Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

A NEW GENERALIZED PROJECTION METHOD OF STRONGLY SUB-FEASIBLE DIRECTIONS FOR GENERAL CONSTRAINED OPTIMIZATION *

Jin-Bao Jian†, Chuan-Hao Guo and Lin-Feng Yang

Abstract: In this paper, a class of optimization problems with equality and inequality constraints is discussed. Firstly, the original problem is transformed into an associated simpler problem with a penalty term and only inequality constraints, then a strongly sub-feasible algorithm is presented. At each iteration of the proposed algorithm, the search direction is generated by only one simple explicit formula of generalized projection. Under some mild assumptions, the globally and the strongly convergent properties are obtained. Another important feature of the proposed algorithm is that the iteration points can enter into the feasible region of the equivalent problem after finite iterations. Finally, some preliminary numerical results are reported.

Key words: *general constrained optimization; generalized projection; method of strongly sub-feasible directions; global and strong convergence*

Mathematics Subject Classification: *90C30, 65K10*

1 Introduction

In this work, we consider the following general constrained optimization problems with nonlinear inequality and equality constraints

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_j(x) \le 0, \quad j \in L_1 = \{1, \ldots, m'\}, \\ & f_j(x) = 0, \quad j \in L_2 = \{m' + 1, \ldots, m\}, \end{array} \tag{1.1}$$

where $x \in R^n$ and functions $f_j : R^n \to R$ $(j = 0, 1, \ldots, m)$ are continuously differentiable. We denote the feasible set of the problem (1.1) as follows: $X = \{x \in R^n : f_j(x) \le 0, j \in L_1;\ f_j(x) = 0, j \in L_2\}$.

It is well-known that the method of feasible directions (MFD) is a kind of efficient methods for solving inequality constrained optimization problems (Refs. [2], [7], [10], [12], [13]), but it is difficult for MFD to deal with optimization problems with nonlinear equality constraints directly.

*Project supported by the National Natural Science Foundation (No. 10771040) of China and Guangxi Province Science Foundation (No. 0832052).

†Corresponding author.

In order to employ the MFD to solve optimization problems with equality constraints, Mayne and Polak [11] transformed the problem (1.1) into the following optimization problem with only inequality constraints

$$\begin{aligned} \min \quad & F(x;c) \triangleq f_0(x) - c \sum_{j \in L_2} f_j(x) \\ \text{s.t.} \quad & f_j(x) \le 0, \quad j \in L \triangleq L_1 \cup L_2, \end{aligned} \tag{1.2}$$

where parameter $c > 0$, which is updated by a simple procedure. Especially, if $L_2 = \emptyset$, define $F(x;c) = f_0(x)$. The feasible set of the problem (1.2) is denoted by $X^+ = \{x \in R^n : f_j(x) \le 0,\ j \in L\}$.

Mayne and Polak [11] showed that the simplified problem (1.2) is equivalent to the original problem (1.1) if c is sufficiently large (but finite), and then presented a feasible direction algorithm for solving the problem (1.2), so for the problem (1.1). More details and advantages were discussed by Lawrence and Tits [10], some further applications of this technique are presented by Herkovits [7], and Jian [2] also showed that Mayne and Polak's scheme bring many advantages.

On the other hand, the gradient projection method introduced by Rosen [15] is an important class of MFD for solving inequality constrained optimization. For a given iteration point x^k, the search direction $d(x^k)$ in [15] is generated by a so-called gradient projection operation

$$d(x^k) = -P(x^k)\nabla f_0(x^k), \quad P(x^k) = E_n - A(x^k)(A(x^k)^T A(x^k))^{-1} A(x^k)^T,$$

where E_n is the n order identity matrix and matrix $A(x^k)$ consists of the gradients of the active constraints at x^k. In fact, $-P(x^k)\nabla f_0(x^k)$ is the projection of $-\nabla f_0(x^k)$ onto the null-space $\{d \in R^n : \ A(x^k)^T d = 0\}$. The gradient projection method is further researched and extended to the generalized gradient projection method (GGPM) (Refs. [4], [6], [9]). In most globally convergent GGPMs, the search direction $d(x^k)$ has the form of

$$d(x^k) = -P(x^k)\nabla f_0(x^k) + Q(x^k)^T v(x^k),\ P(x^k) = E_n - A(x^k)Q(x^k),$$
$$Q(x^k) = (A(x^k)^T A(x^k) + D(x^k))^{-1} A(x^k)^T,$$

where matrix $A(x^k)$ consists of part or all gradients of the constraints, $D(x^k)$ is a suitable diagonal and positive semi-definite matrix, and $v(x^k)$ is a suitable vector.

In 1995, Jian [3] improved the combined Phase I-Phase II algorithm [14] and proposed a strongly sub-feasible direction method, which not only unifies automatically the processes of initialization (Phase I) and optimization (Phase II), but also guarantees that the number of the functions satisfying inequality constraints is nondecreasing. Based on the strongly sub-feasible method and the generalized projection technique, Jian and Zhang [6] presented a new algorithm with arbitrary initial point. Compare with algorithm in [3], this algorithm possesses not only global convergence but also strong convergence.

In this paper, motivated by the techniques in [6, 11] and the generalized projection method, we present a new strongly sub-feasible direction algorithm for solving general nonlinearly constrained optimization problems. The main features of the proposed algorithm are summarized as follows:

- the objective function of the simplified problem is used directly as the merit function;
- the initial point is arbitrary;
- the parameter is adjusted automatically only for a finite number of times;
- the iteration points always enter into the feasible set X^+ after finite iterations;
- at each iteration, the search direction is generated by only one explicit formula of generalized projection and the line search technique is some different from others.

2 Description of Algorithm

For convenience of presentation, we use the following notation throughout this paper with $x \in R^n$ and parameter $p > 0$

$$\left\{\begin{array}{l} g_j(x) = \nabla f_j(x),\ j \in \{0\} \cup L; \\ I^-(x) = \{j \in L : f_j(x) \le 0\},\quad I^+(x) = \{j \in L : f_j(x) > 0\}; \\ \psi(x) = \max\{0, f_j(x),\ j \in L\} = \max\{0, f_j(x),\ j \in I^+(x)\}; \\ L_1^-(x) = \{j \in L_1 :\ f_j(x) = 0\},\ L_1^+(x) = \{j \in L_1 \cap I^+(x) :\ f_j(x) = \psi(x)\}. \end{array}\right. \tag{2.1}$$

$$D_j(x) = \left\{\begin{array}{ll} (-f_j(x))^p, & j \in I^-(x) \cap L_1; \\ (\psi(x) - f_j(x))^p, & j \in I^+(x) \cap L_1; \\ 0, & j \in L_2. \end{array}\right. \tag{2.2}$$

The following basic assumption is assumed to be satisfied.

Assumption A1 *The gradient vectors* $\{g_j(x) : D_j(x) = 0\}$ *are linearly independent.*

It is easy to see that Assumption A1 can be reduced to the usual linear independence whenever $x \in X^+$.

For a given arbitrary iteration point $x^k \in R^n$ and parameter $c_k > 0$, in order to generate the search direction, we define generalized projection matrix $P(x^k)$, multiplier functions $\pi(x^k)$ and $\pi(x^k; c_k)$ as follows:

$$\left\{\begin{array}{l} D_k = D(x^k) = \mathrm{diag}(D_j(x^k), j \in L),\ A_k = A(x^k) = (g_j(x^k), j \in L); \\ P_k = P(x^k) = E_n - A_k Q_k,\ Q_k = Q(x^k) = (A_k^T A_k + D_k)^{-1} A_k^T; \\ \hat{\pi}^k \triangleq \pi(x^k; c_k) = -Q_k \nabla F(x^k; c_k),\ \pi^k = \pi(x^k) = (\pi_j(x^k), j \in L) = -Q_k g_0(x^k). \end{array}\right. \tag{2.3}$$

It is not difficult to know that

$$\hat{\pi}_j^k = \pi_j(x^k; c_k) = \left\{\begin{array}{ll} \pi_j(x^k) = \pi_j^k, & j \in L_1; \\ \pi_j(x^k) + c_k = \pi_j^k + c_k, & j \in L_2. \end{array}\right. \tag{2.4}$$

From [11], we know that the simplified problem (1.2) is equivalent to the original problem (1.1) when c_k is sufficiently large (but finite). In order to check that whether the current iteration point x^k is a KKT point of the problem (1.1) or not, we introduce the following identifying function $\rho(x^k)$:

$$\rho_k = \rho(x^k) = \frac{\|P_k g_0(x^k)\|^2 + \omega(x^k) + \psi(x^k)}{1 + |\varpi^T \hat{\pi}^k|}, \tag{2.5}$$

where $\varpi = (1, \ldots, 1)^T \in R^m$ and

$$\begin{array}{rl} \omega(x^k) & = \sum\limits_{j \in L_1} \max\{-\pi_j^k,\ \pi_j^k D_j(x^k)\} + \sum\limits_{j \in L_2, f_j(x^k) > 0} \hat{\pi}_j^k (\psi(x^k) - f_j(x^k)) \\ & + \sum\limits_{j \in L_2, f_j(x^k) \le 0} (-\hat{\pi}_j^k f_j(x^k)). \end{array} \tag{2.6}$$

The following lemma gives some properties of the formulas defined above.

Lemma 2.1. *Suppose that Assumption A1 holds and iteration point* $x^k \in R^n$. *Then*
(i) *The matrix* $(A_k^T A_k + D_k)$ *is nonsingular and positive definite.*
(ii) $A_k^T P_k = D_k Q_k,\ A_k^T Q_k^T = E_n - D_k (A_k^T A_k + D_k)^{-1}$.
(iii) *If* $D_j(x^k) = 0$, *then* $g_j(x^k)^T P_k = 0$ *and* $g_j(x^k)^T Q_k^T = \varpi_j^T$, *where* $\varpi_j = (0, \ldots, 0, 1_{jth}, 0, \ldots, 0)^T \in R^m$.
(iv) $g_0(x^k)^T P_k g_0(x^k) = \|P_k g_0(x^k)\|^2 + \sum\limits_{j \in L} (\pi_j^k)^2 D_j^k$.
(v) *If* $c_k > |\pi_j^k|,\ \forall j \in L_2$, *then* x^k *is a KKT point of the problem* (1.1) *if and only if* $\rho_k = 0$.

Proof. The proof of (i), (ii), (iii) and (iv) is similar to Lemma 1 in [6], and the proof of (v) is similar to Lemma 2.2 in [5] and Lemma 2 in [6], thus they are all omitted here. □

For a current iteration point $x^k \in R^n$ and $\xi \geq 0$, we construct the search direction d^k as follows:

$$d^k = d(x^k) = \rho_k^\xi[-P_k g_0(x^k) + Q_k^T v^k] - r\psi(x^k)Q_k^T \varpi, \tag{2.7}$$

with

$$v_j^k = v_j(x^k) = \begin{cases} -1-\rho_k, & \text{if } j \in L_1,\ \pi_j^k < 0; \\ D_j(x^k) - \rho_k, & \text{if } j \in L_1,\ \pi_j^k \geq 0; \\ -f_j(x^k) - \rho_k, & \text{if } j \in L_2,\ f_j(x^k) \leq 0; \\ \psi(x^k) - f_j(x^k) - \rho_k, & \text{if } j \in L_2,\ f_j(x^k) > 0. \end{cases} \tag{2.8}$$

Based on Lemma 2.1 and (2.7)-(2.8), we show some properties of d^k in the following lemma.

Lemma 2.2. *Let $x^k \in R^n$. Then* (i) $\nabla F(x^k; c_k) d^k \leq \psi(x^k)\rho_k^\xi - \rho_k^{\xi+1} + r\psi(x^k)\varpi^T\hat{\pi}^k$, *and* (ii) $g_j(x^k)^T d^k \leq -\rho_k^{1+\xi} - r\psi(x^k)$, $\forall j \in \{j \in L:\ f_j(x^k) = 0$ *or* $f_j(x^k) = \psi(x^k)\}$.

Proof. (i) From (2.7) and Lemma 2.1(iii, iv) as well as (2.4), we get

$$\begin{aligned}
\nabla F(x^k; c_k)^T d^k &= (g_0(x^k) - c_k \sum_{j\in L_2} g_j(x^k))^T \left\{\rho_k^\xi[-P_k g_0(x^k) + Q_k^T v^k] - r\psi(x^k)Q_k^T\varpi\right\} \\
&= \rho_k^\xi \Bigg\{-\|P_k g_0(x^k)\|^2 - \sum_{j\in L}(\pi_j^k)^2 D_j(x^k) - (\pi^k)^T v^k \\
&\quad + c_k \sum_{j\in L_2} g_j(x^k)^T P_k g_0(x^k) - c_k \sum_{j\in L_2} g_j(x^k)^T Q_k^T v^k \Bigg\} \\
&\quad + r\psi(x^k)\varpi^T\pi^k + rc_k\psi(x^k)\sum_{j\in L_2} g_j(x^k)^T Q_k^T\varpi \\
&= \rho_k^\xi \Bigg\{-||P_k g_0(x^k)||^2 - \sum_{j\in L}(\pi_j^k)^2 D_j(x^k) - \sum_{j\in L_1}\pi_j^k v_j^k - \sum_{j\in L_2}(\pi_j^k + c_k)v_j^k \\
&\quad + c_k \sum_{j\in L_2} g_j(x^k)^T P_k g_0(x^k)\Bigg\} + r\psi(x^k)\varpi^T\hat{\pi}^k.
\end{aligned}$$

So, in view of (2.8) and (2.4), we further have

$$\begin{aligned}
\nabla F(x^k; c_k)^T d^k &= \rho_k^\xi \Bigg\{-\|P_k g_0(x^k)\|^2 - \sum_{j\in L}(\pi_j^k)^2 D_j(x^k) + \rho_k \sum_{j\in L_1}\pi_j^k \\
&\quad - \Bigg(\sum_{j\in L_1, \pi_j^k<0}(-\pi_j^k) + \sum_{j\in L_1,\pi_j^k\geq 0}\pi_j^k D_j(x^k)\Bigg) + \sum_{j\in L_2, f_j(x^k)\leq 0}\hat{\pi}_j^k f_j(x^k) \\
&\quad + \rho_k \sum_{j\in L_2}\hat{\pi}_j^k - \sum_{j\in L_2, f_j(x^k)>0}\hat{\pi}_j^k(\psi(x^k) - f_j(x^k))\Bigg\} + r\psi(x^k)\varpi^T\hat{\pi}^k \\
&= \rho_k^\xi \Bigg\{-\|P_k g_0(x^k)\|^2 - \sum_{j\in L}(\pi_j^k)^2 D_j(x^k) + \rho_k(\sum_{j\in L_1}\pi_j^k + \sum_{j\in L_2}\hat{\pi}_j^k) \\
&\quad - \Bigg(\sum_{j\in L_1,\pi_j^k<0}(-\pi_j^k) + \sum_{j\in L_1,\pi_j^k\geq 0}\pi_j^k D_j(x^k) + \sum_{j\in L_2, f_j(x^k)\leq 0}\hat{\pi}_j^k(-f_j(x^k)) \\
&\quad + \sum_{j\in L_2, f_j(x^k)>0}\hat{\pi}_j^k(\psi(x^k) - f_j(x^k))\Bigg)\Bigg\} + r\psi(x^k)\varpi^T\hat{\pi}^k.
\end{aligned}$$

Therefore, according to (2.5)-(2.6), we obtain

$$\begin{aligned}\nabla F(x^k;c_k)^T d^k &\le \rho_k^\xi \left\{ -\|P_k g_0(x^k)\|^2 - \sum_{j\in L} (\pi_j^k)^2 D_j(x^k) + \rho_k |\varpi^T \hat{\pi}^k| - \omega(x^k) \right\} \\ &\quad + r\psi(x^k)\varpi^T\hat{\pi}^k \\ &\le \rho_k^\xi(\psi(x^k)-\rho_k) + r\psi(x^k)\varpi^T\hat{\pi}^k.\end{aligned}$$

(ii) From (2.2)–(2.3) and Lemma 2.1(ii), it follows that

$$\begin{aligned}A_k^T d^k &= \rho_k^\xi \left\{ -A_k^T P_k g_0(x^k) + A_k^T Q_k^T v^k \right\} - r\psi_k A_k^T Q_k^T \varpi \\ &= \rho_k^\xi \left\{ -D_k Q_k g_0(x^k) + v^k - D_k(A_k^T A_k + D_k)^{-1} v^k \right\} \\ &\quad - r\psi(x^k)\varpi + r\psi(x^k) D_k (A_k^T A_k + D_k)^{-1}\varpi.\end{aligned} \tag{2.9}$$

For $j \in \{j \in L:\ f_j(x^k) = 0 \ \text{ or } \ f_j(x^k) = \psi(x^k)\}$, the formula (2.2) shows that $D_j^k = 0$, furthermore, (2.8) implies that $v_j^k \le -\rho_k$. So (2.9) gives that $g_j(x^k)^T d^k = \rho_k^\xi v_j^k - r\psi(x^k) \le -\rho_k^{1+\xi} - r\psi(x^k)$. □

Based on the analysis above and the search direction d^k defined by (2.7), we can describe our algorithm as follows.

Algorithm A Parameters: $r,\xi \ge 0,\ \alpha,\beta \in (0,1),\ p, c_{-1}, \gamma, r_0 > 0$. Data: $x^0 \in R^n$. Set $k := 0$.

Step 1 (Update parameter c_k): Compute c_k by

$$c_k = \begin{cases} \max\{s_k,\ c_{k-1}+\gamma\}, & \text{if } s_k > c_{k-1}; \\ c_{k-1}, & \text{if } s_k \le c_{k-1}, \end{cases} \quad s_k = \max\{|\pi_j^k|,\ j \in L_2\} + r_0. \tag{2.10}$$

Step 2 (Yield search direction): Compute ρ_k and d^k according to (2.2)–(2.8). If $\rho_k = 0$, then stop and x^k is a KKT point of the problem (1.1); otherwise, go to Step 3.

Step 3 (Do line search): If $\psi(x^k) > 0$ and $f_j(x^k + d^k) \le 0,$ for all $j \in L$, then let $\lambda_k = 1$ and go to Step 4; otherwise, compute the step-size λ_k which is the first value of λ in the sequence $\{1, \beta, \beta^2, \dots\}$ that satisfies the following inequalities:

$$\delta_k F(x^k+\lambda d^k; c_k) \le \delta_k\{F(x^k;c_k) + \alpha\lambda\nabla F(x^k;c_k)^T d^k + (1-\alpha)\lambda\psi(x^k)(\rho_k^\xi + r\varpi^T\hat{\pi}^k)\}, \tag{2.11}$$

$$f_j(x^k+\lambda d^k) \le \max\{0, \psi(x^k) - \alpha\lambda(\rho_k^{1+\xi} + r\psi(x^k))\}, \quad \forall j \in I_k^+ \triangleq I^+(x^k), \tag{2.12}$$

$$f_j(x^k+\lambda d^k) \le 0, \quad \forall j \in I_k^- \triangleq I^-(x^k), \tag{2.13}$$

where $\delta_k = 1$ if $\psi(x^k) = 0$ and $\delta_k \ge 0$ if $\psi(x^k) > 0$.

Step 4 (Update): Let $x^{k+1} = x^k + \lambda_k d^k$, $k := k+1$, and go back to Step 1.

Remark 1 (i) In order to obtain the global convergence of Algorithm A, it is sufficient that the parameters are restricted by $p > 0$, $r \ge 0$ and $\xi \ge 0$. Furthermore, to get strong convergence and ensure the iteration points enter into the feasible set X^+ after finite iterations, the parameters have to be restricted by $p > 1$, $r > 1$ and $\xi > 1$.

(ii) It does not influence any of our analysis if δ_k is always set to be a positive constant, and the line search (2.11) of Algorithm A will be reduced to

$$F(x^k+\lambda d^k; c_k) \le F(x^k;c_k) + \alpha\lambda\nabla F(x^k,c_k)^T d^k + (1-\alpha)\lambda\psi(x^k)(\rho_k^\xi + r\varpi^T\hat{\pi}^k).$$

The purpose that we allow $\delta_k = 0$ in the case of $\psi(x^k) > 0$ (i.e., $x^k \notin X^+$) is to further reduce the computational cost of the line search, because the line search (2.11) does not work in this case, and the number of objective function evaluations will be reduced, then the numerical effect is expected to be improved.

(iii) The line search technique (2.12) is some different from others. It can ensure that the value of λ can be accepted as a step-size and complete the line search as long as $x^k + \lambda d^k \in X^+$, in the case of $\delta_k = 0$.

The lemma given below shows that the proposed algorithm above is well defined.

Lemma 2.3. *Suppose that Assumption A1 holds. Then the line search in Step 3 can be carried out, i.e., the inequalities (2.11)-(2.13) holds for $\lambda > 0$ sufficiently small.*

Proof. (1) Analyze the inequality (2.11): Using Taylor expansion and Lemma 2.2(i), we have

$$\begin{aligned}
a_k(\lambda) &\triangleq \delta_k\{F(x^k+\lambda d^k;c_k) - F(x^k;c_k) - \alpha\lambda\nabla F(x^k;c_k)^T d^k - (1-\alpha)\lambda\psi(x^k)(\rho_k^{\xi} + r\varpi^T\hat{\pi}^k)\}\\
&\le \delta_k\{(1-\alpha)\lambda[\psi(x^k)\rho_k^{\xi} - \rho_k^{\xi+1} + r\psi(x^k)\varpi^T\hat{\pi}^k] - (1-\alpha)\lambda\psi(x^k)(\rho_k^{\xi} + r\varpi^T\hat{\pi}^k) + o(\lambda\|d^k\|)\}\\
&= \delta_k\{-(1-\alpha)\lambda\rho_k^{1+\xi} + o(\lambda\|d^k\|)\}.
\end{aligned}$$

This together with $\alpha \in (0,1)$ and $\rho_k > 0$ implies that $a_k(\lambda) \le 0$ for $\lambda > 0$ sufficiently small.

(2) Analyze the inequality (2.12): For convenience of analysis, we denote

$$b_j^k(\lambda) \triangleq f_j(x^k+\lambda d^k) - \max\{0, \psi(x^k) - \alpha\lambda(\rho_k^{1+\xi} + r\psi(x^k))\},\ j \in I^+(x^k).$$

(2-i) For $j \in I^+(x^k)$ and $f_j(x^k) < \psi(x^k)$, by using Taylor expansion, one gets for $\lambda > 0$ sufficiently small

$$\begin{aligned}
b_j^k(\lambda) &\le f_j(x^k+\lambda d^k) - \psi(x^k) + \alpha\lambda(\rho_k^{1+\xi} + r\psi(x^k))\\
&= f_j(x^k) - \psi(x^k) + \lambda g_j(x^k)^T d^k + \alpha\lambda(\rho_k^{1+\xi} + r\psi(x^k)) + o(\lambda||d^k||)\\
&= f_j(x^k) - \psi(x^k) + O(\lambda||d^k||) \le 0.
\end{aligned}$$

(2-ii) For $j \in I^+(x^k)$ and $f_j(x^k) = \psi(x^k)$, by expanding $f_j(x^k+\lambda d^k)$ at x^k and combining Lemma 2.2(ii), for $\lambda > 0$ sufficiently small, one has

$$\begin{aligned}
b_j^k(\lambda) &\le f_j(x^k+\lambda d^k) - \psi(x^k) + \alpha\lambda(\rho_k^{1+\xi} + r\psi(x^k))\\
&= f_j(x^k) - \psi(x^k) + \lambda g_j(x^k)^T d^k + \alpha\lambda(\rho_k^{1+\xi} + r\psi(x^k)) + o(\lambda||d^k||)\\
&\le -\lambda(\rho_k^{1+\xi} + r\psi(x^k)) + \alpha\lambda(\rho_k^{1+\xi} + r\psi(x^k)) + o(\lambda||d^k||)\\
&= (\alpha-1)\lambda(\rho_k^{1+\xi} + r\psi(x^k)) + o(\lambda||d^k||) \le 0.
\end{aligned}$$

(3) Analyze the inequality (2.13): First, for $j \in I^-(x^k)$ and $f_j(x^k) < 0$, it follows that $f_j(x^k+\lambda d^k) \le 0$ for $\lambda > 0$ sufficiently small, since $f_j(x^k)$ is continuously differentiable.

Second, for $j \in I^-(x^k)$ and $f_j(x^k) = 0$, using Taylor expansion and Lemma 2.2(ii), one gets

$$f_j(x^k+\lambda d^k) = f_j(x^k) + \lambda g_j(x^k)^T d^k + o(\lambda||d^k||) \le -\lambda(\rho_k^{1+\xi} + r\psi(x^k)) + o(\lambda||d^k||).$$

So, $f_j(x^k+\lambda d^k) \le 0$ for $\lambda > 0$ sufficiently small.

Summarizing the analysis above, we conclude that there exits a constant $\bar{\lambda} > 0$ such that the inequalities (2.11)–(2.13) are satisfied for all $\lambda \in (0,\ \bar{\lambda}]$ and Lemma 2.3 follows. ☐

Remark 2 From the line search conditions (2.11)-(2.13), it holds that one of the following two cases must happen:

Case A. There exists an iteration index s such that $\psi(x^s) = 0$, then $\psi(x^k) \equiv 0,\ \delta_k \equiv 1$ and

$$F(x^{k+1}; c_k) \leq F(x^k; c_k) + \alpha\lambda_k \nabla F(x^k; c_k)^T d^k \leq F(x^k; c_k) - \alpha\lambda_k \rho_k^{1+\xi} < F(x^k; c_k),\ \forall k \geq s. \tag{2.14}$$

Case B. For any $k = 0, 1, \ldots,\ \psi(x^k) > 0,\ \delta_k \geq 0$ and

$$\psi(x^k + 1) \leq \psi(x^k) - \alpha\lambda_k(\rho_k^{1+\xi} + r\psi(x^k)) < \psi(x^k). \tag{2.15}$$

According to Remark 2, it is easy to know that the sets $I^+(x^k),\ I^-(x^k),\ L_2^+(x^k) \triangleq \{j \in L_2 :\ f_j(x^k) > 0\}$ and $L_2^-(x^k) = L_2 \backslash L_2^+(x^k)$ are all monotone, furthermore, in view of they are all being subsets of the fixed and finite set L, thus, these subsets can be fixed if k is sufficiently large. For k large enough, we denote

$$I^+(x^k) = I^+,\ I^-(x^k) = I^-,\ L_2^+(x^k) = L_2^+,\ L_2^-(x^k) = L_2^-.$$

3 Global Convergence

Taking into account Lemma 2.1(v) and Step 2 as well as $c_k > |\pi_j^k|$ $(\forall j \in L_2)$, one knows that x^k is a KKT point of the problem (1.1) if Algorithm A stops at $\{x^k\}$. Now, we assume that an infinite sequence of iteration points is yielded by Algorithm A and we will show that each accumulation point x^* of $\{x^k\}$ is a KKT point of the problem (1.1).

Suppose that x^* is a given accumulation point of the sequence $\{x^k\}$. Then there exists an infinite index set K, such that $x^k \xrightarrow{K} x^*$. Denote

$$D_j^* = \begin{cases} (-f_j(x^*))^p, & j \in I^- \cap L_1; \\ (\psi(x^*) - f_j(x^*))^p, & j \in I^+ \cap L_1; \\ 0, & j \in L_2. \end{cases}$$
$$D_* = D(x^*) = \mathrm{diag}(D_j^*,\ j \in L),\ A_* = A(x^*) = (g_j(x^*),\ j \in L).$$

The following basic assumption is necessary to ensure the global convergence of Algorithm A.

Assumption A2 *The sequence $\{x^k\}$ yielded by Algorithm A is bounded.*

Lemma 3.1. *Suppose that Assumptions A1-A2 hold. Then matrix $(A_*^T A_* + D_*)$ is positive definite. Furthermore, there exists a constant $\epsilon > 0$, such that $\|(A_k^T A_k + D_k)^{-1}\| \leq \epsilon$ holds for all k large enough.*

Lemma 3.2. *Suppose that Assumptions A1-A2 hold. Then there exists an integer $k_0 > 0$, such that $c_k \equiv c \triangleq c_{k_0}$ holds for all $k \geq k_0$.*

The proof of the above two lemmas is similar to Lemma 3.1 in [4] and Lemma 3.1 in [5], respectively. Thus, they are all omitted here.

Due to Lemma 3.2, we always assume that $c_k \equiv c$ for all k in the rest analysis. Now, let us further define

$$Q_* = (A_*^T A_* + D_*)^{-1} A_*^T, \quad P_* = E_n - A_* Q_*, \quad \pi^* = -Q_* g_0(x^*), \quad \pi_c^* = -Q_* \nabla F(x^*; c),$$
$$\omega_* = \sum_{j \in L_1} \max\{-\pi_j^*,\ \pi_j^* D_j^*\} + \sum_{j \in L_2^+} (\pi_j^* + c)(\psi(x^*) - f_j(x^*)) + \sum_{j \in L_2^-} (-f_j(x^*))(\pi_j^* + c)\},$$

$$\rho_* = \frac{\|P_* g_0(x^*)\|^2 + \omega_* + \psi(x^*)}{1 + |\varpi^T \pi_c^*|}. \tag{3.1}$$

Then $(\psi(x^k),\ D_k,\ \rho_k) \xrightarrow{K} (\psi(x^*),\ D_*,\ \rho_*)$, $\{v^k\}_K$ and $\{d^k\}_K$ are all bounded.

Lemma 3.3. *If x^* is not a KKT point of the problem* (1.1)*, then $\rho_* > 0$ and $\rho_k \geq 0.5\rho_*$ for $k \in K$ large enough.*

Proof. In view of $\rho_k \xrightarrow{K} \rho_*$, we only need to show $\rho_* > 0$. First, if $\psi(x^*) = 0$, then $D_* = D(x^*),\ \omega_* = \omega(x^*),\ \pi^* = \pi(x^*),\ \rho_* = \rho(x^*)$, from Lemmas 2.1(v) and 3.2, we obtain $\rho_* > 0$ since $c > |\pi_j^*|,\ \forall j \in L_2$. Second, if $\psi(x^*) > 0$, we get $\rho_* > 0$ from (3.1). □

Lemma 3.4. *If x^* is not a KKT point of the problem* (1.1)*, then $\lambda_k \geq \lambda_* \triangleq \inf\{\lambda_k : k \in K\} > 0,\ \forall k \in K$.*

Proof. It is sufficient to show that the inequalities (2.11)–(2.13) hold for $k \in K$ large enough and $\lambda > 0$ small enough.

(1) Analyze the inequality (2.11): Using Taylor expansion, the boundedness of $\{d^k\}_K$, Lemmas 2.1(v) and 3.3, from part (1) in the proof of Lemma 2.3 and $\alpha \in (0,1)$, we have

$$a_k(\lambda) = \delta_k\{-(1-\alpha)\lambda\rho_k^{1+\xi} + o(\lambda)\} \leq \delta_k\{-0.5^{1+\xi}(1-\alpha)\lambda\rho_*^{1+\xi} + o(\lambda)\} \leq 0.$$

Therefore, the inequality (2.15) holds for $k \in K$ large enough and $\lambda > 0$ sufficiently small.

To simply and finish the rest analysis, the relationship

$$g_j(x^k)^T d^k = \rho_k^{\xi} v_j^k - r\psi(x^k) + O(D_j^k),\ \text{if}\ D_j^k \to 0, \tag{3.2}$$

is important, and which can be obtained from (2.9).

(2) Analyze the inequality (2.12): We divide the proof into three cases:

Case (2a): For $j \in I^+$ and $f_j(x^*) < \psi(x^*)$, then we have from Lemma 3.3

$$\begin{aligned}
&\lim_{\lambda\to 0}(f_j(x^k + \lambda d^k) - \max\{0, \psi(x^k) - \alpha\lambda(\rho_k^{1+\xi} + r\psi(x^k))\}) \\
&\leq \lim_{\lambda\to 0}(f_j(x^k + \lambda d^k) - \psi(x^k) + \alpha\lambda(\rho_k^{1+\xi} + r\psi(x^k))) \\
&= f_j(x^k) - \psi(x^k) \leq 0.5(f_j(x^*) - \psi(x^*)) < 0.
\end{aligned} \tag{3.3}$$

Case (2b): For $j \in I^+ \cap L_1$ and $f_j(x^*) = \psi(x^*)$, then $D_j^k \to 0,\ k \in K$. Therefore, using Taylor expansion, we obtain

$$\begin{aligned}
f_j(x^k &+ \lambda d^k) - \max\{0, \psi(x^k) - \alpha\lambda(\rho_k^{1+\xi} + r\psi(x^k))\} \\
&\leq f_j(x^k + \lambda d^k) - \psi(x^k) + \alpha\lambda(\rho_k^{1+\xi} + r\psi(x^k)) \\
&= f_j(x^k) + \lambda g_j(x^k)^T d^k - \psi(x^k) + \alpha\lambda(\rho_k^{1+\xi} + r\psi(x^k)) + o(\lambda) \\
&\leq \alpha\lambda(\rho_k^{1+\xi} + r\psi(x^k)) + \lambda g_j(x^k)^T d^k + o(\lambda).
\end{aligned}$$

On the other hand, from (3.2) and (2.8), one has

$$g_j(x^k)^T d^k = \rho_k^{\xi} v_j^k - r\psi(x^k) + O(D_j^k) \leq -\rho_k^{1+\xi} - r\psi(x^k) + O(D_j^k).$$

So, this together with $\alpha \in (0,1)$ shows that

$$\begin{aligned}
f_j(x^k &+ \lambda d^k) - \max\{0, \psi(x^k) - \alpha\lambda(\rho_k^{1+\xi} + r\psi(x^k))\} \\
&\leq -\lambda\rho_k^{1+\xi} - r\lambda\psi(x^k) + \lambda O(D_j^k) + \alpha\lambda(\rho_k^{1+\xi} + r\psi(x^k)) + o(\lambda) \\
&\leq -\lambda(1-\alpha)\rho_k^{1+\xi} + o(\lambda) \leq -0.5^{1+\xi}\lambda(1-\alpha)\rho_*^{1+\xi} + o(\lambda) \leq 0.
\end{aligned} \tag{3.4}$$

Case (2c): For $j \in I^+ \cap L_2$ and $f_j(x^*) = \psi(x^*)$, then $D_j^k \equiv 0$ and $\psi(x^k) - f_j(x^k) \to 0,\ k \in K$. In view of (3.2) and (2.8), we get,

$$g_j(x^k)^T d^k = \rho_k^\xi v_j^k - r\psi(x^k) = \rho_k^\xi(\psi(x^k) - f_j(x^k)) - \rho_k^{1+\xi} - r\psi(x^k).$$

So, from Lemma 3.3 and $\alpha \in (0,1)$, in the same fashion of Case (2b), we obtain

$$\begin{aligned} f_j(x^k &+\lambda d^k) - \max\{0, \psi(x^k) - \alpha\lambda(\rho_k^{1+\xi} + r\psi(x^k))\} \\ &\le \lambda g_j(x^k)^T d^k + \alpha\lambda(\rho_k^{1+\xi} + r\psi(x^k)) + o(\lambda) \\ &\le -\lambda\rho_k^{1+\xi} - r\lambda\psi(x^k) + \alpha\lambda(\rho_k^{1+\xi} + r\psi(x^k)) + \lambda\rho_k^\xi(\psi(x^k) - f_j(x^k)) + o(\lambda) \\ &\le -\lambda(1-\alpha)\rho_k^{1+\xi} + \lambda\rho_k^\xi(\psi(x^k) - f_j(x^k)) + o(\lambda) \\ &\le -0.5^{1+\xi}\lambda(1-\alpha)\rho_*^{1+\xi} + o(\lambda) \le 0. \end{aligned} \tag{3.5}$$

Combining (3.3), (3.4) and (3.5), we conclude that the inequality (2.12) holds for $k \in K$ large enough and $\lambda > 0$ sufficiently small.

(3) The analysis of inequality (2.13) is similar to the above analysis of inequality (2.12), so, it is omitted here.

Summarizing the discussion above, the proof of Lemma 3.4 is completed. □

Now, based on Lemmas 3.3 and 3.4, we can present the global convergence of Algorithm A.

Theorem 3.5. *Suppose that Assumptions A1-A2 hold. Then Algorithm A either stops at a KKT point of the problem* (1.1) *after finite iterations, or generates an infinite sequence* $\{x^k\}$ *such that each accumulation point* x^* *of* $\{x^k\}$ *is a KKT point of the problem* (1.1).

Proof. In view of $c_k \ge s_k > |\pi_j^k| + \gamma_0, \forall j \in L_2$, therefore, if Algorithm A stops at the k-th iteration, then from Step 2 and Lemma 2.1(v), we know that x^k is a KKT point of the problem (1.1). Now, we suppose that Algorithm A generates an infinite sequence $\{x^k\}$ and x^* is a given accumulation point of it. Suppose by contradiction that x^* is not a KKT point of the problem (1.1). In view of Remark 2, it follows that one of the two following Cases must happen:

(i) If Case A of Remark 2 happens, then $\{F(x^k;c)\}_{k\ge s}$ is decreasing, furthermore, combining $\lim_{k\in K} F(x^k;c) = F(x^*;c)$, it follows that $\lim_{k\to\infty} F(x^k;c) = F(x^*;c)$. Therefore, from relationships (2.14) and Lemmas 3.3, 3.4, we get

$$F(x^{k+1};c) \le F(x^k;c) - \alpha\lambda_k\rho_k^{1+\xi} \le F(x^k;c) - 0.5^{1+\xi}\alpha\lambda_*\rho_*^{1+\xi}, \quad s \le k \in K.$$

So, passing to the limit in the inequality above, it follows that $-\lambda_*\rho_*^{1+\xi} \ge 0$, which contradicts $\lambda_* > 0$ and $\rho_* > 0$.

(ii) If Case B of Remark 2 happens, then $\{\psi(x^k)\}$ is decreasing, and combining $\lim_{k\in K}\psi(x^k) = \psi(x^*)$, it follows that $\lim_{k\to\infty}\psi(x^k) = \psi(x^*)$. Furthermore, from (2.15) and Lemmas 3.3 as well as 3.4, it is easy to get

$$\psi(x^{k+1}) \le \psi(x^k) - \alpha\lambda_k(\rho_k^{1+\xi} + r\psi(x^k)) \le \psi(x^k) - \alpha\lambda_k\rho_k^{1+\xi} \le \psi(x^k) - 0.5^{1+\xi}\alpha\lambda_*\rho_*^{1+\xi}, \quad k \in K.$$

Similarly, taking limit in the inequality above, it follows that $-\lambda_*\rho_*^{1+\xi} \ge 0$, which is a contradiction too. So, we conclude that x^* is a KKT point of the problem (1.1), and the proof is completed. □

4 Strong Convergence

In this section, under an additional suitable assumption, we further show that Algorithm A has the following important features: (1) The algorithm is strongly convergent, i.e., the whole sequence $\{x^k\}$ is convergent; (2) The iteration points always enter into the feasible set X^+ after finite iterations.

Assumption A3 (i) *The sequence $\{x^k\}$ of iteration points yielded by Algorithm A possesses an isolated accumulation point x^*, and* (ii) *the functions f_j $(j \in L)$ are second-order differentiable.*

Now, we establish the following strongly convergent theorem for Algorithm A.

Theorem 4.1. *Suppose that Assumptions A1-A3 hold and the parameter $\xi > 0$. Then* (i) $\lim\limits_{k\to\infty} \psi(x^k) = 0$; (ii) $\lim\limits_{k\to\infty} \|x^{k+1} - x^k\| = 0$; *and* (iii) $\lim\limits_{k\to\infty} x^k = x^*$, *that is, Algorithm A is strongly convergent.*

Proof. (i) By Theorem 3.5, it follows that the given accumulation point x^* is a KKT point of the problem (1.1), so $\psi(x^*) = 0$. On the other hand, the whole sequence $\{\psi(x^k)\}$ is convergent from the monotone properties and boundedness of $\{\psi(x^k)\}$, so $\lim\limits_{k\to\infty} \psi(x^k) = \psi(x^*) = 0$.

(ii) First, we show that $\lim\limits_{k\to\infty} \lambda_k \rho_k^{1+\xi} = 0$. Similarly to the proof of Theorem 3.5, we can also divide the proof into two cases according to Remark 2.

If Case A of Remark 2 occurs, then $\{F(x^k; c)\}_{k\geq s}$ is decreasing and bounded, so $\lim\limits_{k\to\infty} F(x^k; c) = F(x^*; c)$. Therefore, one has $\lim\limits_{k\to\infty} \lambda_k \rho_k^{1+\xi} = 0$ from (2.14).

If Case B of Remark 2 takes place, then $\lim\limits_{k\to\infty} \lambda_k \rho_k^{1+\xi} = 0$ follows from (2.15) and $\lim\limits_{k\to\infty} \psi(x^k) = 0$.

Then, from (2.7) and Lemma 3.1, there exist two positive constants M and N, such that

$$\begin{aligned}
\lim_{k\to\infty} \|x^{k+1} - x^k\| &= \lim_{k\to\infty} \lambda_k \|d^k\| = \lim_{k\to\infty} \lambda_k \|\rho_k^\xi \{-P_k g_0(x^k) + Q_k^T v^k\} - r\psi(x^k) Q_k^T \varpi\| \\
&\leq \lim_{k\to\infty} (M\lambda_k \rho_k^\xi + N\lambda_k \psi(x^k)) \\
&= \lim_{k\to\infty} \{M[(\lambda_k \rho_k^{1+\xi})^\xi \cdot \lambda_k]^{\frac{1}{1+\xi}} + N(\lambda_k \psi(x^k))\} \\
&= 0.
\end{aligned}$$

(iii) Taking into account $\lim\limits_{k\to\infty} \|x^{k+1} - x^k\| = 0$ and Assumption A3(i), it follows that $\lim\limits_{k\to\infty} x^k = x^*$ (see Theorem 4.1 in [5]). □

Lemma 4.2. *Suppose that Assumptions A1-A3 hold and the parameter $\xi > 1$. Then*

$$\lim_{k\to\infty} \rho_k = 0, \ \lim_{k\to\infty} d^k = 0, \ \|d^k\| = O(\rho_k^\xi) + O(\psi(x^k)), \ \|d^k\|^2 = o(\rho_k^{1+\xi}) + o(\psi(x^k)).$$

Proof. First, let ρ_* be defined by (3.1). In view of $\psi(x^*) = 0$ and $x^k \to x^*$, it is easy to get $\rho_k \to \rho_*$ and $\rho_* = \rho(x^*)$. On the other hand, in view of x^* is a KKT point of the problem (1.1), we have $\rho(x^*) = 0$ from Lemma 2.1(v) and $c > |\pi_j^*|$, $\forall j \in L_2$. So, $\lim\limits_{k\to\infty} \rho_k = 0$. Furthermore, from (2.7) and Lemma 3.1, we have $\lim\limits_{k\to\infty} d^k = 0$. Second, The rest two conclusions follow immediately from (2.7) and Lemma 3.1. □

Theorem 4.3. *Suppose that Assumptions A1-A3 hold and the parameters* $\xi, p, r > 1$. *Then,* $f_j(x^k + d^k) \le 0$ $(\forall j \in L)$ *holds for* k *large enough, therefore,* $\psi(x^{k+1}) = 0$, *i.e., the iteration points generated by Algorithm A always enter into the feasible set* X^+ *after finite iterations.*

Proof. Taking into account Step 3 and Cases A as well as B stated in Remark 2, it is sufficient to show that $f_j(x^k+d^k) \le 0$ $(\forall j \in L)$ for k large enough. Denote $L^* = \{j \in L : f_j(x^*) = 0\}$. First, for $j \notin L^*$, that is, $f_j(x^*) < 0$. In view of $(x^k, d^k) \to (x^*, 0)$, we have $f_j(x^k + d^k) \le 0$ for k large enough.

Second, for $j \in L^*$, i.e., $f_j(x^*) = 0$, one knows that $\psi(x^k) - f_j(x^k) \to \psi(x^*) - f_j(x^*) = 0$. We divide the proof into following four cases.

(Case 1) For $j \in L^* \cap I^+ \cap L_1$, from (2.2), we have $D_j^k = (\psi(x^k) - f_j(x^k))^p \to 0$. Therefore, using Taylor expansion and (3.2), (2.8) as well as Lemma 4.2, we get

$$\begin{aligned} f_j(x^k + d^k) &= f_j(x^k) + g_j(x^k)^T d^k + O(\|d^k\|^2) \\ &\le f_j(x^k) + \rho_k^\xi v_j^k - r\psi(x^k) + O(D_j^k) + O(\|d^k\|^2) \\ &\le f_j(x^k) - \rho_k^{1+\xi} - r\psi(x^k) + O(D_j^k) + O(\|d^k\|^2) \\ &= (f_j(x^k) - \psi(x^k)) + o(f_j(x^k) - \psi(x^k)) - \rho_k^{1+\xi} \\ &\quad + o(\rho_k^{1+\xi}) - (r-1)\psi(x^k) + o(\psi(x^k)) \\ &\le 0. \end{aligned}$$

(Case 2) For $j \in L^* \cap I^+ \cap L_2$, according to (2.2) we have $D_j^k \equiv 0$. Using Taylor expansion, in view of (2.8), (3.2) as well as Lemma 4.2, we get

$$\begin{aligned} f_j(x^k + d^k) &= f_j(x^k) + g_j(x^k)^T d^k + O(\|d^k\|^2) \\ &= f_j(x^k) + \rho_k^\xi \{\psi(x^k) - f_j(x^k) - \rho_k\} - r\psi(x^k) + O(\|d^k\|^2) \\ &= (1 - \rho_k^\xi) f_j(x^k) - \rho_k^{1+\xi} - r\psi(x^k) + \rho_k^\xi \psi(x^k) + o(\rho_k^{1+\xi}) + o(\psi(x^k)) \\ &\le -\rho_k^{1+\xi} + o(\rho_k^{1+\xi}) - r\psi(x^k) + o(\psi(x^k)) \le 0. \end{aligned}$$

(Case 3) For $j \in L^* \cap I^- \cap L_1$, in view of (2.2), (2.8), (3.2) and Lemma 4.2, we have $D_j^k = (-f_j(x^k))^p \to 0$, and

$$\begin{aligned} f_j(x^k + d^k) &= f_j(x^k) + g_j(x^k)^T d^k + O(\|d^k\|^2) \\ &= f_j(x^k) - \rho_k^{1+\xi} - r\psi(x^k) + O(D_j^k) + O(\|d^k\|^2) \\ &\le f_j(x^k) - \rho_k^{1+\xi} - r\psi(x^k) + o(-f_j(x^k)) + o(\rho_k^{1+\xi}) + o(\psi(x^k)) \\ &\le 0. \end{aligned}$$

(Case 4) For $j \in L^* \cap I^- \cap L_2$, in view of (2.2), (2.8) and Lemma 4.2, we have $D_j^k \equiv 0$. Using Taylor expansion, from (3.2), one has

$$\begin{aligned} f_j(x^k + d^k) &= f_j(x^k) + g_j(x^k)^T d^k + O(\|d^k\|^2) \\ &= f_j(x^k) + \rho_k^\xi (-f_j(x^k) - \rho_k) - r\psi(x^k) + O(\|d^k\|^2) \\ &\le -\rho_k^{1+\xi} - r\psi(x^k) + o(\psi(x^k)) + o(\rho_k^{1+\xi}) \\ &\le 0. \end{aligned}$$

Summarizing the discussion above, we conclude that $f_j(x^k+d^k) \le 0$ $(\forall j \in L)$ holds for k large enough. Thus, $\psi(x^{k+1}) = 0$ for k sufficiently large. The whole proof is completed. □

5 Numerical Experiments

In this section, the numerical experiments of Algorithm A are implemented in MATLAB 7.0 and performed on a PC with 1.81GHZ CPU and windows XP OS. The preliminary results show that Algorithm A is promising.

During the numerical experiments, the parameters are selected as follows:

$$p = r = \xi = 1.001, \quad \alpha = \beta = r_0 = 0.5, \quad c_{-1} = 1.5, \quad \gamma = 1.$$

Additionally, in order to further show the influence of δ_k on Algorithm A, we report and compare the numerical results of Algorithm A in two cases:

$$\text{Case I}: \ \delta_k \equiv 1; \qquad \text{Case II}: \ \delta_k = \begin{cases} 0, & \text{if } \psi(x^k) > 0; \\ 1, & \text{if } \psi(x^k) = 0. \end{cases} \tag{5.1}$$

Execution is terminated if $||\rho_k|| \le 10^{-3}$ and $\varphi_k = 0$. The tested problems HS7, HS14, HS32, HS63, HS71, HS81, HS107, HS111 and HS113 are selected from [8], and S217, S225, S252, S263, S325 and S388 are selected from [16]. Moreover, the initial points are all infeasible for the problem (1.2). The columns of the following tables have the following meanings:

$n, |L_1|, |L_2|$: n is the number of variables of the problem, $|L_1|$ is the number of inequality constraints and $|L_2|$ is the number of equality constraints;

Case: Cases I and II means that the choice of δ_k according to (5.1);

INO: the number of iterations out of the feasible set;

NII: the number of iterations within the feasible set;

Nf_0: the number of objective function evaluations;

Nf: the number of all constraint functions f_j $(j \in L)$ evaluations;

$Time$: the CPU time (second);

FV: the objective function value at the final iteration point.

The numerical results are reported in Table 1-Table 2. In Table 2, the approximately optimal solutions and objective values yielded by Algorithm A under Case I and Case II are listed, respectively.

According to the numerical results in Table 1 and Table 2, we see that Algorithm A in Case II is better than in Case I, especially for test problems HS81 and HS113.

All of the test problems in Table 1 are small. In order to test the effectiveness of our algorithm, we further test the problem selected from the CUTE collection [1], where the parameters for Algorithm A are selected as $p = r = 1.001$, $\xi = 1.01$, $\alpha = \beta = r_0 = 0.5$, $c_{-1} = 1.5$, $\gamma = 1$, and the execution is terminated if one of the termination criterions is satisfied: the number of iterations less than $N0$ or $||\rho_k|| \le 10^{-3}$. The initial points are all the same and infeasible. The numerical results are given in the following Table 3. The results show that the advantage of our algorithm when applied to solving problems with the large number of constraints.

On the other hand, we compared Algorithm A with the algorithm (denoted by B) in [17], as showed in the following Table 4. The initial points and tested Examples 1-3 are selected the same as in [17]. A(I) and A(II) means that Algorithm A in Case I and Case II, respectively. From Table 4, we can see that INO of Algorithm A is less than Algorithm B when initial points are not feasible, especially for Example 3. The approximately optimal solutions of Algorithm A are superior to Algorithm B.

Finally, we give some explanations about the choice of the parameters for Algorithm A. From Theorems 4.1 and 4.3, we know that the restriction on "$\xi > 0$" is to ensure Algorithm A is strongly convergent, and the restriction on "$\xi, p, r > 1$" is to ensure the iteration points get into the feasible set of the problem (1.2) after finite iterations. From the results of experiment, we also find that the parameter ξ has an influence on INO and NII as well as the numerical effect of Algorithm A. In the following Table 5, the numerical results of the problem HS107 are given under the different value of ξ, and the value of other parameters are the same as above. By the way, we find that the choice of other parameters has few influences on Algorithm A.

Table 1. *Numerical results report I*

Problem	$n, \|L_1\|, \|L_2\|$	Initial point x^0	Case	INO	NII	Nf_0	Nf	$Time$
HS7	2, 0, 1	$(4,2)^T$	I	14	8	22	109	0.03125
			II	14	8	9	109	0.03125
HS14	2, 1, 1	$(-1,-1)^T$	I	2	9	21	61	0.01563
			II	2	9	10	61	0.01563
HS32	3, 4, 1	$(0.5,0.5,0.5)^T$	I	3	40	43	435	0.04688
			II	3	40	41	435	0.04688
HS63	3, 3, 2	$(2.5,2.5,2.5)^T$	I	6	57	64	659	0.07813
			II	6	57	58	659	0.07813
HS71	4, 9, 1	$(3,4,2,4)^T$	I	3	19	24	481	0.04688
			II	3	19	20	489	0.04688
HS81	5, 10, 3	$(0,1,2,-2,-2)^T$	I	5	90	100	2643	0.34375
			II	4	81	82	2350	0.28125
HS107	9, 14, 0	$(5,\ldots,5)^T \in R^9$	I	7	53	61	5118	0.43750
			II	7	53	54	5118	0.48438
HS111	10, 20, 3	$(-0.5,\ldots,-0.5)^T \in R^{10}$	I	32	0	6	879	0.17188
			II	32	0	1	879	0.15625
HS113	10, 8, 0	$(9,\ldots,9)^T \in R^{10}$	I	23	11	65	978	0.09375
			II	4	31	171	642	0.09375
S217	2, 2, 1	$(-14,-1)^T$	I	19	29	40	309	0.04688
			II	19	29	30	309	0.04688
S225	2, 5, 0	$(2,5)^T$	I	1	20	23	251	0.03125
			II	1	30	53	868	0.07813
S252	3, 1, 1	$(-4,-8,-12)^T$	I	5	18	38	682	0.06250
			II	11	11	12	382	0.04688
S263	4, 2, 2	$(2,3,4,5)^T$	I	4	67	73	633	0.09375
			II	2	68	70	596	0.07813
S325	2, 2, 1	$(-6,2)^T$	I	3	70	74	456	0.07813
			II	3	70	71	456	0.07813
S388	15, 15, 0	$(1,1,1,0,1,$ $1,-1,1,0,1,$ $1,-1,1,0,1)^T$	I	2	85	88	8167	0.71875
			II	2	85	86	8167	0.76563

Table 2. *Numerical results report II*

Problem	Case	Approximately optimal solutions x^*	FV
HS7	I	$(-0.0534830568, 1.7291663148)^T$	$-1.726309961e+000$
	II	$(-0.0534830568, 1.7291663148)^T$	$-1.726309961e+000$
HS14	I	$(0.5687519350, 0.9254296615)^T$	$2.054031759e+000$
	II	$(0.5687519350, 0.9254296615)^T$	$2.054031759e+0000$
HS32	I	$(0.0371682801, 0.0075368556, 0.8969216752)^T$	$9.187879743e-001$
	II	$(0.0371682801, 0.0075368556, 0.8969216752)^T$	$9.187879743e-001$
HS63	I	$(3.5096586061, 0.2138185678, 3.5527528278)^T$	$9.617494273e+002$
	II	$(3.5096586061, 0.2138185678, 3.5527528278)^T$	$9.617494273e+002$
HS71	I	$(1.0059662906, 4.7402123667, 3.8175943055, 1.3831840154)^T$	$1.712497702e+001$
	II	$(1.0059662906, 4.7402123667, 3.8175943055, 1.3831840154)^T$	$1.712497702e+001$
HS81	I	$(-1.7117738550,$ $1.5797011174, 1.8384235442, -0.7726784340, -0.7725588868)^T$	$4.871255970e-002$
	II	$(-1.7332849499,$ $1.6033758012, 1.8004232295, -0.7589316072, -0.7792018683)^T$	$4.823824606e-002$
HS107	I	$(0.6684192680,$ $1.0231853764, 6.0736815151, 6.6540714895, 1.0786697141,$ $1.0881020368, 1.0225367370, 6.3890582958, 5.9319489086)^T$	$5.064389012e+003$
	II	$(0.6684192680,$ $1.0231853764, 6.0736815151, 6.6540714895, 1.0786697141,$ $1.0881020368, 1.0225367370, 6.3890582958, 5.9319489086)^T$	$5.064389012e+003$
HS111	I	$(-4.9007503864, -1.4977323296,$ $-0.3152789659, -4.7133495087, -0.7626199765, -4.8511504922,$ $-4.1268224932, -4.3445837123, -4.1380146813, -2.5506786613)^T$	$-4.520251036e+001$
	II	$(-4.9007503864, -1.4977323296,$ $-0.3152789659, -4.7133495087, -0.7626199765, -4.8511504922,$ $-4.1268224932, -4.3445837123, -4.1380146813, -2.5506786613)^T$	$-4.520251036e+001$
HS113	I	$(2.1746118403, 2.3382248588,$ $8.7632872772, 5.0919639082, 0.9907591989, 1.4664379848,$ $1.3477440525, 9.8437157499, 8.2600367376, 8.3248468009)^T$	$2.471145167e+001$
	I	$(2.1670578232, 2.3566298822,$ $8.7608601887, 5.0979658915, 0.9847258061, 1.4490308374,$ $1.3330475138, 9.8319523537, 8.2505676306, 8.3505841589)^T$	$2.473088409e+001$
S225	I	$(1.0018163479, 1.0020915800)^T$	$2.007823530e+000$
	I	$(2.6186092112, -1.6185752767)^T$	$9.476900127e+000$
S252	I	$(-7.5741433214, 50.3931658879, -2.5635328262)^T$	$4.937854686e+001$
	II	$(-7.1121058935, 50.6827952699, -2.4720431355)^T$	$6.682121813e-001$
S263	I	$(-0.1831041780, 0.0137339367, 0.2965821990, 0.2964327918)^T$	$1.831041780e-001$
	II	$(-0.1834079571, 0.0137749445, 0.2969584566, 0.2967796402)^T$	$1.834079571e-001$
S325	I	$(-2.3651245341, -1.8305786602)^T$	$3.763235402e+000$
	II	$(-2.3651245341, -1.8305786602)^T$	$3.763235402e+000$
S388	I	$(0.6948086601, 1.3427163786, 1.4264216459,$ $0.6820384904, 0.8325599882, 1.1838462680, -1.0407442392,$ $1.0470380672, -0.1263270590, 1.1094736897, 1.0352987701,$ $-1.1668929508, 0.6327427239, -0.3063730142, 0.8363253852)^T$	$-5.792050513e+003$
	II	$(0.6948086601, 1.3427163786, 1.4264216459,$ $0.6820384904, 0.8325599882, 1.1838462680, -1.0407442392,$ $1.0470380672, -0.1263270590, 1.1094736897, 1.0352987701,$ $-1.1668929508, 0.6327427239, -0.3063730142, 0.8363253852)^T$	$-5.792050513e+003$

Table 3. *Numerical results report for Svanberg problems*

Problem	$n, \|L_1\|, \|L_2\|$	Initial point x^0	Case	Nf_0	Nf	FV	$Time$	ρ_k	$N0$
Svanberg-10	10,30,0	$(0.5, 0.5, \ldots, 0.5)^T$	I	204	576	15.7418	0.9531	0.0093	200
			II	202	607	15.7436	0.9063	0.0145	200
Svanberg-20	20,60,0	$(0.5, 0.5, \ldots, 0.5)^T$	I	203	803	32.4507	1.8281	0.0277	200
			II	204	785	32.4507	1.7344	0.0305	200
Svanberg-30	30,90,0	$(0.5, 0.5, \ldots, 0.5)^T$	I	205	1003	49.1956	3.0156	0.1188	200
			II	203	1000	49.2113	2.8594	0.1220	200
Svanberg-40	40,120,0	$(0.5, 0.5, \ldots, 0.5)^T$	I	205	996	66.0257	4.1563	0.1368	200
			II	210	1010	66.0247	3.9844	0.1324	200
Svanberg-50	50,150,0	$(0.5, 0.5, \ldots, 0.5)^T$	I	214	1034	82.8483	5.7813	0.3332	200
			II	214	1034	82.8483	5.3438	0.3322	200
Svanberg-70	70,210,0	$(0.5, 0.5, \ldots, 0.5)^T$	I	308	1609	116.3661	14.4688	0.1748	300
			II	313	1606	116.4184	13.3906	0.3022	300
Svanberg-80	80,240,0	$(0.5, 0.5, \ldots, 0.5)^T$	I	309	1645	133.2518	18.2500	0.2575	300
			II	309	1645	133.2518	17.0156	0.2575	300
Svanberg-100	100,300,0	$(0.5, 0.5, \ldots, 0.5)^T$	I	409	2346	167.0583	36.7031	0.2202	400
			II	415	2315	167.0137	34.9844	0.2094	400
Svanberg-120	120,360,0	$(0.5, 0.5, \ldots, 0.5)^T$	I	524	2998	200.6651	65.9219	0.2341	500
			II	678	2975	200.6742	63.8125	0.2751	500
Svanberg-150	150,450,0	$(0.5, 0.5, \ldots, 0.5)^T$	I	700	3013	251.3467	107.2188	0.2985	500
			II	859	2990	251.3972	105.5781	0.3091	500

Table 4. *The comparison of numerical results between algorithms A and B*

Problem	$n, \|L_1\|, \|L_2\|$	Initial point x^0	Algorithm	INO	NII	x^*	FV	$Time$
Example 1	2, 3, 0	$(8,8)^T$	A(I)	0	15	$(0.5002, 0.2500)^T$	0.5003	0.0313
			A(II)	0	15	$(0.5002, 0.2500)^T$	0.5003	0.0313
			B	0	10	$(0.5191, 0.2405)^T$	0.5010	0.0499
		$(0,1)^T$	A(I)	1	13	$(0.4991, 0.2502)^T$	0.4995	0.0469
			A(II)	1	13	$(0.4991, 0.2502)^T$	0.4995	0.0469
			B	1	12	$(0.5239, 0.2382)^T$	0.5016	0.0499
		$(-11,2)^T$	A(I)	4	18	$(0.5012, 0.2489)^T$	0.4990	0.0313
			A(II)	4	18	$(0.5012, 0.2489)^T$	0.4990	0.0313
			B	5	14	$(0.5196, 0.2403)^T$	0.5010	0.0000
Example 2	2, 2, 0	$(0,0)^T$	A(I)	0	10	$(0.9978, 0.9988)^T$	1.0047	0.0156
			A(II)	0	10	$(0.9978, 0.9988)^T$	1.0047	0.0156
			B	0	4	$(0.9924, 0.9960)^T$	1.0151	0.0000
		$(-1,-1)^T$	A(I)	1	14	$(0.9978, 0.9988)^T$	1.0047	0.0469
			A(II)	1	14	$(0.9978, 0.9988)^T$	1.0047	0.0469
			B	1	4	$(0.9873, 0.9927)^T$	1.0255	0.0000
		$(1,-1)^T$	A(I)	1	8	$(0.9978, 0.9989)^T$	1.0047	0.0156
			A(II)	1	8	$(0.9978, 0.9988)^T$	1.0047	0.0156
			B	1	3	$(0.9925, 0.9971)^T$	1.0148	0.0000
Example 3	2, 2, 0	$(2,0)^T$	A(I)	3	10	$(0.9925, 0.0001)^T$	1.0001	0.0156
			A(II)	3	10	$(0.9925, 0.0001)^T$	1.0001	0.0156
			B	16	18	$(0.9662, 0.0030)^T$	1.0011	0.0500
		$(0,2)^T$	A(I)	2	15	$(0.9677, 0.0003)^T$	1.0010	0.0312
			A(II)	2	15	$(0.9677, 0.0003)^T$	1.0010	0.0312
			B	4	12	$(0.9318, -0.0034)^T$	1.0046	0.5000

Table 5. *The numerical results for the problem HS107 with different* ξ

ξ	Case	INO	NII	Nf_0	Nf	FV	ρ_k
1.001	I	7	53	61	5118	5.064389012e+003	0.0086386372
	II	7	53	54	5118	5.064389012e+003	0.0086386372
1.01	I	5	75	81	6923	5.064709691e+003	0.0088909245
	II	5	75	76	6923	5.064709691e+003	0.0088909245
1.1	I	100	0	100	6697	5.056168491e+003	7.2216953251e-004
	II	100	0	1	6697	5.056168491e+003	7.2216953251e-004
1.2	I	6	60	67	5406	5.064541747e+003	0.0087362419
	II	6	60	61	5406	5.064541747e+003	0.0087362419
1.0001	I	9	53	63	5439	5.064761883e+003	0.0089633417
	II	9	53	54	5439	5.064761883e+003	0.0089633417
1.00001	I	10	51	62	5416	5.064780905e+003	0.0089709173
	II	10	51	52	5416	5.064780905e+003	0.0089709173
1.5	I	7	60	68	5418	5.064365286e+003	0.0086244254
	II	7	60	61	5418	5.064365286e+003	0.0086244254
0	I	6	75	82	7765	5.064545528e+003	0.0086965515
	II	6	75	76	7765	5.064545528e+003	0.0086965515
0.5	I	5	74	80	7223	5.064714624e+003	0.0089308238
	II	5	74	75	7223	5.064714624e+003	0.0089308238

References

[1] I. Bongartz, A.R. Conn, H.I. Gould, PU.L. Toint, CUTE: Constrained and unconstrained testing environment, Association for Computing Machinery, *Trasactions on Mathematical Software* 21 (1995) 123–160.

[2] J.B. Jian, A superlinearly convergent feasible descent algorithm for nonlinear optimization, *Journal of Mathematics (PRC)* 15 (1995) 319–326.

[3] J.B. Jian, Strong combined Phase I-Phase II methods of sub-feasible directions, *Mathematics in Economics (PRC)* 12 (1995) 64–70.

[4] J.B. Jian, W.X. Chen and X.Y. Ke, Finitely convergent ϵ-generalized projection algorithm for nonlinear systems, *Journal of Mathematical Analysis and Applications* 332 (2007) 1445–145.

[5] J.B. Jian, Q.J. Xu and D.L. Han, A strongly conbergent norm-relaxed method of strongly sub-feasible direction for optimization with nonlinear equality and inequality constraints, *Applied Mathematics and Computation* 182 (2006) 854–870.

[6] J.B. Jian and K.C. Zhang, Strongly subfeasible direction method with strong convergence for inequality constrained optimization, *Journal of Xi'an Jiaotong University (PRC)* 33 (1999) 88–103

[7] J. Herskovits, A two-stage feasible directions algorithm for nonlinear constrained optimization, *Mathematical Programming* 36 (1986) 19–38.

[8] W. Hock and K. Schittkowski, *Tests Examples for Nonlinear Programming Codes*, Lecture Notes in Economics and Mathematical Systems, vol.187, Springer-Verlag, Berlin, New York, 1981.

[9] Y.L. Lai, Z.Y. Gao and G.P. He, A generalized gradient projection method for nonlinear optimization, *Science in China (Series A)*, 9 (1992) 916–924.

[10] C.T. Lawrence and A.L. Tits, Nonlinear equality constraints in feasible sequential quadratic programming, *Optimization Methods and Software* 6 (1996) 252–282.

[11] D.Q. Mayne, E. Polak, Feasible direction algorithm for optimization problems with equality and inequality constrains, *Mathematical Programming* 11 (1976) 67–80.

[12] E.R. Panier and A.L. Tits, A superlinearly convergence feasible method for the solution of inequality constrained optimization, *SIAM Journal on Control and Optimization* 25 (1987) 934–950.

[13] E.R. Panier and A.L. Tits, On combining feasibility, descent and superlinear convergence in equality constrained optimization, *Mathematical Programming* 59 (1993) 261–276.

[14] E. Polak, R. Trhan and D.Q. Mayne, Combined Phase I-Phaes II methods of feasible directions, *Mathematical Programming* 17 (1979) 61–73.

[15] J.B. Rosen, The gradient projection method for nonlinear programming, part I, linear constraints, *SIAM Journal on Applied Mathematics* 8 (1960) 181–217.

[16] K. Schittkowski, *More Test Examples for Nonlinear Programming Codes*, Sping-Verlag, 1987.

[17] Q.Y. Sun, Generalized super-memory gradient projection method with arbitrary initial point and conjugate gradient scalar for nonlinear programming with nonlinear inequality constraints, *Mathamatica Numerica Sinica (PRC)* 26 (2004) 401–412.

Manuscript received 15 January 2009
revised 25 April 2009
accepted for publication 25 April 2009

Jin-Bao Jian
College of Mathematics and Information Science, Guangxi University
Nanning, Guangxi 530004, P.R. China
E-mail address: `jianjb@gxu.edu.cn`

Chuan-Hao Guo
College of Mathematics and Information Science, Guangxi University
Nanning, Guangxi 530004, P.R. China
E-mail address: `guo-ch@163.com`

Lin-Feng Yang
College of Electrical Engineering, Guangxi University
Nanning, Guangxi 530004, P.R. China
E-mail address: `ylf@gxu.edu.cn`