A generalized super-memory gradient projection method of strongly sub-feasible directions with strong convergence for nonlinear inequality constrained optimization

Jin-Bao Jian\textsuperscript{a,∗}, You-Fang Zeng\textsuperscript{b,a}, Chun-Ming Tang\textsuperscript{b,a}

\textsuperscript{a} College of Mathematics and Information Science, Guangxi University, Nanning, 530004, PR China
\textsuperscript{b} College of Sciences, Shanghai University, Shanghai, 200444, PR China

Received 26 July 2006; received in revised form 13 December 2006; accepted 22 January 2007

Abstract

In this work, combining the properties of the generalized super-memory gradient projection methods with the ideas of the strongly sub-feasible directions methods, we present a new algorithm with strong convergence for nonlinear inequality constrained optimization. At each iteration, the proposed algorithm can sufficiently use the information of the previous $t$ steps’ iterations to generate a new iterative point. Particularly, the intervals of parameters in the super-memory gradient projection direction are adjustable. The main properties of the new algorithm are described as follows: (i) the improving super-memory gradient projection direction is a combination of the generalized gradient projection and the $t$ steps’ super-memory gradients, which include both the previous $t$ steps’ search directions and gradients; moreover, only the gradients associated with a generalized active constrained set are dealt with rather than the gradients of all constraints; (ii) the initial point can be chosen arbitrarily, and at each iteration, the number of the functions satisfying the inequality constraints is nondecreasing. Especially, once a feasible iteration is obtained, then the subsequent iterations are also feasible; (iii) under suitable assumptions, it possesses global and strong convergence. Finally, some preliminary numerical results show that the proposed algorithm is promising.

Keywords: Inequality constrained optimization; Super-memory gradient method; Strongly sub-feasible directions; Global convergence; Strong convergence

1. Introduction

In this paper, we consider the nonlinear inequality constrained optimization problem as follows:

\begin{equation}
\begin{aligned}
(P) \quad & \min f(x) \\
& \text{s.t. } g_j(x) \leq 0, \quad j \in J = \{1, 2, \ldots, m\},
\end{aligned}
\end{equation}

∗Project supported by the National Natural Science Foundation (No. 10261001), Guangxi Science Foundation (Nos. 0236001, 0640001), Guangxi University Key Program for Science and Technology Research (No. 2005ZD02) and Guangxi University Research Foundation (No. X061057) of China.

∗ Corresponding author. Fax: +86 0771 3232084.
E-mail address: jianjb@gxu.edu.cn (J.-B. Jian).
where \( x \in \mathbb{R}^n \) and \( f(x), g_j(x) \ (j \in J) : \mathbb{R}^n \to \mathbb{R}^1 \) are all smooth functions. We denote the feasible set for the problem (P) by

\[
X = \{ x \in \mathbb{R}^n \mid g_j(x) \leq 0, \ j \in J \}.
\]

A given point \( x \) is said to be feasible if it satisfies all constraints of the problem (P), i.e., \( x \in X \).

The Gradient projection method (GPM for short), which was first developed by Rosen in 1960s [1,2], is one of the early important methods of feasible directions for solving the problem (P). Thereafter, many authors (see, e.g., [3–5]) further researched and improved the GPM, especially in the case of nonlinear constraints. In the recent two decades, a new kind of projection type method collectively called generalized gradient projection methods (GGPM) have been proposed, see, e.g., [6–8].

On the other hand, the super-memory gradient method, a generalization of the conjugate gradient method, is one of the most effective methods for solving unconstrained optimization problems, see, e.g., [9–16]. The super-memory gradient method can sufficiently use the information of the previous \( t \) steps’ iterations to generate the next iterative point. As proposed in Ref. [10], the next iteration is obtained by

\[
x^{k+1} = x^k + \Delta x^k \quad \text{with} \quad \Delta x^k = -\alpha^k p^k + \sum_{r=1}^t \beta^k_r \Delta x^{k-r},
\]

where \( p^k \) is a given vector such that \( (p^k)^T \nabla f(x^k) \neq 0 \) and parameters \( \alpha^k, \beta^k_r \ (r = 1, 2, \ldots, t) \) satisfying

\[
f_2 \left( x^k - \alpha^k p^k + \sum_{r=1}^t \beta^k_r \Delta x^{k-r} \right) = \min_{\alpha, \beta^1, \ldots, \beta^t} f_2 \left( x^k - \alpha p^k + \sum_{r=1}^t \beta^r \Delta x^{k-r} \right).
\]

Besides, in Ref. [12], Shi constructed the search direction \( d^k \) by

\[
d^k = \begin{cases} 
-\nabla f(x^k), & \text{if } k = 0; \\
-\nabla f(x^{k-1}) + \beta_k \nabla f(x^{k-1}), & \text{if } k \geq 1,
\end{cases}
\]

where \( \beta_k \in \{ (-\infty, b_k], \theta_k = 0; \ [-a_k, b_k], 0 < \theta_k < \pi; \ [-a_k, \infty), \theta_k = \pi, \} \)

and \( \theta_k \) is an angle between \( \nabla f(x^k) \) and \( \nabla f(x^{k-1}) \), \( a_k \) and \( b_k \) satisfy

\[
a_k (1 - \cos \theta_k) = \frac{\| \nabla f(x^k) \|}{\| \nabla f(x^{k-1}) \|}, \quad b_k (1 + \cos \theta_k) = \frac{\| \nabla f(x^{k-1}) \|}{\| \nabla f(x^{k-1}) \|}.
\]

To extend the super-memory gradient method for unconstrained optimization to constrained optimization, combining the GGPM, a kind of so-called generalized super-memory gradient projection method (GSM-GPM) was proposed, see, e.g., [15,16]. The method in Ref. [15] iterates in the feasible region \( X \), so it must begin with a feasible initial point. However, it is not usually easy to compute a feasible initial point. In order to overcome this shortcoming, Sun [16] further presented a GSM-GPM, which could start with an arbitrary initial point by using the generalized projection technique. The search direction \( d^k \) in Ref. [16] was yielded by

\[
d^k = \bar{S}_k - \rho(x^k) B(x^k)^T e,
\]

with

\[
\bar{S}_k = \bar{S}_k^2 + B(x^k)^T v(x^k), \quad S_k^2 = P(x^k) \left( -\nabla f(x^k) + \sum_{r=1}^2 \beta^k_r d^{k-r} \right),
\]

\[
\rho(x^k) = \frac{-\nabla f(x^k)^T \bar{S}_k^T + \psi(x^k)}{\sigma_1 \| a(x^k)^T e \| + \sigma_2}, \quad e = (1, 1, \ldots, 1)^T \in \mathbb{R}^m,
\]

where the parameters \( \sigma_1 > 1, \sigma_2 > 0 \) and the generalized projection matrix \( P(x^k) \) is computed by

\[
P(x^k) = E - N(x^k) B(x^k), \quad B(x^k) = (N(x^k)^T N(x^k) - H(x^k))^{-1} N(x^k)^T, \quad N(x^k) = (\nabla g_j(x^k), \ j \in J), \quad H(x^k) = \text{diag}(g_j(x^k) - \psi(x^k), \ j \in J), \quad \psi(x^k) = \max\{0; g_j(x^k), \ j \in J\}.
\]
In Ref. [16], the step-length \( \lambda_k \) is chosen as the first value of \( \lambda \) in the sequence \( \{1, \frac{1}{\rho}, \frac{1}{\rho^2}, \ldots \} \) (\( \rho > 1 \)) satisfying:

(a) in the case of \( \psi(x^k) = 0 \), i.e., \( x^k \in X \),
\[
\begin{align*}
  f(x^k + \lambda d^k) &\leq f(x^k) - \mu \lambda \nabla f(x^k)^T d^k, \\
  g_j(x^k + \lambda d^k) &\leq 0, \quad \forall j \in J.
\end{align*}
\]

(b) in the case of \( \psi(x^k) > 0 \), i.e., \( x^k \not\in X \), \( \psi(x^k + \lambda d^k) \leq \psi(x^k) + \mu \lambda D_d \psi(x^k) \), where \( D_d \psi(x^k) \) denotes the direction derivative of \( \psi(x^k) \).

We can see, in case (b) above, that the line search cannot effectively control the increase of the objective function \( f \).

For the case of an infeasible initial point, Jian [17] presented and researched a new class of methods called strongly sub-feasible directions methods (SSFDM). Further researches on SSFDM can be found in [18–21]. The main properties of SSFDM are shown below.

(i) the initial point can be chosen arbitrarily, and it can unify automatically the operations of initialization (Phase I) and optimization (Phase II).

(ii) it can guarantee that the inequality constraints, which hold at the previous iteration, also hold at the subsequent iterative points, that is, if \( g_j(x^k) \leq 0 \), then \( g_j(x^j) \leq 0 \) for all \( k \geq s \). So the number of the functions satisfying the inequality constraints is nondecreasing.

In this paper, based on the ideas of GSM-GPM and SSFDM, we present a new GSM-GPM–SSFDM for nonlinear programming problem (P). The main features of the proposed algorithm are summarized as follows:

- it possesses all the properties of SSFDM;
- it can sufficiently use the information of the previous \( t \) steps’ iterations, i.e., search directions \( d^{k-1}, d^{k-2}, \ldots, d^{k-t} \) and gradients \( \nabla f(x^{k-1}), \nabla f(x^{k-2}), \ldots, \nabla f(x^{k-t}) \), to generate the search direction;
- the intervals of the parameters in the super-memory gradient projection direction are symmetric and adjustable;
- besides the global convergence, strong convergence is obtained under some suitable assumptions.

This paper is organized as follows: in the next section, our algorithm is proposed and the relative properties are discussed in detail. In Sections 3 and 4, we analyze the global convergence and the strong convergence of the algorithm, respectively. Some preliminary numerical tests are reported in Section 5.

2. Description of algorithm

For convenience of presentation, for a given point \( x^k \in \mathbb{R}^n \), we introduce and use the following notations throughout this paper:

\[
J_k^- \triangleq J_-(x^k) = \{ j \in J \mid g_j(x^k) \leq 0 \}, \quad J_k^+ \triangleq J_+(x^k) = \{ j \in J \mid g_j(x^k) > 0 \},
\]

\[
\psi(x^k) = \max\{0; g_j(x^k), j \in J\} = \max\{0; g_j(x^k), j \in J_k^+\}.
\]  

(2.1)  
(2.2)

Furthermore, we introduce two approximate active constrained sets (called “working sets” too): the \( \delta_k \)-active constrained set \( I_-(x^k, \delta_k) \) and \( \varepsilon_k \)-active constrained set \( I_+(x^k, \varepsilon_k) \) as follows:

\[
I_k^- \triangleq I_-(x^k, \delta_k) = \{ j \in J_k^- \mid -\delta_k \leq g_j(x^k) \leq 0 \},
\]

\[
I_k^+ \triangleq I_+(x^k, \varepsilon_k) = \{ j \in J_k^+ \mid 0 \leq \psi(x^k) - g_j(x^k) \leq \varepsilon_k \},
\]  

(2.3)

where \( \delta_k \geq 0 \) and \( \varepsilon_k \geq 0 \). Moreover, we call set \( I(x^k, \varepsilon_k, \delta_k) \) defined by

\[
I_k \triangleq I(x^k, \varepsilon_k, \delta_k) = I_k^+ \cup I_k^-,
\]

(2.4)
a generalized \( (\varepsilon_k, \delta_k) \)-active constrained set at point \( x^k \).

In order to use the generalized projection technique to construct an improving direction, the following assumptions for the problem (P) are necessary.
Assumption A1. Functions \( f, g_j \ (j \in J) \) are all continuously differentiable.

Assumption A2. The gradient vectors \( \{\nabla g_j(x^k)\}, \ j \in I(x^k, 0, 0) \) are linearly independent for each point \( x^k \) in \( \mathbb{R}^n \).

Basing on the working sets above, we define the following matrices and vector:

\[
H_k \triangleq H(x^k) = \text{diag}(H_j(x^k), \ j \in I^k), \quad H_j(x^k) = \begin{cases} \psi(x^k) - g_j(x^k), & j \in I^k; \\ -g_j(x^k), & j \in I^c, \end{cases}
\]  

(2.5)

\[
N_k \triangleq N(x^k) = (\nabla g_j(x^k), \ j \in I^k), \quad B_k \triangleq B(x^k) = (N_k^T N_k + H_k)^{-1} N_k^T,
\]  

(2.6)

\[
u^k \triangleq u(x^k) = (u(x^k), \ j \in I^k)^T = -B_k \nabla f(x^k), \quad P_k \triangleq P(x^k) = E - N_k B_k,
\]  

(2.7)

where \( E \) is an \( n \)-rank unit matrix, and \( P_k \) is called a generalized projection matrix at point \( x^k \).

To show that the formulas above are well defined, we present the following lemma, and its proof can be finished by the definitions of the related vectors and matrices (it also can be seen in Lemma 1 and formulas (11) and (12) of Ref. [20]), so we omit it here.

Lemma 2.1. Suppose that Assumptions A1 and A2 hold and let \( x^k \in \mathbb{R}^n \). If a diagonal matrix \( H = \text{diag}(H_j, \ j \in I^k) \) satisfying \( H_j \geq 0, j \in I^k \) and \( H_j > 0, j \in I^k \setminus I(x^k, 0, 0) \), then the matrix \((N_k^T N_k + H_k)^{-1}\) is positive definite, and so is the matrix \((N_k^T N_k + H_k)^{-1}\). Furthermore, the following equalities hold

\[
z^T P_k \tilde{z} = \|P_k \tilde{z}\|^2 + \sum_{j \in I^k} H_j(x^k) \tilde{z}_j^2, \quad \tilde{z} = B_k \tilde{z}, \ \forall \tilde{z} \in \mathbb{R}^n.
\]  

(2.8)

\[
N_k^T B_k^T = E - H_k(N_k^T N_k + H_k)^{-1}, \quad N_k^T P_k = H_k B_k.
\]  

(2.9)

Suppose \( x^k \in \mathbb{R}^n \) is a given iteration. Based on the idea of super-memory gradient projection method, we introduce a \( t \) steps’ super-memory gradient projection direction \( S_t^k \) by

\[
S_t^k = P_k \left( -\nabla f(x^k) + \sum_{r=1}^t \left( \alpha_k^r \nabla f(x^{k-r}) + \beta_k^r d^{k-r} \right) \right),
\]  

(2.10)

where \( t \) is a given positive integer, and \( d^{k-r} \) is the search direction associated with the iteration \( x^{k-r} \). If \(-t \leq k - r \leq -1\), we always assume that \( \nabla f(x^{k-r}) = 0, d^{k-r} = 0 \) and the associated \( \alpha_k^r, \beta_k^r \) are arbitrarily constants. Denote sets

\[
T = \{1, 2, \ldots, t\}, \quad T_{\alpha}^k = \{r \in T \mid \nabla f(x^{k-r}) \neq 0\}, \quad T_{\beta}^k = \{r \in T \mid d^{k-r} \neq 0\}.
\]  

(2.11)

In order to obtain the “descent” property of the direction \( S_t^k \), we expect to choose the intervals of parameters \( \alpha_k^r, \beta_k^r (r \in T) \) such that

\[
\nabla f(x^k)^T S_t^k \leq -\gamma \|P_k \nabla f(x^k)\|^2,
\]  

(2.12)

where \( \gamma > 0 \) is a constant to be estimated.

Now, we analyze the intervals of \( \alpha_k^r, \beta_k^r \), which can ensure relationship (2.12) holds.

First, combining with (2.10) and (2.8), one has

\[
\nabla f(x^k)^T S_t^k = -\nabla f(x^k)^T P_k \nabla f(x^k) + \sum_{r=1}^t \left( \alpha_k^r \nabla f(x^{k-r})^T P_k \nabla f(x^{k-r}) + \beta_k^r \nabla f(x^{k-r})^T P_k d^{k-r} \right)
\leq -\|P_k \nabla f(x^k)\|^2 + \sum_{r=1}^t |\alpha_k^r| \cdot \|P_k \nabla f(x^k)\| \cdot \|\nabla f(x^{k-r})\| + \sum_{r=1}^t |\beta_k^r| \cdot \|P_k \nabla f(x^k)\| \cdot \|d^{k-r}\|
\leq -\|P_k \nabla f(x^k)\|^2 + \|P_k \nabla f(x^k)\| \cdot \sum_{r=1}^t \left( |\alpha_k^r| \cdot \|\nabla f(x^{k-r})\| + |\beta_k^r| \cdot \|d^{k-r}\| \right).
\]
Therefore, we can get a sufficient condition for (2.12) as follows:

\[-\|P_k \nabla f(x^k)\|^2 + \|P_k \nabla f(x^k)\| \cdot \sum_{r=1}^t \left( |\alpha'_k| \cdot \|\nabla f(x^{k-r})\| + |\beta'_k| \cdot \|d^{k-r}\| \right) \leq -\gamma \|P_k \nabla f(x^k)\|^2.\]

The inequality above can be rewritten as

\[\sum_{r \in T^k_d} |\alpha'_k| \cdot \|\nabla f(x^{k-r})\| + \sum_{r \in T^k_b} |\beta'_k| \cdot \|d^{k-r}\| \leq (1 - \gamma) \cdot \|P_k \nabla f(x^k)\| \]

\[= \sum_{r \in T^k_d} \left( \frac{(1 - \gamma)\epsilon}{|T^k_d|} \cdot \frac{\|P_k \nabla f(x^k)\|}{\|\nabla f(x^{k-r})\|} \cdot \|\nabla f(x^{k-r})\| \right) \]

\[+ \sum_{r \in T^k_b} \left( \frac{(1 - \gamma)(1 - \epsilon)}{|T^k_b|} \cdot \frac{\|P_k \nabla f(x^k)\|}{\|d^{k-r}\|} \cdot \|d^{k-r}\| \right), \tag{2.13}\]

where parameter \(\epsilon \in [0, 1]\), \(|T^k_d|\) and \(|T^k_b|\) denote the numbers of elements of the sets \(T^k_d\) and \(T^k_b\), respectively. Hence, formula \(2.13\) leads us to choose \(\gamma \in (0, 1)\) and the intervals of parameters \(\alpha'_k, \beta'_k\) by

\[
|\alpha'_k| \begin{cases} \leq \frac{(1 - \gamma)\epsilon}{|T^k_d|} \cdot \|P_k \nabla f(x^k)\| / \|\nabla f(x^{k-r})\|, & \text{if } r \in T^k_d, \\ 0, & \text{if } r \in T \setminus T^k_d, \end{cases} \tag{2.14}
\]

\[
|\beta'_k| \begin{cases} \leq \frac{(1 - \gamma)(1 - \epsilon)}{|T^k_b|} \cdot \|P_k \nabla f(x^k)\| / \|d^{k-r}\|, & \text{if } r \in T^k_b, \\ 0, & \text{if } r \in T \setminus T^k_b. \end{cases}
\]

Certainly, \(\alpha'_k (r \in T \setminus T^k_d)\) and \(\beta'_k (r \in T \setminus T^k_b)\) also can be chosen as arbitrary constants.

From the discussion above, we can obtain the “descent” property of the direction \(S^k_t\), as shown in the lemma below.

**Lemma 2.2.** Suppose that \(x^k\) is a current iterative point, parameters \(\gamma \in (0, 1)\), \(\epsilon \in [0, 1]\) and \(\alpha^k\) as well as \(\beta^k\) satisfy \(2.14\). Then

\[
\nabla f(x^k)^T S^k_t \leq -\gamma \|P_k \nabla f(x^k)\|^2. \tag{2.15}
\]

With regard to the boundary of \(S^k_t\), we have the following lemma.

**Lemma 2.3.** Under the assumptions of **Lemma 2.2**, it follows that

\[
\|S^k_t\| \leq (2 - \gamma) \|P_k \nabla f(x^k)\|, \tag{2.16}
\]

\[
\left\| -\nabla f(x^k) + \sum_{r=1}^t \left( \alpha'_k \nabla f(x^{k-r}) + \beta'_k d^{k-r} \right) \right\| \leq \|\nabla f(x^k)\| + (1 - \gamma) \|P_k \nabla f(x^k)\|. \tag{2.17}
\]

**Proof.** At first, we show \(\|S^k_t\| \leq (2 - \gamma) \|P_k \nabla f(x^k)\|\). One has, from \(2.10\) and \(2.14\),

\[
\|S^k_t\| = \left\| P_k \left( -\nabla f(x^k) + \sum_{r=1}^t \left( \alpha'_k \nabla f(x^{k-r}) + \beta'_k d^{k-r} \right) \right) \right\| \leq \|P_k \nabla f(x^k)\| + \sum_{r=1}^t |\alpha'_k| \cdot \|P_k \nabla f(x^{k-r})\| + \sum_{r=1}^t |\beta'_k| \cdot \|P_k d^{k-r}\| \leq \|P_k \nabla f(x^k)\| + \sum_{r \in T^k_d} \left( \frac{(1 - \gamma)\epsilon}{|T^k_d|} \cdot \frac{\|P_k \nabla f(x^k)\|}{\|\nabla f(x^{k-r})\|} \cdot \|\nabla f(x^{k-r})\| \right).
\]
and

Prove the sufficient condition

Hence

Proof. Theorem 2.1.

Finally, from (2.14), one has

Finally, (2.14), one has

\[
\| - \nabla f(x^k) + \sum_{r=1}^{r} (\alpha_k \nabla f(x^{k-r}) + \beta_k d^{k-r}) \| \leq \| \nabla f(x^k) \| + \sum_{r \in T_\beta} (1 - \gamma) \cdot \| \nabla f(x^{k-r}) \| \cdot \| d^{k-r} \|
\]

\[
= \| \nabla f(x^k) \| + (1 - \gamma) \| P_k \nabla f(x^k) \|. \quad \square
\]

To judge whether the current iterative point \( x^k \) is a KKT point or not, we introduce an optimal controlling function \( \rho(x^k) \) by:

\[
\rho(x^k) = \frac{\gamma \| P_k \nabla f(x^k) \|^2 + \omega(x^k) + \psi(x^k)}{1 + \| u(x^k)^T e^k \|}, \quad (2.18)
\]

where

\[
\omega(x^k) = \max_{j \in I_k} \{ -u_j(x^k), u_j(x^k) H_j(x^k) \}, \quad e^k = (1, \ldots, 1)^T \in R^{|I_k|}. \quad (2.19)
\]

From the definition of \( \omega(x^k) \), we can conclude that \( \omega(x^k) \geq 0 \) and \( \omega(x^k) \neq 0 \) if and only if \( u_j(x^k) = 0 \) or \( u_j(x^k) H_j(x^k) = 0 \) for each \( j \in I_k \). Combining with \( \psi(x^k) \geq 0 \) as well as (2.18), we get \( \rho(x^k) \geq 0 \) and \( \rho(x^k) = 0 \) if and only if \( P_k \nabla f(x^k) = 0 \), \( \omega(x^k) = 0 \) and \( \psi(x^k) = 0 \).

Theorem 2.1. The current iterative point \( x^k \) is a KKT point of the problem (P) if and only if \( \rho(x^k) = 0 \).

Proof. Prove the sufficient condition. Suppose that \( \rho(x^k) = 0 \). Then \( P_k \nabla f(x^k) = 0 \), \( \omega(x^k) = 0 \) and \( \psi(x^k) = 0 \). From \( \psi(x^k) = 0 \), we know \( g_j(x^k) \leq 0 \), \( j \in J \), that is, \( x^k \in X \). Combining (2.7) as well as \( P_k \nabla f(x^k) = 0 \), one has

\[
0 = P_k \nabla f(x^k) = (E - N_k B_k) \nabla f(x^k) = \nabla f(x^k) + N_k u^k.
\]

On the other hand, from (2.9), (2.7) and \( P_k \nabla f(x^k) = 0 \), we get

\[
0 = N_k^T P_k \nabla f(x^k) = H_k B_k \nabla f(x^k) = -H_k u^k.
\]

Hence \( -H_j(x^k) u_j(x^k) = g_j(x^k) u_j(x^k) = 0 \), \( j \in I_k^c \). Let \( u_j(x^k) = 0 \), \( j \in J \setminus I_k^c \), then \( g_j(x^k) u_j(x^k) = 0 \), \( j \in J \).

Furthermore, we suppose by contradiction that there exists \( j_0 \in I_k^c \) such that \( u_{j_0}(x^k) < 0 \). Then, from the definition of \( \omega(x^k) \) and \( u_j(x^k) H_j(x^k) = 0 \), \( j \in I_k^c \), we have \( \omega(x^k) > 0 \). It is contrary to \( \omega(x^k) = 0 \). So, \( u_j(x^k) \geq 0 \), \( j \in J \).
Theorem 2.1 and (2.19) is completed. From (2.19), we have, from (2.9) of Lemma 2.1, that

$$P_k \nabla f(x^k) = -P_k N_k \sigma^k = -B_k^T H_k \sigma^k = 0,$$

and $\rho(x^k) = 0$ by (2.18). The whole proof of Theorem 2.1 is completed. \Box

When $x^k$ is not a KKT point, we need to find an improving search direction at $x^k$. First, we consider whether $S^k$ is a suitable one. From (2.15), it follows that \( \nabla f(x^k)^T S^k = 0 \) when $P_k \nabla f(x^k) = 0$. So, in this case, we cannot decide whether $S^k$ is a descent direction or not. Furthermore, even if $S^k$ is a descent direction, it is not necessarily “feasible” because

$$\nabla g_j(x^k)^T S^k = 0, \quad j \in I(x^k, 0, 0)$$

follows from

$$N_k^T S^k = H_k B_k \left( -\nabla f(x^k) + \sum_{r=1}^l (\alpha^r_k \nabla f(x^{k-r}) + \beta^r_k d^{k-r}) \right)$$

and $H_j(x^k) = 0$ for $j \in I(x^k, 0, 0)$. Therefore, based on Theorem 2.1 and the properties (2.15) and (2.20) of $S^k$, in order to yield an improving direction, we update the direction $S^k$ as follows:

$$d^k = \rho(x^k)^{\xi} (S^k + B_k^T v^k),$$

where $\xi > 0$ is a constant and

$$v^k = (v_j(x^k), j \in I^k), \quad v_j(x^k) = \begin{cases} -1 - \rho(x^k), & \text{if } u_j(x^k) < 0, \ j \in I^k; \\ H_j(x^k) - \rho(x^k), & \text{if } u_j(x^k) \geq 0, \ j \in I^k. \end{cases}$$

The following lemma will show that $d^k$ given by (2.22) is indeed an improving direction.

Lemma 2.4. For the iterative point $x^k$, one has

$$\nabla f(x^k)^T d^k \leq \rho(x^k)^{\xi+1} \psi(x^k) - \rho(x^k)^{\xi+1}, \quad \nabla g_j(x^k)^T d^k \leq -\rho(x^k)^{\xi+1}, \quad j \in I(x^k, 0, 0).$$

Proof. From (2.22), (2.15), (2.7) and (2.23), we have

$$\nabla f(x^k)^T d^k = \rho(x^k)^{\xi} \nabla f(x^k)^T (S^k + B_k^T v^k) \leq \rho(x^k)^{\xi} \left( -\gamma \| P_k \nabla f(x^k) \|^2 - u(x^k)^T v^k \right)$$

$$= \rho(x^k)^{\xi} \left[ -\gamma \| P_k \nabla f(x^k) \|^2 + \sum_{j \in I^k, u_j(x^k) < 0} u_j(x^k) (1 + \rho(x^k)) \right.$$}

$$-\sum_{j \in I^k, u_j(x^k) > 0} u_j(x^k) (H_j(x^k) - \rho(x^k)) \right].$$
This inequality, together with (2.19) and (2.18), further shows that
\[
\nabla f(x^k)^T d^k \leq \rho(x^k) \left[ -\gamma \| P_k \nabla f(x^k) \|^2 + \rho(x^k) \sum_{j \in I^k} u_j(x^k) \right. \\
\left. - \left( \sum_{j \in I^k, u_j(x^k) < 0} (-u_j(x^k)) + \sum_{j \in I^k, u_j(x^k) \geq 0} u_j(x^k)H_j(x^k) \right) \right].
\]

Noting that $H_j(x^k) = 0$ for $j \in I(x^k, 0, 0)$, from (2.25) and (2.23), one has
\[
\nabla g_j(x^k)^T d^k = \rho(x^k)^\xi v_j(x^k) \leq -\rho(x^k)^\xi + 1, \quad j \in I(x^k, 0, 0).
\]

Then the whole proof is completed. $\square$

**Remark 1.** (1) If $x^k \notin X$, then $\psi(x^k) > 0$ and $\rho(x^k) > 0$ from (2.2) and (2.18). Taking account of (2.24), we know that $d^k$ is a feasible direction of set $A = \{ y \mid g_j(y) \leq 0, j \in J_-(x^k) \}$ at point $x^k \in A$, and the increasing speed of the objective function $f$ along $d^k$ at $x^k$ is restricted by the positive term $\rho(x^k)^\xi \psi(x^k)$.

(2) If $x^k \in X$ but is not a KKT point of the problem (P), then $\psi(x^k) = 0$ and $\rho(x^k) > 0$ from (2.2), (2.18) and Theorem 2.1. From (2.24), we get $\nabla f(x^k)^T d^k \leq -\rho(x^k)^\xi + 1 < 0$. This shows that $d^k$ is a descent direction. On the other hand, taking account of (2.24), one has $\nabla g_j(x^k)^T d^k \leq -\rho(x^k)^\xi + 1 < 0$, $j \in I_-(x^k, 0)$. That is, $d^k$ is a feasible direction. Therefore, $d^k$ is a feasible descent direction.

For any $x \in R^n \setminus X$ and vector $d$, $D_d \psi(x) = \lim_{\lambda \to 0^+} \frac{\psi(x + \lambda d) - \psi(x)}{\lambda}$ means the direction derivative of $\psi(x)$. It is not difficult to get
\[
D_d \psi(x) = \max_{j \in I_+(x, 0)} \{ \nabla g_j(x)^T d \},
\]
and this together with (2.24) shows that
\[
D_d^k \psi(x^k) = \max_{j \in I_+(x^k, 0)} \{ \nabla g_j(x^k)^T d^k \} \leq -\rho(x^k)^\xi + 1 < 0, \quad \text{if } x^k \notin X.
\]

Therefore, it is easy to show that
\[
\psi(x^k + \lambda d^k) \leq \psi(x^k) - \mu \lambda \rho(x^k)^\xi + 1, \quad \text{if } x^k \notin X
\]
holds for $\lambda > 0$ small enough and $\mu \in (0, 1)$.

Based on the discussion above, now we can give the details of our algorithm as follows.

**Algorithm**

*Parameters: $\gamma, \mu, \epsilon \in (0, 1), \epsilon \in [0, 1], \mu > 1 - \epsilon, \xi > 0, t > 1.$*

*Data: $x^0 \in R^n, d^{-1} = d^{-2} = \cdots = d^{-t} = 0, \nabla f(x^{-1}) = \nabla f(x^{-2}) = \cdots = \nabla f(x^{-t}) = 0.$*

**Step 0. (Initialization)** Let $k := 0.$

**Step 1.** For the iteration $x^k$, compute $J^k, J^k_+, \psi(x^k)$ by (2.1) and (2.2), choose parameters $\epsilon_k > 0, \delta_k > 0$, and yield index sets $I^k, I^k_+$ and $I^k$ by (2.3) and (2.4).
Step 2. Compute $H_k$, $N_k$, $B_k$, $d_k$, $P_k$, $\omega(x^k)$ and $\rho(x^k)$ by (2.5)–(2.7), (2.19) and (2.18). If $\rho(x^k) = 0$, then $x^k$ is a KKT point of the problem (P), stop; otherwise, go to Step 3.

Step 3. (Generate search direction) Determine $\alpha_k^j$ and $\beta_k^j$ according to (2.14). Compute $S_k^T$, $v^k$ and $d_k^k$ by (2.10), (2.23) and (2.22).

Step 4. (Perform line search) Compute the step size $\lambda_k$, the first number $\lambda$ of the sequence $\{1, \varepsilon^2, \varepsilon^3, \ldots\}$ satisfying:

$$
\begin{align*}
\sigma_k f(x^k + \lambda d_k^k) &\leq \sigma_k \left( f(x^k) + \mu \lambda \nabla f(x^k)^T d_k^k + \bar{\mu} \lambda \rho(x^k)^{\xi}\psi(x^k) \right), \\
g_j(x^k + \lambda d_k^k) &\leq \psi(x^k) - \mu \lambda \rho(x^k)^{\xi+1}, \quad \forall j \in J_k^*, \\
g_j(x^k + \lambda d_k^k) &\leq 0, \quad \forall j \in J_k^\alpha,
\end{align*}
$$

where parameter $\sigma_k \begin{cases} 1, \text{if } x^k \in X; \\ 0, \text{if } x^k \notin X. \end{cases}$

Step 5. (Updates) Set $x^{k+1} = x^k + \lambda_k d_k$ and $k := k + 1$, go back to Step 1.

Remark 2. For $k \geq t$, from Step 2 of the algorithm, if we have deduced that $x^k$ is not a KKT point, this means that $x^l$ ($l \leq k$) are all not either, and thus $x^{k-r}$ ($r \in T$) are not KKT points. So from Theorem 2.1 and formula (2.22) of $d_k$, we know $d^{k-r} \neq 0$. Combining with the definitions of $T_\beta^k$ and $|T_\beta^k|$, it is obvious that if $x^k$ is not a KKT point, then $d^{k-r} \neq 0$ and $|T_\beta^k| = t$.

Remark 3. The constraint (2.29) at Step 4 shows that the objective value $f$ may not be monotone when $x^k \notin X$, and it is monotone decreasing whenever $x^k \in X$. Furthermore, in the case of the iteration $x^k$ being infeasible, if one sets $\sigma_k = 0$, then the search condition (2.29) automatically holds; that is, this inequality can be gotten rid of, and the computation cost is further reduced, and if one sets $\sigma_k > 0$, the search condition (2.29) can restrict the increase of the objective function $f$.

Remark 4. The constraint (2.31) at Step 4 guarantees that the inequality constraints, which hold at the previous iteration, also hold at the subsequent iterative points; that is, if $g_j(x^k) \leq 0$, then $g_j(x^k) \leq 0$ for all $k \geq s$. So the number of the functions satisfying the inequality constraints is nondecreasing. Hence, once the iterative point $x^s$ is feasible, then the subsequent iterations $x^k (k \geq s)$ are all feasible.

In order to show that the algorithm is well defined, we introduce the following lemma.

Lemma 2.5. The line search at Step 4 can be carried out, that is, the inequalities (2.29)–(2.31) hold for $\lambda > 0$ sufficiently small.

Proof. In view of Step 2 and (2.18), we know that $\rho(x^k) > 0$. Now we prove that the inequalities (2.29)–(2.31) hold for $\lambda > 0$ small enough.

Analyze the inequality (2.29): using Taylor expansion, combining with (2.24), one has

$$
\begin{align*}
\sigma_k \left( f(x^k + \lambda d_k^k) - f(x^k) - \mu \lambda \nabla f(x^k)^T d_k^k - \bar{\mu} \lambda \rho(x^k)^{\xi}\psi(x^k) \right) \\
= \sigma_k \left( (1 - \mu)\lambda \nabla f(x^k)^T d_k^k - \bar{\mu} \lambda \rho(x^k)^{\xi}\psi(x^k) + o(\lambda) \right) \\
\leq \sigma_k \left( (1 - \mu)\lambda \rho(x^k)^{\xi}\psi(x^k) - \rho(x^k)^{\xi+1} - \bar{\mu} \lambda \rho(x^k)^{\xi}\psi(x^k) + o(\lambda) \right) \\
= \sigma_k \left( (1 - \mu - \bar{\mu}) \lambda \rho(x^k)^{\xi}\psi(x^k) - (1 - \mu) \lambda \rho(x^k)^{\xi+1} + o(\lambda) \right),
\end{align*}
$$

Therefore, in view of $(1 - \mu - \bar{\mu}) \leq 0$, the inequality (2.29) holds for $\lambda > 0$ small enough.

Analyze the inequality (2.30): from (2.2) and (2.28), it is obvious that

$$
g_j(x^k + \lambda d_k^k) \leq \psi(x^k + \lambda d_k^k) \leq \psi(x^k) - \mu \lambda \rho(x^k)^{\xi+1}, \quad j \in J_k^\alpha
$$

holds for $\lambda > 0$ small enough.
Analyse the inequality (2.31): using Taylor expansion, one has
\[
g_j(x^k + \lambda d^k) = g_j(x^k) + \lambda \nabla g_j(x^k)^T d^k + o(\lambda). \tag{2.33}
\]
(i) if \( j \in J^k_h \) and \( g_j(x^k) = 0 \), from (2.24), we obtain that \( \nabla g_j(x^k)^T d^k \leq -\rho(x^k)^{\xi+1} < 0 \). So, by (2.33), one has
\[
g_j(x^k + \lambda d^k) \leq -\lambda \rho(x^k)^{\xi+1} + o(\lambda) \leq 0.
\]
(ii) if \( j \in J^k_h \) and \( g_j(x^k) < 0 \), by (2.33), one has
\[
g_j(x^k + \lambda d^k) = g_j(x^k) + O(\lambda) \leq 0.
\]
Therefore, the inequality (2.31) holds for \( \lambda > 0 \) small enough.

Summarizing the analysis above, we can conclude that the inequalities (2.29)–(2.31) hold for \( \lambda > 0 \) small enough. In other words, the algorithm is well defined. \( \square \)

From the procedure of the proposed algorithm, one of the following two cases must occur.

Case I: there exists an iteration index \( s \) such that \( \psi(x^s) = 0 \). So \( \psi(x^k) = 0 \) holds as well for all \( k \geq s \);

Case II: \( \psi(x^k) > 0 \) and \( \psi(x^{k+1}) < \psi(x^k) \) for any \( k = 0, 1, 2, \ldots \).

3. Global convergence analysis

In this section, we analyze the global convergence of the proposed algorithm. From Theorem 2.1 and Step 3 of our algorithm, we know that if the algorithm stops at a point \( x^k \), then \( x^k \) is a KKT point. Thus, in the following discussion, we always assume that the algorithm generates an infinite iterative sequence \( \{x^k\} \) of points, and our goal is to prove that any accumulation point \( x^* \) of \( \{x^k\} \) is a KKT point of the problem (P) under some suitable conditions. For this purpose, the following constraint on the parameters \( \varepsilon_k \) and \( \delta_k \) is necessary.

Assumption A3. There exist constants \( \delta > 0 \) and \( \varepsilon > 0 \), such that \( \delta_k \geq \delta \) and \( \varepsilon_k \geq \varepsilon \) for \( k \) large enough.

Remark 5. Here, we give two simple choices of \( \delta_k \) and \( \varepsilon_k \) as follows, and anyone can satisfy Assumption A3:

1. let \( \varepsilon_k \equiv \varepsilon \) and \( \delta_k \equiv \delta, k = 0, 1, 2, \ldots \), where \( \varepsilon \) and \( \delta \) are sufficiently small positive constants;

2. let \( \varepsilon_k = \varepsilon + \max\{0; \psi(x^k) - g_j(x^k), j \in J^k_h \} \) and \( \delta_k = \delta + \max\{0; -g_j(x^k), j \in J^k_h \}, k = 0, 1, 2, \ldots \). In this case, \( I^k_k \equiv J^k_h, J^k_k \equiv J^k_h, I^k_k \equiv J^k \).

In the rest of this section, we suppose that \( x^k \) is a given accumulation point of \( \{x^k\} \). More specifically, we assume \( x^k \rightarrow x^*, k \in K_0 \). Noting that \( J^k_h, J^k_k, I^k_h, I^k_k \) are all subsets of the fixed and finite set \( J \), we can first choose an infinite index subset \( K_1 \subseteq K_0 \) in which the sets \( J^k_h, k \in K_1 \) can be fixed; i.e., there exists a subset \( J_\infty \) (independent of \( k \)) of \( J \) such that \( J^k_h \equiv J_\infty \) for all \( k \in K_1 \). Subsequently, we can find another infinite index subset \( K_2 \subseteq K_1 \) such that \( J^k_h, k \in K_2 \) can also be fixed, say \( J^k_h \equiv J_h, k \in K_2 \); also similarly we can find a \( K_3 \subseteq K_2 \) such that \( I^k_k \equiv I_k, k \in K_3 \). Finally, we are able to find an infinite subset \( K \subseteq K_3 \subseteq K_2 \subseteq K_1 \subseteq K_0 \) such that
\[
J^k_k \equiv J_\infty, \quad J^k_h \equiv J_h, \quad I^k_k \equiv I_k, \quad I^k_h \equiv I_h, \quad J^k \equiv I \triangleq I_+ \cup I_-, \quad x^k \rightarrow x^*, \quad k \in K. \tag{3.1}
\]

Similarly to the definitions of \( N_k \) and \( H_k \), we define \( N_* \) and \( H^* \) by
\[
N_* = (\nabla g_j(x^*), j \in I), \quad H^* = \text{diag}(H^*_j, j \in I), \quad H^*_j = \begin{cases} \psi(x^*) - g_j(x^*), & j \in I_+; \\ -g_j(x^*), & j \in I_- \end{cases}. \tag{3.2}
\]

Then, under Assumption A1, we have
\[
\{\psi(x^k), H_k, N_k\} \rightarrow \{\psi(x^*), H^*, N_*\}, \quad k \in K.
\]

Lemma 3.1. Suppose that Assumptions A1 and A2 hold. Then \( H^*_j \geq 0 \) for \( j \in I \) and \( H^*_j > 0 \) for \( j \in I \setminus I(x^*, 0, 0) \), so the matrix \((N^*_1 N_* + H^*)\) is positive definite.
Proof. Obviously, \( H_j^* \geq 0 \) for \( j \in I \). If \( j \in I_+ \setminus I(x^*, 0, 0) \), then \( g_j(x^k) > 0 \), \( g_j(x^*) \neq 0 \) and \( \psi(x^*) - g_j(x^*) \neq 0 \). Hence \( g_j(x^k) \xrightarrow{k \in K} g_j(x^*) > 0 \); thus \( H_j^* = \psi(x^*) - g_j(x^*) > 0 \). If \( j \in I_- \setminus I(x^*, 0, 0) \), then \( g_j(x^k) \leq 0 \), \( g_j(x^*) \neq 0 \). Hence \( g_j(x^k) \xrightarrow{k \in K} g_j(x^*) < 0 \); thus \( H_j^* = -g_j(x^*) > 0 \). So, the proof of the first part is finished. Further, the positive definite property of matrix \( (N_s^TN_s + H^*) \) follows immediately from Lemma 2.1. □

Based on Lemma 3.1, similarly to (2.6), (2.7), (2.19) and (2.18), we can define \( B_s, u^*, P_s, \omega_s \) and \( \rho_s \) as follows:

\[
B_s = (N_s^TN_s + H^*)^{-1}N_s^T, \quad u^* = (u^*_j, j \in I) = -B_s\nabla f(x^*), \quad P_s = E - N_sB_s, \tag{3.3}
\]

\[
\omega^* = \sum_{j \in J} \max\{-u_j^*, u_j^*H_j^*\}, \quad e = (1, \ldots, 1)^T \in \mathbb{R}^{|I|}, \tag{3.4}
\]

\[
\rho_s = \frac{\gamma \|P_s\nabla f(x^*)\|^2 + \omega^* + \psi(x^*)}{1 + |(u^*)^Te|}. \tag{3.5}
\]

Therefore,

\[
\{B_k, u^k, P_k, \omega(x^k), \rho(x^k)\} \to \{B_s, u^*, P_s, \omega_s, \rho_s\}, \quad k \to K.
\]

In view of the definition of \( u^k \), we know that sequence \( \{u^k\}_K \) is bounded, and that there exists an infinite index set \( K_1 \subseteq K \), such that \( \{u^k\} \xrightarrow{k \in K_1} u^* = (u^*_j, j \in I) \). On the other hand, from Lemma 2.3, we know that the sequences \( \{S^k\}_K \) and \( \{-\nabla f(x^k) + \sum_{r=1}^I (\alpha^r_k \nabla f(x^{k-r}) + \beta^r_k d^{k-r})\}_K \) are bounded. So, we can take a subset \( K_2 \subseteq K_1 \) such that

\[
\begin{align*}
S^k \to S^*, & \quad -\nabla f(x^k) + \sum_{r=1}^I (\alpha^r_k \nabla f(x^{k-r}) + \beta^r_k d^{k-r}) \to \Delta^*, \\
d^k \to \rho_s \xi S^*_i + B^Tv^* \triangleq d^*, & \quad k \to K_2 \subseteq K.
\end{align*}
\]

Lemma 3.2. If \( x^* \) is not a KKT point of the problem \((P)\), then \( \rho_s > 0 \), and yet \( \rho(x^k) \geq \frac{1}{2} \rho_s \) for \( k \to K_2 \) large enough.

Proof. Firstly, if \( \psi(x^*) = 0 \), i.e., \( x^* \in X \), then \( H^* = H(x^*) \) and \( \rho_s = \rho(x^*) \). Hence, \( \rho_s > 0 \) by Theorem 2.1. Secondly, if \( \psi(x^*) > 0 \), i.e., \( x^* \notin X \), then \( \rho_s > 0 \) follows from (3.5). Therefore, \( \rho_s > 0 \) and \( \rho(x^k) \geq \frac{1}{2} \rho_s \) for \( k \to K_2 \) large enough, since \( \rho(x^k) \xrightarrow{k \in K_2} \rho_s \). □

Lemma 3.3. If \( x^* \) is not a KKT point of the problem \((P)\), then there exists a constant \( \bar{\lambda} > 0 \) such that \( \lambda_k \geq \bar{\lambda} \) for \( k \to K_2 \) large enough.

Proof. Since \( x^* \) is not a KKT point of the problem \((P)\), from Lemma 3.2, we know that \( \rho_s > 0 \) and \( \rho(x^k) \geq \frac{1}{2} \rho_s \) for \( k \to K_2 \) large enough. Now we prove that the line search inequalities (2.29)–(2.31) hold for \( \lambda > 0 \) small enough and all \( k \to K_2 \) large enough.

Analyze the inequality (2.29): taking account of \( d^k \xrightarrow{k \in K_2} d^* \), similarly to the proof of Lemma 2.5, using (2.32) and Lemma 3.2, one gets

\[
\begin{align*}
\sigma_k \left( f(x^k + \lambda d^k) - f(x^k) - \mu \lambda \nabla f(x^k)^T d^k - \bar{\mu} \lambda \rho(x^k) \psi(x^k) \right) & \\
\leq \sigma_k \left( (1 - \mu - \bar{\mu}) \lambda \rho(x^k) \psi(x^k) - (1 - \mu) \lambda \rho(x^k) \psi(x^k) + o(\lambda) \right) & \\
\leq \sigma_k \left( \frac{1}{2\lambda} (1 - \mu) \lambda \rho_s (\lambda) + o(\lambda) \right),
\end{align*}
\]

This implies that the inequality (2.29) holds for \( k \to K_2 \) large enough and \( \lambda > 0 \) small enough.

Analyze the inequality (2.30): two cases are considered.

Case A: For \( j \in I_+ \) and \( g_j(x^*) < \psi(x^*) \), we have

\[
\lim_{k \to K_2} \left( \frac{g_j(x^k) - \psi(x^k)}{x^k} \right) = g_j(x^*) - \psi(x^*) < 0,
\]
so, \(g_j(x^k) - \psi(x^k) \leq \frac{1}{2}(g_j(x^*) - \psi(x^*)) < 0\) for \(k \in K_2\) large enough. Thus,

\[
\lim_{k \to 0} \left( g_j(x^k + \lambda d^k) - \psi(x^k) + \mu \lambda \rho(x^k)^{\xi+1} \right) = g_j(x^k) - \psi(x^k) \leq \frac{1}{2}(g_j(x^*) - \psi(x^*)) < 0.
\]

Therefore, the inequality (2.30) holds for \(k \in K_2\) large enough and \(\lambda > 0\) small enough.

**Case B:** For \(j \in I_-\) and \(g_j(x^*) = \psi(x^*)\), by contradiction, suppose \(j \not\in I_+\). In view of \(\psi(x^*) - g_j(x^*) > \epsilon \geq \epsilon > 0\), we know \(\psi(x^*) - g_j(x^*) \geq \epsilon > 0\), this contradicts \(\psi(x^*) - g_j(x^*) = 0\). So we can conclude \(j \in I_+\). Furthermore, passing to the limit in (2.25) for \(k \to \infty\), one has

\[
N_*^T d^* = \rho_*^\xi \left( H^* B_* \Delta^* + \left( E - H^* (N_*^T N_* + H^*)^{-1} \right) v^* \right).
\]  

(3.7)

Noting that \(H_j^* = \psi(x^*) - g_j(x^*) = 0\) since \(j \in I_+\), we get from (3.7) and (2.24)

\[
\nabla g_j(x^*)^T d^* = \rho_*^\xi v_j^* = \rho_*^\xi \lim_{k \to K_2} v_j(x^k)
\]

\[
\leq \rho_*^\xi \lim_{k \to K_2} \max \{-1 - \rho(x^k), H_j(x^k) - \rho(x^k)\} = -\rho_*^{\xi+1} < 0.
\]

Again,

\[
\nabla g_j(x^k)^T d^k + \mu \rho(x^k)^{\xi+1} \to \nabla g_j(x^*)^T d^* + \mu \rho_*^{\xi+1} < -(1 - \mu) \rho_*^{\xi+1} < 0,
\]

Therefore, \(\nabla g_j(x^k)^T d^k + \mu \rho(x^k)^{\xi+1} \leq -\frac{1-\mu}{2} \rho_*^{\xi+1}\). Using Taylor expansion, combining with (3.7), one has for \(\lambda > 0\) small enough and \(k \in K_2\) large enough,

\[
g_j(x^k + \lambda d^k) - \psi(x^k) + \mu \lambda \rho(x^k)^{\xi+1} = g_j(x^k) - \psi(x^k) + \lambda \left( \nabla g_j(x^k)^T d^k + \mu \rho(x^k)^{\xi+1} \right) + o(\lambda)
\]

\[
\leq -\frac{1-\mu}{2} \lambda \rho_*^{\xi+1} + o(\lambda) \leq 0.
\]

Hence, for \(k \in K_2\) large enough and \(\lambda > 0\) small enough,

\[
g_j(x^k + \lambda d^k) - \psi(x^k) + \mu \lambda \rho(x^k)^{\xi+1} \leq 0, \quad j \in I_+\) and \(g_j(x^*) = \psi(x^*)\).
\]

So, the inequality (2.30) holds by the discussion of Cases A and B above.

**Analyze the inequality (2.31):** two cases are considered respectively.

**Case C:** For \(j \in J_-\) and \(g_j(x^*) < 0\), we have \(\lim_{k \to K_2} g_j(x^k) = g_j(x^*) < 0\). Therefore, \(g_j(x^k) \leq \frac{1}{2} g_j(x^*) < 0\) for \(k \in K_2\) large enough. Thus

\[
g_j(x^k + \lambda d^k) = g_j(x^k) + O(\lambda) \leq \frac{1}{2} g_j(x^*) + O(\lambda) \leq 0
\]

holds for \(k \in K_2\) large enough and \(\lambda > 0\) small enough.

**Case D:** If \(j \in J_-\) and \(g_j(x^*) = 0\), then it is easy to show that \(j \in I_-\) by **Assumption A3**. Therefore, we have \(H_j^* = -g_j(x^*) = 0\) from (3.2). Further, in the same fashion as in Case B above, we get from (3.7):

\[
\nabla g_j(x^*)^T d^* = -\rho_*^{\xi+1} < 0.
\]

Again, \(\nabla g_j(x^k)^T d^k \to \nabla g_j(x^*)^T d^*\), therefore, \(\nabla g_j(x^k)^T d^k \leq -\frac{1}{2} \rho_*^{\xi+1}\). So, using Taylor expansion, one has

\[
g_j(x^k + \lambda d^k) = g_j(x^k) + \lambda \nabla g_j(x^k)^T d^k + o(\lambda) \leq -\frac{\lambda}{2} \rho_*^{\xi+1} + o(\lambda) \leq 0
\]

holds for \(k \in K_2\) large enough and \(\lambda > 0\) small enough.

Hence, the inequality (2.31) holds for \(k \in K_2\) large enough and \(\lambda > 0\) small enough.

Summarizing the analysis above, we can conclude that there exists \(\tilde{\lambda} > 0\) such that \(\lambda_k \geq \tilde{\lambda}\) for all \(k \in K_2\) large enough. □
Theorem 3.1. If Assumptions A1–A3 hold, then the algorithm either stops in a finite number of steps with a KKT point $x^*$, or generates an infinite sequence $\{x^k\}$ of points such that each accumulation point $x^*$ of $\{x^k\}$ is a KKT point of the problem (P).

Proof. If the algorithm stops at $x^k$, then $\rho(x^k) = 0$ and $x^k$ is a KKT point of the problem (P) from Theorem 2.1. Now suppose that an infinite iterative sequence $\{x^k\}$ is generated by the algorithm and $x^*$ is a given limit point of $\{x^k\}$. Let infinite index sets $K$ and $K_2$ satisfy (3.1) and (3.6). We will show that $x^*$ is a KKT point of the problem (P). By contradiction, suppose that $x^*$ is not a KKT point of the problem (P). We should use Lemmas 3.1 and 3.2 to bring a contradiction. Two cases below are considered respectively.

Case I: In this case, there exists an iteration index $s$ such that $x^s \in X$, i.e., $\psi(x^s) = 0(J^s_+ = \emptyset)$. Therefore, $\psi(x^* - x^s) = 0$. Hence, we have from (2.29) and (2.24):

$$f(x^{k+1}) \leq f(x^k) + \mu \lambda_k \nabla f(x^k)^T d^k \leq f(x^k) - \mu \lambda_k \rho(x^k) \xi^k, \quad \forall k \geq s.$$  

This implies that $\{f(x^k)\}_{k \geq s}$ is decreasing. Further, combining $\lim_{k \in K_2} f(x^k) = f(x^*)$, we get $\lim_{k \to \infty} f(x^k) = f(x^*)$. Now, using the inequality above and combining Lemmas 3.1 and 3.2, one gets

$$0 = \lim_{k \in K_2} (f(x^{k+1}) - f(x^k)) \leq \lim_{k \in K_2} (-\mu \lambda_k \rho(x^k) \xi_{k+1}) \leq -\mu \bar{\lambda} \rho^*_{k+1},$$

this contradicts $\rho_* > 0$ and $\bar{\lambda} > 0$.

Case II: $\psi(x^k)$ is a KKT point of the problem (P) from Theorem 3.1. Next, under the given conditions, we will discuss the strong convergence of the proposed algorithm.

4. Strong convergence

In this section, we begin with the following Assumption A4. Next, under the given conditions, we will discuss the strong convergence of the proposed algorithm.

Assumption A4. The iterative sequence $\{x^k\}$ of points generated by the proposed algorithm is bounded.

Lemma 4.1. Suppose that Assumptions A1–A4 hold. Then there exists a constant $\tilde{\mu} > 0$ such that $\|N_k^T N_k + H_k^{-1}\| \leq \tilde{\mu}$ for $k = 0, 1, 2, \ldots$.

Proof. By contradiction, suppose that there exists an infinite index subset $K$ such that

$$\|N_k^T N_k + H_k^{-1}\| \to \infty, \quad k \in K, \quad k \to \infty. \quad (4.1)$$

By Assumption A4, we know the iterative sequence $\{x^k\}_{k \in K}$ of points is bounded; then there exists an infinite index subset $K_1 \subseteq K$ such that

$$x^k \to x^*, \quad k \in K_1, \quad k \to \infty.$$  

Thus, there exists an infinite index subset $K_2 \subseteq K_1$ such that (3.1) is satisfied by setting $K = K_2$, which implies that $(N_k^T N_k + H_k)^{-1}$, $k \in K_2$ is continuous. Let $H^*$ and $N_*$ be the matrices defined by (3.2). Then the matrix $(N_*^T N_* + H^*)^{-1}$ is positive definite by Lemma 3.1, and so is the matrix $(N_k^T N_k + H_k)^{-1}$. It then follows from (2.5), (2.6) and (3.2) that

$$\lim_{k \in K_2} \|N_k^T N_k + H_k^{-1}\| = \|N_*^T N_* + H^*\|^{-1}.$$
However, from (4.1), we know that $\lim_{k \in K} \| (N^T_k N_k + H_k)^{-1} \| = \infty$. This is a contradiction, since $(N^T_k N_k + H^*)^{-1}$ is a fixed positive definite matrix, and the proof is finished. \hfill \Box

**Theorem 4.1.** Suppose that Assumptions A1–A4 hold. Then (1) there exist two constants $c_0 > 0$, $c = c_0(1 + \frac{1}{\xi})$ such that the search direction $d^k$ satisfies

$$
\rho(x^k)^{\xi} \geq c_0 \| d^k \|; \quad \nabla f(x^k)^T d^k \leq -c \| d^k \|^{(1 + \frac{1}{\xi})}, \quad \text{if } \psi(x^k) = 0. \quad (4.2)
$$

(2) $\lim_{k \to \infty} \| x^{k+1} - x^k \| = 0$.

(3) If $\{x^k\}$ possesses an isolate accumulation $x^*$ and $\lim_{k \to \infty} \| x^{k+1} - x^k \| = 0$, then $\lim_{k \to \infty} x^k = x^*$; that is, the algorithm is strongly convergent.

**Proof.** (1) From Lemma 4.1, we know $\{B_k\}$ is bounded. Combining with the boundedness of $S^k_f$ (see Lemma 2.3) and $v^k$, we can conclude that there exists a constant $c_0 > 0$ such that

$$
\| d^k \| = \| \rho(x^k)^{\xi} (S^k_f + B_k^T v^k) \| \leq \rho(x^k)^{\xi} (\| S^k_f \| + \| B_k^T v^k \|) \leq \frac{1}{c_0} \rho(x^k)^{\xi},
$$

that is, $\rho(x^k)^{\xi} \geq c_0 \| d^k \|$. Furthermore, since $\psi(x^k) = 0$, by (2.24), we know

$$
\nabla f(x^k)^T d^k \leq -\rho(x^k)^{\xi+1} \leq -c \| d^k \|^{(1 + \frac{1}{\xi})}.
$$

(2) Case I: In this case, there exists an iteration index $s$ such that $x^s \in X$, i.e., $\psi(x^s) = 0$ ($J_+^s = \emptyset$). So $\psi(x^k) = 0$, $J_+^s = \emptyset$ hold too for all $k \geq s$. Combining the discussion of Theorem 3.1 with (4.2), we have

$$
0 = \lim_{k \to \infty} (f(x^{k+1}) - f(x^k)) \leq \lim_{k \to \infty} (\mu \lambda_k \nabla f(x^k)^T d^k) \leq \lim_{k \to \infty} (-c \mu \lambda_k \| d^k \|^{(1 + \frac{1}{\xi})}).
$$

Thus, $\lim_{k \to \infty} \lambda_k \| d^k \|^{(1 + \frac{1}{\xi})} = 0$, and further, $\lim_{k \to \infty} \lambda_k \| d^k \| = 0$ and

$$
\lim_{k \to \infty} \| x^{k+1} - x^k \| = \lim_{k \to \infty} \lambda_k \| d^k \| = 0.
$$

Case II: $\psi(x^k) > 0$ and $\psi(x^{k+1}) < \psi(x^k)$ for any $k = 1, 2, \ldots$. Combining the discussion of Theorem 3.1 with (4.2), one has

$$
0 = \lim_{k \to \infty} (\psi(x^{k+1}) - \psi(x^k)) \leq \lim_{k \to \infty} (-\mu \lambda_k \rho(x^k)^{\xi+1}) \leq \lim_{k \to \infty} (-c \mu \lambda_k \| d^k \|^{(1 + \frac{1}{\xi})}).
$$

This implies that $\lim_{k \to \infty} \lambda_k \| d^k \|^{(1 + \frac{1}{\xi})} = 0$. Furthermore, it follows that $\lim_{k \to \infty} \lambda_k \| d^k \| = 0$ and

$$
\lim_{k \to \infty} \| x^{k+1} - x^k \| = \lim_{k \to \infty} \lambda_k \| d^k \| = 0.
$$

(3) Under the conditions that $\lim_{k \to \infty} \| x^{k+1} - x^k \| = 0$ and $\{x^k\}$ possesses an isolate accumulation $x^*$, it is a well known result that $\lim_{k \to \infty} x^k = x^*$ (or see Proposition 4.1. in Ref. [22]).

Hence, the whole proof of Theorem 4.1 is completed. \hfill \Box

**5. Numerical results**

In this section, based on the proposed algorithm, we test some problems taken from [16,23]. The numerical experiments were implemented on MATLAB 6.5, under Windows XP. The preliminary numerical results show that the proposed algorithm is promising.

During the numerical experiments, the parameters are selected as follows: $\gamma = 0.5, \varnothing = 0.5, \epsilon = 0, 0.5$ or 1, $\mu = 0.1, \bar{\mu} = 1, \xi = 0.1, t = 3, \varepsilon_k \equiv 0.1, \delta_k \equiv 0.1$ and

$$
\alpha_k^{\xi} = \begin{cases} 
   \frac{(1 - \gamma)\epsilon}{6} \cdot \frac{\| P_k \nabla f(x^k) \|}{\| \nabla f(x^k) \|}, & \text{if } r \in T_{\alpha}^k; \\
   0, & \text{if } r \notin T_{\alpha}^k 
\end{cases}
$$

$$
\beta_k^{\xi} = \begin{cases} 
   \frac{(1 - \gamma)(1 - \epsilon)}{6} \cdot \frac{\| P_k \nabla f(x^k) \|}{\| d^{k-r} \|}, & \text{if } r \in T_{\beta}^k; \\
   0, & \text{if } r \notin T_{\beta}^k 
\end{cases}
$$
Table 1
Comparison of the numerical results between algorithms A and B

<table>
<thead>
<tr>
<th>Prob</th>
<th>$n$, $m$</th>
<th>Initial point</th>
<th>$\sigma_k$</th>
<th>algorithm</th>
<th>$\epsilon$</th>
<th>$N_1$</th>
<th>$N_f$</th>
<th>$x^*$</th>
<th>$f(x^*)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example 1</td>
<td>2, 3</td>
<td>$(8, 8)^T \in X$</td>
<td>A</td>
<td>0 + 10</td>
<td>(0.5191, 0.2405)$^T$</td>
<td>8</td>
<td>5010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (i) B</td>
<td>0</td>
<td>0 + 12</td>
<td>13</td>
<td>(0.5001, 0.2500)$^T$</td>
<td>5001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (i) B</td>
<td>0.5</td>
<td>0 + 12</td>
<td>13</td>
<td>(0.5005, 0.2498)$^T$</td>
<td>5001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (i) B</td>
<td>1</td>
<td>0 + 15</td>
<td>18</td>
<td>(0.4999, 0.2501)$^T$</td>
<td>5001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (ii) B</td>
<td>0</td>
<td>0 + 12</td>
<td>12</td>
<td>(0.5001, 0.2500)$^T$</td>
<td>5001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (ii) B</td>
<td>0.5</td>
<td>0 + 12</td>
<td>12</td>
<td>(0.5005, 0.2498)$^T$</td>
<td>5001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (ii) B</td>
<td>1</td>
<td>0 + 15</td>
<td>17</td>
<td>(0.4999, 0.2501)$^T$</td>
<td>5001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2, 3</td>
<td>$(0, 1)^T \notin X$</td>
<td>A</td>
<td>1 + 11</td>
<td>(0.5239, 0.2382)$^T$</td>
<td>5016</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (i) B</td>
<td>0</td>
<td>2 + 6</td>
<td>14</td>
<td>(0.5003, 0.2499)$^T$</td>
<td>5002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (i) B</td>
<td>0.5</td>
<td>2 + 6</td>
<td>13</td>
<td>(0.5002, 0.2500)$^T$</td>
<td>5002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (i) B</td>
<td>1</td>
<td>2 + 5</td>
<td>12</td>
<td>(0.4993, 0.2504)$^T$</td>
<td>5001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (ii) B</td>
<td>0</td>
<td>1 + 11</td>
<td>11</td>
<td>(0.5001, 0.2500)$^T$</td>
<td>5001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (ii) B</td>
<td>0.5</td>
<td>1 + 8</td>
<td>8</td>
<td>(0.5003, 0.2499)$^T$</td>
<td>5002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (ii) B</td>
<td>1</td>
<td>1 + 12</td>
<td>12</td>
<td>(0.4995, 0.2503)$^T$</td>
<td>5002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2, 3</td>
<td>$(-11, 2)^T \notin X$</td>
<td>A</td>
<td>5 + 9</td>
<td>(0.5196, 0.2403)$^T$</td>
<td>5010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (i) B</td>
<td>0</td>
<td>10 + 10</td>
<td>21</td>
<td>(0.5001, 0.2500)$^T$</td>
<td>5001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (i) B</td>
<td>0.5</td>
<td>10 + 8</td>
<td>19</td>
<td>(0.5025, 0.2488)$^T$</td>
<td>5001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (i) B</td>
<td>1</td>
<td>11 + 12</td>
<td>24</td>
<td>(0.5011, 0.2495)$^T$</td>
<td>5001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (ii) B</td>
<td>0</td>
<td>10 + 10</td>
<td>10</td>
<td>(0.5001, 0.2500)$^T$</td>
<td>5001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (ii) B</td>
<td>0.5</td>
<td>10 + 8</td>
<td>8</td>
<td>(0.5025, 0.2488)$^T$</td>
<td>5001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (ii) B</td>
<td>1</td>
<td>11 + 12</td>
<td>12</td>
<td>(0.5011, 0.2495)$^T$</td>
<td>5001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 2</td>
<td>2, 2</td>
<td>$(0, 0)^T \in X$</td>
<td>A</td>
<td>0 + 4</td>
<td>(0.9924, 0.9960)$^T$</td>
<td>1.0151</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (i) B</td>
<td>0</td>
<td>0 + 10</td>
<td>11</td>
<td>(0.9999, 0.9999)$^T$</td>
<td>1.0003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (i) B</td>
<td>0.5</td>
<td>0 + 11</td>
<td>12</td>
<td>(0.9999, 0.9999)$^T$</td>
<td>1.0002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (i) B</td>
<td>1</td>
<td>0 + 10</td>
<td>11</td>
<td>(0.9998, 0.9999)$^T$</td>
<td>1.0003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (ii) B</td>
<td>0</td>
<td>0 + 10</td>
<td>10</td>
<td>(0.9999, 0.9999)$^T$</td>
<td>1.0003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (ii) B</td>
<td>0.5</td>
<td>0 + 11</td>
<td>11</td>
<td>(0.9999, 0.9999)$^T$</td>
<td>1.0002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (ii) B</td>
<td>1</td>
<td>0 + 10</td>
<td>10</td>
<td>(0.9998, 0.9999)$^T$</td>
<td>1.0003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2, 2</td>
<td>$(-1, -1)^T \notin X$</td>
<td>A</td>
<td>1 + 4</td>
<td>(0.9873, 0.9927)$^T$</td>
<td>1.0255</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (i) B</td>
<td>0</td>
<td>1 + 12</td>
<td>14</td>
<td>(0.9999, 0.9999)$^T$</td>
<td>1.0002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (i) B</td>
<td>0.5</td>
<td>1 + 10</td>
<td>12</td>
<td>(0.9999, 0.9999)$^T$</td>
<td>1.0002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (ii) B</td>
<td>0</td>
<td>1 + 10</td>
<td>12</td>
<td>(0.9999, 1.0000)$^T$</td>
<td>1.0002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (ii) B</td>
<td>0.5</td>
<td>1 + 12</td>
<td>12</td>
<td>(0.9999, 0.9999)$^T$</td>
<td>1.0002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (ii) B</td>
<td>1</td>
<td>1 + 10</td>
<td>10</td>
<td>(0.9999, 1.0000)$^T$</td>
<td>1.0002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2, 2</td>
<td>$(1, -1)^T \notin X$</td>
<td>A</td>
<td>1 + 3</td>
<td>(0.9925, 0.9971)$^T$</td>
<td>1.0148</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (i) B</td>
<td>0</td>
<td>1 + 11</td>
<td>13</td>
<td>(0.9999, 0.9999)$^T$</td>
<td>1.0002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (i) B</td>
<td>0.5</td>
<td>1 + 9</td>
<td>11</td>
<td>(0.9998, 0.9999)$^T$</td>
<td>1.0003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (i) B</td>
<td>1</td>
<td>1 + 9</td>
<td>11</td>
<td>(0.9999, 0.9999)$^T$</td>
<td>1.0002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (ii) B</td>
<td>0</td>
<td>1 + 11</td>
<td>11</td>
<td>(0.9999, 0.9999)$^T$</td>
<td>1.0002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (ii) B</td>
<td>0.5</td>
<td>1 + 9</td>
<td>9</td>
<td>(0.9998, 0.9999)$^T$</td>
<td>1.0003</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (ii) B</td>
<td>1</td>
<td>1 + 9</td>
<td>9</td>
<td>(0.9999, 0.9999)$^T$</td>
<td>1.0002</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Example 3</td>
<td>2, 2</td>
<td>$(0, 0)^T \in X$</td>
<td>A</td>
<td>0 + 3</td>
<td>(0.8910, 0.0000)$^T$</td>
<td>1.0118</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (i) B</td>
<td>0</td>
<td>0 + 3</td>
<td>4</td>
<td>(0.9938, 0.0000)$^T$</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (i) B</td>
<td>0.5</td>
<td>0 + 4</td>
<td>5</td>
<td>(0.9951, 0.0000)$^T$</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (i) B</td>
<td>1</td>
<td>0 + 4</td>
<td>5</td>
<td>(0.9941, 0.0000)$^T$</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (ii) B</td>
<td>0</td>
<td>0 + 3</td>
<td>3</td>
<td>(0.9938, 0.0000)$^T$</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (ii) B</td>
<td>0.5</td>
<td>0 + 4</td>
<td>4</td>
<td>(0.9951, 0.0000)$^T$</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (ii) B</td>
<td>1</td>
<td>0 + 4</td>
<td>4</td>
<td>(0.9941, 0.0000)$^T$</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2, 2</td>
<td>$(2, 0)^T \notin X$</td>
<td>A</td>
<td>16 + 2</td>
<td>(0.9662, 0.0030)$^T$</td>
<td>1.0011</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (i) B</td>
<td>0</td>
<td>2 + 4</td>
<td>7</td>
<td>(0.9954, 0.0000)$^T$</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (i) B</td>
<td>0.5</td>
<td>2 + 3</td>
<td>6</td>
<td>(0.9931, 0.0000)$^T$</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Case (i) B</td>
<td>1</td>
<td>2 + 5</td>
<td>8</td>
<td>(0.9953, 0.0000)$^T$</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(continued on next page)
Table 1 (continued)

<table>
<thead>
<tr>
<th>Prob</th>
<th>n, m</th>
<th>Initial point</th>
<th>σ_k</th>
<th>algorithm</th>
<th>ϵ</th>
<th>Ni</th>
<th>Nf</th>
<th>x*</th>
<th>f(x*)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2, 2</td>
<td>(0, 2)^T \notin X</td>
<td>Case (ii) B</td>
<td>0 2 + 4</td>
<td>(0.9954, 0.0000)^T</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Case (ii) B</td>
<td>0.5 2 + 3</td>
<td>(0.9931, 0.0000)^T</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Case (ii) B</td>
<td>1 2 + 5</td>
<td>(0.9953, 0.0000)^T</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>A</td>
<td>4 + 8</td>
<td>(0.9318, -0.0034)^T</td>
<td>1.0046</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Case (i) B</td>
<td>0 6 + 6</td>
<td>(0.9949, 0.0001)^T</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Case (i) B</td>
<td>0.5 7 + 4</td>
<td>(0.9934, -0.0004)^T</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Case (i) B</td>
<td>1 7 + 5</td>
<td>(0.9952, -0.0005)^T</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Case (ii) B</td>
<td>0 6 + 6</td>
<td>(0.9949, 0.0001)^T</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Case (ii) B</td>
<td>0.5 7 + 6</td>
<td>(0.9951, 0.0000)^T</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Case (ii) B</td>
<td>1 7 + 5</td>
<td>(0.9952, -0.0005)^T</td>
<td>1.0000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Besides, we discuss two cases on parameter σ_k: Case (i) σ_k = 1 for all x^k; and Case (ii) σ_k \[ \begin{align*} & = 1, \text{ if } x^k \in X; \\ & = 0, \text{ if } x^k \notin X. \end{align*} \] Execution is terminated if ρ(x^k) \leq 10^{-4}.

The tested problems, from Examples 1–3 in Table 1, are selected from Ref. [16] (2004). In Tables 2 and 3, the typical test problems are selected from Ref. [23] (1981), which are not tested by the method in Ref. [16] and the feasible and infeasible starting points are selected at will. The columns of Tables 1–3 have the following meanings: the Prob column displays the tested problems; n and m are the number of variables and constraints of the tested problems; Ni and Nf give the number of iterations and objective function evaluations, respectively; x* and f(x*) denote the approximate solution and objective value, respectively.

Especially, the Ni columns of Tables 1 and 3 are displayed as the sum of two numbers: the former indicates the number of iterations required in Phase I (generate a feasible point); and the latter indicates the number of iterations required in Phase II (perform feasible descent method to get an optimal solution). As a result, the sum is the total number of iterations. For example, “2 + 5” means that after two iterations, the algorithm generates a feasible point (i.e. \( \psi(x^2) = 0 \)), and after another five iterations the algorithm produces an approximate solution \( x^* \).

In addition, we compared our algorithm (denoted by B) with the algorithm (denoted by A) in [16], as shown in Table 1. And the same initial points as in [16] were selected.
Table 3
Numerical results of algorithm B for infeasible initial points

<table>
<thead>
<tr>
<th>Prog</th>
<th>n, m</th>
<th>Initial point</th>
<th>$\sigma_k$</th>
<th>$\epsilon$</th>
<th>$N_k$</th>
<th>$N_f$</th>
<th>$x^*$</th>
<th>$f(x^*)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS12</td>
<td>2, 1</td>
<td>$(3, 3)^T$</td>
<td>Case (i)</td>
<td>0</td>
<td>23 + 0</td>
<td>24</td>
<td>(2.0000, 3.0000)$^T$</td>
<td>-30.0002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (i)</td>
<td>0.5</td>
<td>33 + 0</td>
<td>34</td>
<td>(2.0000, 3.0000)$^T$</td>
<td>-30.0003</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (i)</td>
<td>1</td>
<td>27 + 0</td>
<td>28</td>
<td>(2.0000, 3.0000)$^T$</td>
<td>-30.0003</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (ii)</td>
<td>0</td>
<td>23 + 0</td>
<td>0</td>
<td>(2.0000, 3.0000)$^T$</td>
<td>-30.0002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (ii)</td>
<td>0.5</td>
<td>33 + 0</td>
<td>0</td>
<td>(2.0000, 3.0000)$^T$</td>
<td>-30.0003</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (ii)</td>
<td>1</td>
<td>27 + 0</td>
<td>0</td>
<td>(2.0000, 3.0000)$^T$</td>
<td>-30.0003</td>
</tr>
<tr>
<td></td>
<td>2, 1</td>
<td>$(6, 6)^T$</td>
<td>Case (i)</td>
<td>0</td>
<td>26 + 0</td>
<td>27</td>
<td>(2.0000, 3.0000)$^T$</td>
<td>-30.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (i)</td>
<td>0.5</td>
<td>22 + 0</td>
<td>23</td>
<td>(2.0000, 3.0000)$^T$</td>
<td>-30.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (i)</td>
<td>1</td>
<td>27 + 0</td>
<td>28</td>
<td>(2.0001, 2.9999)$^T$</td>
<td>-30.0002</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (ii)</td>
<td>0</td>
<td>26 + 0</td>
<td>0</td>
<td>(2.0000, 3.0000)$^T$</td>
<td>-30.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (ii)</td>
<td>0.5</td>
<td>22 + 0</td>
<td>0</td>
<td>(2.0000, 3.0000)$^T$</td>
<td>-30.0001</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (ii)</td>
<td>1</td>
<td>27 + 0</td>
<td>0</td>
<td>(2.0001, 2.9999)$^T$</td>
<td>-30.0002</td>
</tr>
<tr>
<td>HS29</td>
<td>3, 1</td>
<td>$(3, 4, 2)^T$</td>
<td>Case (i)</td>
<td>0</td>
<td>66 + 0</td>
<td>67</td>
<td>(3.9999, 2.8281, 2.0003)$^T$</td>
<td>-22.6275</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (i)</td>
<td>0.5</td>
<td>46 + 0</td>
<td>47</td>
<td>(4.0000, 2.8284, 2.0000)$^T$</td>
<td>-22.6278</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (i)</td>
<td>1</td>
<td>94 + 0</td>
<td>95</td>
<td>(4.0000, 2.8284, 2.0000)$^T$</td>
<td>-22.6275</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (i)</td>
<td>0</td>
<td>66 + 0</td>
<td>0</td>
<td>(3.9999, 2.8281, 2.0003)$^T$</td>
<td>-22.6275</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (i)</td>
<td>0.5</td>
<td>46 + 0</td>
<td>0</td>
<td>(4.0000, 2.8284, 2.0000)$^T$</td>
<td>-22.6278</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (i)</td>
<td>1</td>
<td>94 + 0</td>
<td>0</td>
<td>(4.0000, 2.8284, 2.0000)$^T$</td>
<td>-22.6275</td>
</tr>
<tr>
<td>HS31</td>
<td>3, 1</td>
<td>$(2, 4, 2)^T$</td>
<td>Case (i)</td>
<td>0</td>
<td>17 + 0</td>
<td>30</td>
<td>(0.5773, 1.7325, -0.0002)$^T$</td>
<td>6.0005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (i)</td>
<td>0.5</td>
<td>3 + 17</td>
<td>38</td>
<td>(0.5775, 1.7317, 0.0001)$^T$</td>
<td>6.0003</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (i)</td>
<td>1</td>
<td>3 + 17</td>
<td>44</td>
<td>(0.5773, 1.7322, 0.0003)$^T$</td>
<td>6.0004</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (i)</td>
<td>0</td>
<td>3 + 14</td>
<td>26</td>
<td>(0.5773, 1.7325, -0.0002)$^T$</td>
<td>6.0005</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (i)</td>
<td>0.5</td>
<td>3 + 17</td>
<td>34</td>
<td>(0.5775, 1.7317, 0.0001)$^T$</td>
<td>6.0003</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (i)</td>
<td>1</td>
<td>3 + 17</td>
<td>40</td>
<td>(0.5773, 1.7322, 0.0003)$^T$</td>
<td>6.0004</td>
</tr>
<tr>
<td>HS100</td>
<td>7, 4</td>
<td>$(1.1, 2.1, 0.1, 4.1, 0.1, 1.1, 1.1)^T$</td>
<td>Case (i)</td>
<td>0</td>
<td>3 + 74</td>
<td>217</td>
<td>(2.3305, 1.9514, -0.4775, 4.3656, -0.6247, 1.0381, 1.5942)$^T$</td>
<td>680.6302</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (i)</td>
<td>0.5</td>
<td>2 + 61</td>
<td>100</td>
<td>(2.3305, 1.9515, -0.4775, 4.3656, -0.6247, 1.0381, 1.5942)$^T$</td>
<td>680.6302</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (i)</td>
<td>1</td>
<td>2 + 75</td>
<td>120</td>
<td>(2.3305, 1.9514, -0.4775, 4.3656, -0.6245, 1.0381, 1.5942)$^T$</td>
<td>680.6302</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (ii)</td>
<td>0</td>
<td>3 + 74</td>
<td>213</td>
<td>(2.3305, 1.9514, -0.4775, 4.3656, -0.6247, 1.0381, 1.5942)$^T$</td>
<td>680.6302</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (ii)</td>
<td>0.5</td>
<td>2 + 61</td>
<td>97</td>
<td>(2.3305, 1.9515, -0.4775, 4.3656, -0.6247, 1.0381, 1.5942)$^T$</td>
<td>680.6302</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (ii)</td>
<td>1</td>
<td>2 + 75</td>
<td>117</td>
<td>(2.3305, 1.9515, -0.4775, 4.3656, -0.6245, 1.0381, 1.5942)$^T$</td>
<td>680.6302</td>
</tr>
<tr>
<td>HS35</td>
<td>3, 1</td>
<td>$(1, 1, 1)^T$</td>
<td>Case (i)</td>
<td>0</td>
<td>1 + 10</td>
<td>15</td>
<td>(1.3337, 0.7776, 0.4442)$^T$</td>
<td>0.1112</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (i)</td>
<td>0.5</td>
<td>1 + 12</td>
<td>16</td>
<td>(1.3335, 0.7778, 0.4441)$^T$</td>
<td>0.1112</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (i)</td>
<td>1</td>
<td>1 + 11</td>
<td>15</td>
<td>(1.3329, 0.7768, 0.4449)$^T$</td>
<td>0.1112</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (i)</td>
<td>0</td>
<td>1 + 10</td>
<td>13</td>
<td>(1.3337, 0.7776, 0.4442)$^T$</td>
<td>0.1112</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (ii)</td>
<td>0.5</td>
<td>1 + 12</td>
<td>14</td>
<td>(1.3335, 0.7778, 0.4441)$^T$</td>
<td>0.1112</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case (ii)</td>
<td>1</td>
<td>1 + 11</td>
<td>13</td>
<td>(1.3329, 0.7768, 0.4449)$^T$</td>
<td>0.1112</td>
</tr>
</tbody>
</table>
From Table 1, we can see that the accuracy of the algorithm B is obviously superior to that of the algorithm A. Except for the problem HS29 in the case of the infeasible initial point (see Table 3), there are only slight differences in the numerical results among $\epsilon$ equals 0, 0.5 and 1 in Tables 1–3. These show that the change of the proposed algorithm is minor for $\epsilon \in [0, 1]$ to certain degree, and we can choose the value of $\epsilon$ based on practical requirements.

References