Sequential System of Equation Method for Switching Constrained Optimization Problems

Yuting Sheng, Minglei Fang, Defeng Ding, and Ao Xu

Abstract—**The switching constrained optimization problem is a new class of constrained optimization problem proposed in recent years. However, its special constraints make the commonly used constraint specifications unsatisfactory. For this reason, the method for the sequential systems of linear equations (SSLE) is applied to solve the problem. In each iteration, the algorithm needs to solve a system of four linear equations with the same coefficient matrices, which reduces the amount of computation compared with the sequential quadratic programming algorithm. Moreover, it is proven that the limit point of the sequence generated by using the new algorithm is the Karush-Kuhn-Tucker point of the problem. Finally, it is shown through numerical results that the SSLE method is feasible for dealing with this type of problem.**

Index Terms—**Switching constrained, Sequential systems of linear equations, KKT point, Global convergence.**

I. Introduction

IN this study, the following general nonlinear programming problems are considered:

$$\begin{cases} \min \quad f(x) \\ \quad \text{s.t. } g_i(x) \le 0,\ i = 1, ...p, \\ \qquad\quad h_j(x) = 0,\ j = 1, ...q, \\ G_t(x)H_t(x) = 0,\ t = 1, ...l. \end{cases}$$

For the convenience of description, $I = \{i|i = 1, ...p\}$, $L = \{j|j = 1, ...q\}$, and $T = \{t|t = 1, ...l\}$; where, the functions $f(x) : \mathbb{R}^n \to \mathbb{R}^n$ are continuously differentiable and $g_i(x)$, $h_j(x)$, $G_t(x)$, $H_t(x)$: $\mathbb{R}^n \to \mathbb{R}^n$ are also continuously differentiable.

For any fixed t, at least one of $G_t(x)$ and $H_t(x)$ is zero. This constraint is called the switching constraint and denoted as the switching constrained optimization (mathematical program with switching constraints, MPSC) problem. This problem model was proposed and systematically studied by Mehlitz. Further, Mehlitz [1] pointed out that problems such as discretization of optimal control problems, either-or constraint optimization, 0-1 programming, and other problems could be solved by transforming them into MPSC problems. However, owing to its special constraint conditions, the commonly used constraint norms, such as Linear Independence Constraint Qualification (LICQ) and Mangasarian-Fromovitz Constraint Qualification (MFCQ), are not satisfied at the feasible point of the MPSC problem; thus, it cannot be regarded as a general nonlinear programming operation [2–4]. Luo et al. [5] proposed a relaxation method for solving the MPSC. Achtziger et al. [6] studied the second-order optimality conditions of MPSC, modified the standard Abadie constraint, and proved that the modified constraints could hold under relatively weak assumptions. Liang and Ye [7] presented research results on mathematical program with disjunctive constraints and optimality conditions, applied them to the MPSC problems, and listed two types of sufficient conditions for the local error bounds and exact penalty results for MPSC. Luo et al. [8] proposed a Wolfe-type dyadic model for the MPSC problem by using dyadic theory. Li and Guo [9] proposed the weakest constraint condition for Mordukhovich-stationarity of the MPSC problem at the local minima, and finally discussed the relationship among the MPSC customized constraints.

Sequential Quadratic Programming (SQP) algorithms exhibit good superlinear convergence and are therefore considered to be one of the most effective methods for solving nonlinear programming problems [10–12]. However, most SQP algorithms bring forward two serious drawbacks: (1) One or more quadratic programming sub-problems need to be solved in each iteration to obtain the search direction. This is computationally intensive and makes it difficult to utilize good sparsity and symmetry in the computation of the sub-problems. (2) The algorithm requires the sub-problems to be solvable at each iteration [13, 14]. Therefore, the following system of linear equations is utilized instead of the quadratic programming sub-problems.

In 1988, Panier et al. [15] proposed a feasible QP-free algorithm that was not able to solve quadratic programming sub-problems; since then, extensive research efforts have been devoted to the study on QP-free algorithms. These algorithms overcome the difficulty of solving sub-problems without solutions by replacing one or more quadratic programs in an SQP method with a number of systems of linear equations having the same coefficient matrices. They take full advantage of some of the benefits of solving systems of linear equations. These algorithms offer the advantages of less iteration time, smaller storage, and faster convergence, and they can be used to solve large-scale nonlinear optimization problems [16]. However, in addition to iteratively solving two linear equations in each step, Panier's algorithm also requires solving a quadratic programming sub-problem, which still requires a large amount of computation. Moreover, to make the algorithm converge to the Karush-Kuhn-Tucker (KKT) point, it must be further assumed that the number

Manuscript received May 8, 2024; revised September 30, 2024. This work was partly supported by the Key Program of the University Natural Science Research Fund of Anhui Province under Grant KJ2021 A0451 and Anhui Provincial Natural Science Foundation under Grant No. 2008085MA01.

Y. T. Sheng is a postgraduate student at the School of Mathematics and Big Data, Anhui University of Science and Technology, Huainan, 232001, P. R. China (e-mail: 805015354@qq.com).

M. L. Fang is a professor at the School of Mathematics and Big Data, Anhui University of Science and Technology, Huainan, 232001, P. R. China (corresponding author to provide e-mail: fmlmath@aust.edu.cn).

D. F. Ding is a postgraduate student at the School of Mathematics and Big Data, Anhui University of Science and Technology, Huainan, 232001, P. R. China (e-mail:2021201353@aust.edu.cn).

A. Xu is a postgraduate student at the School of Mathematics and Big Data, Anhui University of Science and Technology, Huainan, 232001, P. R. China (e-mail: 2734825169@qq.com.)

of stabilization points is limited. To overcome the above-mentioned shortcomings, Gao Ziyou et al. [17] proposed a new algorithm in 1994. This algorithm only needs to solve three systems of linear equations with the same coefficients in each iteration, which improves computational efficiency. The one-step superlinear convergence of the algorithm is also proven, which completely eliminates the difficulty of solving quadratic programming sub-problems. Therefore, herein, it is denoted as the Sequential Systems of Linear Equations (SSLE).

This study improves on a prior algorithm from the literature [12], and proposes the SSLE algorithm for solving switching constraints optimization problems. Compared with the SQP algorithm, the new algorithm offers the following advantages: (1) Each iteration only requires solving four sets of linear equations with the same coefficients. Thus the new algorithm is QP-free, and involves reduced computational complexity. (2) The iteration points produced by the new algorithm are valid. Based on this, global convergence and the algorithm are proven to be benign. Finally, the algorithm is feasible when combined with the results of numerical experiments.

II. A New Class of SSLE Algorithms

The sequence $\{d_k^0\}$ is formed from the subsequent linear system:

$$B_k d_k^0 + \nabla f(x_k) + \sum_{i=1}^{p} \lambda_{k,i}^0 \nabla g_i(x_k) + \sum_{j=1}^{q} \lambda_{k,j}^0 \nabla h_j(x_k) + \sum_{t=1}^{l} \lambda_{k,t}^0 (\nabla G_t(x_k) H_t(x_k) + \nabla H_t(x_k) G_t(x_k)) = 0, \quad (1)$$

$$\mu_{k,i} \nabla g_i(x_k)^{\mathrm{T}} d_k^0 + \lambda_{k,i}^0 g_i(x_k) = 0,\ i \in I, \quad (2)$$

$$\nabla h_j(x_k)^{\mathrm{T}} d_k^0 + h_j(x_k) = 0,\ j \in L, \quad (3)$$

$$(\nabla G_t(x_k) H_t(x_k) + \nabla H_t(x_k) G_t(x_k))^{\mathrm{T}} d_k^0 + G_t(x_k) H_t(x_k) = 0,\ t \in T, \quad (4)$$

where B_k is an approximation of the Lagrangian function $L(x, \lambda_i, \lambda_j, \lambda_t)$,

$$L(x, \lambda_i, \lambda_j, \lambda_t) = f(x) + \sum_{i=1}^{p} \lambda_i g_i(x) + \sum_{j=1}^{q} \lambda_j h_j(x) + \sum_{t=1}^{l} \lambda_t G_t(x) H_t(x).$$

Where x_k is the estimated value of x^*, $x_k + d_k^0$ denotes the next estimated value, μ_k represents the current estimate of the multiplier vector related to x^*, while λ_k^0 denotes the subsequent estimated value. According to the analysis in Theorem 4.6 [15], the expression $x_{k+1} = x_k + d_k^0$ exhibits superlinear convergence; therefore, d_k^0 is chosen as the initial direction. However, a negative number v_i is added to the right of system (2) since d_k^0 may be zero in some iteration of KKT points that are not MPSC problems. Moreover, to maintain the convergence of $\{x_k\}$, v_i must approach zero faster than d_k^0. Therefore, let $v_i = (\lambda_{k,i}^0)^3$ or 0, so that after obtaining d_k^0, a new direction d_k^1 can be obtained by solving the linear system mentioned below.

The sequence $\{d_k^1\}$ is formed from the subsequent linear system:

$$B_k d_k^1 + \nabla f(x_k) + \sum_{i=1}^{p} \lambda_{k,i}^1 \nabla g_i(x_k) + \sum_{j=1}^{q} \lambda_{k,j}^1 \nabla h_j(x_k) + \sum_{t=1}^{l} \lambda_{k,t}^1 (\nabla G_t(x_k) H_t(x_k) + \nabla H_t(x_k) G_t(x_k)) = 0, \quad (5a)$$

$$\mu_{k,i} \nabla g_i(x_k)^{\mathrm{T}} d_k^1 + \lambda_{k,i}^1 g_i(x_k) = \mu_i v_{k,i}, \quad (5b)$$

$$\nabla h_j(x_k)^{\mathrm{T}} d_k^1 + h_j(x_k) = 0, \quad (5c)$$

$$(\nabla G_t(x_k) H_t(x_k) + \nabla H_t(x_k) G_t(x_k))^{\mathrm{T}} d_k^1 + G_t(x_k) H_t(x_k) = 0, \quad (5d)$$

The following Lemma 5 shows that d_k^1 is a strict descent direction of the penalty function $W_r(x)$. However, if one wants each of $g_i(x)$, $h_j(x)$, $G_t(x)$, and $H_t(x)$ to be very close to 0, then the above-mentioned system makes d_k^1 tend toward the direction where the feasible set is tangent, as follows:

$$X = \{x \mid g_i(x) \le 0, i \in I\}.$$

This may lead to step size collapse due to the need for feasibility of all iterations. Therefore, search direction d_k is used as follows.

The sequence $\{d_k\}$ is formed from the subsequent linear system:

$$B_k(d_k - d_k^1) + \sum_{i=1}^{p} (\lambda_{k,i} - \lambda_{k,i}^1) \nabla g_i(x_k) + \sum_{t=1}^{l} (\lambda_{k,t} - \lambda_{k,t}^1)(\nabla G_t(x_k) H_t(x_k) + \nabla H_t(x_k) G_t(x_k)) + \sum_{j=1}^{q} (\lambda_{k,j} - \lambda_{k,j}^1) \nabla h_j(x_k) = 0, \quad (6a)$$

$$\mu_{k,i} \nabla g_i(x_k)^{\mathrm{T}} (d_k - d_k^1) + (\lambda_{k,i} - \lambda_{k,i}^1) g_i(x_k) = -\rho_k \|d_k^1\|^\eta \mu_{k,i}, \quad (6b)$$

$$\nabla h_j(x_k)^{\mathrm{T}} (d_k - d_k^1) = 0, \quad (6c)$$

$$(\nabla G_t(x_k) H_t(x_k) + \nabla H_t(x_k) G_t(x_k))^{\mathrm{T}} (d_k - d_k^1) = 0, \quad (6d)$$

where ρ_k is a specific positive number, and the basic convergence of sequence $\{x_k\}$ remains unchanged by this search direction. However, as shown in prior study [18] when using exact penalty functions in line search SQP iterations, the unit step size might fall short of being close enough to the solution to be satisfactory. Therefore, to avoid this situation, it is necessary to solve a new linear system.

The sequence $\{\hat{d}_k\}$ is formed from the subsequent linear system:

$$B_k(\hat{d}_k - d_k) + \sum_{i=1}^{p} (\hat{\lambda}_{k,i} - \lambda_{k,i}) \nabla g_i(x_k) + \sum_{t=1}^{l} (\hat{\lambda}_{k,t} - \lambda_{k,t})(\nabla G_t(x_k) H_t(x_k) + \nabla H_t(x_k) G_t(x_k)) + \sum_{j=1}^{q} (\hat{\lambda}_{k,j} - \lambda_{k,j}) \nabla h_j(x_k) = 0, \quad (7a)$$

$$\mu_{k,i} \nabla g_i(x_k)^{\mathrm{T}} (\hat{d}_k - d_k) + (\hat{\lambda}_{k,i} - \lambda_{k,i}) g_i(x_k) = -\phi_k + \mu_{k,i} q_{k,i}, \quad (7b)$$

$$\nabla h_j(x_k)^{\mathrm{T}}(\hat{d}_k - d_k) + h_j(x_k + d_k) = 0, \tag{7c}$$

$$(\nabla G_t(x_k)H_t(x_k) + \nabla H_t(x_k)G_t(x_k))^{\mathrm{T}}(\hat{d}_k - d_k) + G_k(x_k + d_k)H_k(x_k + d_k) = 0. \tag{7d}$$

Where $q_{k,i} = g_i(x_k + d_k)$ or 0; and ϕ_k is a positive number that can ensure further bending of d_k, thus obtaining the feasibility of $x_k + d_k$. The algorithm, global convergence, and numerical experiments are detailed below.

In this study, the symbols given below are used.

$$X = \{x \mid g_i(x) \le 0, i \in I\},$$
$$G = \mathrm{diag}[g_1(x), ...g_p(x), 0, ...0] \in E^{(p+q+l)\times(p+q+l)},$$
$$A = [\nabla g_1(x), ...\nabla g_p(x), \nabla h_1(x), ...\nabla h_q(x), (\nabla G_1(x)H_1(x) + \nabla H_1(x)G_1(x)), ...(\nabla G_l(x)H_l(x) + \nabla H_l(x)G_l(x))],$$
$$g^L(x) = (0, ...0, h_1(x), ...h_q(x), G_1(x)H_1(x), ...G_l(x) H_l(x))^{\mathrm{T}} \in E^{(p+q+l)},$$

$e_1 = (1, ...1) \in E^p, \quad e_2 = (0, ...0) \in E^{q+l}$,
then,

$$e = \begin{pmatrix} e_1 \\ e_2 \end{pmatrix} \in E^{p+q+l}.$$

The following penalty function is used:

$$W_r(x) = f(x) + r\sum_{j=1}^{q}|h_j(x)| + r\sum_{t=1}^{l}|G_t(x)H_t(x)|. \tag{8}$$

Where r is the penalty parameter. A convex function $\overline{W}_r(x, d)$ about d is defined as follows:

$$\begin{aligned}\overline{W}_r(x) =& f(x) + \nabla f(x)^T d + r\sum_{j=1}^{q}|h_j(x) + \nabla h_j(x)^T d| \\ &+ r\sum_{t=1}^{l}|G_t(x)H_t(x) + (\nabla G_t(x)H_t(x) \\ &+ \nabla H_t(x)G_t(x))^T d|.\end{aligned} \tag{9}$$

Herein, a set on x is defined as follows: $J_0(x) = \{i | g_i(x) = 0,\ i \in I\}$. In this article, $||\cdot||$ represents the Euclidean norm. In prior study [2], it was pointed out that the stationary point (KKT point) of the MPSC problem is:

$$\begin{aligned}&\nabla f(x^*) + \sum_{J_0}\lambda_i\nabla g_i(x^*) + \sum_{j}\lambda_j\nabla h_j(x^*) \\ &+ \sum_{t}\lambda_t(\nabla G_t(x^*)H_t(x^*) + \nabla H_t(x^*)G_t(x^*)), \\ &\lambda_i \ge 0,\ \lambda_i(x^*)g_i(x^*) = 0, \\ &h_j(x^*) = 0,\ G_t(x^*)H_t(x^*) = 0.\end{aligned}$$

Algorithm SSLE

Input Set parameters $\alpha \in (0, 0.5)$, $\beta \in (0, 1)$, $\theta \in (0, 1)$, $\eta > 2$, $\tau \in (2, 3)$, $\gamma \in (0, 1)$, $\overline{\mu} > 0$, select initial value $x_0 \in X$, $B_0 \in E^{n\times n}$ is a symmetric positive definite matrix, $0 < \mu_{0,i} \le \overline{\mu}$, $k = 0$;

Step 1 Computation of a search direction

1.1 Let (d_k^0, λ_k^0) be the solution of the following linear equation:

$$F(x_k, B_k, \mu_k)\begin{bmatrix} d \\ \lambda \end{bmatrix} = -\begin{bmatrix} \nabla f(x_k) \\ g^L(x_k) \end{bmatrix},$$

where,

$$F(x, B, \mu) = \begin{bmatrix} B & A \\ MA^T & G \end{bmatrix}, \tag{10}$$

$$M = \mathrm{diag}(\mu_1, ...\mu_p, 1, ...1) \in E^{(p+q+l)\times(p+q+l)}, \tag{11}$$

If $d_k^0 = 0$, and $\lambda_{k,i}^0 \ge 0$, then stop.

1.2 Let (d_k^1, λ_k^1) be the solution of the following linear equation:

$$F(x_k, B_k, \mu_k)\begin{bmatrix} d \\ \lambda \end{bmatrix} = \begin{bmatrix} -\nabla f(x_k) \\ -g^L(x_k) + M_k v_k \end{bmatrix},$$

where,

$$v_k = (v_{k,1}, ...v_{k,p+q+l})^{\mathrm{T}}, \tag{12a}$$

$$v_{k,m} = \begin{cases} (\lambda_{k,i}^0)^3, & \text{if } \lambda_{k,i}^0 \le 0, \\ 0, & \text{otherwise.} \end{cases} \tag{12b}$$

1.3 Let (d_k, λ_k) be the solution of the following linear equation:

$$F(x_k, B_k, \mu_k)\begin{bmatrix} d - d_k^1 \\ \lambda - \lambda_k^1 \end{bmatrix} = -\rho_k||d_k^1||^\eta\begin{bmatrix} 0 \\ M_k e \end{bmatrix},$$

where,

$$\rho_k = \frac{[(\theta - 1)(\overline{W}_r(x_k; d_k^1) - \overline{W}_r(x_k; 0))]}{[|(\overline{\lambda_k})^T e| \cdot ||d_k^1||^\eta + 1]}. \tag{13}$$

Note: $\overline{\lambda_k}$ is the solution to the equation presented below.

$$F(x_k, B_k, \mu_k)\begin{bmatrix} d \\ \lambda \end{bmatrix} = \begin{bmatrix} -\nabla f(x_k) \\ 0 \end{bmatrix}.$$

1.4 Let $(\hat{d}_k, \hat{\lambda}_k)$ be the solution of the following linear equation:

$$F(x_k, B_k, \mu_k)\begin{bmatrix} d - d_k \\ \lambda - \lambda_k \end{bmatrix} = -\begin{bmatrix} 0 \\ \phi_k e + q_k \end{bmatrix},$$

where,

$$q_k = (q_{k,1}, ...q_{k,p+q+l}), \tag{14a}$$

$$q_{k,m} = \begin{cases} \mu_{k,i}g_i(x_k + d_k), & m \in I_k, \\ 0, & m \in I/I_k, \\ h_j(x_k + d_k), & m \in L, \\ G_t(x_k + d_k)H_t(x_k + d_k) & m \in T. \end{cases} \tag{14b}$$

$$I_k = \{i \in I \mid -\lambda_{k,i}^0 \le g_i(x_k)\}, \tag{15}$$

$$\phi_k = \max\left\{||d_k||^\tau, \max\left\{|\mu_{k,i/\lambda_{k,i}-1}|^\gamma \cdot ||d_k||^2\right\}\right\} \tag{16}$$

If $I_k = \emptyset$ or $||\hat{d}_k - d_k|| > ||d_k||$, set $\hat{d}_k = d_k$.

Step 2 Let t_k satisfy

$$\begin{aligned}&W_r(x_k + td_k + t^2(\hat{d}_k - d_k)) \\ &\le W_r(x) + \alpha t[\overline{W}_r(x_k; d_k) - \overline{W}_r(x_k; 0)],\end{aligned} \tag{17}$$

$$g_i(x_k + td_k + t^2(\hat{d}_k - d_k)) \le 0, \tag{18}$$

Step 3 Let B_{k+1} be a new approximation of B_k,

$$\mu_{k+1,i} = \min\left\{\max\left\{\lambda_{k,i}^0, ||d_k||, \overline{\mu}\right\}\right\},$$

$$x_{k+1} = x_k + t_k d_k + t_k^2(\hat{d}_k - d_k),$$

$k = k + 1$, return to Step 1.

III. Global Convergence of the New SSLE Algorithm

The following assumptions are made to prove the global convergence of algorithm:

Assumption 1: $X = \{x|g_i(x) \leq 0,\ i \in I\}$ is not empty and compact.

Assumption 2: The functions $f(x)$, $g_i(x)$, $h_j(x)$, $G_t(x)$ and $H_t(x)$, are continuously differentiable.

Assumption 3: For $\forall x \in X$, the vectors $\nabla g_i(x), i \in J_0(x)$, $\nabla h_j(x)$, $j \in L$, $\nabla G_t(x)$, $t \in T$, and $\nabla H_t(x)$, $t \in T$, are linearly independent.

Assumption 4: There exists a constant $\kappa > 0$ such that for all k and for all $y \in E^n$, the inequality $(1/\kappa)||y||^2 \leq y^{\mathrm{T}} B_k y \leq \kappa ||y||^2$ holds.

Assumption 5: The sequences $\{\lambda_k^0\}$ and $\{\lambda_k^1\}$ are both bounded, and the penalty parameter r is sufficiently large to satisfy the following relationships:

$$
\begin{aligned}
&r > \sup_k \{\max_{j\in L} \{|2\lambda_{k,j}^0 - \lambda_{k,j}^1|\}\},\\
&r > \sup_k \{\max_{t\in T} \{|2\lambda_{k,t}^0 - \lambda_{k,t}^1|\}\};\\
&r > \sup_k \{\max_{j\in L} \{|\lambda_{k,j}^0|\}\},\\
&r > \sup_k \{\max_{k\in T} \{|\lambda_{k,t}^0|\}\}.
\end{aligned}
$$

Lemma 1: Given any $x \in X$, any positive definite matrix B and any non-negative vector $\mu \in E^m$, such that $\mu_i > 0$, for $\forall i \in J_0(x)$, then $F(x, B, \mu)$ is nonsingular and

$$
F(x,B,\mu)^{-1} = \begin{bmatrix} P & Z \\ MZ^{\mathrm{T}} & -D^{-1} \end{bmatrix},
$$

where,

$$
\begin{aligned}
&D = MA^{\mathrm{T}}B^{-1}A - G,\\
&Z = B^{-1}AD^{-1},\\
&P = B^{-1} - B^{-1}AMZ^{\mathrm{T}}.
\end{aligned}
$$

Lemma 2: Let $x_k \in X$ for $k \in N^+$, and assume that for some subsets $K, \{x_k\}_K \to x^*$. Let $\mu_k \in E^m, \mu_k > 0, \{\mu_k\}_K \to \mu^*$ and $\mu_i^* > 0, \forall i \in J_0(x)$. If $\{B_k\}_K$ is a positive definite matrix that satisfies Assumption 4, then $\{||F(x_k, B_k, \mu_k)^{-1}||\}_K$ is bounded.

Lemma 3: Assume $x \in X$, and B be a positive definite matrix. If $\mu > 0$, then $M^{-1}D = A^{\mathrm{T}}B^{-1}A - M^{-1}G$ is positively definite, and $y^{\mathrm{T}}Py \geq ||H^{1/2}Py||^2$ holds for $\forall y \in E^n$.

From Lemma 1, if $\mu_k > 0$ holds for each k, (LS1)-(LS4) have unique solutions. Thus, it follows from (LS1) to (LS3) that

$$
\begin{aligned}
&d_k^0 = -P_k \nabla f(x_k) - Z_k g^L(x_k),\\
&\lambda_k^0 = -M_k Z_k^{\mathrm{T}} \nabla f(x_k) + D_k^{-1} g^L(x_k).
\end{aligned} \tag{19}
$$

$$
d_k^1 = d_k^0 + Z_k M_k v_k, \lambda_k^1 = \lambda_k^0 - D_k^{-1} M_k v_k. \tag{20}
$$

$$
\begin{aligned}
&d_k = d_k^1 - \rho_k ||d_k^1||^\eta Z_k M_k e,\\
&\lambda_k = \lambda_k^1 + \rho_k ||d_k^1||^\eta D_k^{-1} M_k e.
\end{aligned} \tag{21}
$$

$$
\overline{\lambda}_k = -M_k Z_k^{\mathrm{T}} \nabla f(x_k). \tag{22}
$$

Lemma 4: If Algorithm stops at a point $x_k \in X$ such that $d_k^0 = 0,\ \lambda_{k,i}^0 \geq 0,\ i \in I$, x_k is the KKT point of the MPSC problem.

Lemma 5: If $x_k \in X$ is not a KKT point and $\mu_k > 0$, then d_k^1 satisfies

1) $\overline{W}_r(x_k; d_k^1) - \overline{W}_r(x_k; 0) < 0$,
2) $\nabla g_i(x_k)^{\mathrm{T}} d_k^1 \leq 0,\ i \in J_0(x_k)$.

Proof.

1) From (9) and (LS2), we have

$$
\begin{aligned}
&\overline{W}_r(x_k; d_k^1)\\
&= f(x_k) + \nabla f(x_k)^{\mathrm{T}} d_k^1 + r\sum_{j=1}^{q} |h_j(x_k)\\
&\quad + \nabla h_j(x_k)^{\mathrm{T}} d_k^1| + r\sum_{t=1}^{l} |G_t(x_k)H_t(x_k)\\
&\quad + \nabla(G_t(x_k)H_t(x_k) + \nabla H_t(x_k)G_t(x_k))^{\mathrm{T}} d_k^1|.\\
&\overline{W}_r(x_k; 0)\\
&= f(x_k) + r\sum_{j=1}^{q} |h_j(x_k)| + r\sum_{t=1}^{l} |G_t(x_k)H_t(x_k)|.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\overline{W}_r(x_k; d_k^1) - \overline{W}_r(x_k; 0)\\
&= \nabla f(x_k)^{\mathrm{T}} d_k^1 + r\sum_{j=1}^{q} |\nabla h_j(x_k)^{\mathrm{T}} d_k^1|\\
&\quad + r\sum_{t=1}^{l} |(\nabla G_t(x_k)H_t(x_k) + G_t(x_k)\nabla H_t(x_k))^{\mathrm{T}} d_k^1|\\
&= \nabla f(x_k)^{\mathrm{T}} d_k^1 - r\sum_{j=1}^{q} |h_j(x_k)|\\
&\quad + r\sum_{t=1}^{l} |G_t(x_k)H_t(x_k)|.
\end{aligned} \tag{23}
$$

Equations (19)-(21) show that

$$
\begin{aligned}
&\overline{W}_r(x_k; d_k^1) - \overline{W}_r(x_k; 0)\\
&= \nabla f(x_k)^{\mathrm{T}} d_k^0 + \nabla f(x_k)^{\mathrm{T}} Z_k M_k v_k - r\sum_{j=1}^{q} |h_j(x_k)|\\
&\quad - r\sum_{t=1}^{l} |G_t(x_k)H_t(x_k)|\\
&= -(d_k^0)^{\mathrm{T}} B_k d_k^0 - (\lambda_k^0)^{\mathrm{T}} A_k^{\mathrm{T}} d_k^0 + (-\lambda_k^0 + D_k^{-1} g^L(x_k))^{\mathrm{T}}\\
&\quad v_k - r\sum_{j=1}^{q} |h_j(x_k)| - r\sum_{t=1}^{l} |G_t(x_k)H_t(x_k)|\\
&= -(d_k^0)^{\mathrm{T}} B_k d_k^0 - (\lambda_k^0)^{\mathrm{T}} (-M_k^{-1} g^L(x_k)\\
&\quad - M_k^{-1} G_k \lambda_k^0) + (-\lambda_k^0 + D_k^{-1} g^L(x_k))^{\mathrm{T}} v_k\\
&\quad - r\sum_{j=1}^{q} |h_j(x_k)| - r\sum_{t=1}^{l} |G_t(x_k)H_t(x_k)|.
\end{aligned}
$$

Since $M_k g^L(x_k) = g^L(x_k)$ and $M_k^{-1} G_k$ is negative definite, then from (20) and Assumptions 4 and 5, it can be obtained that

$$\begin{aligned}
&\overline{W}_r(x_k;d_k^1)-\overline{W}_r(x_k;0)\\
&\le -(1/\kappa)||d_k^0||^2+(\lambda_k^0)^{\mathrm{T}}g^L(x_k)\\
&\quad -(\lambda_k^0)^{\mathrm{T}}v_k+(g^L(x_k))^{\mathrm{T}}M_kD_k^{-\mathrm{T}}v_k\\
&\quad -r\sum_{j=1}^{q}|h_j(x_k)|-r\sum_{t=1}^{l}|G_t(x_k)H_t(x_k)|\\
&=-(1/\kappa)||d_k^0||^2-(\lambda_k^0)^{\mathrm{T}}v_k\\
&\quad +(g^L(x_k))^{\mathrm{T}}(\lambda_k^0+M_kD_k^{-\mathrm{T}}v_k)\\
&\quad -r\sum_{j=1}^{q}|h_j(x_k)|-r\sum_{t=1}^{l}|G_t(x_k)H_t(x_k)|\\
&=-(1/\kappa)||d_k^0||^2-(\lambda_k^0)^{\mathrm{T}}v_k+(g^L(x_k))^{\mathrm{T}}(2\lambda_k^0-\lambda_k^1)\\
&\quad -r\sum_{j=1}^{q}|h_j(x_k)|-r\sum_{t=1}^{l}|G_t(x_k)H_t(x_k)|\\
&\le -(1/\kappa)||d_k^0||^2-\sum_{i\in I,\lambda_{k,i}^0\le 0}(\lambda_{k,i}^0)^4\\
&\quad -(r-\max\{|2\lambda_{k,j}^0-\lambda_{k,j}^1|\})(\sum_{j=1}^{q}|h_j(x_k)|)\\
&\quad -(r-\max\{|2\lambda_{k,t}^0-\lambda_{k,t}^1|\})(\sum_{t=1}^{l}|G_t(x_k)H_t(x_k)|)\\
&<0. \qquad (24)
\end{aligned}$$

2) From (LS2), when $i\in J_0(x_k)$, then

$$\nabla g_i(x_k)^{\mathrm{T}}d_k^1=-(\lambda_{k,i}^1/\mu_{k,i})g_i(x_k)+v_{k,i}\le 0. \quad (25)$$

□

Lemma 6: If $x_k\in X$ is not the KKT point of the MPSC problem, $\mu_k>0$, then d_k satisfies

1) $\overline{W}_r(x_k;d_k)-\overline{W}_r(x_k;0)\le \theta(\overline{W}_r(x_k;d_k^1)-\overline{W}_r(x_k;0))<0$.
2) $\nabla g_i(x_k)^{\mathrm{T}}d_k<0,\ i\in J_0(x_k)$.
3) $d_k\ne 0,\ \mu_{k+1}>0$.

Proof.

1) From Equations (9), (21), and (23), since $(\nabla G_t(x_k)H_t(x_k)+G_t(x_k)\nabla H_t(x_k))^{\mathrm{T}}(d_k-d_k^1)=0$, $(\nabla h_j(x_k))^{\mathrm{T}}(d_k-d_k^1)=0$, then

$$\begin{aligned}
&\overline{W}_r(x_k;d_k)-\overline{W}_r(x_k;0)\\
&=\nabla f(x_k)^{\mathrm{T}}d_k-r\sum_{j=1}^{q}|h_j(x_k)|-r\sum_{t=1}^{l}|G_t(x_k)H_t(x_k)|\\
&=\nabla f(x_k)^{\mathrm{T}}d_k^1-r\sum_{j=1}^{q}|h_j(x_k)|-r\sum_{t=1}^{l}|G_t(x_k)H_t(x_k)|\\
&\quad -\rho_k||d_k^1||^\eta\nabla f(x_k)^{\mathrm{T}}Z_kM_ke\\
&=\overline{W}_r(x_k;d_k^1)-\overline{W}_r(x_k;0)\\
&\quad -\rho_k||d_k^1||^\eta\nabla f(x_k)^{\mathrm{T}}Z_kM_ke. \qquad (26)
\end{aligned}$$

According to the definitions of $\overline{\lambda}_k$ and ρ_k,

$$\begin{aligned}
&-\rho_k||d_k^1||^\eta\nabla f(x_k)^{\mathrm{T}}Z_kM_ke\\
&\le(\theta-1)(\overline{W}_r(x_k;d_k^1)-\overline{W}_r(x_k;0)).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\overline{W}_r(x_k;d_k)-\overline{W}_r(x_k;0)\\
&\le\overline{W}_r(x_k;d_k^1)-\overline{W}_r(x_k;0)+(\theta-1)\overline{W}_r(x_k;d_k^1)\\
&\quad -(\theta-1)\overline{W}_r(x_k;0)\\
&\le\theta(\overline{W}_r(x_k;d_k^1)-\overline{W}_r(x_k;0))<0. \qquad (27)
\end{aligned}$$

2) From (LS3), it can be obtained that

$$M_kA_k^{\mathrm{T}}\left(d_k-d_k^1\right)+G_k\left(\lambda_k-\lambda_k^1\right)=-\rho_k||d_k^1||^\eta M_ke.$$

Therefore,

$$M_kA_k^{\mathrm{T}}d_k=M_kA_k^{\mathrm{T}}d_k^1-G_k\left(\lambda_k-\lambda_k^1\right)-\rho_k||d_k^1||^\eta M_ke,$$

that is,

$$A_k^{\mathrm{T}}d_k=A_k^{\mathrm{T}}d_k^1-M_k^{-1}G_k\left(\lambda_k-\lambda_k^1\right)-\rho_k||d_k^1||^\eta e.$$

Therefore,

$$\mu_{k,i}\nabla g_i(x_k)^{\mathrm{T}}(d_k-d_k^1)=-(\lambda_{k,i}-\lambda_{k,i}^1)g_i(x_k)-\rho_k||d_k^1||^\eta\mu_{k,i}.$$

$$\nabla g_i(x_k)^{\mathrm{T}}(d_k^-d_k^1)=\frac{-(\lambda_{k,i}-\lambda_{k,i}^1)}{\mu_{k,i}}g_i(x_k)-\rho_k||d_k^1||^\eta.$$

and from (25), it can be known that

$$\nabla g_i(x_k)^{\mathrm{T}}d_k=\nabla g_i(x_k)^{\mathrm{T}}d_k^1-\rho_k||d_k^1||^\eta. \quad (28)$$

For $\forall i\in J_0(x_k)$, $\nabla g_i(x_k)^{\mathrm{T}}d_k<0, i\in J_0(x_k)$.

3) Herein, it can be derived from (2) and the definitions of μ_{k+1}.

□

Lemma 7: Algorithm is well defined.

Proof. Given a constant k such that $0<t<1$ let $\varpi_k=x_k+td_k+t^2(\hat{d}_k-d_k)$, it follows that

$$\begin{aligned}
&W_r(\varpi_k)-W_r(x_k)\\
&=f(\varpi_k)+r\sum_{j=1}^{q}|h_j(\varpi_k)|+r\sum_{t=1}^{l}|G_t(\varpi_k)H_t(\varpi_k)|\\
&\quad -f(x_k)-r\sum_{j=1}^{q}|h_j(x_k)|-r\sum_{t=1}^{l}|G_t(x_k)H_t(x_k)|\\
&=f(\varpi_k)-f(x_k)+r\sum_{j=1}^{q}(|h_j(\varpi_k)|-|h_j(x_k)|)\\
&\quad +r\sum_{t=1}^{l}(|G_t(\varpi_k)H_t(\varpi_k)|-|G_t(x_k)H_t(x_k)|)\\
&=f(x_k)+\nabla f(x_k)^{\mathrm{T}}(\varpi_k)-f(x_k)\\
&\quad +r\sum_{j=1}^{q}(|h_j(x_k)+\nabla h_j(x_k)^{\mathrm{T}}(\varpi_k)|-|h_j(x_k)|)\\
&\quad +r\sum_{t=1}^{l}(|G_t(x_k)H_t(x_k)+(\nabla G_t(x_k)H_t(x_k)\\
&\quad +G_t(x_k)\nabla H_t(x_k))^{\mathrm{T}}(\varpi_k)|)-|G_t(x_k)H_t(x_k)|)\\
&=t\nabla f(x_k)^{\mathrm{T}}d_k+r\sum_{j=1}^{q}(|h_j(x_k)+t\nabla h_j(x_k)^{\mathrm{T}}d_k|\\
&\quad -|h_j(x_k)|)+r\sum_{t=1}^{l}(|G_t(x_k)H_t(x_k)\\
&\quad +G_t(x_k)\nabla H_t(x_k))^{\mathrm{T}}d_k|-|G_t(x_k)H_t(x_k)|)+o(t)\\
&=(\overline{W}_r(x_k;td_k)-\overline{W}_r(x_k;0))+o(t). \qquad (29)
\end{aligned}$$

Moreover, since $\overline{W}_r(x_k;d)$ is a convex function about d, it follows that

$$\overline{W}_r(x_k;td_k)-\overline{W}_r(x_k;0)\le t(\overline{W}_r(x_k;d_k)-\overline{W}_r(x_k;0)). \tag{30}$$

From (27), when t is small enough, then

$$\begin{aligned}&W_r(x_k+td_k+t^2(\hat{d}_k-d_k))-W_r(x_k)\\&\le \alpha t(\overline{W}_r(x_k;d_k)-W_r(x_k;0)).\end{aligned}$$

Moreover, by Lemma 6 (2), (18) always holds when t is small enough, and the conclusion is thus proven.

□

Lemma 8: The sequences $\{x_k\}$ and $\{d_k^0\}$ are bounded.

Proof. The boundedness of $\{x_k\}$ follows directly from Assumption 1. Following (LS1), Assumptions 4 and 5, and $M_k^{-1}G_k$ is negative definite:

$$\begin{aligned}&(1/\kappa)||d_k^0||^2-||\nabla f(x_k)||\cdot||d_k^0||\\&\le \left(d_k^0\right)^{\mathrm{T}}B_kd_k^0+\nabla f(x_k)^{\mathrm{T}}d_k^0\\&=-(\lambda_k^0)^{\mathrm{T}}A_k^{\mathrm{T}}d_k^0\\&=(\lambda_k^0)^{\mathrm{T}}M_k^{-1}(G_k\lambda_k^0+g^L(x_k))\\&\le\sum_{j=1}^{q}\lambda_{k,j}^0h_j(x_k)+\sum_{t=1}^{l}\lambda_{k,t}^0G_t(x_k)H_t(x_k)\\&\le r\sum_{j=1}^{q}|h_j(x_k)|+\sum_{t=1}^{l}|G_t(x_k)H_t(x_k)|.\end{aligned}$$

Thus, $\{d_k^0\}$ is bounded.

□

Lemma 9: Let x^* be the limit point of the sequence $\{x_k\}$ generated by Algorithm, which is $\{x_k\}_K\to x^*$. If $\{d_k\}_K\to 0$, then x^* is the KKT point of the MPSC problem, and $\{\lambda_k^0\}_K\to\lambda^*$, where λ^* denotes the unique Lagrange multiplier vector related to x^*. Furthermore, if $\lambda_i^*\le\bar{\mu},\ i\in I$, there is $\{\mu_{k+1,i}\}_K\to\lambda_i^*,\ i\in I$.

Proof. From (24) and (27):

$$\begin{aligned}&0\leftarrow\overline{W}_r(x_k;d_k)-\overline{W}_r(x_k;0)\\&\le\theta(\overline{W}_r(x_k;d_k^1)-\overline{W}_r(x_k;0))\\&\le\theta(-\frac{1}{\kappa}||d_k^0||^2-\sum_{\lambda_{k,i}^0\le 0}(\lambda_{k,i}^0)^4\\&-(r-\max\{|2\lambda_{k,j}^0-\lambda_{k,j}^1|\})(\sum|h_j(x_k)|)\\&-(r-\max\{|2\lambda_{k,t}^0-\lambda_{k,t}^1|\})(\sum|G_t(x_k)H_t(x_k)|))\\&<0\end{aligned}$$

for $k\in K,k\to\infty$. □

Lemma 10: Let x^* be the limit point of the sequence $\{x_k\}$, which is $\{x_k\}_K\to x^*$. If $\inf\{||d_{k-1}||\}_K=0$, then x^* is the KKT point of the MPSC problem.

Lemma 11: If $\{x_k\}_K\to x^*$, $\{\mu_k\}_K\to\mu^*$, and $\mu_i^*>0$, hold for $\forall i\in J_0(x)$, then $\{d_k^1\}_K$ and $\{\hat{d}_k-d_k\}_K$ are all bounded.

Proof. Similar to Lemma 8, it can be concluded that $\{d_k^1\}_K$ is bounded, and from Step 1.4 of Algorithm, it can be inferred that $\{\hat{d}_k-d_k\}_K$ is bounded.

□

Lemma 12: Let x^* be the limit point of the sequence $\{x_k\}$ generated by Algorithm, which is $\{x_k\}_K\to x^*$, if $\inf\{||d_{k-1}||\}_K>0$, then $\{d_k^0\}_K\to 0$.

Proof. Assuming there exists a $\bar{d}>0$ such that

$$x_k\to x^*,k\in K,k\to\infty.$$

$$||d_k^0||\ge\bar{d},\forall k\in K.$$

Then, from (24), it is ensured that there exists a number $\underline{d}>0$, an infinite subset $K'\subset K$ such that

$$||d_k^1||\ge\underline{d},\forall k\in K'.$$

For $\forall\underline{d}>0$, if there is a k that makes $||d_k^1||<\underline{d}$, then there exists a subset $K''\subset K$ that makes $\{d_k^0\}_{K''}\to 0$. Then, as shown in (24), $\{d_k^0\}_{K''}\to 0$ contradicts the hypothesis.

From the boundedness of $\{||B_k||\}$ and $\{\mu_k\}$, it can be assumed that

$$\{B_k\}_{K'}\to B^*,\{\mu_k\}_{K'}\to\mu^*.$$

Based on the definition of μ_{k+1}, the assumptions stated in the Lemma, it follows that all the components of μ^* are strictly positive. Therefore, it follows from Lemma 2 and 11 that $\{d_k\}_{K'}$ is bounded.

Next, it must first be proven that step t_k and $k\in K'$ obtained from line search are bounded away from zero; that is,

$$\exists\bar{t}>0,\ \text{s.t.}\quad t_k\ge\bar{t},\ \forall k\in K'. \tag{31}$$

Since (d_k,λ_k) and (d_k^1,λ_k^1) are solutions of (LS2) and (LS3), respectively,
thus,

$$\begin{aligned}A_k^{\mathrm{T}}d_k&=A_k^{\mathrm{T}}d_k^1-M_k^{-1}G_k\left(\lambda_k-\lambda_k^1\right)-\rho_k||d_k^1||^\eta e\\&=v_k-M_k^{-1}g^L(x_k)-M_k^{-1}G_k\lambda_k-\rho_k||d_k^1||^\eta e.\end{aligned} \tag{32}$$

Therefore, for $\forall i\in I$, it follows that

$$\begin{aligned}&\nabla g_i(x_k)^{\mathrm{T}}d_k\\&=v_{k,i}-\mu_{k,i}^{-1}g_i(x_k)\lambda_{k,i}-\rho_k||d_k^1||^\eta\\&=v_{k,i}-\mu_{k,i}^{-1}g_i(x_k)\lambda_{k,i}\\&\quad-\{\frac{(\theta-1)(\bar{W}_r(x_k;d_k^1)-\bar{W}_r(x_k;0))}{[|(\bar{\lambda}_k)^{\mathrm{T}}e|\cdot||d_k^1||^\eta+1]}\}||d_k^1||^\eta.\end{aligned} \tag{33}$$

Moreover, $|(\bar{\lambda}_k)^{\mathrm{T}}e|=|\nabla f(x_k)Z_kM_ke|$ and $\{d_k^1\}_{K'}$ are bounded.

Thus, it follows from (24) that

$$\overline{W}_r(x_k;d_k^1)-\overline{W}_r(x_k;0)\le-(1/\kappa)\bar{d}^2<0. \tag{34}$$

Therefore, according to the definition of v_k, there exists $\delta_1>0$, with the following inequalities:

$$\nabla g_i(x_k)^{\mathrm{T}}d_k\le-\delta_1,\ i\in J_0(x^*), \tag{35}$$

$$g_i(x_k)\le-\delta_1,\ i\in I\backslash J_0(x^*), \tag{36}$$

which hold for all sufficiently large $k\in K'$.

Then, let $\varpi_k=x_k+td_k+t^2(\hat{d}_k-d_k)$, from the constant equation

$$\begin{aligned}g_i(\varpi_k)=g_i(x_k)+\int_0^1 t\nabla(g_i(x_k+t\xi d_k+t^2\xi^2(\hat{d}_k-d_k)))^{\mathrm{T}}\\(d_k+2t\xi(\hat{d}_k-d_k))d\xi.\end{aligned}$$

It can be seen that, when k is large enough, $k \in K'$, and $i \in J_0(x^*)$, $\xi_k = x_k + t\xi d_k + t^2\xi^2(\hat{d}_k - d_k)$

$$\begin{aligned}
&g_i(\varpi_k) - g_i(x_k) \\
&= t(\int_0^1 [\nabla g_i \xi_k{}^{\mathrm{T}}(d_k + 2t\xi(\hat{d}_k - d_k)) - \nabla g_i(x_k)^{\mathrm{T}}]d\xi \\
&\quad + \nabla g_i(x_k)^{\mathrm{T}} d_k) \\
&\leq t(\sup_{0\leq\xi\leq 1}\{||\nabla g_i(\xi_k)) - \nabla g_i(x_k)|| \cdot ||d_k||\} \\
&\quad + 2t \sup_{0\leq\xi\leq 1}\left\{||\nabla g_i(\xi_k))|| \cdot ||\hat{d}_k - d_k||\right\} - \delta_1).
\end{aligned}$$

Since $g_i(x) \in C^1$, the existence of $t_j > 0$, which is independent of k, makes k large enough for, $t \in [0, t_j]$, $k \in K'$, and

$$g_i(x_k + td_k + t^2(\hat{d}_k - d_k)) \leq 0,\ i \in J_0(x^*).$$

Simultaneously, $g_i(x)$ is continuous, and the existence of $t_j > 0$, independent of k, makes k large enough for, $t \in [0, t_j]$, $k \in K'$, and

$$g_i(x_k + td_k + t^2(\hat{d}_k - d_k)) \leq -\delta_1/2,\ i \in I\backslash J_0(x^*).$$

Furthermore, it follows from (29) that

$$\begin{aligned}
&W_r(x_k + td_k + t^2(\hat{d}_k - d_k)) - W_r(x_k) \\
&= (\overline{W}_r(x_k; td_k) - \overline{W}_r(x_k; 0)) + o(t^2 \cdot ||\hat{d}_k - d_k||) \\
&\quad + o(t^2 \cdot ||d_k + t(\hat{d}_k - d_k)||^2).
\end{aligned}$$

Since $|(\bar{\lambda}_k)^{\mathrm{T}} e|$, $\{d_k^1\}_{K'}$, and $\{d_k\}_{K'}$ are bounded, for any point $t \in [0, 1]$, it follows from Lemma 11 that both $||d_k + t(\hat{d}_k - d_k)||$ and $||\hat{d}_k - d_k||$ are bounded. Moreover, based on (27), (30) and (34) we can conclude that there is a $t_0 > 0$, such that for $t \in [0, t_0]$, $k \in K'$, and sufficiently large k, it follows that

$$\begin{aligned}
&W_r(x_k + td_k + t^2(\hat{d}_k - d_k)) \\
&\leq W_r(x_k) + \alpha t\left[\overline{W}_r(x_k; d_k) - \overline{W}_r(x_k; 0)\right].
\end{aligned}$$

The proof of Equation (31) is thus completed.

Let $\bar{t} = \min\{t_0, t_i\}$, for $k \in K'$, when k is large enough, there is

$$W_r(x_{k+1}) \leq W_r(x_k) - (1/\kappa)\,\theta\alpha\bar{t}\bar{d}^2. \tag{37}$$

Moreover, since $\{W_r(x_k)\}$ is monotonically decreasing sequence and $W_r(x)$ is continuous, it follows that $W_r(x_k) \rightarrow W_r(x^*)$, $k \rightarrow \infty$. Thus, the assumption does not hold, which contradicts Equation (37); that is, Lemma 12 is proven.

□

Theorem 1: If the algorithm stops at the KKT point of the MPSC problem or generates an infinite sequence $\{x_k\}$, then its limit point is the KKT point of the MPSC problem.

Proof. Assuming that $\{x_k\}$ represents an infinite sequence and $\{x_k\}_K \rightarrow x^*$. From Lemma 10, only the following situation needs to be considered, which is $\inf\{||d_{k-1}||\}_K > 0$.

At this point, from Lemma 12, $\{d_k^0\}_K \rightarrow 0$, $\{h_j(x_k)\}_K \rightarrow 0$, $j \in L$ and $\{G_t(x_k)H_t(x_k)\}_K \rightarrow 0$, $t \in T$ from (LS1). According to Lemma 12, it can be assumed that all components of $\{B_k\}_K \rightarrow B^*$, $\{\mu_k\}_K \rightarrow \mu^*$ and μ^* are strictly positive.

First, if $\{\overline{W}_r(x_k; d_k^1) - \overline{W}_r(x_k; 0)\}_K \rightarrow 0$, then, due to $\{d_k^0\}_K \rightarrow 0$, $\{|h_j(x_k)|\}_K \rightarrow 0$, and $\{|G_t(x_k)H_t(x_k)|\}_K \rightarrow 0$, it is available from (LS1) that

$$\sum_{i\in I, \lambda_{k,i}^0 \leq 0} (\lambda_{k,i}^0)^4 \rightarrow 0.$$

Therefore, from Lemma 8, there exists some infinite subset $K' \subseteq K$ such that for $\forall i \in I$,

$$\{\lambda_{k,i}^0\}_{K'} \rightarrow \lambda_i^* \geq 0. \tag{38}$$

Then, from (LS1), it follows that

$$\begin{aligned}
&\nabla f(x^*) + \sum_{J_0} \lambda_i \nabla g_i(x^*) + \sum_j \lambda_j \nabla h_j(x^*) \\
&+ \sum_t \lambda_t(\nabla G_t(x^*)H_t(x^*) + \nabla H_t(x^*)G_t(x^*)), \\
&\lambda_i \geq 0,\ \lambda_i(x^*)g_i(x^*) = 0, \\
&h_j(x^*) = 0,\ G_t(x^*)H_t(x^*) = 0.
\end{aligned}$$

Then, x^* is the KKT point of the MPSC problem.

Now suppose there exists a positive number $\xi_1 > 0$ such that,

$$\overline{W}_r(x_k; d_k^1) - \overline{W}_r(x_k; 0) \leq -\xi_1 < 0,\ ||d_k^1|| \geq \xi_1 > 0,$$

hold for $k \in K$.

Through a method comparable to that found in Lemma 12, there exists a $\underline{t}$ such that for $k \in K$, when k is large enough,

$$W_r(x_{k+1}) \leq W_r(x_k) - \theta\alpha\underline{t}\xi_1.$$

This contradicts monotonic decrease in $W_r(x_k)$ and $W_r(x_k) \rightarrow W_r(x^*)$, and thus the theorem is proven.

□

IV. Numerical Results

This study uses Python as the main processing tool during the numerical experiment, and the testing environment is Intel (R) Pentium (R) Silver N5000 CPU @ 1.10GHz 1.10GHz. The experimental parameters used are:

$\alpha = 1/10,\ r = 10000,\ \beta = 1/2,\ \theta = 0.99,\ \tau = 2.99,\ \eta = 2.1,\ \bar{\mu} = 1, \gamma = 1/10,$

The following examples aid in analyzing the results of numerical experiments on the algorithm of the system of sequential equations via using three optimization problems with switching constraints.

Example 1

$$\begin{aligned}
\min\quad & (x_1 - 2)^2 + (x_2 - 1)^2 + (x_3 - 2)^2, \\
\text{s.t.}\quad & x_1^2 + x_2^2 + x_3^2 \leq 3, \\
& x_3 \leq 1, \\
& (x_1 - x_2^2)\ (x_2 - x_1^2) = 0.
\end{aligned}$$

Among them, (1,1,1) is the global optimal solution.

Example 2

$$\begin{aligned}
\min\quad & x_1 + x_2^2, \\
\text{s.t.}\quad & -x_1 + x_2 \leq 0, \\
& x_1 x_2 = 0.
\end{aligned}$$

Among them, (0,0) is the global optimal solution.

Example 3

$$\begin{aligned}
\min\quad & x_1 + x_2, \\
\text{s.t.}\quad & x_1^2 - x_2 \leq 0, \\
& x_1 x_2 = 0.
\end{aligned}$$

TABLE I: Experimental Results of MPSC Problems

example	x^*	NF	x'	NF$'$	NI
1	(1,1,1)	2	(1.000 002 69 , 0.999 983 12 , 0.999 999 98)	1.999 994 66	12
2	(0,0)	0	$(9.793 \times 10^{-7}, 0)$	9.793×10^{-7}	16
3	(0,0)	0	$(9.255 \times 10^{-7}, 0)$	9.255×10^{-7}	15

Among them, (0,0) is the global optimal solution.

Numerical experiments were conducted on the above-mentioned examples, and the corresponding results are listed in Table 1. In Table 1, the first column represents each example, x^* represents the exact solution of the problem, NF represents the function calculations, x' represents the approximate solution of the problem, NF$'$ indicates the approximate value of the problem, and NI represents the number of iterations.

Table 1 demonstrates that the approximate solution of the problem could be successfully obtained through the sequence equation system algorithm. It indicates that the sequence systems of linear equations algorithm are feasible for addressing MPSC problems.

V. Conclusions

In this study, the methods for solving optimization problems with switching constraints were studied. First, based on the algorithm in a literature report, the SSLE algorithm for the MPSC problem was designed. In each iteration, only four equations with the same coefficient matrix were solved to generate the main search direction. Then, by using reasonable assumptions, the global convergence of the new algorithm was proven. The limit point of the sequence generated by the new algorithm was the KKT point of the problem. Finally, the results of numerical experiments indicate that the newly proposed algorithm effectively solves MPSC problems. This study provides a new approach for solving MPSC problems. However, whether the corresponding conclusions and convergence effects can be obtained under weaker conditions remains an area for further investigation.

References

[1] P. Mehlitz, “Stationarity conditions and constraint qualifications for mathematical programs with switching constraints,” Mathematical Programming, vol. 181, no. 1, pp. 149-186, 2019.

[2] M. L. Luo, G. X. Li, Y. Q. Huang, and L. Y. Liu, “SQP Methods for Mathematical Programs With Switching Constraints,” Applied Mathematics and Mechanics, vol. 43, no. 7, pp. 792-801, 2022.

[3] Xiang-Jing Liu, Jian-Ke Zhang, and Jun-Feng Chen, “A Modified Smoothing Newton Method for Solving Weighted Complementarity Problems with a Nonmonotone Line Search,” Engineering Letters, vol. 31, no. 4, pp1684-1690, 2023

[4] De-Feng Ding, Ming-Lei Fang, Min Wang, and Yu-Ting Sheng, “A Modulus-Based Shamanskii-Like Levenberg-Marquardt Method for Solving Nonlinear Complementary Problems,” IAENG International Journal of Applied Mathematics, vol. 54, no. 3, pp411-416, 2024

[5] Z. Q. Luo, J. S. Pang, and D. Ralph, Mathematical programs with equilibrium constraints, First ed. Cambridge: Cambridge University Press, 1996.

[6] Achtziger, Wolfgang, Kanzow, and Christian, “Mathematical programs with vanishing constraints: optimality conditions and constraint qualifications,” Mathematical Programming, vol. 114, no. 1, pp. 69-99, 2008.

[7] Y. C. Liang and J. J. Ye, “New optimality conditions and exact penalty for mathematical programs with switching constraints,” Journal of Optimization Theory and Applications, vol. 190, no. 1, pp. 1-31, 2021.

[8] M. L. Luo, G. X. Li, and C. Wu, “Duality Problem of Mathematical Programs with Swtiching constraints,” Journal of Mathematics, vol. 43, no. 4, pp. 347-355, 2023.

[9] G. X. Li and L. Guo, “Mordukhovich stationarity for mathematical programs with switching constraints under weak constraint qualifications,” Optimization, vol. 72, no. 7, pp. 1817-1838, 2023.

[10] J. B. Jian, “Two Extension Models of SQP and SSLE Algorithms for Optimization and Their Superlinear and Quadratic Convergence,” Applied Mathematics A Journal of Chinese Universities(Ser.A), vol. 16, no. 4, pp. 435-444, 2001.

[11] Z. Y. Gao, F. Wu, and Y. L. Lai, “A Superlinearly Convergent Algorithm of the Sequential Systems of Linear Equations for Nonlinear Optimization Problems,” Chinese Science Bulletin, vol. 39, no. 23, pp. 1946-1951, 1994.

[12] Z. Y. Gao, G. P. He, and F. Wu, “Sequential systems of linear equations algorithm for nonlinear optimization problems with general constraints,” Journal of Optimization Theory and Applications, vol. 95, no. 2, pp. 371-397, 1997.

[13] W. L. Liao, Q. L. Zhang, and H. L. Meng, “An SQP Algorithm for Structural Topology Optimization Based on Majorization-Minimization Method,” Applied Sciences, vol. 12, no. 13, pp. 6304-6304, 2022.

[14] R. Yousefpour and E. Jafari, “An SQP method for minimization of locally Lipschitz functions with nonlinear constraints,” Optimization, vol. 68, no. 4, pp. 731-751, 2019.

[15] E. R. Panier, “Tits A L, Herskovits J N. A QP-free, globally convergent, locally superlinearly convergent algorithm for inequality constrained optimization,” SIAM Journal on Control and Optimization, vol. 26, no. 4, pp. 788-811, 1988.

[16] K. Su, L. L. Ren, and X. H. Ren, “A modified nonmonotone QP-free method without penalty function or filter,” Journal of Intelligent Fuzzy Systems, vol. 38, no. 3, pp. 3331-3342, 2019.

[17] Z. Y. Gao, F. Wu, and Y. L. Lai, “A Super Linear Convergent Sequential Equation System Method for Nonlinear Optimization,” Chinese Science Bulletin, vol. 39, no. 9, pp. 774-777, 1994.

[18] N. Maratos, Exact Penalty Function Algorithms for Finite-Dimensional and Optimization Problems, PhD Thesis, Imperial College of Science and Technology, London, England, 1978.