(x_k) - g_k \| h_{m_k} \| + \kappa g_k \| h_{m_k} \| \| h_k \| + \kappa g_k \| h_{m_k} \| \Delta_k \| g_k \|
\]

which implies that the inequality (20) holds with \( \kappa_2 = \kappa g_k \Delta_{\text{max}} + \kappa g_k \kappa_k \).

**Lemma 3** Suppose that (A1)–(A5), the error bounds (10)–(12) and the fact that \( \Delta_k \leq \Delta_{\text{max}} \) hold. If \( \| \nabla f^T h_k \| \neq 0 \), then step 3 of Algorithm 1 will stop in a finite number of improvement steps.

**Proof** Now we should prove that \( \| \nabla f^T h_k \| \) must be zero if the loop of Algorithm 2 is infinite.

In fact, there are two cases could cause Algorithm 2 to be implemented. One is that \( m_k \) is not fully quadratic, the other is that the radius \( \Delta_k \leq \iota \| g^T h_{m_k} \| \). Then set \( m_k^{(0)} = m_k \), and improve the model to be fully quadratic on \( B(x_k, \Delta_k) \), which denoted by \( m_k^{(1)} \). If \( (g^T h_{m_k})^{(1)} \) of \( m_k^{(1)} \) satisfies the inequality \( \iota \| (g^T h_{m_k})^{(1)} \| \geq \Delta_k \), Algorithm 2 stops with \( \Delta_k = \Delta_k \leq \iota \| (g^T h_{m_k})^{(1)} \| \).

Otherwise, \( \iota \| (g^T h_{m_k})^{(1)} \| < \Delta_k \) holds. Algorithm 2 will improve the model on \( B(x_k, \omega \Delta_k) \) and the resulting model is denoted by \( m_k^{(2)} \). If \( m_k^{(2)} \) satisfies \( \iota \| (g^T h_{m_k})^{(2)} \| \geq \omega \Delta_k \), the procedure stops. If not, the radius should be multiplied by \( \omega \) and Algorithm 2 will improve the model on \( B(x_k, \omega^2 \Delta_k) \), and go on.

The only case for Algorithm 2 to be infinite is if

\[
\iota \| (g^T h_{m_k})^{(i)} \| < \omega^{i-1} \Delta_k \quad \text{for all } i \geq 1.
\]

It implies

\[
\lim_{i \to +\infty} \| (g^T h_{m_k})^{(i)} \| = 0.
\]

By the bound (20) \( \| \nabla f^T h_k \| - (g^T h_{m_k})^{(i)} \| \leq \kappa_2 \omega^{i-1} \alpha_k \Delta_k \) for all \( i \geq 1 \), we obtain

\[
\| \nabla f^T h_k \| \leq \| \nabla f^T h_k \| - (g^T h_{m_k})^{(i)} \| + \| (g^T h_{m_k})^{(i)} \|
\]

\[
\leq \kappa_2 \omega^{j-1} \alpha_k \Delta_k + \frac{\omega^{j-1}}{l} \Delta_k
\]

\[
\leq \kappa_2 \omega^{j-1} \Delta_k + \frac{\omega^{j-1}}{l} \Delta_k
\]

\[
\leq \left( \kappa_2 + \frac{1}{l} \right) \omega^{j-1} \Delta_k.
\]

By the choice of \( \omega \in (0, 1) \) the above inequality means that \( \| \nabla f^T h_k \| = 0 \). The conclusion shows us step 3 will stop in a finite number of improvements.

**3.2 Sufficiently descent property**

In order to guarantee the global convergence property of the proposed algorithm, it is necessary to show that a sufficiently descent condition is satisfied at the \( k \)th iteration. We
obtained in [6] if step \( p_k \) is the optimal point of the trust-region subproblem (7), there is a constant \( \kappa_3 > 0 \) such that
\[
\frac{g_k^T p_k}{\|g_k^T h_{mk}\|_2} \leq -\kappa_3 \min \left\{ \frac{\|g_k^T h_{mk}\|_2}{\|M_{mk}\|}, \frac{\|\nabla f(x_k)\|_2}{2\alpha_k (\kappa_1 - 1)\|g_k^T h_{mk}\|_2} \right\}.
\] (21)

**Lemma 4** Suppose that (A1)–(A5) and the error bounds (10)–(12) hold, \( p_k \) is the solution of the trust-region subproblem (7). Then there must exist an appropriate \( \alpha_k > 0 \) which satisfied inequalities (13a).

**Proof** We start by considering the maximal step-length along the trust-region subproblem descent direction that preserves sufficient feasibility in the sense of the (13a). Successively using the mean value theorem and (11), we obviously obtain
\[
f(x_k) - f(x_k + \alpha p_k) = -\alpha \nabla f(x_k)^T p_k - \frac{1}{2} \alpha^2 p_k^T \nabla^2 f(\xi_k) p_k \geq -\kappa_{eg} \alpha^3 \Delta_k^3 - \alpha g_k^T p_k - \frac{1}{2} \alpha^2 p_k^T \nabla^2 f(\xi_k) p_k
\]
\[
= -\kappa_{eg} \alpha_1 g_k^T p_k - \kappa_{eg} \alpha_3 \Delta_k^3 + \alpha (\kappa_1 - 1) g_k^T p_k - \frac{1}{2} \alpha^2 p_k^T \nabla^2 f(\xi_k) p_k,
\] (22)

where \( \xi_k \in (x_k, x_k + s_k) \).

There are two cases that may be considered. The first is \( p_k^T \nabla^2 f(\xi_k) p_k \leq 0 \). By canceling the last term of Eqs. (22), (21), \( \frac{\|g_k^T h_{mk}\|_2}{\|M_{mk}\|} \leq \Delta_k \) for large enough \( k \) and the fact that \( \Delta_k \leq \Delta_{\text{max}} \), it is thus easy to see that there exists an \( \alpha^* = \frac{\kappa_3 (1 - \kappa_1)}{\kappa_{eg} \Delta_{\text{max}}} \frac{\|g_k^T h_{mk}\|_2}{\|M_{mk}\|} > 0 \) such that (13a) holds. The second case is \( p_k^T \nabla^2 f(\xi_k) p_k > 0 \). Using the Cauchy–Schwarz inequality and the fact that \( \alpha_k \in (0, 1) \) and \( \Delta_k \leq \Delta_{\text{max}} \), we deduce that
\[
f(x_k) - f(x_k + \alpha p_k) + \alpha \kappa_1 g_k^T p_k \geq -\kappa_{eg} \alpha_3 \Delta_k^3 + \alpha (\kappa_1 - 1) g_k^T p_k - \frac{1}{2} \alpha^2 p_k^T \nabla^2 f(\xi_k) p_k
\]
\[
\geq -\kappa_{eg} \alpha_3 \Delta_{\text{max}}^3 + \alpha (\kappa_1 - 1) g_k^T p_k - \frac{1}{2} \alpha^2 \kappa_{H_f} \Delta_k^2
\]
\[
= \alpha \left[ \left( -\kappa_{eg} \Delta_{\text{max}} - \frac{1}{2} \kappa_{H_f} \right) \Delta_k^2 + (\kappa_1 - 1) g_k^T p_k \right]
\]
\[
\geq 0,
\]
when \( \alpha^* = \frac{\kappa_3 (1 - \kappa_1) \kappa_{H_f} \Delta_{\text{max}}}{\kappa_{eg} \Delta_{\text{max}}^2 + \frac{1}{2} \kappa_{H_f}} > 0 \). Thus the final conclusion obtained.  

We therefore see that it is reasonable to design line-search step criterion in step 5, which provided us a nonincreasing sequence \( \{f(x_k)\} \).

**Lemma 5** Let step \( p_k \) be the solution of the trust-region subproblem (7). Suppose that (A1)–(A5) hold. Then there exists a positive constant \( \kappa_4 \) such that step \( p_k \) satisfies the following
sufficiently descent condition:

$$\text{Pred}(s_k) \geq \kappa_4 \alpha_k \theta_k \|g_k^T h_{m_k}\|^{\frac{1}{2}} \min \left\{ \Delta_k, \frac{\|g_k^T h_{m_k}\|^{\frac{1}{2}}}{\|M_{m_k}\|} \right\},$$  \hspace{1cm} (23)

for all $g_k, h_{m_k}, \|M_{m_k}\|, \text{ and } \Delta_k$.

**Proof** Combining now (7), (17), Lemma 4, $\theta_k \in (\theta_0, 1]$ and the fact that $\alpha_k \leq 1$, we get

$$\text{Pred}(s_k) = m_k - m(x_k + s_k)$$

$$= -\alpha_k \theta_k g_k^T p_k - \frac{\alpha_k^2 \theta_k^2}{2} p_k^T M_{m_k} p_k$$

$$\geq \frac{1}{2} \kappa_4 \alpha_k \theta_k \|g_k^T h_{m_k}\|^{\frac{1}{2}} \min \left\{ \Delta_k, \frac{\|g_k^T h_{m_k}\|^{\frac{1}{2}}}{\|M_{m_k}\|} \right\}.$$  \hspace{1cm} (23)

3.3 **Global convergence**

Every iteration point in the $k + 1$th iteration will be chosen on the region $B(x_k, \alpha_k \Delta_k)$. Following the lemma one first shows that the current iteration must be successful if $\alpha_k \Delta_k$ is small enough.

**Lemma 6** Suppose that (A1)–(A5) and the error bounds (10)–(12) hold. $m_k$ is fully quadratic on $B(x_k, \Delta_k)$, $\|g_k^T h_{m_k}\| \neq 0$ and

$$\alpha_k \Delta_k \leq \Delta_k \leq \min \left\{ \frac{1}{\kappa_{Hm} \kappa_{m_m}}, \frac{\kappa_4(1 - \eta_2)}{2 \kappa_{Hm} \Delta_{m_m}} \right\} \|g_k^T h_{m_k}\|^{\frac{1}{2}},$$

where $\kappa_{m_m}$ is the bound of $C_{m_m}$, for all $x \in L(x_0)$. Then the $k$th iteration is successful.

**Proof** We notice that, for all $k$ and the model function $m_k$, one has $f(x_k) = m(x_k)$. Let $M_{k} = \begin{bmatrix} H_{k} & 0 \\ 0 & C_{k} \end{bmatrix}$, from (16) and (A3), we know that $\|M_{m_k}\| \leq \kappa_{Hm} \kappa_{m_m}$. Thus combining $\Delta_k \leq \|g_k^T h_{m_k}\|^{\frac{1}{2}} \kappa_{m_m}^{-\frac{1}{2}}$ with the sufficient decrease condition (23), we immediately get

$$\text{Pred}(s_k) \geq \kappa_4 \alpha_k \theta_k \|g_k^T h_{m_k}\|^{\frac{1}{2}} \min \left\{ \Delta_k, \frac{\|g_k^T h_{m_k}\|^{\frac{1}{2}}}{\|M_{m_k}\|} \right\} \geq \kappa_4 \alpha_k \theta_k \|g_k^T h_{m_k}\|^{\frac{1}{2}} \Delta_k.$$  \hspace{1cm} (23)

Using Eqs. (12), (23), the fact that $\alpha_k \in (0, 1]$ and $\theta_k \in (0, 1]$, we have

$$|\rho_k - 1| = \left| \frac{f(x_k) - f(x_{k+1})}{m(x_k) - m(x_{k+1})} - 1 \right|$$
\[
\begin{align*}
\leq & \quad \left| \frac{f(x_k) - m(x_k)}{m(x_k) - m(x_{k+1})} \right| + \left| \frac{m(x_{k+1}) - f(x_{k+1})}{m(x_{k}) - m(x_{k+1})} \right| \\
& \quad \frac{f(x_{k+1}) - m(x_{k+1})}{m(x_{k}) - m(x_{k+1})} \\
\leq & \quad 2 \kappa \alpha \delta^3 \Delta_k^3 \\
& \quad \frac{\kappa \alpha \delta_k \|g_k^T h_{mk}\|}{\kappa \|g_k^T h_{mk}\|} \Delta_k \\
& \quad \frac{\Delta_{\max}^2}{\kappa \|g_k^T h_{mk}\|} \Delta_k \\
& \quad 1 - \eta_1.
\end{align*}
\]

Thus \( \rho_k \geq \eta_1 \) and the iteration is successful. \( \square \)

**Lemma 7** Suppose that (A1)–(A5) and the error bounds (10)–(12) hold. If the number of successful iteration is finite, then

\[
\lim_{k \to +\infty} \|\nabla f(x_k)^T h_k\| = 0.
\]

**Proof** We consider that all the model-improving iterations before \( m_k \) becomes fully quadratic are less than a constant \( N \). Suppose that the current iteration is an iteration after a successful one. It means that an infinite number of iterations are acceptable or not nice. In these two cases, \( \Delta_k \) is shrinking. Furthermore, \( \Delta_k \) is reduced by a factor \( \xi \) at least once every \( N \) iterations, which implies \( \Delta_k \to 0 \).

For the \( j \)th iteration, we denote the \( i \)th iteration after \( j \) by the index \( i_j \), then

\[
\|x_j - x_{i_j}\| \leq N \Delta_j \to 0, \quad j \to +\infty.
\]

Using the triangle inequality, we obtain

\[
\|\nabla f(x_j)^T h_{i_j}\| \leq \|\nabla f(x_j)^T h_{i_j} - \nabla f(x_j)^T h_{j}\| + \|\nabla f(x_j)^T h_{j} + \nabla f(x_j)^T h_{i_j}\| \\
+ \|\nabla f(x_j)^T h_{i_j} - g_{i_j}^T h_{i_j} - g_{i_j}^T h_{i_j} + g_{i_j}^T h_{i_j}\|.
\]

The following work is to show that all these terms on the right-hand side are converging to zero. Because of the Lipschitz continuity of \( \nabla f \) and the fact that \( \|x_{i_j} - x_j\| \to 0 \) the first and second terms converge to zero. The inequalities (10) and (11) imply the third and fourth terms on the right-hand side are converging to zero. According to Lemma 3, if \( \|g_k^T h_{mk}\| \to 0 \) for small enough \( \Delta_{i_j}, i_j \) would be a successful iteration, which yield a contradiction. Thus the last term converges to zero. \( \square \)

**Lemma 8** Suppose that (A1)–(A5), the error bounds (10)–(12) and (23) hold. Suppose furthermore that the strict complementarity of the problem (1) holds. Then

\[
\lim_{k \to +\infty} \|g_k^T h_{mk}\| = 0.
\]

**Proof** The key is that we may find a contradiction with the fact that \( \{f(x_k)\} \) is a nonincreasing bounded sequence unless \( x_k \) is a stationary point. We thus have to verify that there exists some \( \epsilon > 0 \) such that \( \{f(x_k)\} \) is not convergent under the assumption of \( \|g_k^T h_{mk}\| \geq \epsilon^2 \).
We observe from (13a), Lemma 4 and (21) that
\[
\begin{align*}
 f(x_k) - f(x_k + \alpha_k p_k) & \geq -\alpha_k \kappa_1 g^T_k p_k \\
 & \geq \alpha_k \kappa_1 \kappa_3 \left\| g^T_k h_{m_k} \right\|^2 \min \left\{ \Delta_k, \frac{\| g^T_k h_{m_k} \| \| g^T_k h_{m_k} \|}{\| M_{m_k} \|} \right\} \\
 & \geq \alpha_k \kappa_1 \kappa_3 \epsilon \min \left\{ \Delta_k, \frac{\epsilon}{\| M_{m_k} \|} \right\} \to 0.
\end{align*}
\]

Thus from (24), two cases should be considered next, that is,
\[
\liminf_{k \to \infty} \alpha_k \neq 0 \tag{25}
\]
and
\[
\lim_{k \to \infty} \Delta_k \neq 0. \tag{26}
\]

We now start the proof of (25). On one hand, \(\alpha_k\) is accepted by (13b) the boundary of inequality constraints along \(p_k\). From Eq. (15)
\[
\Gamma_k = \min \left\{ -\frac{a_i^T x_k - b_i}{a_i^T p_k} \mid \frac{a_i^T x_k - b_i}{a_i^T p_k} > 0, i = 1, 2, \ldots, m \right\},
\]
with \(\alpha_k = +\infty\) if \(- (a_i^T x_k - b_i)/(a_i^T p_k) \leq 0\) for all \(i = 1, 2, \ldots, m\), \(\hat{p}_k = D_k^{-\frac{1}{2}} A p_k\) and (17), we know that there exists \(\hat{\lambda}_{m_{k+1}}\) such that
\[
\alpha_k = - \frac{a_i^T x_k - b_i}{a_i^T p_k} \geq \frac{\nu_{m_k} + |\lambda^i_{m_{k+1}}|}{\lambda^i_{m_{k+1}}} \geq \frac{\nu_{m_k} + |\lambda^i_{m_{k+1}}|}{\| \lambda_{m_{k+1}} \|_\infty}. \tag{27}
\]
From (17), we have
\[
\begin{bmatrix}
 A^T \\
 -D_k^{-\frac{1}{2}}
\end{bmatrix} \hat{\lambda}_{m_{k+1}} = - \begin{bmatrix} g_k \\
 0
\end{bmatrix} + (M_{m_k} + \nu_{m_k} I) \begin{bmatrix} p_k \\
 \hat{p}_k
\end{bmatrix}.
\]
Since \([A - D_k^{-\frac{1}{2}}]^T\] is full rank for all \(x \in \mathcal{L}(x_0)\), \(\lambda_{m_k}\) is bounded and \(m(x)\) is twice continuously differentiable, there exist \(\kappa_5 > 0\) and \(\kappa_6 > 0\) such that
\[
\| \lambda_{m_{k+1}} \|_\infty \leq \kappa_5 + (\kappa_6 + \nu_{m_k}) \Delta_k.
\]
Using the fact that $v_{mk} (\Delta_k - \| (p_k; \hat{p}_k) \|) = 0$ and taking the norm to both sides of (17), we deduce that

\[ v_{mk} \Delta_k = v_{mk} \left\| \left( p_k \right)_{\hat{p}_k} \right\| \]

\[ \geq \left( \left\| g_k - A^T \lambda_{mk} \right\|^2 + \left\| D_k^{1/2} \lambda_{mk} \right\|^2 \right)^{1/2} - \| M_{mk} \| \left\| (p_k; \hat{p}_k) \right\| \]

\[ = \left\| g_k^T h_{mk} \right\|^{1/2} - \| M_{mk} \| \left\| (p_k; \hat{p}_k) \right\|. \]

And noting $\left\| (p_k; \hat{p}_k) \right\| \leq \Delta_k$, we can obtain

\[ v_{mk} \geq \frac{\left\| g_k^T h_{mk} \right\|^{1/2}}{\Delta_k} - \| M_{mk} \|. \]

Combining the assumption $\| g_k^T h_{mk} \| > \epsilon^2$ with $\Delta_k \rightarrow 0$ deduced from (24), it is clear from the fact $\| M_{mk} \| \leq \kappa \lambda_m \kappa H_m$ that, for $\forall k$,

\[ \lim_{k \rightarrow \infty} v_{mk} = +\infty. \]

Thus (27) implies that

\[ \lim_{k \rightarrow \infty} \alpha_k \neq 0. \]

Furthermore, $\Delta_k \rightarrow 0$ means that $\lim_{k \rightarrow \infty} \| p_k \| = 0$, from which we deduce that, for some $0 < \theta_0 < 1$ and $\theta_k - 1 = O(\| p_k \|^2)$, the strictly feasible stepsize $\theta_k \in (\theta_0, 1) \rightarrow 1$. From the above, we have already seen that (25) holds in the case that $\alpha_k$ is determined by (13b).

There is another case that $\alpha_k$ is determined by (13a). In this case, we are able to verify that $\alpha_k = 1$ is acceptable when $k$ sufficiently large. If not,

\[ \kappa g_k^T p_k < f(x_k + p_k) - f(x_k) \]

must hold. Applying the Taylor series, (10)–(11), (A3) and the fact that $\Delta_k \leq \Delta_{\max}$, we deduce that

\[ \kappa g_k^T p_k < f(x_k + p_k) - f(x_k) = \nabla f(x_k)^T p_k + \frac{1}{2} p_k^T \nabla^2 f(\xi_k) p_k \]

\[ \leq \left( \kappa g \Delta_{\max} + \frac{1}{2} \kappa_{eh} \Delta_{\max} + \frac{1}{2} \kappa H_m \right) \Delta_k^2 + g_k^T p_k, \]

where $\xi_k \in (x_k, x_k + s_k)$. This inequality is equivalent to the form of

\[ (1 - \kappa) g_k^T p_k + \left( \kappa g \Delta_{\max} + \frac{1}{2} \kappa_{eh} \Delta_{\max} + \frac{1}{2} \kappa H_m \right) \Delta_k^2 > 0. \]

Moreover, (21) and $\| g_k^T h_{mk} \| \geq \epsilon^2$ imply that

\[ -(1 - \kappa) \epsilon^2 \min \left\{ k_k, \frac{\epsilon}{\kappa H_m} \right\} + \left( \kappa g \Delta_{\max} + \frac{1}{2} \kappa_{eh} \Delta_{\max} + \frac{1}{2} \kappa H_m \right) \Delta_k^2 > 0. \]
Thus if $\Delta_k \leq \frac{2(1-\kappa_1)\epsilon \kappa_2}{2\kappa_2 \Delta_{\max} + \kappa_3 \Delta_{\max} + \kappa H_{\max}} \leq \frac{\epsilon}{\kappa H_{\max}}$, we deduce from the inequality

$$
\Delta_k \left[ \Delta_k \left( \kappa_2 \Delta_{\max} + \frac{1}{2} \kappa_3 \Delta_{\max} + \frac{1}{2} \kappa \frac{H_{\max}}{\Delta_{\max}} \right) - (1 - \kappa_1) \kappa_3 \epsilon \right] > 0
$$

that

$$
\Delta_k > \frac{2(1 - \kappa_1)\kappa_3 \epsilon}{2\kappa_2 \Delta_{\max} + \kappa_3 \Delta_{\max} + \kappa H_{\max}}.
$$

Clearly, a contradiction appears here. It implies that $\alpha_k = 1$ for $k$ sufficiently large. Therefore (25) always holds.

On the other hand, we should prove that (26) is true. From step 3 of Algorithm 1, we know that

$$
\Delta_k \geq \iota \frac{\|g^T h_{m_k}\|}{\Delta_{\max}}.
$$

By the assumption that $\|g^T h_{m_k}\| \geq \epsilon$, we obtain

$$
\Delta_k \geq \epsilon.
$$

Whenever $\Delta_k$ falls below a constant $\bar{\kappa}_7$ given by

$$
\bar{\kappa}_7 = \min \left\{ \frac{\epsilon}{\kappa H_{\max}}, \frac{\kappa_2 (1 - \eta_1)}{2 \kappa_2 \Delta_{\max}} \right\},
$$

the $k$th iteration is either successful or model-improving, and hence from step 9, we are able to deduce both that $\Delta_{k+1} \geq \Delta_k$ and $\Delta_{k+1} \geq \zeta \Delta_k$. Combining with the rules of step 9 we conclude that $\Delta_{k+1} \geq \min \{\epsilon, \xi, \bar{\kappa}_7\} = \kappa_7$. It means that $\Delta_k \to 0$, if $\|g^T h_{m_k}\| \geq \epsilon^2$.

In conclusion, the sequence $\{f(x_k)\}$ is not convergent if we suppose that $\|g^T h_{m_k}\| \geq \epsilon^2$, which contradicts the fact that $\{f(x_k)\}$ is a nonincreasing bounded sequence. It implies that

$$
\liminf_{k \to +\infty} \|g^T h_{m_k}\| = 0. \quad \Box
$$

Lemma 9 For any subsequence $\{k_i\}$ such that

$$
\lim_{i \to +\infty} \|g^T h_{m_{k_i}}\| = 0, \quad (28)
$$

we also have

$$
\lim_{i \to +\infty} \|\nabla f_{k_i}^T h_{k_i}\| = 0. \quad (29)
$$

Proof First, we note that, by (28), $\|g^T h_{m_{k_i}}\| \leq \epsilon$ when $i$ sufficiently large. Thus the criticality step ensures that the model $m_{k_i}$ is a fully quadratic function on the ball $B(x_{k_i}, \Delta_{k_i})$, with $\Delta_{k_i} \leq \epsilon \|g^T h_{m_{k_i}}\|$ for all $i$ sufficiently large (if $\|\nabla f_{k_i}^T h_{k_i}\| \neq 0$). Then, using the bound (20) on the error between the terminal conditions of function and model, we have

$$
\|\nabla f_{k_i}^T h_{k_i} - g^T h_{m_{k_i}}\| \leq \kappa_2 \alpha_{k_i} \Delta_{k_i} \leq \kappa_2 \alpha_{k_i} \|g^T h_{m_{k_i}}\| \leq \kappa_2 \epsilon \|g^T h_{m_{k_i}}\|.
$$
As a consequence, we have
\[
\| \nabla f_k^T h_{h_k} \| = \| \nabla f_k^T h_{h_k} - g_k^T h_{m_{k_j}} \| + \| g_k^T h_{m_{k_j}} \| \leq (\kappa_2 r_{k_i^0} + 1) \| g_k^T h_{m_{k_j}} \|
\]
for all \( i \) sufficiently large. But \( \| g_k^T h_{m_{k_j}} \| \to 0 \) implies (29) holds.

Then we obtain the global convergence derived from Lemmas 8 and 9.

**Theorem 1** Suppose that (A1)–(A5), the error bounds (10)–(12) and (23) hold. Suppose furthermore that the strict complementarity of the problem (1) holds. Let \( \{x_k\} \subset \mathbb{R}^n \) be sequence generated by Algorithm 1. Then
\[
\lim_{k \to +\infty} \| \nabla f_k^T h_{h_k} \| = 0.
\]

The above theorem shows us there exists a limit point that is first-order critical. In fact, we are able to prove that all limit points of the sequence of iterations are first-order critical.

**Theorem 2** Suppose that (A1)–(A5), the error bounds (10)–(12) and (23) hold. Suppose furthermore that the strict complementarity of the problem (1) holds. Let \( \{x_k\} \subset \mathbb{R}^n \) be sequence generated by Algorithm 1. Then
\[
\lim_{k \to +\infty} \| \nabla f_k^T h_{h_k} \| = 0.
\]

**Proof** We first obtained from Lemma 7 that the theorem holds in the case when \( S \) is finite. Hence, we will assume that \( S \) is infinite. For the purpose of deriving a contradiction, we suppose that there exists a subsequence \( \{k_i\} \) of successful or acceptable iterations such that
\[
\| \nabla f_{k_i}^T h_{h_{k_i}} \| \geq \epsilon_1^2 > 0
\]
for some \( \epsilon_1 > 0 \) and for all \( i \). Then, because of Lemma 9, we obtain
\[
\| g_{k_i}^T h_{m_{k_i}} \| \geq \epsilon_2^2 > 0
\]
for some \( \epsilon_2 > 0 \) and for all \( i \) sufficiently large. Without loss of generality, we pick \( \epsilon_2 \) such that
\[
\epsilon_2^2 \leq \min \left\{ \frac{\epsilon_1^2}{2(2 + \kappa_{\text{ed}})}, \epsilon \right\}.
\]
Lemma 8 then ensures the existence, for each \( \{k_i\} \) in the subsequence, of a first iteration \( \ell_i > k_i \) such that \( \| g_{\ell_i}^T h_{m_{\ell_i}} \| < \epsilon_2^2 \). By removing elements from \( \{k_i\} \), without loss of generality and without a change of notation, we thus see that there exists another subsequence indexed by \( \{\ell_i\} \) such that
\[
\| g_{\ell_i}^T h_{m_{\ell_i}} \| \geq \epsilon_2^2 \quad \text{for} \quad k_i \leq k \leq \ell_i \quad \text{and} \quad \| g_{\ell_i}^T h_{m_{\ell_i}} \| < \epsilon_2^2,
\]
for sufficiently large \( i \), with inequality (30) being retained.
We now restrict our attention to the set $\mathcal{K}$ corresponding to the subsequence of iterations whose indices are in the set

$$\bigcup_{i \in N_0} \{ k \in N_0 : k_i \leq k \leq \ell_i \},$$

where $k_i$ and $\ell_i$ belong to the two subsequences given above in (31).

We know that $\| g_k^T h_{m_k} \| \geq \epsilon^2_k$ for $k \in \mathcal{K}$. From Lemma 8 we have $\lim_{k \to +\infty} \alpha_k \Delta_k = 0$ and by Lemma 5 we conclude that for any large enough $k \in \mathcal{K}$ the iteration $k$ is either successful if the model is fully quadratic or model-improving otherwise. Moreover, for each $k \in \mathcal{K} \cap S$ we have

$$f(x_k) - f(x_k + s_k) \geq \eta_1 \left[ m(x_k) - m(x_k + h_k) \right] \geq \eta_1 K_4 \alpha_k \theta_k \| g_k^T h_{m_k} \| \frac{1}{2} \min \left\{ \Delta_k, \frac{1}{\| M_{m_k} \|} \right\},$$

and, for any such $k$ large enough, $\Delta_k \leq \frac{\epsilon^2_k}{\eta_1 K_4 \theta_k}$. Hence, we have $\alpha_k \theta_k \Delta_k \leq \frac{(f(x_k) - f(x_k + s_k))}{\eta_1 K_4 \epsilon_k^2}$ for $k \in \mathcal{K} \cap S$ sufficiently large. Since for any $k \in \mathcal{K}$ large enough the iteration is either successful or model-improving and since for a model-improving iteration $x_{k+1} = x_k + s_k$, we have, for all $i$ sufficiently large,

$$\| x_{k_i} - x_{\ell_i} \| \leq \sum_{j=k_i}^{\ell_i-1} \| x_j - x_{j+1} \| \leq \sum_{j=k_i}^{\ell_i-1} \alpha_j \theta_j \Delta_j \leq \frac{1}{\eta_1 K_4 \epsilon_k^2} \left[ f(x_k) - f(x_{\ell_i}) \right].$$

Because the sequence $\{ f(x_k) \}$ is bounded below and monotonic decreasing, we see that the right-hand side of this inequality must converge to zero, and we therefore obtain

$$\lim_{i \to +\infty} \| x_{k_i} - x_{\ell_i} \| = 0.$$

Now,

$$\| \nabla f(x_k)^T h_{f_{k_i}} \| \leq \| \nabla f(x_k)^T h_{f_{k_i}} - \nabla f(x_{\ell_i})^T h_{f_{\ell_i}} \| + \| \nabla f(x_{\ell_i})^T h_{f_{\ell_i}} - g_{\ell_i}^T h_{m_{\ell_i}} \| + \| g_{\ell_i}^T h_{m_{\ell_i}} \|.$$

Since $\nabla f$ is Lipschitz continuity, we see that the first term of the above inequality $\| \nabla f(x_k)^T h_{f_{k_i}} - \nabla f(x_{\ell_i})^T h_{f_{\ell_i}} \| \to 0$ and is bounded by $\epsilon^2_k$ for $i$ sufficiently large. Equation (32) implies the third term $\| g_{\ell_i}^T h_{m_{\ell_i}} \| \leq \epsilon^2_k$. From (31) we see that $m_{\ell_i}$ is a fully quadratic function on $B(x_{\ell_i}, \| g_{\ell_i}^T h_{m_{\ell_i}} \|)$. Using (11) and (32), we deduce that $\| \nabla f(x_{\ell_i})^T h_{f_{\ell_i}} - g_{\ell_i}^T h_{m_{\ell_i}} \| \leq \kappa_y \epsilon_k^2$ for $i$ sufficiently large. Combining with these bounds we obtain the consequence that

$$\| \nabla f_{f_{k_i}}^T h_{f_{k_i}} \| \leq (2 + \kappa_y \epsilon_k^2) \epsilon^2_k \leq \frac{1}{2} \epsilon^2_k$$

for $i$ large enough. This result contradicts (30), which implies the initial assumption is false and the theorem follows. □
3.4 Local convergence

Having proved the global convergence, we now focus on the speed of the local convergence. For this motivation, more acceptable assumptions are given as follows.

Assumptions

(A6) $x_*$ is the solution of problem (1), which satisfies the strong second-order sufficient condition, that is, let the columns of $Z_*$ denote an orthogonal basis for the null space of $[A - D_1^2]$, then there exists $\sigma > 0$ such that

$$d^T(Z_*M_*Z_*)d \geq \sigma \|d\|^2, \quad \forall d.$$  (33)

(A7) Let

$$\lim_{k \to \infty} \frac{\|(M_{mk} - M_f)Zkp_k\|}{\|p_k\|} = 0.$$  (34)

This means that for large $k$

$$p_k^T(Z_k^TM_{mk}Z_k)p_k = p_k^T(Z_k^TM_fZ_k)p_k + o(\|p_k\|^2).$$

Theorem 3 Suppose that (A1)–(A7), the error bounds (10)–(12) and (23) hold. \{x_k\} is a sequence generated by Algorithm 1. Suppose furthermore that the strict complementarity of the problem (1) holds. Then, for sufficiently large $k$, the stepsize $\alpha_k \equiv 1$ and there exists $\hat{\Delta} > 0$ such that

$$\Delta_k \geq \Delta_k' \geq \hat{\Delta}, \quad \forall k \geq K',$$ where $K'$ is a large enough index.

Proof According to the algorithm, the stepsize $\alpha_k$ is given in (15)

$$\Gamma_k = \min \left\{ -\frac{a_i^T x_k - b_i}{a_i^T p_k} \mid -\frac{a_i^T x_k - b_i}{a_i^T p_k} > 0, i = 1, 2, \ldots, m \right\}.$$ From (17), there exists $\lambda_{mk+1}$ such that

$$a_i^T p_k = (a_i^T p_k - b_i)^\frac{1}{2} \hat{p'}_k = -\frac{(a_i^T p_k - b_i)\lambda_{mk+1}^i}{v_{mk} + |\lambda_{mk+1}^i|},$$  (35)

where $\hat{p'}_k$ and $\lambda_{mk+1}^i$ are the $i$th component of the vectors $\hat{p}_k$ and $\lambda_{mk+1}$, respectively.

If $\|p_k\| < \Delta_k$, then $v_{mk} = 0$. Since the strict complementarity of the problem (1) holds at every limit point of \{x_k\}, i.e., $|\lambda_{mk+1}^i|/|a_i^T x_k - b_i| > 0$, for all large $k$, $\lambda_{mk+1} = \lambda_{mk+1}^N > 0$ when $v_{mk} = 0$. So, $\lambda_{mk+1}^N > 0$. From (35), it is clear that $\lim_{k \to \infty} \alpha_k = 1$.

If $\|p_k\| = \Delta_k \to 0$, then $v_{mk+1} \to \infty$. From (35),

$$\alpha_k = -\frac{a_i^T x - b_j}{a_i^T p_k} \geq \frac{v_{mk} + |\lambda_{mk+1}^i|}{|\lambda_{mk+1}^i|} \geq \frac{v_{mk} + |\lambda_{mk+1}^i|}{\|\lambda_{mk+1}^i\|_{\infty}} \to \infty.$$  

From the above, we have found that if $\|g_k^T h_{mk}\| \geq \varepsilon^2$ holds and $\Delta_k \to 0$, we conclude that $\lim_{k \to \infty} \alpha_k = +\infty$, and $\lim_{k \to \infty} \theta_k = 1$. 

Further, by the condition on the strictly feasible stepsize \( \theta_k - 1 = O(\|p_k\|) \), and \( \lim_{k \to \infty} p_k = 0 \), we have \( \lim_{k \to \infty} \theta_k = 1 \).

We can obtain from above that \( \lim_{k \to \infty} \Gamma_k = +\infty \) when \( \alpha_k \) is given in (15) along \( p_k \). It means that if \( \alpha_k \) is determined by (13b), \( \alpha_k \equiv 1 \) for sufficiently large \( k \). Thus

\[
f(x_k + p_k) = f(x_k) + \nabla f(x_k)^T p_k + \frac{1}{2} p_k^T H_k p_k + o(\|p_k\|^2) \\
= f(x_k) + \kappa_1 g_k^T p_k + \left( \frac{1}{2} - \kappa_1 \right) g_k^T p_k + \nabla f_k^T p_k - g_k^T p_k \\
+ \frac{1}{2} (g_k^T p_k + p_k^T H_{k_{m_k}} p_k) + \frac{1}{2} p_k^T (H_{k_{m_k}} - H_{m_k}) p_k + o(\|p_k\|^2).
\]

The error bound (11) shows us \( (g_k - \nabla f_k)^T p_k = o(\|p_k\|^2) \). Hence we see from (36) that \( f(x_k + p_k) \leq f(x_k) + \kappa_1 g_k^T p_k \) at the \( k \)th iteration.

Combining with the fact that \( p_k^T A^T D_k^{-1} C_{m_k} A p_k \to 0 \), we know that \( x_{k+1} = x_k + p_k \). So

\[
\left| f(x_k) - f(x_k + p_k) - m(x_k) + m(x_k + p_k) \right| \\
= \left| \left[ g_k^T p_k + \frac{1}{2} p_k^T M_{m_k} p_k \right] - \left[ \nabla f(x_k)^T p_k + \frac{1}{2} p_k^T H_k p_k + o(\|p_k\|^2) \right] \right| \\
= \left| (g_k - \nabla f_k)^T p_k + \frac{1}{2} p_k^T (H_{m_k} - H_k) p_k + o(\|p_k\|^2) \right| \\
\overset{(10)/(11)}{=} o(\|p_k\|^2).
\]

By assumptions (A1)–(A7), we can obtain

\[
\rho_k - 1 = \frac{f(x_k) - f(x_k + p_k) + m(x_k) - m(x_k + p_k)}{\text{Pred}(p_k)} \\
= \frac{\left[ g_k^T p_k + \frac{1}{2} p_k^T M_{m_k} p_k \right] - \left[ \nabla f(x_k)^T p_k + \frac{1}{2} p_k^T H_k p_k + o(\|p_k\|^2) \right]}{\text{Pred}(p_k)} \\
= \frac{o(\|p_k\|^2)}{\text{Pred}(p_k)}.
\]

By (16) and (17), we get

\[
g_k^T p_k = \begin{bmatrix} g_k^T \\
0 \end{bmatrix} + \begin{bmatrix} A^T \\
-D_k^T \end{bmatrix}_{m_k+1} \hat{\lambda}_{m_k+1} \left[ 
\begin{array}{c}
p_k \\
\hat{p}_k \end{array} \right] \\
= -\left[ p_k^T, \hat{p}_k^T \right] (M_{m_k} + M_{m_k} I) \left[ 
\begin{array}{c}
p_k \\
\hat{p}_k \end{array} \right] \\
\leq -\left[ p_k^T, \hat{p}_k^T \right] M_{m_k} \left[ 
\begin{array}{c}
p_k \\
\hat{p}_k \end{array} \right].
\]

Let the columns of \( Z_k \) denote an orthogonal basis for the null space of \( [A \quad -D_k^T] \). We get \( g_k^T p_k \leq -[p_k^T, \hat{p}_k^T] M_{m_k} \left[ 
\begin{array}{c}
p_k \\
\hat{p}_k \end{array} \right] = -p_k^T Z_k^T M_{m_k} Z_k p_k \). Therefore, from (33)–(34), we see that for
all large $k$
\[ g_k^T p_k \leq -\frac{\sigma}{2} \|p_k\|^2 + o(\|p_k\|^2). \]

Hence, one has
\begin{align*}
\text{Pred}(p_k) &= -g_k^T p_k - \frac{1}{2} p_k^T M_k p_k \\
&= -g_k^T p_k - \frac{1}{2} p_k^T (H_k + A_k D_k^{-1} C_k A_k) p_k \\
&= -g_k^T p_k - \frac{1}{2} p_k^T M_k p_k + o(\|p_k\|^2) \\
&\geq -\frac{\sigma}{4} \|p_k\|^2 + o(\|p_k\|^2).
\end{align*}

(38)

For a similar proof, we can obtain $p_k \to 0$. Combining (37) with (38), one has the fact that $\rho_k \to 1$. Hence there exists $\hat{\Delta} > 0$ such that when $\|p_k\| \leq \hat{\Delta}$, $\rho_k \geq \eta_2$, and $\Delta_{k+1} \geq \Delta_k$. As $p_k \to 0$, there exists an index $K'$ such that $\|p_k\| \leq \hat{\Delta}$ whenever $k \geq K'$. Thus, the conclusion holds.

Theorem 3 implies that the local convergence rate of Algorithm 1 depends on the Hessian at $x_*$ and the local convergence rate of $p_k$. Meanwhile, if $p_k$ is a quasi-Newton step, for sufficiently large $k$, the sequence $\{x_k\}$ will reach a superlinear local convergence rate to the optimal point $x_*$.

4 Numerical experiments
We now demonstrate the experimental performance of the proposed derivative-free trust-region method.

Environment: The algorithms are written in Matlab R2009a and run on a PC with 2.66 GHz Intel(R) Core(TM)2 Quad CPU and 4 G DDR2.

Initialization: The values $\Delta_0 = 2$, $\eta_0 = 0.25$, $\eta_1 = 0.75$, $\zeta = 0.5$, $\xi = 1.5$, $\iota = 0.5$, $\beta = 0.25$, $\alpha = 0.2$, $\varepsilon = 10^{-8}$ and $\omega = 0.3$ are used. $\Delta_{\max}$ is equal to 4, 6, 8, respectively.

Termination criteria: $\|g_k^T h_k\| \leq \varepsilon$.

Problems: We first test 20 linear inequality constrained optimization problems (listed in Table 1) from Test Examples for Nonlinear Programming Codes [15, 16]. It is worth noting that the assumptions (A2)–(A5) play very important roles in the theoretical proof. Here (A2) is a general assumption in the optimization problem and (A5) can be satisfied.

| No. | Problem  | Dim   | $x_0$         | No. | Problem  | Dim   | $x_0$         |
|-----|----------|-------|---------------|-----|----------|-------|---------------|
| 1   | HS21     | 2     | [-1, -1]      | 2   | HS24     | 2     | [1, 0.5]      |
| 3   | HS25     | 3     | [100, 12.5, 3]| 4   | HS35     | 3     | [0.5, 0.5, 0.5]| 
| 5   | HS36     | 3     | [10, 10, 10]  | 6   | HS37     | 3     | [10, 10, 10]  |
| 7   | HS44     | 4     | [0, 0, 0, 0]  | 8   | HS45     | 5     | [2, 2, 2, 2]  |
| 9   | HS76     | 4     | [0.5, 0.5, 0.5, 0.5] | 10  | HS224    | 2     | [0.1, 0.1]    |
| 11  | HS231    | 2     | [-1.2, 1]     | 12  | HS232    | 2     | [2, 0.5]      |
| 13  | HS224    | 2     | [0.1, 0.1]    | 14  | HS232    | 2     | [2, 0.5]      |
| 15  | HS250    | 3     | [10, 10, 10]  | 16  | HS251    | 3     | [10, 10, 10]  |
| 17  | HS253    | 3     | [0.2, 0.2]    | 18  | HS268    | 5     | [1, 1, …, 1] |
| 19  | HS331    | 2     | [0.5, 0.5, 0.5] | 20  | HS340    | 3     | [1, 1, 1]     |
if the iteration points are not optimal. According to the definitions of error bounds in our algorithm, the gradient (or Hessian) of the model function must be bounded if there exists a constant such that the gradient (or Hessian) norm of the objective function is bounded. Therefore, most of the above test problems satisfy the assumptions (A2)–(A5).

For example (HS21)

\[
\begin{align*}
\min & \quad f(x) = 0.01x_1^2 + x_2^2 - 100 \\
\text{s.t.} & \quad 10x_1 - x_2 - 10 \geq 0, \\
& \quad 2 \leq x_1 \leq 50, \\
& \quad -50 \leq x_2 \leq 50; \\
\end{align*}
\]

\[
\|\nabla f(x)\| = \left\| \begin{bmatrix} 0.02x_1 \\ 2x_2 \end{bmatrix} \right\| \leq 980.0005, \quad \|\nabla^2 f(x)\| = \left\| \begin{bmatrix} 0.02 & 0 \\ 0 & 2 \end{bmatrix} \right\| = 2.
\]

Of course, we will use the level set to limit the bound of \(\|\nabla f(x)\|\) during program execution, which will be much smaller than this value. Even if the boundedness of the gradient and of the Hessian of the objective functions cannot be satisfied at the same time, at least the boundedness within the level set can be guaranteed.

We use the tool of Dolan and Moré [17] to analyze the efficiency of the given algorithm. Figures 1 and 2 show that Algorithm 1 is feasible and has the robust property.

Furthermore we test five simple linear inequality constrained optimization problems from [16] and compare the experiment results of different trust-region radius upper bound
Table 2 shows the experiment results, where $nf$ represents the number of function evaluations, $n$ is the dimension of the test problems and $F$ means the algorithm terminated in the case that the iteration number exceeds the maximum number. The CPU times of the test problems are reported. Table 2 indicates that Algorithm 1 is executable to reach optimal point. The choice of $\Delta_{\text{max}} = 6$ is made to enable us to carry out more gratifying results. But the results show that the number of iterations maybe higher than any other derivative-based algorithms. The reason we think is that the derivatives of most of the test problems we chose are available and a derivative-free technique may increase the number of executions; then higher iteration numbers are necessary.

5 Conclusions

In this paper, we propose an affine-scaling derivative-free method for linear inequality constrained optimizations.

(1) This algorithm is mainly designed to solve the unavailable derivatives optimization problems in engineering. The proposed algorithm adopts interior backtracking technique and possesses the trust-region property.

(2) The global convergence is proved by using the definition of fully quadratic. It shows that the iteration points generated by the proposed algorithm could converge to the optimal points of (1). Meanwhile, we get the result that the local convergence rate of the proposed algorithm depends on $p_k$. If $p_k$ becomes the quasi-Newton step, then the sequence $x_k$ generated by the algorithm converges to $x_*$ superlinearly.

(3) The preliminary numerical experiments verify the new algorithm we proposed is feasible and effective for solving unavailable-derivative linear inequality constrained optimization problems.

Acknowledgements
This work is supported by the National Science Foundation of China under Grant No. 11626037, 13th five-year Science and Technology Project of Education Department of Jilin Province under Grant No. JJKH20170036KJ, the PhD Start-up Fund of Natural Science Foundation of Beihua University and Youth Training Project Foundation of Beihua University.

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
All authors contributed equally and significantly in writing this article. All authors read and approved the final manuscript.

Author details
1 School of Mathematics and Statistics, Beihua University, Jilin, PR. China. 2 School of Information Technology and Media, Beihua University, Jilin, PR. China.
References

1. Kanzow, C., Klug, A.: An interior-point affine-scaling trust-region method for semismooth equations with box constraints. Comput. Optim. Appl. 37(3), 329–353 (2007)
2. Kanzow, C., Klug, A.: On affine-scaling interior-point Newton methods for nonlinear minimization with bound constraints. Comput. Optim. Appl. 35(2), 177–197 (2006)
3. Heinenschloss, M., Ulbrich, M., Ulbrich, S.: Superlinear and quadratic convergence of affine-scaling interior-point Newton methods for problems with simple bounds without strict complementarity assumption. Math. Program. 86(1), 615–635 (1999)
4. Liuzzi, G., Lucidi, S., Sciandrone, M.: Sequential penalty derivative-free methods for nonlinear constrained optimization. SIAM J. Optim. 20(5), 2614–2635 (2010)
5. Coleman, T.F., Li, Y.: A trust region and affine scaling interior point method for nonconvex minimization with linear inequality constraints. Math. Program. 88(1), 1–31 (1997)
6. Zhu, D.: A new affine scaling interior point algorithm for nonlinear optimization subject to linear equality and inequality constraints. J. Comput. Appl. Math. 161(1), 1–25 (2003)
7. Sahu, D.R., Yao, J.C.: A generalized hybrid steepest descent method and applications. J. Nonlinear Var. Anal. 1, 111–126 (2017)
8. Gibali, A.: Two simple relaxed perturbed extragradient methods for solving variational inequalities in Euclidean spaces. J. Nonlinear Var. Anal. 2, 49–61 (2018)
9. Zhang, H., Conn, A.R., Scheinberg, K.: A derivative-free algorithm for least-squares minimization. SIAM J. Optim. 20(6), 3555–3576 (2010)
10. Zhang, H., Conn, A.R.: On the local convergence of a derivative-free algorithm for least-squares minimization. Comput. Optim. Appl. 51(2), 461–507 (2012)
11. Liuzzi, G., Lucidi, S., Rinaldi, F.: A derivative-free approach to constrained multiobjective nonsmooth optimization. SIAM J. Optim. 26(4), 2744–2774 (2016)
12. Tung, L.T.: Higher-order contingent derivative of perturbation maps in multiobjective optimization. J. Nonlinear Funct. Anal. 2015, 19 (2015)
13. Conn, A.R., Scheinberg, K., Vicente, L.N.: Global convergence of general derivative-free trust-region algorithms to first- and second-order critical points. SIAM J. Optim. 20(1), 387–415 (2006)
14. Jing, G., Zhu, D.: An affine scaling derivative-free trust region method with interior backtracking technique for bounded-constrained nonlinear programming. J. Syst. Sci. Complex. 27(3), 537–564 (2014)
15. Hock, W., Schittkowski, K.: Test Examples for Nonlinear Programming Codes. Springer, Bayreuth (1987)
16. Schittkowski, K.: More test examples for nonlinear programming codes (1987)
17. Dolan, E.D., Moré, J.J.: Benchmarking optimization software with performance profiles. Math. Program. 91(2), 201–213 (2002)