OPTIMAL SWITCHING SIGNAL DESIGN WITH A COST ON SWITCHING ACTION

WEI XU, LIYING YU AND GUI-HUA LIN*
School of Management, Shanghai University, Shanghai, China

ZHI GUO FENG
Faculty of Mathematics and Computer Science, Guangdong Ocean University
Zhanjiang, Guangdong, China
College of Mathematics Science, Chongqing Normal University
Chongqing, China

(Communicated by Bin Li)

Abstract. In this paper, we consider a particular class of optimal switching problem for the linear-quadratic switched system in discrete time, where an optimal switching sequence is designed to minimize the quadratic performance index of the system with a switching cost. This is a challenging issue and studied only by few papers. First, we introduce a total variation function with respect to the switching sequence to measure the volatile switching action. In order to restrain the switching magnitude, it is added to the cost functional as a penalty. Then, the particular optimal switching problem is formulated. With the positive semi-definiteness of matrices, we construct a series of exact lower bounds of the cost functional at each time and the branch and bound method is applied to search all global optimal solutions. For the comparison between different global optimization methods, some numerical examples are given to show the efficiency of our proposed method.

1. Introduction. In practical applications, the dynamics of some hybrid system may be classified as a switched system, in which a group of subsystems are included and only one of them is active at each time. For example, the product process switches between the batch mode and the feeding mode in the microbial fermentation [17], the voltage is operated by using capacitors and switches in the DC/DC power converter [19] and some other engineering problems in [4, 5, 13, 14], and so on. Therefore, as a particular class of hybrid system, the switched system has been a hot and challenging research topic in recent several decades.

Based on the structural characteristic of the switched system, its state can be controlled by some external inputs and a switching law, which schedules these active subsystems at different times. Thus, a variety of optimal control problems mathematicians and engineers face are produced. Since there are many factors in the switching law, such as the switching frequency, the switching sequence and the switching time, researchers usually only consider the continuous event where for

2020 Mathematics Subject Classification. 49M25.
Key words and phrases. Global optimal solution, optimal switching problem, switching cost, branch and bound.

* Corresponding author: Gui-Hua Lin.
a predetermined switching sequence, the optimal inputs and the optimal switching time are obtained. In [16, 12, 26], some numerical techniques based on the gradient-based method and its variants are proposed. But in practice, the switching sequence may not be known in advance. Then the optimal switching problem is very important and becomes more practical in switched systems. In order to deal with the discrete event in continuous time, the two-stage optimization method is proposed in [21], where the switching sequence is obtained by the dynamic programming method in the second stage. Furthermore, these subsystems can be merged in [2, 27] into a large dynamic system with some weighting coefficients. Based on the relaxation method, the switching sequence can be obtained by solving a standard nonlinear programming problem. In [1], a mode-insertion algorithm is presented to update the switching sequence. Then the method is extended by [22, 3] such that it can efficiently determine the switching time and the switching sequence simultaneously. As another class of switched system, the discrete-time switched system has also been extensively studied in many fields. We refer readers to [10, 20, 25] for more details. Since the gradient information does not exist, the switching sequence generally can only be obtained by the enumeration or the dynamic programming method. However, the computational complexity may grow exponentially as the number of switching times increases. Thus some special optimization techniques, such as the relaxation method [11], the discrete filled function method [6], the branch and bound method [7, 23, 24] and the semi-definite programming method [8, 15], are explored.

Moreover, it is undesirable and difficult to be implemented for the controller when the control inputs and the switching sequence fluctuate frequently. In order to restrain the volatile changes, a switching penalty is added in [8] into the cost functional and the optimal control problem can be handled in discrete time. In [18, 9], the changing cost of control inputs based on the gradient-based optimization method is studied in continuous time.

In this paper, we consider a discrete-time autonomous switched system, where these subsystems are linear and the cost functional is given by a quadratic performance index and a total variation function with respect to the state and the switching sequence, respectively. The branch and bound method is applied to solve this particular optimal switching problem. Different from [23] where the regularization parameter approaches to zero in the lower bound, we design an exact lower bound such that it is always performable for any nonnegative regularization parameter. Then, we need not to evaluate the sufficient small parameter any more. Some numerical experiments show that the exact lower bound is more stable to search all global optimal solutions in the branch and bound strategy.

The rest of this paper is organized as follows. In Section 2, we present the optimal switching problem of the discrete-time switched system with a switching cost and some total variation expressions with respect to the switching sequence in different practices. In Section 3, the exact lower bound of the cost functional is constructed. Based on lower bounds at different times, we propose an exact branch and bound method to search all global optimal switching sequences. Compared with other global optimization methods, two numerical experiments are tested in Section 4. In Section 5, we conclude this paper.

2. Problem formulation. Let’s assume that there are $M$ subsystems denoted by an index set $\mathcal{M} = \{1, 2, \ldots, M\}$ in the switched system and only one of them is
active at each time. Then we formulate the linear discrete-time switched system as

\[
\begin{cases}
  x(t+1) = A_{\sigma(t)}(t)x(t), \\
  x(0) = x_0.
\end{cases}
\]

(1)

Here, for each time, \(x(t) \in \mathbb{R}^n\) is the state, the subscript \(\sigma(t) \in \mathcal{M}\) represents the switching signal which is in charge of activating the corresponding subsystem. For any \(m \in \mathcal{M}\), \(\sigma(t) = m\) means that the \(m\)th subsystem is active at time \(t\) and \(A_m(t) \in \mathbb{R}^{n \times n}\) is a time-varying matrix. \(x_0 \in \mathbb{R}^n\) is a given initial state.

For the dynamics (1), we aim to find a switching sequence composed of all switching signals such that the system achieves the optimal performance in a given time horizon. However, frequent and dramatic changes among these active subsystems can generate a volatile switching sequence. It is difficult to be implemented in practice and some potential risks may be caused. In order to restrain the switching vibration, we introduce a total variation function with respect to the switching signals such that the system achieves the optimal performance in a given time horizon. We describe its total variation function by

\[
\sigma = \sum_{t=1}^{T-1} |\sigma(t-1) - \sigma(t)|
\]

to measure the change magnitude of all switching signals.

Now let us consider the following linear-quadratic optimal switching problem with a cost on switching action

\[
\begin{align*}
\min_{\sigma \in \mathcal{M}^T} & \quad J(\sigma) = x^T(T)\Phi x(T) + \sum_{t=0}^{T-1} x^T(t)Q_{\sigma(t)}(t)x(t) + \alpha \bigg( \sum_{t=1}^{T-1} |\sigma(t-1) - \sigma(t)| \bigg) \\
\text{s.t.} & \quad x(t+1) = A_{\sigma(t)}(t)x(t), \quad t = 0, 1, \ldots, T-1, \\
& \quad x(0) = x_0,
\end{align*}
\]

(2)

where the performance matrix of the \(m\)th subsystem is denoted by \(Q_m(t) \in \mathbb{R}^{n \times n}\). For any time \(t \in \mathcal{T}\) and \(m = 1, 2, \ldots, M\), these performance matrices are all time-varying positive semi-definite. \(\Phi \in \mathbb{R}^{n \times n}\) is the given final performance matrix of the system, which is also positive semi-definite. In the cost functional (2), \(\alpha > 0\) is the penalty parameter which penalizes the change of switching signals.

Since the number of all switching scenarios is finite at each time, the total variation function \(\bigvee_{t=1}^{T-1} \sigma\) is bounded. Its lower bound and upper bound are 0 and \((T-1)(M-1)\), respectively. If the switching signal \(\sigma(t)\) is monotonous with respect to \(t\), such as \(\sigma(0) \leq \sigma(1) \leq \ldots \leq \sigma(T-1)\), the total variation function is \(\bigvee_{t=1}^{T-1} \sigma = |\sigma(0) - \sigma(T-1)|\). In particular, when these switching signals are constant, we have \(\bigvee_{t=1}^{T-1} \sigma = 0\). The more frequently the system switches, the greater the total variation function is. So the switching frequency and the switching magnitude can be restrained by minimizing the total variation term in the cost functional (2).

**Remark 2.1.** For the switching sequence, when we just consider the switching frequency and do not care about the switching magnitude, the total variation function...
can be defined by
\[
\sum_{t=1}^{T-1} \chi(\sigma(t-1), \sigma(t)),
\]
where $\chi: \mathcal{M}^2 \to \{0,1\}$ is the indicator function as
\[
\chi(a,b) = \begin{cases} 
0, & a = b, \\
1, & a \neq b.
\end{cases}
\]
This kind of total variation function has been discussed in [8]. We can find that the value of the total variation function directly shows the number of switchings.

3. **Method analysis.** Since the switching sequence is a set of discrete values and the gradient-based method can not be applied in the discrete-time switched system. Moreover, there is a penalty with respect to the switching sequence $\sigma$ in the cost functional, the relaxation method may not be executable. This is a typical NP-compete problem, the global optimal solution is generally obtained by the enumeration. As a simplified tree search, the branch and bound method is a considerable optimization technique to solve the discrete optimization problem. However, it is necessary that we must construct a lower bound as a rule to prune branches at each time first.

In the switching sequence $\sigma = (\sigma(0), \sigma(1), \ldots, \sigma(T-1)) \in \mathcal{M}^T$, we assume that its first $\tau$ switching signals are known, where $1 \leq \tau \leq T-1$. That is, $\sigma(0), \sigma(1), \ldots, \sigma(\tau-1)$ are fixed. Then in these active subsystems, the performance matrices $Q_{\sigma(0)}(0), Q_{\sigma(1)}(1), \ldots, Q_{\sigma(\tau-1)}(\tau-1)$ are given. Based on the switched dynamics (1), these states $x(1), x(2), \ldots, x(\tau)$ are also determined. For the remainder of the switching sequence, $\sigma(\tau), \sigma(\tau+1), \ldots, \sigma(T-1)$ are not confirmed which can take any values from the index set $\mathcal{M}$. Thus we divide the cost functional (2) into two parts such that

\[
J(\sigma | \sigma(0), \sigma(1), \ldots, \sigma(\tau-1))
= \left( \sum_{t=1}^{\tau-1} x^T(t)Q_{\sigma(t)}(t)x(t) + \alpha \sum_{t=1}^{\tau-1} \chi(t) \right) + \\
\left( x^T(\tau)\Phi x(\tau) + \sum_{t=\tau}^{T-1} x^T(t)Q_{\sigma(t)}(t)x(t) + \alpha \sum_{t=\tau}^{T-1} \chi(t) \right)
= \left( \sum_{t=0}^{\tau-1} x^T(t)Q_{\sigma(t)}(t)x(t) + \alpha \sum_{t=1}^{\tau-1} \chi(t) \right)
+ \\
\left( x^T(\tau)\Phi x(\tau) + \sum_{t=\tau}^{T-1} \left( x^T(t)Q_{\sigma(t)}(t)x(t) + \alpha |\sigma(t-1) - \sigma(t)| \right) \right).
\]

For the cost functional (3), the total variation function is denoted by zero when $\tau = 1$ in the first part, while the second part is undetermined. Then, we need to construct its lower bound for all feasible switching signals $\sigma(\tau), \sigma(\tau+1), \ldots, \sigma(T-1) \in \mathcal{M}$ at current time $t = \tau - 1$.

For this, we analyze some properties of the positive semi-definite matrix. In [23], a simple way is proposed to construct a greater diagonal matrix for the symmetric matrix. In fact, the conclusion is also established without the symmetry. Now we have the following theorem and refer readers to the proof of Theorem 1 in [23] for details.
Theorem 3.1. For any matrix \( P = [P_{ij}] \in \mathbb{R}^{n \times n} \), there exists a diagonal matrix \( \text{diag}[\lambda_i] \in \mathbb{R}^{n \times n} \) such that

\[
\lambda_i = \sum_{j=1}^{n} |P_{ij}|, \quad \forall \ i = 1, 2, \ldots, n.
\]

Then we have \( P \preceq \text{diag}[\lambda_i] \). That is, \( \text{diag}[\lambda_i] - P \in \mathbb{R}^{n \times n} \) is positive semi-definite.

Corollary 3.1. Given an invertible matrix \( \Psi = [\Psi_{ij}] \in \mathbb{R}^{n \times n} \), we can obtain a positive definite matrix \( \Gamma \in \mathbb{R}^{n \times n} \) such that \( \Gamma \preceq \Psi \).

Proof. We denote the inverse of the matrix \( \Psi \) by \( \hat{\Psi} \in \mathbb{R}^{n \times n} \). Clearly, for any \( i = 1, 2, \ldots, n \), there exists a subscript \( j \in \{1, 2, \ldots, n\} \) such that \( \hat{\Psi}_{ij} \neq 0 \).

Then based on Theorem 3.1, we can obtain a diagonal matrix \( \text{diag}[\hat{\lambda_i}] \in \mathbb{R}^{n \times n} \) satisfying the relation \( \hat{\Psi} \preceq \text{diag}[\hat{\lambda_i}] \), where

\[
\hat{\lambda}_i = \sum_{j=1}^{n} |\hat{\Psi}_{ij}| > 0, \quad \forall \ i = 1, 2, \ldots, n.
\]

Hence, let \( \Gamma = \text{diag}[\frac{1}{\hat{\lambda}_i}] \), it holds that \( 0 < \Gamma \preceq \hat{\Psi}^{-1} = \Psi \).

This completes the proof. \( \square \)

Let us consider the undetermined part of the cost functional (3). For any time \( \tau + 1 \leq t \leq T - 1 \), we have

\[
x^T(t)Q_{\sigma(t)}(t)x(t)
= x^T(\tau)A^T_{\sigma(\tau)}(\tau) \cdots A^T_{\sigma(t-1)}(t-1)Q_{\sigma(t)}(t)A_{\sigma(t-1)}(t-1) \cdots A_{\sigma(\tau)}(\tau)x(\tau).
\]

Based on the positive semi-definiteness of the matrix, we give the following theorem.

Theorem 3.2. Assume that there are \( M \) time-varying positive semi-definite matrices \( Q_1(t), Q_2(t), \ldots, Q_M(t) \in \mathbb{R}^{n \times n} \) and \( M \) time-varying real matrices \( A_1(t), A_2(t), \ldots, A_M(t) \in \mathbb{R}^{n \times n} \). For the time \( \tau \leq t \leq T - 1 \), we can obtain a smallest diagonal matrix \( \text{diag}[\eta^{\varepsilon,\tau}(t)] \in \mathbb{R}^{n \times n} \) such that, for any \( \sigma(\tau), \sigma(t+1), \ldots, \sigma(t) \in \mathcal{M} \),

\[
\text{diag}[\eta^{\varepsilon,\tau}(t)]
\preceq \begin{cases} \quad \left\{ \begin{array}{ll} Q_{\sigma(t)}(\tau), & \text{if } t = \tau, \\ A^T_{\sigma(\tau)}(\tau) \cdots A^T_{\sigma(t-1)}(t-1)Q_{\sigma(t)}(t)A_{\sigma(t-1)}(t-1) \cdots A_{\sigma(\tau)}(\tau), & \text{otherwise}, \end{array} \right. \end{cases}
\]

where \( \tau \geq 1 \) and \( \varepsilon > 0 \) is a regularization parameter.

Proof. For any time \( \tau \leq t \leq T - 1 \) and \( \sigma(t) \in \mathcal{M} \), \( Q_{\sigma(t)}(t) \) is a positive semi-definite matrix. We introduce a regularization parameter \( \varepsilon > 0 \) to construct a positive definite matrix \( Q^\varepsilon_{\sigma(t)}(t) = Q_{\sigma(t)}(t) + \varepsilon I \), where \( I \in \mathbb{R}^{n \times n} \) is the identity matrix. Clearly, \( Q^\varepsilon_{\sigma(t)}(t) \) is an invertible matrix. According to Corollary 3.1, we can obtain a diagonal matrix \( \text{diag}[\lambda^{\varepsilon,t,\sigma(t)}_i(t)] \in \mathbb{R}^{n \times n} \) at time \( t \), such that

\[
\text{diag}[\lambda^{\varepsilon,t,\sigma(t)}_i(t)] \preceq Q^\varepsilon_{\sigma(t)}(t).
\]

Let \( \sigma(t) = 1, 2, \ldots, M \) and denote

\[
\lambda^{\varepsilon,t}_i(t) = \min \{ \lambda^{\varepsilon,t,1}_i(t), \lambda^{\varepsilon,t,2}_i(t), \ldots, \lambda^{\varepsilon,t,M}_i(t) \}, \quad \forall \ i = 1, 2, \ldots, n,
\]

a smallest diagonal matrix \( \text{diag}[\lambda^{\varepsilon,t}_i(t)] \in \mathbb{R}^{n \times n} \) is obtained, which satisfies the following relation

\[
0 < \text{diag}[\lambda^{\varepsilon,t}_i(t)] \preceq \text{diag}[\lambda^{\varepsilon,t,\sigma(t)}_i(t)] \preceq Q^\varepsilon_{\sigma(t)}(t), \quad \forall \ \sigma(t) \in \mathcal{M}.
\]
We introduce a diagonal matrix \( \text{diag}[\mu^t_i(t)] = I \) at time \( t \). Then for any \( \sigma(t) \in \mathcal{M} \), it always holds that

\[
\text{diag}[\lambda^{\varepsilon,t}_i(t)] - \varepsilon \text{diag}[\mu^t_i(t)] \preceq Q_{\sigma(t)}(t), \quad \tau \leq t \leq T - 1.
\]

Thus when time \( t = \tau \), let

\[
\text{diag}[\eta^{\varepsilon,\tau}_i(\tau)] = \text{diag}[\lambda^{\varepsilon,\tau}_i(\tau)] - \varepsilon \text{diag}[\mu^\tau_i(\tau)],
\]

we have

\[
\text{diag}[\eta^{\varepsilon,\tau}_i(\tau)] \preceq Q_{\sigma(\tau)}(\tau), \quad \forall \sigma(\tau) \in \mathcal{M}.
\]

When time \( t > \tau \), let us continue to consider the following formulation

\[
A_{\sigma(t-1)}^T(t-1)\text{diag}[\lambda^{\varepsilon,t}_i(t)]A_{\sigma(t-1)}(t-1), \quad \forall \sigma(t-1) \in \mathcal{M},
\]

which is also a positive semi-definite matrix, then we have

\[
A_{\sigma(t-1)}^T(t-1)\text{diag}[\lambda^{\varepsilon,t}_i(t)]A_{\sigma(t-1)}(t-1) + \varepsilon I
\]

\[
\leq A_{\sigma(t-1)}^T(t-1)Q_{\sigma(t)}(t)A_{\sigma(t-1)}(t-1) + \varepsilon I
\]

\[
= A_{\sigma(t-1)}^T(t-1)Q_{\sigma(t)}(t)A_{\sigma(t-1)}(t-1) + \varepsilon (A_{\sigma(t-1)}^T(t-1)A_{\sigma(t-1)}(t-1) + I).
\]

Based on Theorem 3.1 and Corollary 3.1, we can obtain two diagonal matrices \( \text{diag}[\mu^{t-1,\sigma(t-1)}_i(t)] \) and \( \text{diag}[\lambda^{\varepsilon,t-1,\sigma(t-1)}_i(t)] \), respectively, such that

\[
\text{diag}[\mu^{t-1,\sigma(t-1)}_i(t)] \geq A_{\sigma(t-1)}^T(t-1)A_{\sigma(t-1)}(t-1) + I,
\]

\[
\text{diag}[\lambda^{\varepsilon,t-1,\sigma(t-1)}_i(t)] \leq A_{\sigma(t-1)}^T(t-1)\text{diag}[\lambda^{\varepsilon,t}_i(t)]A_{\sigma(t-1)}(t-1) + \varepsilon I.
\]

Let \( \sigma(t-1) = 1, 2, \ldots, M \), the largest and smallest diagonal matrices \( \text{diag}[\mu^{t-1,1}_i(t)] \) and \( \text{diag}[\lambda^{\varepsilon,t-1,1}_i(t)] \) can be constructed, where their diagonal elements \( \mu^{t-1,1}_i(t) \) and \( \lambda^{\varepsilon,t-1,1}_i(t) \) satisfy, respectively,

\[
\mu^{t-1,1}_i(t) = \max\{\mu^{t-1,1}_i(t), \mu^{t-1,2}_i(t), \ldots, \mu^{t-1,M}_i(t)\}, \quad \forall i = 1, 2, \ldots, n,
\]

\[
\lambda^{\varepsilon,t-1,1}_i(t) = \min\{\lambda^{\varepsilon,t-1,1}_i(t), \lambda^{\varepsilon,t-1,2}_i(t), \ldots, \lambda^{\varepsilon,t-1,M}_i(t)\}, \quad \forall i = 1, 2, \ldots, n.
\]

Then for any \( \sigma(t-1), \sigma(t) \in \mathcal{M} \),

\[
\text{diag}[\lambda^{\varepsilon,t-1}_i(t)]
\]

\[
\succeq \text{diag}[\lambda^{\varepsilon,t-1,\sigma(t-1)}_i(t)]
\]

\[
\succeq A_{\sigma(t-1)}^T(t-1)\text{diag}[\lambda^{\varepsilon,t}_i(t)]A_{\sigma(t-1)}(t-1) + \varepsilon I
\]

\[
\succeq A_{\sigma(t-1)}^T(t-1)Q_{\sigma(t)}(t)A_{\sigma(t-1)}(t-1) + \varepsilon (A_{\sigma(t-1)}^T(t-1)A_{\sigma(t-1)}(t-1) + I)
\]

\[
\succeq A_{\sigma(t-1)}^T(t-1)Q_{\sigma(t)}(t)A_{\sigma(t-1)}(t-1) + \varepsilon \text{diag}[\mu^{t-1,\sigma(t-1)}_i(t)]
\]

\[
\succeq A_{\sigma(t-1)}^T(t-1)Q_{\sigma(t)}(t)A_{\sigma(t-1)}(t-1) + \varepsilon \text{diag}[\mu^{t-1}_i(t)].
\]

By the recursion, we can obtain a series of largest and smallest diagonal matrices \( \text{diag}[\mu^{k}_i(t)] \in \mathbb{R}^{n \times n} \) and \( \text{diag}[\lambda^{\varepsilon,k}_i(t)] \in \mathbb{R}^{n \times n} \), respectively, where \( k = \)
we can also construct a diagonal matrix \( \text{diag}[\lambda_i^{\varepsilon,k}(t)] \)
\[ \preceq \text{diag}[\lambda_i^{\varepsilon,k,\sigma(k)}(t)] \]
\[ \preceq A_{\sigma(k)}^T(t) \text{diag}[\lambda_i^{\varepsilon,k+1}(t)] A_{\sigma(k)}(t) + \varepsilon I \]
\[ \preceq A_{\sigma(k)}^T(t) \cdots A_{\sigma(t-1)}^T(t-1) Q_{\sigma(t)}(t) A_{\sigma(t-1)}(t-1) \cdots A_{\sigma(k)}(t-1) \]
\[ + \varepsilon \text{diag}[\mu_i^{\varepsilon,\sigma(k)}(t)] \]
\[ \preceq A_{\sigma(k)}^T(t) \cdots A_{\sigma(t-1)}^T(t-1) Q_{\sigma(t)}(t) A_{\sigma(t-1)}(t-1) \cdots A_{\sigma(k)}(t-1) \]
\[ + \varepsilon \text{diag}[\mu_i^{\varepsilon,\sigma(k)}(t)]. \]
That is, for any \( \sigma(k), \sigma(k+1), \ldots, \sigma(t) \in \mathcal{M}, \tau \leq k \leq t-2, \) we have
\[ \text{diag}[\lambda_i^{\varepsilon,k}(t)] - \varepsilon \text{diag}[\mu_i^{\varepsilon}(t)] \]
\[ \preceq A_{\sigma(k)}^T(t) \cdots A_{\sigma(t-1)}^T(t-1) Q_{\sigma(t)}(t) A_{\sigma(t-1)}(t-1) \cdots A_{\sigma(k)}(t-1) \]
So we let \( \text{diag}[\eta_{i}^{\varepsilon,\tau}(t)] = \text{diag}[\lambda_i^{\varepsilon,\tau}(t)] - \varepsilon \text{diag}[\mu_i^{\varepsilon}(t)], \)
\[ \text{diag}[\eta_{i}^{\varepsilon,\tau}(t)] \leq A_{\sigma(\tau)}^T(\tau) \cdots A_{\sigma(t-1)}^T(t-1) Q_{\sigma(t)}(t) A_{\sigma(t-1)}(t-1) \cdots A_{\sigma(\tau)}(\tau). \]
This completes the proof.

Hence, we have for any \( \sigma(\tau), \sigma(\tau+1), \ldots, \sigma(t) \in \mathcal{M}, \)
\[ x^T(\tau) \text{diag}[\eta_{i}^{\varepsilon,\tau}(t)] x(\tau) \leq x^T(t) Q_{\sigma(t)}(t) x(t), \quad t = \tau, \tau+1, \ldots, T-1. \]

Similarly, for the final performance cost
\[ x^T(T) \Phi x(T) \]
\[ = x^T(\tau) A_{\sigma(\tau)}^T(\tau) \cdots A_{\sigma(T-1)}^T(T-1) \Phi A_{\sigma(T-1)}(T-1) \cdots A_{\sigma(\tau)}(\tau) x(\tau), \]
we can also construct a diagonal matrix \( \text{diag}[\eta_{i}^{\varepsilon,\tau}(T)] \in \mathbb{R}^{n \times n} \) such that
\[ x^T(\tau) \text{diag}[\eta_{i}^{\varepsilon,\tau}(T)] x(\tau) \leq x^T(T) \Phi x(T). \]

Since the lower bound of the total variation function is zero, we obtain an exact lower bound of the cost functional denoted by \( L(\sigma(0), \sigma(1), \ldots, \sigma(T-1)) \) at current time \( t = \tau - 1 \) such that, for any \( \sigma(\tau), \sigma(\tau+1), \ldots, \sigma(T-1) \in \mathcal{M}, \)
\[ L(\sigma(0), \sigma(1), \ldots, \sigma(\tau-1)) \]
\[ = \sum_{t=0}^{\tau-1} x^T(t) Q_{\sigma(t)}(t) x(t) + \alpha \sqrt{\sigma} + \sum_{t=\tau}^{T} x^T(\tau) \text{diag}[\eta_{i}^{\varepsilon,\tau}(t)] x(\tau) \]
\[ \leq J(\sigma|\sigma(0), \sigma(1), \ldots, \sigma(\tau-1)), \]
where the regularization parameter is \( \varepsilon > 0. \)

As an upper bound, we denote the current optimal value by \( J^* \) and compare it with the lower bound \( L(\sigma(0), \sigma(1), \ldots, \sigma(\tau-1)). \) If \( L(\sigma(0), \sigma(1), \ldots, \sigma(\tau-1)) > J^*, \) then
\[ J^* < L(\sigma(0), \sigma(1), \ldots, \sigma(\tau-1)) \leq J(\sigma|\sigma(0), \sigma(1), \ldots, \sigma(\tau-1)). \]
That is, for any \( \sigma(\tau), \sigma(\tau + 1), \ldots, \sigma(T - 1) \in M \) in the switching sequence, we can not find a better switching sequence \( \sigma \) with the first \( \tau \) fixed switching signals \( \sigma(0), \sigma(1), \ldots, \sigma(\tau - 1) \) to improve the current optimal value. Thus these branches \( (\sigma(0), \sigma(1), \ldots, \sigma(\tau - 1), *, \ldots, *) \in M^T \) can be pruned, where the symbol * means any value in the set \( M \). This simplifies the tree-search process.

Based on the analysis above, a series of exact lower bounds can be obtained at different times. Then the branch and bound method can be applied. We summarize its searching process in Algorithm 1.

**Algorithm 1** The exact branch and bound method for the optimal switching problem with a cost on switching action.

**Input:** The terminal time \( T \); The number of subsystems \( M \); The regularization parameter \( \varepsilon > 0 \); The penalty parameter \( \alpha > 0 \); The initial state \( x(0) = x_0 \in \mathbb{R}^n \); The current optimal value \( J^* = +\infty \); Denote these subsystems by \((A_1(t), Q_1(t)), (A_2(t), Q_2(t)), \ldots, (A_M(t), Q_M(t))\); Set \( \tau := 1 \).

**Output:** The optimal solution \( \sigma^* \) and the optimal value \( J^* \).

1. Let \( \sigma(\tau - 1) = 1, 2, \ldots, M \); Compute the first determined part of the cost functional as follows

   \[
   Ld(\tau - 1) = \sum_{t=0}^{\tau-1} x^T(t)Q_{\sigma(t)}(t)x(t) + \alpha \bigvee_{t=1}^{\tau-1} \sigma.
   \]

   With the state \( x(\tau) = A_{\sigma(\tau-1)}(\tau - 1)x(\tau - 1) \), we obtain the lower bound of the second undetermined part of the cost functional and denote it by

   \[
   Lf(x(\tau), \tau) = \sum_{t=\tau}^{T} x^T(\tau)\text{diag}[\eta^T(\tau)]x(\tau).
   \]

   Then the exact lower bound of the cost functional at current time \( t = \tau - 1 \) is formulated by \( L(\sigma(0), \ldots, \sigma(\tau - 1)) = Ld(\tau - 1) + Lf(x(\tau), \tau) \).

2. Sort these lower bounds \( L(\sigma(0), \ldots, 1), \ldots, L(\sigma(0), \ldots, M) \) by the ascending rule such that \( L(\sigma(0), \ldots, v_{v_1}) \leq \ldots \leq L(\sigma(0), \ldots, v_{v_M}) \). Set \( k_{\tau-1} = 1 \).

3. Determine the switching signal by \( \sigma(\tau - 1) = v_{k_{\tau-1}} \). If \( L(\sigma(0), \ldots, v_{k_{\tau-1}}) > J^* \) or \( k_{\tau-1} = M \), break and set \( \tau := \tau - 1 \). Go to Step 5. Else, set \( \tau := \tau + 1 \).

4. If \( \tau = T \), let \( \sigma(\tau - 1) = 1, 2, \ldots, M \). Compute the cost functional value \( J(\sigma) \).

   When \( J(\sigma) \leq J^* \), update the current optimal value and save the corresponding optimal solution by \( J^* := J \) and \( \sigma^* := \sigma \), respectively. Set \( \tau := \tau - 1 \). Else, go to Step 1.

5. If \( \tau = 0 \), terminate and return \( \sigma^* \) and \( J^* \). Else, set \( k_{\tau-1} := k_{\tau-1} + 1 \). Go to Step 3.

Now we give the convergence property of the algorithm to state that the exact branch and bound method can search the global optimal solution of the optimal switching problem.

Since the total number of feasible solutions is \( M^T \), the algorithm can surely terminate in finite time. In the exact branch and bound method, \( M \) lower bounds are computed for the cost functional at each time. Without loss of generality, we assume that the first \( \tau \) switching signals have been fixed temporarily. Let us consider the unknown switching signal \( \sigma(\tau) \), where \( 1 \leq \tau \leq T - 1 \). We first computer \( M \) lower bounds and sort them by the ascending rule at time \( t = \tau \), such that

\[
L(\sigma(0), \ldots, \sigma(\tau-1), v_1) \leq L(\sigma(0), \ldots, \sigma(\tau-1), v_2) \leq \ldots \leq L(\sigma(0), \ldots, \sigma(\tau-1), v_M).
\]
Then let \( k = 1 \) and denote \( \sigma(\tau) = v_k, k \in \mathcal{M} \). If

\[
L(\sigma(0), \ldots, \sigma(\tau - 1), v_k) \leq J^*,
\]

some potential better solutions should be explored in the following undetermined switching signals. So we continue to consider the next unknown switching signals and repeat the searching rule above. Until all feasible scenarios in the latter switching signals have been searched, the current optimal value can be obtained by \( J^* = \min \{ J(\sigma), J^* \} \). Then we turn to the next branch with the switching signal \( \sigma(\tau) = v_{k+1} \) to explore the potential better solution.

Otherwise, the condition (5) is not satisfied. We begin to consider the previous time to update the switching signal \( \sigma(\tau - 1) \), because at time \( t = \tau \), we have

\[
J^* < L(\sigma(0), \ldots, \sigma(\tau - 1), v_k) \leq L(\sigma(0), \ldots, \sigma(\tau - 1), v_{k+1})
\]

\[
\leq \cdots \leq L(\sigma(0), \ldots, \sigma(\tau - 1), v_M).
\]

By the relation (4), all these feasible solutions \((\sigma(0), \ldots, \sigma(\tau - 1), v_k, *, \ldots, *)\), \((\sigma(0), \ldots, \sigma(\tau - 1), v_{k+1}, *, \ldots, *)\), \((\sigma(0), \ldots, \sigma(\tau - 1), v_M, *, \ldots, *)\) are pruned. There is no potential better solution in the following switching signals any more.

So when the condition (5) is not satisfied or the current switching signal turns to \( \sigma(\tau - 1) = v_M \), all these branches are considered at time \( t = \tau \). We go back to the previous time to update the temporarily fixed switching signal \( \sigma(\tau - 1) \) with \( v_{k+1} \).

Ultimately, the algorithm returns to \( \sigma(0) \). By the ascending rule, for the case \( k_0 = M \) or there exists a number \( k_0 \in \mathcal{M} \) such that

\[
J^* < L(v_{k_0}) \leq L(v_{k_0+1}) \leq \cdots \leq L(v_M),
\]

the algorithm can stop. Then all scenarios have been considered and the search tree is implemented completely. That means the stored current optimal solution is the global optimal solution.

**Remark 3.1.** From Algorithm 1, the lower bound of the cost functional plays an important role. When the lower bound is constructed too small, a plenty of branches can not be pruned effectively. The worst case is that the lower bound does not prune any branches in the tree search and the searching effect becomes the enumeration. In our algorithm, the lower bound depends on

\[
Lf(x(\tau), \tau) = \sum_{t=\tau}^{T} x^T(\tau) \text{diag}[\eta_{i,\tau}(t)] x(\tau)
\]

\[
= \sum_{t=\tau}^{T} x^T(\tau) (\text{diag}[\lambda_{i,\tau}(t)] - \epsilon \text{diag}[\mu_{i,\tau}(t)]) x(\tau).
\]

It is an exact lower bound of the undetermined part in the cost functional. For more details, we give Algorithm 2 to obtain \( Lf(x(\tau), \tau) \) as a subprogram in Algorithm 1. However, in [23], the lower bound of the undetermined part is given by

\[
\overline{L}f(x(\tau), \tau) = \sum_{t=\tau}^{T} x^T(\tau) \text{diag}[\lambda_{i,\tau}(t)] x(\tau).
\]
Algorithm 2 The lower bound \( Lf(x(\tau), \tau) \) of the undetermined part in the cost functional at Period \( \tau \).

**Input:** The terminal time \( T \); The current period \( \tau \); The number of subsystems \( M \); The regularization parameter \( \varepsilon > 0 \); The current state \( x(\tau) \in \mathbb{R}^n \); The final performance matrix \( \Phi \in \mathbb{R}^{n \times n} \); Denote these subsystems by \((A_1(t), Q_1(t)), (A_2(t), Q_2(t)), \ldots, (A_M(t), Q_M(t))\); Let \( G = U = V = 0 \in \mathbb{R}^{n \times n} \); \( Lf := 0 \).

**Output:** The lower bound \( Lf \) of the undetermined part.

1. for \( t_1 = \tau \) to \( T \) do
2. \( s = 1; \)
3. if \( t_1 < T \) then
4. \( \Psi^m = [Q_m(t_1) + \varepsilon I]^{-1}, m = 1, 2, \ldots, M; \)
5. for \( i = 1 \) to \( n \) do
6. \( g_m = \sum_{j=1}^{n} |\Psi^m_{ij}|, m = 1, 2, \ldots, M; \)
7. \( G_{ii} = \sum_{j=1}^{n} |\Psi_{ij}|; \)
8. end for
9. else
10. \( \Psi = [\Phi + \varepsilon I]^{-1}; \)
11. for \( i = 1 \) to \( n \) do
12. \( G_{ii} = \sum_{j=1}^{n} |\Psi_{ij}|; \)
13. end for
14. \( Y_s = I; \)
15. for \( t_2 = t_1 - 1 \) to \( \tau \) do
16. \( \hat{A}^m = A_m^T Y_s A_m + I, \hat{A}^m = [A_m^T \Gamma_s A_m + \varepsilon I]^{-1}, m = 1, 2, \ldots, M; \)
17. for \( i = 1 \) to \( n \) do
18. \( u_m = \sum_{j=1}^{n} |\hat{A}^m_{ij}|, v_m = \sum_{j=1}^{n} |\hat{A}^m_{ij}|, m = 1, 2, \ldots, M; \)
19. \( U_{ii} = \max\{u_1, u_2, \ldots, u_M\}; V_{ii} = \max\{v_1, v_2, \ldots, v_M\}; \)
20. end for
21. \( s = s + 1; \)
22. \( Y_s = U; \)
23. end for
24. \( \Theta_s = \Gamma_s - \varepsilon Y_s; \)
25. end for
26. for \( s = 1 \) to \( T - \tau + 1 \) do
27. \( Lf = Lf + x^T(\tau) \Theta_s x(\tau); \)
28. end for

Then we have

\[
J(\sigma|\sigma(0), \sigma(1), \ldots, \sigma(\tau - 1)) - (Ld(\tau - 1) + \tilde{L}f(x(\tau), \tau)) \leq \varepsilon \left( \sum_{t=\tau}^{T} x^T(\tau) \text{diag}[\mu^*_t(t)] x(\tau) \right).
\]

So the optimal solution obtained by \([23]\) is a global \( \varepsilon \)-optimal solution. Only when the regularization parameter \( \varepsilon \) approaches to zero, the global optimal solution can be searched. But it is obscure to define such a sufficient small number in practical
problems. In our paper, the algorithm can be implemented for any regularization parameter \( \varepsilon > 0 \). Therefore, it shows the stability of our algorithm to obtain the global optimal solution.

4. Numerical example. In order to verify the efficiency of the proposed method, two numerical experiments are tested. All codes are edited in MATLAB 2012a and implemented on a notebook with a 2.50 GHz CPU and 4G RAM.

**Example 4.1.** Assume that there are 4 time-invariant subsystems in the switched system. We denote them as follows.

\[
(A_1, Q_1) = \begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 1 & 0
\end{pmatrix}, \quad \begin{pmatrix}
2 & 0 & -1 \\
0 & 1 & 0 \\
-1 & 0 & 1
\end{pmatrix},
\]

\[
(A_2, Q_2) = \begin{pmatrix}
0 & 1 & -1 \\
1 & 0 & 0 \\
0 & 0 & 1
\end{pmatrix}, \quad \begin{pmatrix}
1 & -1 & 0 \\
-1 & 2 & 0 \\
0 & 0 & 1
\end{pmatrix},
\]

\[
(A_3, Q_3) = \begin{pmatrix}
0 & 1 & 0 \\
-1 & 0 & 0 \\
1 & 0 & 1
\end{pmatrix}, \quad \begin{pmatrix}
2 & 0 & 0 \\
0 & 1 & -1 \\
0 & -1 & 2
\end{pmatrix},
\]

\[
(A_4, Q_4) = \begin{pmatrix}
-1 & 0 & 0 \\
0 & 0 & 1 \\
1 & 1 & 0
\end{pmatrix}, \quad \begin{pmatrix}
1 & 0 & -1 \\
0 & 2 & 0 \\
-1 & 0 & 2
\end{pmatrix}.
\]

Let the initial state be \( x_0 = [2 \ 1 \ 3]^T \), the terminal time be \( T = 10 \) and the final performance matrix be

\[
\Phi = \begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{pmatrix}.
\]

Then we consider the following optimal switching problem

\[
\min_{\sigma} J(\sigma) = x^T(10)x(10) + \sum_{t=0}^{9} x^T(t)Q_{\sigma(t)}x(t) + \alpha \sum_{t=1}^{9} |\sigma(t-1) - \sigma(t)|
\]

s.t. \( x(t+1) = A_{\sigma(t)}x(t), \quad t = 0, 1, \ldots, 9, \)

\[
x(0) = [2 \ 1 \ 3]^T,
\]

where \( \sigma = (\sigma(0), \ldots, \sigma(9)) \in \{1, 2, 3, 4\}^{10} \) and \( \alpha \geq 0 \) is a given penalty parameter.

In order to find the global optimal solution under different penalty conditions, we enumerate all feasible switching sequences. On average, it takes about 396.98 seconds to search \( 4^{10} \) scenarios for each case. Figure 1 shows all global optimal solutions under different penalty parameters and Table 1 lists more consequences.

From Figure 1, when the switching cost is not considered, there are two global optimal solutions \((1 \ 4 \ 2 \ 3 \ 1 \ 4 \ 2 \ 3 \ 2)\) and \((1 \ 4 \ 2 \ 1 \ 3 \ 1 \ 4 \ 2 \ 1 \ 1)\). Obviously, the latter is better because the change among subsystems is more stable. When we penalize the switchings with \( \alpha = 0.5 \), two global optimal solutions \((1 \ 4 \ 2 \ 1 \ 3 \ 1 \ 4 \ 2 \ 1 \ 1)\) and \((1 \ 1 \ 3 \ 4 \ 4 \ 4 \ 2 \ 1 \ 1 \ 1)\) are obtained. Although the former achieves a better performance 40 than the latter which is 45 in Table 1, the switching cost of the former is greater than the latter, where their switching costs are 8 and 3, respectively. So their cost functional values are the same. With the penalty parameter increases, the switching frequency and the switching amplitude reduce gradually.
Figure 1. Global optimal switching sequences under different penalty parameters in Example 4.1.

Table 1. Global optimal solutions under different penalty parameters in Example 4.1.

| α  | Global optimal solution σ* | Switching times | Performance index | Switching cost | Optimal functional value J* |
|----|-----------------------------|-----------------|-------------------|----------------|-----------------------------|
| 0  | (1, 4, 2, 1, 3, 1, 4, 2, 1) | 9               | 40                | 0              | 40                          |
| 0.5| (1, 4, 2, 1, 4, 2, 1, 1)    | 8               | 40                | 0              | 40                          |
| 1  | (1, 4, 2, 1, 4, 2, 1, 1)    | 4               | 45                | 3              | 48                          |
| 2  | (2, 3, 2, 2, 2, 2, 2, 1)    | 3               | 50                | 6              | 51                          |
| 5  | (1, 1, 2, 2, 2, 2, 2)       | 1               | 50                | 6              | 56                          |
| 10 | (1, 1, 2, 2, 2, 2, 2)       | 1               | 50                | 6              | 64                          |
### Table 2. Two kinds of B&B methods with $\epsilon = 10^{-5}$ in Example 4.1.

| $\alpha$ | \(\sigma^*\) | Approximate B&B method | \(J^*\) | Searching times | Time   | Exact B&B method | \(\sigma^*\) | \(J^*\) | Searching times | Time   |
|----------|----------------|-------------------------|--------|-----------------|--------|-----------------|----------------|--------|-----------------|--------|
| 0        | (1 4 2 1 3 1 4 2 3 2) | 40                      | 116    | 0.0826s         |        | (1 4 2 1 3 1 4 2 3 2) | 40                      | 128    | 0.1055s         |        |
| 0.5      | (1 4 2 1 3 1 4 2 1 1) | 48                      | 144    | 0.1275s         |        | (1 4 2 1 3 1 4 2 1 1) | 48                      | 152    | 0.1310s         |        |
| 1        | (1 1 3 4 4 4 2 1 1 1) | 51                      | 376    | 0.2371s         |        | (1 1 3 4 4 4 2 1 1 1) | 51                      | 392    | 0.2481s         |        |
| 2        | (2 3 2 2 2 2 2 1 1 1) | 56                      | 440    | 0.2799s         |        | (2 3 2 2 2 2 2 1 1 1) | 56                      | 444    | 0.2806s         |        |
| 5        | (1 1 1 2 2 2 2 2 2 2) | 64                      | 360    | 0.2280s         |        | (1 1 1 2 2 2 2 2 2 2) | 64                      | 372    | 0.2346s         |        |

### Table 3. Two kinds of B&B methods with $\epsilon = 1$ in Example 4.1.

| $\alpha$ | \(\sigma^*\) | Approximate B&B method | \(J^*\) | Searching times | Time   | Exact B&B method | \(\sigma^*\) | \(J^*\) | Searching times | Time   |
|----------|----------------|-------------------------|--------|-----------------|--------|-----------------|----------------|--------|-----------------|--------|
| 0        | (1 4 2 1 1 1 1 3 1) | 55                      | 8      | 0.0115s         |        | (1 4 2 1 3 1 4 2 3 2) | 40                      | 2796   | 10.4855s        |        |
| 0.5      | (1 4 2 1 1 1 1 1 3 1) | 60                      | 8      | 0.0115s         |        | (1 4 2 1 3 1 4 2 1 1) | 48                      | 2948   | 10.7044s        |        |
| 1        | (1 1 3 4 4 4 2 1 1 1) | 51                      | 4      | 0.0091s         |        | (1 1 3 4 4 4 2 1 1 1) | 51                      | 2768   | 10.4657s        |        |
| 2        | (1 1 3 4 4 4 2 1 1 1) | 57                      | 12     | 0.1379s         |        | (2 3 2 2 2 2 1 1 1 1) | 56                      | 2504   | 10.3680s        |        |
| 5        | (1 1 3 4 4 4 4 3 3 3) | 70                      | 4      | 0.0091s         |        | (1 1 1 2 2 2 2 2 2 2) | 64                      | 2028   | 9.6311s         |        |
Now we apply two kinds of branch and bound methods to solve the same problem, where we call the branch and bound method with the $\varepsilon$-approximate lower bound mentioned in [23] as the approximate B&B method, and our proposed method as the exact B&B method. For comparison, we test them in the same environment via adjusting the penalty parameter and the regularization parameter. Then the results are shown in Table 2 and Table 3, respectively.

For a sufficiently small regularization parameter, such as $\varepsilon = 10^{-5}$, both the approximate B&B method and the exact B&B method can obtain efficiently all global optimal solutions under different penalty parameters. Since the lower bound constructed by the approximate B&B method is larger than that of the proposed method, its searching efficiency can be better, while the advantage is not remarkable.

When the regularization parameter is $\varepsilon = 1$, we observe that the approximate B&B method can not always effectively search the global optimal solution in Table 3. That is, the approximate lower bound is larger than some local optimal values. So these branches containing the global optimal solutions are pruned unexpectedly. In the same environment, our exact lower bound can still be implemented efficiently and all global optimal solutions can also be obtained. It shows that our exact B&B method is more stable to search all global optimal solutions in the time-invariant switched system.

**Example 4.2.** Let us consider the time-varying switched system, where 4 subsystems are denoted by

\[
\begin{align*}
(A_1(t), Q_1(t)) &= \left( \begin{bmatrix} \sin t & 1 \\ -1 & 0.5t \end{bmatrix}, \begin{bmatrix} 0.5t & 0 \\ 0 & 2 \end{bmatrix} \right), \\
(A_2(t), Q_2(t)) &= \left( \begin{bmatrix} t & \cos t \\ 1 & -1 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 0.5t \end{bmatrix} \right), \\
(A_3(t), Q_3(t)) &= \left( \begin{bmatrix} -0.5t & \cos t \\ \sin t & 1 \end{bmatrix}, \begin{bmatrix} 1 + \sin t & 0 \\ 0 & 1 \end{bmatrix} \right), \\
(A_4(t), Q_4(t)) &= \left( \begin{bmatrix} \sin t & -1 \\ -1 & \cos t \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 + \cos t \end{bmatrix} \right).
\end{align*}
\]

The initial state, the final performance matrix and the terminal time are given, respectively, by

\[
\begin{align*}
x_0 &= \begin{bmatrix} 1 \\ 1 \end{bmatrix}, & \Phi &= \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}, & T = 10.
\end{align*}
\]

In this switched system, we focus on another case of the switching cost, where only the switching frequency is considered and the switching magnitude is ignored. Then the optimal switching problem is formulated by

\[
\begin{align*}
\min_{\sigma} & \quad J(\sigma) = x^T(10)\Phi x(10) + \sum_{t=0}^{9} x^T(t)Q_{\sigma(t)}(t)x(t) + \alpha \sum_{t=1}^{9} \chi(\sigma(t-1), \sigma(t)) \\
\text{s.t.} & \quad x(t+1) = A_{\sigma(t)}(t)x(t), \quad t = 0, 1, \ldots, 9, \\
& \quad x(0) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}.
\end{align*}
\]
where $\sigma = (\sigma(0), \sigma(1), \ldots, \sigma(9)) \in \{1, 2, 3, 4\}^{10}$ and $\alpha \geq 0$ is a given penalty parameter. For each time $t = 1, 2, \ldots, 9$, the indicator function
\[
\chi(\sigma(t-1), \sigma(t)) = \begin{cases} 
0, & \sigma(t-1) = \sigma(t), \\
1, & \sigma(t-1) \neq \sigma(t).
\end{cases}
\]
Under different penalty parameters, we enumerate all scenarios of switching sequences to find the global optimal solution. It takes about 428.6297 seconds to search $4^{10}$ feasible solutions for each case and the search results are shown in Figure 2 and Table 4. We observe that the more times the system switches, the better its performance is. When we increase the penalty parameter for the switchings, the switching frequency is restrained gradually. First, if we do not consider the switching cost, the optimal switching sequence is obtained as $(3 \ 4 \ 2 \ 3 \ 4 \ 1 \ 3 \ 3 \ 4)$. When the penalty parameter is $\alpha = 10$, the optimal switching sequence stays at the fourth subsystem. That is, the switching cost is larger than the improvement of the performance caused by any change among subsystems.

For comparison, we implement the approximate B&B method and the exact B&B method to verify the searching efficiency under different regularization parameters. First the regularization parameter is given by $\varepsilon = 10^{-20}$, we list the results in Table 5. It shows that both two methods are very efficient to search the global optimal solutions. Then we further adjust the regularization parameter as $\varepsilon = 1$ and list the results in Table 6. The approximate B&B method is not stable to search the global optimal solution under different penalty parameters. However, our exact B&B method is still executable and efficient in the time-varying switched system.

5. Conclusion. In this paper, we consider a class of optimal switching problem with a switching cost. Since the switching sequence is composed of a set of discrete variables and there exists a changing cost with respect to the switching sequence in the cost functional, some existing optimization methods may not be implemented efficiently. By constructing an exact lower bound of the cost functional at each time, we improve the branch and bound method such that the algorithm can exactly search all global optimal solutions for any given penalty parameter $\alpha \geq 0$ and the regularization parameter $\varepsilon > 0$. In order to compare with some other global optimization methods, two numerical experiments are given. The results show the efficiency and stability of our proposed method.

Acknowledgments. This work is supported by National Natural Science Foundation of China (Nos. 71831008, 11671250, 11431004, 61673078); Humanity and Social Science Foundation of Ministry of Education of China (No. 15YJA630034); the program for scientific research start-up funds of Guangdong Ocean University; the grant of Chongqing Science and Technology Commission (No. cstc2017jcyjAX0161); the grant of Chongqing Normal University (No. 17XLB010).
Figure 2. Global optimal switching sequences under different penalty parameters in Example 4.2.

Table 4. Global optimal solutions under different penalty parameters in Example 4.2.

| α  | Global optimal solution σ* | Switching times | Performance index J* | Switching cost J \*σ | Optimal functional value J* |
|----|---------------------------|-----------------|-----------------------|---------------------|-----------------------------|
| 0  | (3 4 2 3 1 4 4 4 4 4)     | 7               | 5.1452                | 0                   | 5.1452                      |
| 0.1| (3 4 2 4 4 4 4 4 4 4)     | 5               | 5.2376                | 0.5                 | 5.7876                      |
| 0.5| (2 3 3 3 4 4 4 4 4 4)     | 3               | 5.6759                | 1.5                 | 7.1759                      |
| 2  | (2 2 2 4 4 4 4 4 4 4)     | 2               | 7.7973                | 6                   | 11.7973                     |
| 5  | (2 3 3 4 4 4 4 4 4 4)     | 2               | 7.7973                | 10                  | 18.3901                     |
| 10 | (4 4 4 4 4 4 4 4 4 4)     | 0               | 18.5961               | 0                   | 18.5961                     |
Table 5. Two kinds of B&B methods with $\varepsilon = 10^{-20}$ in Example 4.2.

| $\alpha$ | $\sigma^*$ | $J^*$ | Searching times | Time | $\sigma^*$ | $J^*$ | Searching times | Time |
|----------|------------|-------|----------------|------|------------|-------|----------------|------|
| 0        | (3 4 2 3 4 1 1 3 3 4) | 5.1452 | 164            | 0.5458s | (3 4 2 3 4 1 1 3 3 4) | 5.1452 | 168            | 0.6742s |
| 0.1      | (3 4 2 4 1 4 4 4 4 4) | 5.7376 | 108            | 0.6379s | (3 4 2 4 1 4 4 4 4 4) | 5.7376 | 64             | 0.7968s |
| 0.5      | (2 3 3 2 4 4 4 4 4 4) | 7.1759 | 96             | 0.7146s | (2 3 3 2 4 4 4 4 4 4) | 7.1759 | 120            | 0.9573s |
| 2        | (2 3 3 2 4 4 4 4 4 4) | 11.6759 | 132            | 1.5675s | (2 3 3 2 4 4 4 4 4 4) | 11.6759 | 172            | 1.7968s |
| 5        | (2 3 3 3 4 4 4 4 4 4) | 17.7973 | 36             | 1.2749s | (2 3 3 3 4 4 4 4 4 4) | 17.7973 | 36             | 1.6238s |
| 10       | (4 4 4 4 4 4 4 4 4 4) | 18.5961 | 24             | 0.8772s | (4 4 4 4 4 4 4 4 4 4) | 18.5961 | 68             | 1.2210s |

Table 6. Two kinds of B&B methods with $\varepsilon = 1$ in Example 4.2.

| $\alpha$ | $\sigma^*$ | $J^*$ | Searching times | Time | $\sigma^*$ | $J^*$ | Searching times | Time |
|----------|------------|-------|----------------|------|------------|-------|----------------|------|
| 0        | (2 3 3 2 4 1 3 4 4 4) | 5.5347 | 12             | 0.0679s | (3 4 2 3 4 1 1 3 3 4) | 5.1452 | 660            | 85.3819s |
| 0.1      | (2 3 3 2 4 1 4 4 4 4) | 6.1436 | 4              | 0.0615s | (3 4 2 4 1 4 4 4 4 4) | 5.7376 | 568            | 86.7247s |
| 0.5      | (2 3 3 2 4 4 4 4 4 4) | 7.1759 | 4              | 0.0618s | (2 3 3 2 4 4 4 4 4 4) | 7.1759 | 508            | 87.3288s |
| 2        | (2 3 3 2 4 4 4 4 4 4) | 11.6759 | 24             | 0.1131s | (2 3 3 2 4 4 4 4 4 4) | 11.6759 | 408            | 86.0455s |
| 5        | (2 3 3 3 4 4 4 4 4 4) | 17.7973 | 4              | 0.0797s | (2 3 3 3 4 4 4 4 4 4) | 17.7973 | 372            | 84.8442s |
| 10       | (2 3 3 3 4 4 4 4 4 4) | 27.7973 | 4              | 0.2317s | (4 4 4 4 4 4 4 4 4 4) | 18.5961 | 332            | 83.2846s |
REFERENCES

[1] H. Axelsson, Y. Wardi, M. Egerstedt and E. I. Verriest, Gradient descent approach to optimal mode scheduling in hybrid dynamical systems, Journal of Optimization Theory and Applications, 136 (2008), 167–186.

[2] S. C. Bengea and R. A. DeCarlo, Optimal control of switching systems, Automatica, 41 (2005), 11–27.

[3] T. M. Caldwell and T. D. Murphey, Projection-based iterative mode scheduling for switched systems, Nonlinear Analysis: Hybrid Systems, 21 (2016), 59–83.

[4] Z. G. Feng, K. L. Teo and V. Rehbock, Hybrid method for a general optimal sensor scheduling problem in discrete time, Automatica, 44 (2008), 1295–1303.

[5] Z. G. Feng, K. L. Teo and V. Rehbock, Optimal sensor scheduling in continuous time, Dynamic Systems and Applications, 17 (2008), 331–350.

[6] Z. G. Feng, K. L. Teo and V. Rehbock, A discrete filled function method for the optimal control of switched systems in discrete time, Optimal Control Applications and Methods, 30 (2009), 585–593.

[7] Z. G. Feng, K. L. Teo and Y. Zhao, Branch and bound method for sensor scheduling in discrete time, Journal of Industrial and Management Optimization, 1 (2005), 499–512.

[8] J. Gao and D. Li, Linear-quadatic switching control with switching cost, Automatica, 48 (2012), 1138–1143.

[9] Z. Gong, C. Liu and Y. Wang, Optimal control of switched systems with multiple time-delays and a cost on changing control, Journal of Industrial and Management Optimization, 14 (2018), 183–198.

[10] D. Görse, M. Izák and S. Liu, Optimal control and scheduling of switching systems, IEEE Transactions on Automatic Control, 56 (2011), 135–140.

[11] J. F. He, W. Xu, Z. G. Feng and X. Yang, On the global optimal solution for linear quadratic problems of switched system, Journal of Industrial and Management Optimization, 15 (2019), 817–832.

[12] M. Kamgarpour and C. Tomlin, On optimal control of non-autonomous switched systems with a fixed mode sequence, Journal of Global Optimization, 48 (2012), 1177–1181.

[13] B. Li and Y. Rong, Joint transceiver optimization for wireless information and energy transfer in nonregenerative MIMO relay systems, IEEE Transactions on Vehicular Technology, 67 (2018), 8348–8362.

[14] B. Li and Y. Rong, AF MIMO relay systems with wireless powered relay node and direct link, IEEE Transactions on Communications, 66 (2018), 1508–1519.

[15] B. Li, Y. Rong, J. Sun and K. L. Teo, A distributionally robust linear receiver design for multi-access space-time block coded MIMO systems, IEEE Transactions on Wireless Communications, 16 (2017), 464–474.

[16] R. Li, K. L. Teo, K. H. Wong and G. R. Duan, Control parametrization enhancing transform for optimal control of switched systems, Mathematical and Computer Modelling, 43 (2006), 1393–1403.

[17] C. Liu, Z. Gong, K. L. Teo, J. Sun and L. Caccetta, Robust multi-objective optimal switching control arising in 1,3-propanediol microbial fed-batch process, Nonlinear Analysis: Hybrid Systems, 25 (2017), 1–20.

[18] R. Loxton, Q. Lin and K. L. Teo, Minimizing control variation in nonlinear optimal control, Automatica, 49 (2013), 2652–2664.

[19] R. C. Loxton, K. L. Teo, V. Rehbock and W. K. Ling, Optimal switching instants for a switched-capacitor DC/DC power converter, Automatica, 45 (2009), 973–980.

[20] W. Lu, P. Zhu and S. Ferrari, A hybrid-adaptive dynamic programming approach for the model-free control of nonlinear switched systems, IEEE Transactions on Automatic Control, 61 (2016), 3203–3208.

[21] C. Seatzu, D. Corona, A. Giua and A. Bemporad, Optimal control of continuous-time switched affine systems, IEEE Transactions on Automatic Control, 51 (2006), 726–741.

[22] Y. Wardi, M. Egerstedt and M. Hale, Switched-mode systems: Gradient-descent algorithms with Armijo step sizes, Discrete Event Dynamic Systems, 25 (2015), 571–599.

[23] W. Xu, Z. G. Feng, J. W. Peng and K. F. C. Yiu, Optimal switching for linear quadratic problem of switched systems in discrete time, Automatica, 78 (2017), 185–193.

[24] W. Xu, Z. G. Feng, G. H. Lin and L. Yu, Optimal scheduling of discrete-time switched linear systems, IMA Journal of Mathematical Control and Information, (2018).
[25] H. Yan, Y. Sun and Y. Zhu, A linear-quadratic control problem of uncertain discrete-time switched systems, *Journal of Industrial and Management Optimization*, 13 (2017), 267–282.

[26] F. Yang, K. L. Teo, R. Loxton, V. Rehbock, B. Li, C. Yu and L. Jennings, Visual miser: an efficient user-friendly visual program for solving optimal control problems, *Journal of Industrial and Management Optimization*, 12 (2016), 781–810.

[27] J. Zhai, T. Niu, J. Ye and E. Feng, Optimal control of nonlinear switched system with mixed constraints and its parallel optimization algorithm, *Nonlinear Analysis: Hybrid Systems*, 25 (2017), 21–40.

Received October 2018; revised January 2019.

E-mail address: xuwei19951@qq.com
E-mail address: yly_shu@126.com
E-mail address: guihualin@shu.edu.cn
E-mail address: 18281102@qq.com