Switching Time and Parameter Optimization in Nonlinear Switched Systems with Multiple Time-Delays

Chongyang Liu
Ryan Loxton, Curtin University of Technology
Kok Lay Teo, Curtin University of Technology
Switching Time and Parameter Optimization in Nonlinear Switched Systems with Multiple Time-Delays

Chongyang Liu · Ryan Loxton · Kok Lay Teo

Received: 28 July 2013 / Accepted: 24 January 2014 / Published online: 25 March 2014
© Springer Science+Business Media New York 2014

Abstract In this paper, we consider a dynamic optimization problem involving a general switched system that evolves by switching between several subsystems of nonlinear delay-differential equations. The optimization variables in this system consist of: (1) the times at which the subsystem switches occur; and (2) a set of system parameters that influence the subsystem dynamics. We first establish the existence of the partial derivatives of the system state with respect to both the switching times and the system parameters. Then, on the basis of this result, we show that the gradient of the cost function can be computed by solving the state system forward in time followed by a costate system backward in time. This gradient computation procedure can be combined with any gradient-based optimization method to determine the optimal switching times and parameters. We propose an effective optimization algorithm based on this idea. Finally, we consider three numerical examples, one involving the 1,3-propanediol fed-batch production process, to illustrate the effectiveness and applicability of the proposed algorithm.

Communicated by Alexander S. Strekalovsky.

C. Liu
School of Mathematics and Information Science, Shandong Institute of Business and Technology, Yantai, China
e-mail: liu_chongyang@yahoo.com

R. Loxton · K. L. Teo (✉)
Department of Mathematics and Statistics, Curtin University, Perth, Australia
e-mail: k.l.teo@curtin.edu.au

R. Loxton
e-mail: r.loxton@curtin.edu.au

R. Loxton
Institute of Cyber-Systems and Control, Zhejiang University, Hangzhou, China
Keywords  Switched system · Time-delay system · Switching times · Nonlinear optimization

Mathematics Subject Classification  49M07 · 49M37 · 65K10

1 Introduction

A hybrid system is a dynamic system that exhibits both continuous and discrete characteristics [1]. The continuous characteristics of a hybrid system are typically modelled by difference or differential equations, while the discrete characteristics are modelled by discrete logic or discrete events [2]. Because of their flexibility, hybrid systems are often capable of accomplishing control objectives beyond the reach of conventional control systems [3,4]. Thus, hybrid systems and hybrid control have become hot research topics over the past decade. In this paper, we are concerned with a special class of hybrid systems called switched systems. Such systems consist of a number of distinct subsystems, with a switching law governing when the system switches from one subsystem to another. Switched systems arise in many real-world applications, including locomotives [5], biochemical reactors [6], switched-capacitor DC–DC power converters [7], and hybrid power systems [8].

The optimization of switched systems is an important and challenging research topic for applied mathematicians [9–14]. In the most general case, determining an optimal operation strategy for a switched system involves determining an optimal continuous input function (a continuous optimization variable) and an optimal switching sequence (a discrete optimization variable). Many recent works in the literature are based on the assumption that the switching sequence is fixed and the subsystem switching instants are decision variables to be optimized. Relevant literature includes [15], in which linear switched systems with pre-fixed switching sequence are investigated over an infinite time horizon, and [16,17], in which computational approaches are developed based on the partial derivatives of the cost function with respect to the switching times. More recently in [18], a new method for calculating the partial derivatives of the cost function is developed for switched systems in which the state experiences instantaneous jumps at the switching times. This new method is based on the time-scaling transformation described in [7,19].

The vast majority of optimization techniques for switched systems, including those mentioned above, are restricted to switched systems without time-delays. However, time-delays are common in practical engineering systems [20]. Indeed, switched systems with time-delays have various applications in areas such as power systems [21] and network control systems [22]. The presence of delays in a switched system complicates the search for an optimal operation policy. In particular, the time-scaling transformation mentioned above, a powerful tool for solving switched system optimization problems [7,13,18,19], is not applicable to switched systems with time-delays [23]. In fact, optimization techniques for switched systems with time-delays are scarce in the literature. Necessary conditions for determining optimal switching times and/or optimal impulse magnitudes for such systems are derived in [24–26] via classical variational techniques. However, these analytical results are only applicable to sepa-
rable systems with a single delay. Reference [27] presents an effective optimization algorithm for switched systems with time-delays, which is based on a parameterization scheme in which the switching instants are expressed in terms of the subsystem durations. However, this algorithm has three limitations: (1) it is only applicable to switched systems with a single delay and no system parameters; (2) it involves integrating a large number of auxiliary differential equations (one auxiliary system for each subsystem); and (3) it is based on the assumption that the duration between any two adjacent switching times is always larger than the time-delay.

The purpose of this paper is to develop a new computational method that does not suffer from these drawbacks. We consider a general nonlinear switched system with multiple time-delays and multiple system parameters. Each subsystem in this switched system is described by a set of nonlinear delay-differential equations, and the switching times and system parameters are optimization variables to be determined. The optimization problem involves choosing these optimization variables to minimize a given cost function. We first investigate the existence of the partial derivatives of the system state with respect to the system parameters and the switching times. In particular, we will show that the left and right partial derivatives of the system state with respect to each switching time exist and are equal at all time points except for the switching point in question. We then derive the gradient of the cost function and show that this gradient can be computed by solving the original state system forward in time, followed by an auxiliary system—called the costate system—backward in time. This is different from the method in [27], which involves solving multiple auxiliary systems forward in time. The advantage of our new approach is that we only require one auxiliary system, whereas the method in [27] requires many such systems. Moreover, our new algorithm caters for more general switched systems with multiple delays and multiple system parameters. We conclude the paper by validating our new algorithm on three numerical examples.

2 Problem Formulation

Consider the following switched system with $p$ subsystems and $m$ time-delays:

$$
\begin{align*}
\dot{x}(t) &= f^i(t, x(t), x(t - \alpha_1), \ldots, x(t - \alpha_m), \zeta), \quad t \in [\tau_{i-1}, \tau_i], \quad i = 1, 2, \ldots, p, \\
x(t) &= \phi(t, \zeta), \quad t \leq 0,
\end{align*}
$$

where $\tau_0 = 0$ is the initial time; $\tau_p = T > 0$ is a given terminal time; $\tau_i, i = 1, 2, \ldots, p - 1$, are switching times; $x(t) = (x_1(t), x_2(t), \ldots, x_n(t))^\top \in \mathbb{R}^n$ is the state vector; $\alpha_j, j = 1, 2, \ldots, m$, are given time-delays; $\zeta = (\zeta_1, \zeta_2, \ldots, \zeta_v)^\top \in \mathbb{R}^v$ is a vector of system parameters; and $f^i : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^{nm} \times \mathbb{R}^v \to \mathbb{R}^n, i = 1, 2, \ldots, p$, and $\phi : \mathbb{R} \times \mathbb{R}^v \to \mathbb{R}^n$ are given functions.

Equation (1a) expresses the dynamics of each subsystem in terms of the current time, the current state, the delayed state, and the parameter vector. Note that the time-delays and system parameters may be different for each subsystem. For example,
subsystem 1 may only involve $\zeta_1$ and $x(t - \alpha_1)$, subsystem 2 may only involve $\zeta_2$ and $x(t - \alpha_2)$, and so on.

Define

$$\tilde{x}(t) := (x(t - \alpha_1)^\top, x(t - \alpha_2)^\top, \ldots, x(t - \alpha_m)^\top)^\top.$$  

We will use this notation throughout the paper.

System (1) is controlled by manipulating the switching times and the system parameters. The switching times must satisfy the following constraints:

$$\tau_i - \tau_{i-1} \geq \Delta_i, \quad i = 1, 2, \ldots, p,$$  

where $\Delta_i > 0$ is the minimum duration of the $i$th subsystem. Any vector $\tau := (\tau_1, \tau_2, \ldots, \tau_{p-1})^\top$ satisfying (2) is called a feasible switching time vector. Let $T$ denote the set of all such feasible switching time vectors.

Define

$$Z := \{(\zeta_1, \zeta_2, \ldots, \zeta_v)^\top \in \mathbb{R}^v : a_q \leq \zeta_q \leq b_q, \quad q = 1, 2, \ldots, v\},$$  

where $a_q$ and $b_q$ are given constants such that $a_q \leq b_q$. Clearly, $Z$ is a compact and convex subset of $\mathbb{R}^v$. Any vector $\zeta \in Z$ is called a feasible parameter vector. Accordingly, any pair $(\tau, \zeta) \in T \times Z$ is called a feasible pair for system (1).

We assume throughout this paper that the following conditions are satisfied.

**Assumption 2.1** The functions $f^i, i = 1, 2, \ldots, p$, are continuously differentiable. Moreover, the function $\phi$ is twice continuously differentiable.

**Assumption 2.2** There exists a positive real number $L_1 > 0$ such that, for each $i = 1, 2, \ldots, p$,

$$|f^i(t, x, \tilde{x}, \zeta)| \leq L_1(1 + |x| + |\tilde{x}|), \quad (t, x, \tilde{x}, \zeta) \in [0, T] \times \mathbb{R}^n \times \mathbb{R}^{nm} \times Z,$$

where $|\cdot|$ denotes the Euclidean norm.

Assumptions 2.1–2.2 ensure that the switched system (1) has a unique solution $x(\cdot|\tau, \zeta)$ corresponding to each pair $(\tau, \zeta) \in T \times Z$ [28]. This solution is called the state trajectory.

We suppose that the total system cost can be expressed as a given function of the system parameters and the final state reached by the system. Accordingly, we define the following cost function:

$$J(\tau, \zeta) := \Phi(x(T|\tau, \zeta), \zeta),$$  

where $\Phi : \mathbb{R}^n \times \mathbb{R}^v \to \mathbb{R}$ is a given continuously differentiable function. Note that we can easily transform an integral running cost into the form of (4) by introducing an additional state variable. For example, consider the following integral term:
\[
\sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_i} L^i(t, x(t), \tilde{x}(t), \zeta) \, dt,
\]
where \( L^i : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^{nm} \times \mathbb{R}^v \rightarrow \mathbb{R}, i = 1, 2, \ldots, p, \) are given continuously differentiable functions. Clearly, this term can be replaced by \( x_{n+1}(T), \) where \( x_{n+1} \) satisfies the dynamics
\[
\dot{x}_{n+1}(t) = L^i(t, x(t), \tilde{x}(t), \zeta), \quad t \in ]\tau_{i-1}, \tau_i], \quad i = 1, 2, \ldots, p,
\]
\[x_{n+1}(t) = 0, \quad t \leq 0.\]

Thus, there is no loss of generality in ignoring integral terms in the cost function (4).

Our objective is to choose the switching time vector \( \tau \in \mathcal{T} \) and the parameter vector \( \zeta \in \mathcal{Z} \) to minimize the cost function (4) subject to the switched system (1). This optimization problem is stated formally as follows.

**Problem (P).** Find a feasible pair \((\tau, \zeta)\) \(\in\) \(\mathcal{T}\) \(\times\) \(\mathcal{Z}\) such that the cost function (4) is minimized.

Problem (P) is an optimization problem involving a nonlinear switched system with multiple time-delays, where the switching times and system parameters are variables to be optimized. It is well known that variable switching times pose a significant challenge for conventional numerical optimization techniques [18, 29, 30]. One of the most widely used methods for overcoming this challenge is the time-scaling transformation described in [7, 13, 18, 19]. This transformation involves mapping the variable switching times to fixed points in a new time horizon. Unfortunately, since the time-scaling transformation is not applicable to delay systems [23], it cannot be used to solve Problem (P). Therefore, a new approach is needed to solve Problem (P). This provides the motivation for the work in this paper.

### 3 State Variation

In essence, Problem (P) is a nonlinear programming problem with decision vectors \( \tau \) and \( \zeta \). The main difficulty with solving this problem is that the cost function (4) is not an explicit function of the decision vectors. Indeed, the decision vectors influence the cost function *implicitly* through the governing switched system (1); changing the switching times and/or system parameters changes the state trajectory, which subsequently changes the value of the cost function. It follows that computing the gradient of (4) is a challenging task.

Our ultimate goal is to develop a computational algorithm for evaluating this gradient. Such an algorithm can then be combined with any standard gradient-based optimization method to solve Problem (P) as a nonlinear programming problem. In this section, as a preliminary step towards determining the gradient of the cost function in Problem (P), we establish the existence of the gradient of the system state.

Let \( \mathcal{S} \) denote the set of all switching time vectors \( \tau = (\tau_1, \tau_2, \ldots, \tau_{p-1})^T \) such that
0 = \tau_0 < \tau_1 < \cdots < \tau_{p-1} < \tau_p = T.

Clearly, \( T \subset S \). We have already mentioned that system (1) admits a unique solution corresponding to each pair in \( T \times Z \). This result can, in fact, be extended to pairs in \( S \times Z \). Let \( x(t|\cdot, \cdot) \) denote the unique solution of (1) corresponding to \( (\tau, \zeta) \in S \times Z \). Furthermore, let \( \tilde{x}(t|\cdot, \cdot) : S \times Z \to \mathbb{R}^n \) denote the function that returns the value of the state vector at time \( t \) corresponding to a given input pair in \( S \times Z \). We will show in this section that the partial derivatives of the function \( x(t|\cdot, \cdot) \) exist on \( S \times Z \). This result will then be exploited in Sect. 4 to derive an effective computational procedure for calculating the gradient of the cost function (4).

The partial derivative of \( x(t|\cdot, \cdot) \) with respect to \( \zeta \) is called the state variation with respect to \( \zeta \). The following result, which can be proved in a similar manner to Theorem 2 in [31], gives a method for determining this state variation. Note that, in this and subsequent sections, we use the notation \( \partial \tilde{x}^j \) to denote differentiation with respect to the delayed state \( x(t-\alpha_j) \).

**Theorem 3.1** Let \( t \in [0, T] \) be a fixed time point. Then,

\[
\frac{\partial x(t|\tau, \zeta)}{\partial \zeta^q} = \Gamma^q(t|\tau, \zeta), \quad q = 1, 2, \ldots, v, \tag{5}
\]

where \( \Gamma^q(t|\tau, \zeta) \) is the solution of the following auxiliary switched system:

\[
\dot{\Gamma}^q(s) = \frac{\partial f^i(s, x(s), \tilde{x}(s), \zeta)}{\partial x} \Gamma^q(s) + \sum_{j=1}^{m} \frac{\partial f^i(s, x(s), \tilde{x}(s), \zeta)}{\partial \tilde{x}^j} \Gamma^q(s-\alpha_j) \nonumber \\
+ \frac{\partial f^i(s, x(s), \tilde{x}(s), \zeta)}{\partial \zeta^q}, \quad s \in [\tau_{i-1}, \tau_i], \quad i = 1, 2, \ldots, p, \tag{6}
\]

with the initial condition

\[
\Gamma^q(s) = \frac{\partial \phi(s, \zeta)}{\partial \zeta^q}, \quad s \leq 0. \tag{7}
\]

Theorem 3.1 gives the state variation with respect to the system parameters. To solve Problem (P), we also need the state variation with respect to the switching times. Unfortunately, unlike the state variation with respect to the system parameters, the state variation with respect to the switching times does not follow easily from known results. In fact, it is well known in the dynamic optimization field that gradient calculations involving variable switching times present major numerical challenges [18]. The remainder of this section is devoted to this important issue.

### 3.1 Preliminaries

Let \( k \in \{1, 2, \ldots, p-1\} \) and \( (\tau, \zeta) \in S \times Z \) be arbitrary but fixed. For notational simplicity, we write \( x(\cdot) \) instead of \( x(t|\tau, \zeta) \) and \( x^\epsilon(\cdot) \) instead of \( x(t|\tau+\epsilon e_k, \zeta) \), where \( e_k \) denotes the \( k \)th unit basis vector in \( \mathbb{R}^{p-1} \).

\( \subseteq \) Springer
Define
\[ \Theta := \tau_{k-1} - \tau_k, \tau_{k+1} - \tau_k. \]

Note that \( \Theta \) is a nonempty and open interval. Clearly,
\[ \epsilon \in \Theta \iff \tau + \epsilon e^k \in S. \]

Now, for each \( \epsilon \in \Theta \), define the following functions:
\[ \varphi^\epsilon (t) := x^\epsilon (t) - x(t), \quad t \leq T, \quad (8) \]
\[ \theta^\epsilon,j (t) := x^\epsilon (t - \alpha_j) - x(t - \alpha_j), \quad t \leq T, \quad j = 1, 2, \ldots, m. \quad (9) \]

Furthermore, let
\[ \theta^\epsilon (t) := (\theta^{\epsilon,1}(t)^\top, \theta^{\epsilon,2}(t)^\top, \ldots, \theta^{\epsilon,m}(t)^\top)^\top \in \mathbb{R}^{nm}, \quad t \leq T, \]
and
\[ \hat{x}^\epsilon (t) := (x^\epsilon (t - \alpha_1)^\top, x^\epsilon (t - \alpha_2)^\top, \ldots, x^\epsilon (t - \alpha_m)^\top)^\top, \quad t \leq T. \]

Obviously, for each \( \epsilon \in \Theta \),
\[ \theta^\epsilon,j (t) = \varphi^\epsilon (t - \alpha_j), \quad t \leq T, \quad j = 1, 2, \ldots, m, \quad (10) \]
and
\[ \varphi^\epsilon (t) = 0, \quad t \leq \min \{ \tau_k + \epsilon, \tau_k \}. \quad (11) \]

The following result can be proved in a similar manner to Lemma 6.4.2 in [32]. Note that Assumption 2.2 plays a key role in the proof.

**Lemma 3.1** There exists a real number \( L_2 > 0 \) such that, for all \( \epsilon \in \Theta \),
\[ |x^\epsilon (t)| \leq L_2, \quad t \in [-\alpha_{\text{max}}, T], \quad (12) \]
where \( \alpha_{\text{max}} := \max_{j \in \{1, 2, \ldots, m\}} \{ \alpha_j \} \).

In the next lemma, we give an important upper bound for the functions defined in (8) and (9).

**Lemma 3.2** There exists a real number \( L_3 > 0 \) such that, for all \( \epsilon \in \Theta \),
\[ |\varphi^\epsilon (t)|, \quad \max_{j \in \{1, 2, \ldots, m\}} |\theta^\epsilon,j (t)| \leq L_3 |\epsilon|, \quad t \in [0, T]. \quad (13) \]
**Proof** Let \( t \in [0, T] \) be arbitrary but fixed. Furthermore, let \( i(s) \in \{1, 2, \ldots, p\} \) denote the active subsystem at time \( s \) corresponding to the switching time vector \( \tau \), and let \( i^\epsilon(s) \) denote the active subsystem at time \( s \) corresponding to the switching time vector \( \tau + \epsilon e^k \). Then,

\[
x(t) = \phi(0, \zeta) + \int_0^t f^{i(s)}(s, x(s), \tilde{x}(s), \zeta) \, ds
\]

and

\[
x^\epsilon(t) = \phi(0, \zeta) + \int_0^t f^{i^\epsilon(s)}(s, x^\epsilon(s), \tilde{x}^\epsilon(s), \zeta) \, ds.
\]

Combining (14) and (15) gives

\[
\varphi^\epsilon(t) = x^\epsilon(t) - x(t)
= \int_0^t \left\{ f^{i^\epsilon(s)}(s, x^\epsilon(s), \tilde{x}^\epsilon(s), \zeta) - f^{i(s)}(s, x(s), \tilde{x}(s), \zeta) \right\} \, ds.
\]

Since \( \epsilon \in \Theta \), there exists an interval \( I_\epsilon \) such that \( |I_\epsilon| = \epsilon \) and

\[i^\epsilon(s) = i(s), \quad s \notin I_\epsilon.\]

Thus, we can rewrite (16) as follows:

\[
\varphi^\epsilon(t) = \int_{[0,t] \cap I_\epsilon} \left\{ f^{i^\epsilon(s)}(s, x^\epsilon(s), \tilde{x}^\epsilon(s), \zeta) - f^{i(s)}(s, x(s), \tilde{x}(s), \zeta) \right\} \, ds
+ \int_{[0,t] \setminus I_\epsilon} \left\{ f^{i(s)}(s, x^\epsilon(s), \tilde{x}^\epsilon(s), \zeta) - f^{i(s)}(s, x(s), \tilde{x}(s), \zeta) \right\} \, ds.
\]

Taking the norm of both sides of (17) yields

\[
|\varphi^\epsilon(t)| \leq \int_0^t |f^{i^\epsilon(s)}(s, x^\epsilon(s), \tilde{x}^\epsilon(s), \zeta) - f^{i(s)}(s, x(s), \tilde{x}(s), \zeta)| \, ds
+ \int_{I_\epsilon} |f^{i(s)}(s, x^\epsilon(s), \tilde{x}^\epsilon(s), \zeta) - f^{i(s)}(s, x(s), \tilde{x}(s), \zeta)| \, ds.
\]

Now, by Assumption 2.1 and (12), there exists a Lipschitz constant \( M_1 > 0 \) such that the first integral term on the right-hand side of (18) can be simplified as follows:
\[
\int_0^t \left| f^i(s, x^\epsilon(s), \tilde{x}^\epsilon(s), \zeta) - f^i(s, x(s), \tilde{x}(s), \zeta) \right| ds \\
\leq \int_0^t M_1 |\phi^\epsilon(s)| ds + \sum_{j=1}^m \int_0^t M_1 |\theta^{\epsilon,j}(s)| ds.
\]

Furthermore, there exists another constant \(M_2 > 0\) such that, for all \(\epsilon \in \Theta\),
\[
\left| f^l(s, x^\epsilon(s), \tilde{x}^\epsilon(s), \zeta) \right| \leq M_2, \quad s \in [0, T], \quad l = 1, 2, \ldots, p.
\]

Thus, the last integral term on the right-hand side of (18) can be simplified as follows:
\[
\int_{I_\epsilon} \left| f^i(s, x^\epsilon(s), \tilde{x}^\epsilon(s), \zeta) - f^i(s, x(s), \tilde{x}(s), \zeta) \right| ds \leq 2M_2|\epsilon|.
\]

Substituting (19) and (20) into (18), and then simplifying using (10) and (11), we obtain
\[
|\phi^\epsilon(t)| \leq 2M_2|\epsilon| + \int_0^t M_1 |\phi^\epsilon(s)| ds + \sum_{j=1}^m \int_0^t M_1 |\phi^\epsilon(s - \alpha_j)| ds \\
\leq 2M_2|\epsilon| + (m + 1)M_1 \int_0^t |\phi^\epsilon(s)| ds.
\]

Thus, by the Gronwall–Bellman Lemma [32], it follows that
\[
|\phi^\epsilon(t)| \leq L_3|\epsilon|,
\]
where \(L_3 := 2M_2 \exp(mM_1T + M_1T)\). Therefore, by (10) and (11),
\[
\max_{j \in \{1,2,\ldots,m\}} |\theta^{\epsilon,j}(t)| = \max_{j \in \{1,2,\ldots,m\}} |\phi^\epsilon(t - \alpha_j)| \leq L_3|\epsilon|.
\]

Since \(t \in [0, T]\) and \(\epsilon \in \Theta\) were chosen arbitrarily, (21) and (22) verify the result.

Now, consider the following auxiliary switched system:
\[
\dot{\Lambda}^k(s) = \frac{\partial f^i(s, x(s), \tilde{x}(s), \zeta)}{\partial x} \Lambda^k(s) + \sum_{j=1}^m \frac{\partial f^i(s, x(s), \tilde{x}(s), \zeta)}{\partial \tilde{x}^j} \Lambda^k(s - \alpha_j), \\
s \in [\tau_{i-1}, \tau_i], \quad i = k + 1, \ldots, p.
\]
with the jump condition
\[ \Lambda^k(\tau_k^+) = f^k(\tau_k, x(\tau_k), \bar{x}(\tau_k), \zeta) - f^{k+1}(\tau_k, x(\tau_k), \bar{x}(\tau_k), \zeta) \] (24)
and the initial condition
\[ \Lambda^k(s) = 0, \quad s < \tau_k. \] (25)

Let \( \Lambda^k(\cdot | \tau, \zeta) \) denote the unique right-continuous solution of (23)–(25). We will show that the state variation with respect to \( \tau_k \) is equal to \( \Lambda^k \) at all time points \( t \neq \tau_k \).

### 3.2 Right State Variation with Respect to the Switching Times

The right state variation at time \( t \) with respect to \( \tau_k \) is defined by
\[ \lim_{\epsilon \downarrow 0} \frac{x(t | \tau + \epsilon \epsilon^k, \zeta) - x(t | \tau, \zeta)}{\epsilon} = \lim_{\epsilon \downarrow 0} \frac{x^\epsilon(t) - x(t)}{\epsilon}. \]

We will now establish the existence of the right state variation.

For each \( \epsilon \in \Theta \cap [0, \infty[ \), define
\[ \rho^\epsilon_k := \int_{\tau_k}^{\tau_k + \epsilon} \{ f^k(s, x^\epsilon(s), \bar{x}^\epsilon(s), \zeta) - f^k(\tau_k, x(\tau_k), \bar{x}(\tau_k), \zeta) \} ds \] (26)
and
\[ \rho^\epsilon_{k+1} := \int_{\tau_k}^{\tau_k + \epsilon} \{ f^{k+1}(s, x^\epsilon(s), \bar{x}^\epsilon(s), \zeta) - f^{k+1}(\tau_k, x(\tau_k), \bar{x}(\tau_k), \zeta) \} ds. \] (27)

We now derive an upper bound for the functions defined in (26) and (27).

**Lemma 3.3** There exists a real number \( L_4 > 0 \) such that, for all \( \epsilon \in \Theta \cap [0, \infty[ \),
\[ |\rho^\epsilon_k|, \ |\rho^\epsilon_{k+1}| \leq L_4 \epsilon^2. \]

**Proof** From the definition of \( \rho^\epsilon_k \) in (26), we have
\[ |\rho^\epsilon_k| \leq \int_{\tau_k}^{\tau_k + \epsilon} \left| f^k(s, x^\epsilon(s), \bar{x}^\epsilon(s), \zeta) - f^k(s, x(s), \bar{x}(s), \zeta) \right| ds \\
+ \int_{\tau_k}^{\tau_k + \epsilon} \left| f^k(s, x(s), \bar{x}(s), \zeta) - f^k(\tau_k, x(\tau_k), \bar{x}(\tau_k), \zeta) \right| ds. \]
Thus, since $f^k$ is Lipschitz continuous,

$$
|\rho_\epsilon^k| \leq \int_{\tau_k}^{\tau_k+\epsilon} M_1 |\varphi^\epsilon(s)|\,ds + \sum_{j=1}^m \int_{\tau_k}^{\tau_k+\epsilon} M_1 |\theta^\epsilon_{j-\epsilon}(s)|\,ds + \int_{\tau_k}^{\tau_k+\epsilon} M_1 |s - \tau_k|\,ds
$$

$$
+ \int_{\tau_k}^{\tau_k+\epsilon} M_1 |x(s) - x(\tau_k)|\,ds + \sum_{j=1}^m \int_{\tau_k}^{\tau_k+\epsilon} M_1 |x(s - \alpha_j) - x(\tau_k - \alpha_j)|\,ds,
$$

where $M_1 > 0$ is the Lipschitz constant defined in the proof of Lemma 3.2. Applying Lemma 3.2, we obtain

$$
|\rho_\epsilon^k| \leq M_1 L_3 \epsilon^2 + mM_1 L_3 \epsilon^2 + M_1 \epsilon^2 + \int_{\tau_k}^{\tau_k+\epsilon} M_1 |x(s) - x(\tau_k)|\,ds
$$

$$
+ \sum_{j=1}^m \int_{\tau_k}^{\tau_k+\epsilon} M_1 |x(s - \alpha_j) - x(\tau_k - \alpha_j)|\,ds. \quad (28)
$$

Now, let $M_2 > 0$ be as defined in the proof of Lemma 3.2. Since $\phi$ is continuously differentiable, we may assume that $M_2$ is such that $|\dot{\phi}(\eta, \zeta)| \leq M_2$ for all $\eta \in [-\alpha_{\max}, 0]$, where $\alpha_{\max} := \max_{j \in \{1, 2, \ldots, m\}} \{\alpha_j\}$. Then clearly,

$$
|\dot{x}(\eta)| \leq M_2, \quad \eta \in [-\alpha_{\max}, T].
$$

Hence, for each $s \in [\tau_k, \tau_k + \epsilon],

$$
|x(s) - x(\tau_k)| \leq \int_{\tau_k}^{s} |\dot{x}(\eta)|\,d\eta \leq \int_{\tau_k}^{\tau_k+\epsilon} |\dot{x}(\eta)|\,d\eta \leq M_2 \epsilon, \quad (29)
$$

and

$$
|x(s - \alpha_j) - x(\tau_k - \alpha_j)| \leq \int_{\tau_k - \alpha_j}^{s-\alpha_j} |\dot{x}(\eta)|\,d\eta \leq \int_{\tau_k - \alpha_j}^{\tau_k+\epsilon-\alpha_j} |\dot{x}(\eta)|\,d\eta \leq M_2 \epsilon. \quad (30)
$$

Substituting (29) and (30) into (28) gives

$$
|\rho_\epsilon^k| \leq M_1 L_3 \epsilon^2 + mM_1 L_3 \epsilon^2 + M_1 \epsilon^2 + M_1 M_2 \epsilon^2 + mM_1 M_2 \epsilon^2. \quad (31)
$$

In a similar manner, it is also possible to show that

$$
|\rho_{k+1}^\epsilon| \leq M_1 L_3 \epsilon^2 + mM_1 L_3 \epsilon^2 + M_1 \epsilon^2 + M_1 M_2 \epsilon^2 + mM_1 M_2 \epsilon^2. \quad (32)
$$
Choosing \( L_4 := M_1 L_3 + M_1 + M_1 M_2 + m M_1 M_2 \) in (31) and (32) completes the proof. \( \square \)

On the basis of Lemmas 3.1–3.3, we now establish the existence of the right state variation with respect to the switching times.

**Theorem 3.2** Let \( k \in \{1, 2, \ldots, p - 1\} \) and \( (\tau, \zeta) \in S \times Z \). Then, for all time points \( t \neq \tau_k \),

\[
\lim_{\epsilon \downarrow 0} \epsilon^{-1} \varphi^{\epsilon}(t) = \Lambda^k(t|\tau, \zeta),
\]

where \( \Lambda^k(\cdot|\tau, \zeta) \) is the unique right-continuous solution of the auxiliary switched system (23)–(25). In addition, for \( t = \tau_k \),

\[
\lim_{\epsilon \downarrow 0} \epsilon^{-1} \varphi^{\epsilon}(\tau_k) = 0.
\]

**Proof** For each \( i = 1, 2, \ldots, p \) and \( \epsilon \in \Theta \), define the following functions:

\[
\tilde{f}^{i, \epsilon}(s, \eta) := f^i(s, x(s) + \eta \varphi^{\epsilon}(s), \tilde{x}(s) + \eta \theta^{\epsilon}(s), \zeta), \quad (s, \eta) \in [\tau_{i-1}, \tau_i] \times [0, 1],
\]

\[
\Delta_1^{i, \epsilon}(s, \eta) := \left\{ \frac{\partial \tilde{f}^{i, \epsilon}(s, \eta)}{\partial x} - \frac{\partial \tilde{f}^{i, \epsilon}(s, 0)}{\partial x} \right\} \varphi^{\epsilon}(s), \quad (s, \eta) \in [\tau_{i-1}, \tau_i] \times [0, 1],
\]

and

\[
\Delta_2^{i, \epsilon}(s, \eta) := \sum_{j=1}^{m} \left\{ \frac{\partial \tilde{f}^{i, \epsilon}(s, \eta)}{\partial \tilde{x}^j} - \frac{\partial \tilde{f}^{i, \epsilon}(s, 0)}{\partial \tilde{x}^j} \right\} \theta^{\epsilon, j}(s), \quad (s, \eta) \in [\tau_{i-1}, \tau_i] \times [0, 1].
\]

By Assumption 2.1 and Lemma 3.1, there exists a real number \( N_1 > 0 \) such that, for all \( \epsilon \in \Theta \),

\[
\left| \frac{\partial \tilde{f}^{i, \epsilon}(s, 0)}{\partial x} \right| \leq N_1, \quad s \in [\tau_{i-1}, \tau_i], \quad i = 1, 2, \ldots, p,
\]

and

\[
\left| \frac{\partial \tilde{f}^{i, \epsilon}(s, 0)}{\partial \tilde{x}^j} \right| \leq N_1, \quad s \in [\tau_{i-1}, \tau_i], \quad i = 1, 2, \ldots, p, \quad j = 1, 2, \ldots, m,
\]

where \( |\cdot| \) denotes the induced matrix norm on \( \mathbb{R}^{n \times n} \). Moreover, by Lemma 3.2, the following limits exist uniformly with respect to \( \eta \in [0, 1] \) and \( s \in [0, T] \):

\[
\lim_{\epsilon \downarrow 0} \{ x(s) + \eta \varphi^{\epsilon}(s) \} = x(s), \quad \lim_{\epsilon \downarrow 0} \{ \tilde{x}(s) + \eta \theta^{\epsilon}(s) \} = \tilde{x}(s).
\]

Thus, it follows from Assumption 2.1 that, for each \( \delta > 0 \), there exists a corresponding \( \epsilon' > 0 \) such that, for all \( \epsilon \in \Theta \) satisfying \( |\epsilon| < \epsilon' \),
\[
\left| \frac{\partial \tilde{f}_{i,\epsilon}(s, \eta)}{\partial x} - \frac{\partial \tilde{f}_{i,\epsilon}(s, 0)}{\partial x} \right| < \delta, \quad (s, \eta) \in [\tau_{i-1}, \tau_i] \times [0, 1], \quad i = 1, 2, \ldots, p. \tag{35}
\]

and
\[
\left| \frac{\partial \tilde{f}_{i,\epsilon}(s, \eta)}{\partial \tilde{x}^j} - \frac{\partial \tilde{f}_{i,\epsilon}(s, 0)}{\partial \tilde{x}^j} \right| < \delta, \quad (s, \eta) \in [\tau_{i-1}, \tau_i] \times [0, 1], \quad i = 1, 2, \ldots, p, \quad j = 1, 2, \ldots, m. \tag{36}
\]

Combining (35) and (36) with Lemma 3.2, we see that, for all \(\epsilon \in \Theta\) with \(|\epsilon| < \epsilon'\),
\[
|\Delta_i^{1,\epsilon}(s, \eta)| \leq \delta L_3|\epsilon|, \quad |\Delta_i^{2,\epsilon}(s, \eta)| \leq \delta m L_3|\epsilon|, \quad i = 1, 2, \ldots, p. \tag{37}
\]

By the chain rule, we obtain
\[
\frac{\partial \tilde{f}_{i,\epsilon}(s, \eta)}{\partial \eta} = \frac{\partial \tilde{f}_{i,\epsilon}(s, \eta)}{\partial x} \varphi^\epsilon(s) + \sum_{j=1}^{m} \frac{\partial \tilde{f}_{i,\epsilon}(s, \eta)}{\partial \tilde{x}^j} \theta^{\epsilon, j}(s). \tag{38}
\]

We can rewrite (38) as follows:
\[
\frac{\partial \tilde{f}_{i,\epsilon}(s, \eta)}{\partial \eta} = \Delta_i^{1,\epsilon}(s, \eta) + \Delta_i^{2,\epsilon}(s, \eta) + \frac{\partial \tilde{f}_{i,\epsilon}(s, 0)}{\partial x} \varphi^\epsilon(s) + \sum_{j=1}^{m} \frac{\partial \tilde{f}_{i,\epsilon}(s, 0)}{\partial \tilde{x}^j} \theta^{\epsilon, j}(s). \tag{39}
\]

Now, let \(t \in [0, T]\) be a fixed time point. We consider two cases: (1) \(t \leq \tau_k\); and (2) \(t > \tau_k\).

Consider case (1). In this case, \(t \leq \tau_k\), and thus it follows from (11) that
\[
\lim_{\epsilon \downarrow 0} \epsilon^{-1} \varphi^\epsilon(t) = 0. \tag{40}
\]

In particular, (40) holds when \(t = \tau_k\), thus proving (34). If \(t < \tau_k\), then in view of (25),
\[
A^k(t) = 0 = \lim_{\epsilon \downarrow 0} \epsilon^{-1} \varphi^\epsilon(t).
\]

Thus, (33) holds for any \(t < \tau_k\).

Now, consider case (2). In this case, there exists an integer \(\varsigma \in \{k + 1, \ldots, p\}\) such that \(t \in [\tau_{\varsigma-1}, \tau_{\varsigma}]\). Let \(\gamma \in [0, t - \tau_k]\) be arbitrary but fixed. Furthermore, choose \(\epsilon \in \Theta\) such that
\[
0 < \epsilon < \min\{\gamma, \epsilon'\}, \tag{41}
\]

where \(\epsilon'\) corresponds to \(\delta = \gamma\) in (35) and (36). Then, \(\tau_k + \epsilon < \tau_{k+1}\) and \(\tau_k + \epsilon < t\). Consequently,
\[
\varphi^\epsilon(t) = \int_{\tau_k}^{\tau_k + \epsilon} \left\{ \bar{f}^{k,\epsilon}(s, 1) - \bar{f}^{k+1,\epsilon}(s, 1) \right\} ds + \sum_{l=k+1}^{\min[\tau_l, t]} \int_{\tau_{l-1}}^{\tau_l} \left\{ \bar{f}^{l,\epsilon}(s, 1) - \bar{f}^{l,\epsilon}(s, 0) \right\} ds.
\]

Using (26) and (27), we obtain
\[
\varphi^\epsilon(t) = \rho^\epsilon_k - \rho^\epsilon_{k+1} + \epsilon \left\{ \bar{f}^{k,\epsilon}(\tau_k, 0) - \bar{f}^{k+1,\epsilon}(\tau_k, 0) \right\}
+ \sum_{l=k+1}^{\min[\tau_l, t]} \int_{\tau_{l-1}}^{\tau_l} \left\{ \bar{f}^{l,\epsilon}(s, 1) - \bar{f}^{l,\epsilon}(s, 0) \right\} ds. \tag{42}
\]

By using the fundamental theorem of calculus, (42) can be written as
\[
\varphi^\epsilon(t) = \rho^\epsilon_k - \rho^\epsilon_{k+1} + \epsilon \left\{ \bar{f}^{k,\epsilon}(\tau_k, 0) - \bar{f}^{k+1,\epsilon}(\tau_k, 0) \right\}
+ \sum_{l=k+1}^{\min[\tau_l, t]} \int_{\tau_{l-1}}^{\tau_l} \left( \int_0^1 \frac{\partial \bar{f}^{l,\epsilon}(s, \eta)}{\partial \eta} d\eta \right) ds. \tag{43}
\]

Substituting (39) into (43) gives
\[
\varphi^\epsilon(t) = \rho^\epsilon_k - \rho^\epsilon_{k+1} + \epsilon \left\{ \bar{f}^{k,\epsilon}(\tau_k, 0) - \bar{f}^{k+1,\epsilon}(\tau_k, 0) \right\}
+ \sum_{l=k+1}^{\min[\tau_l, t]} \int_{\tau_{l-1}}^{\tau_l} \left( \int_0^1 \frac{\partial \bar{f}^{l,\epsilon}(s, \eta)}{\partial \eta} d\eta + \int_0^1 \Delta_{l^2,\epsilon}(s, \eta) d\eta + \int_0^1 \Delta_{l^1,\epsilon}(s, \eta) d\eta + \frac{\partial \bar{f}^{l,\epsilon}(s, 0)}{\partial x} \varphi_s(s) \right)
+ \sum_{j=1}^m \frac{\partial \bar{f}^{l,\epsilon}(s, 0)}{\partial x^j} \theta^\epsilon_j(s) \right) ds. \tag{44}
\]

Next, integrating the auxiliary switched system (23)–(25) yields
\[
\Lambda^k(t) = \bar{f}^{k,\epsilon}(\tau_k, 0) - \bar{f}^{k+1,\epsilon}(\tau_k, 0) + \sum_{l=k+1}^{\min[\tau_l, t]} \int_{\tau_{l-1}}^{\tau_l} \left( \frac{\partial \bar{f}^{l,\epsilon}(s, 0)}{\partial x} \Lambda^k(s) \right)
+ \sum_{j=1}^m \frac{\partial \bar{f}^{l,\epsilon}(s, 0)}{\partial x^j} \Lambda^k(s - \alpha^j) \right) ds. \tag{45}
\]

Multiplying (44) by \(\epsilon^{-1}\), subtracting (45), taking the norm of both sides and then applying (37) with \(\delta = \gamma\), we obtain
\[ |\epsilon^{-1}\varphi^\epsilon(t) - \Lambda^k(t)| \leq \epsilon^{-1}|\rho_k^\epsilon| + \epsilon^{-1}|\rho_{k+1}^\epsilon| + (L_3 T + mL_3 T)\gamma \]
\[ + \int_{\tau_k}^{t} N_1 |\epsilon^{-1}\varphi^\epsilon(s) - \Lambda^k(s)|ds \]
\[ + \sum_{j=1}^{m} \int_{\tau_k - \alpha_j}^{t} N_1 |\epsilon^{-1}\varphi^\epsilon(s - \alpha_j) - \Lambda^k(s - \alpha_j)|ds. \]  
(46)

In view of (11) and (25), the last integral term on the right-hand side of (46) can be simplified as follows:
\[ \sum_{j=1}^{m} \int_{\tau_k}^{t} N_1 |\epsilon^{-1}\varphi^\epsilon(s - \alpha_j) - \Lambda^k(s - \alpha_j)|ds = \sum_{j=1}^{m} \int_{\tau_k - \alpha_j}^{t} N_1 |\epsilon^{-1}\varphi^\epsilon(s) - \Lambda^k(s)|ds \]
\[ \leq \int_{\tau_k}^{t} mN_1 |\epsilon^{-1}\varphi^\epsilon(s) - \Lambda^k(s)|ds. \]  
(47)

Thus, using Lemma 3.3 and (47) to simplify (46), we obtain
\[ |\epsilon^{-1}\varphi^\epsilon(t) - \Lambda^k(t)| \leq 2L_4\epsilon + (L_3 T + mL_3 T)\gamma \]
\[ + (m + 1)N_1 \int_{\tau_k}^{t} |\epsilon^{-1}\varphi^\epsilon(s) - \Lambda^k(s)|ds. \]

Since \( \epsilon < \gamma \) and \( \tau_k + \gamma < t \), this inequality becomes
\[ |\epsilon^{-1}\varphi^\epsilon(t) - \Lambda^k(t)| \leq 2L_4\epsilon + (L_3 T + mL_3 T)\gamma \]
\[ + (m + 1)N_1 \int_{\tau_k}^{\tau_k + \gamma} |\epsilon^{-1}\varphi^\epsilon(s) - \Lambda^k(s)|ds \]
\[ + (m + 1)N_1 \int_{\tau_k + \gamma}^{t} |\epsilon^{-1}\varphi^\epsilon(s) - \Lambda^k(s)|ds. \]  
(48)

Using (13), we have
\[ |\epsilon^{-1}\varphi^\epsilon(s) - \Lambda^k(s)| \leq \epsilon^{-1}|\varphi^\epsilon(s)| + |\Lambda^k(s)| \leq L_3 + N_2, \quad s \in [\tau_k, \tau_k + \gamma], \]  
(49)

where \( N_2 > 0 \) is an upper bound for \( |\Lambda^k| \) (recall that \( \Lambda^k \) is piecewise continuous and therefore bounded on \([0, T]\)). Substituting (49) into (48) gives
\[ |e^{-1}\varphi^\epsilon(t) - \Lambda^k(t)| \leq N_3 \gamma + (m + 1)N_1 \int_{\tau_k + \gamma}^{t} |e^{-1}\varphi^\epsilon(s) - \Lambda^k(s)| \, ds, \]

where \( N_3 := 2L_4 + L_3T + mL_3T + (m + 1)N_1(L_3 + N_2) \). This inequality holds for all \( \epsilon \in \Theta \) satisfying (41), uniformly with respect to \( t \in [\tau_k + \gamma, T] \). Thus, by the Gronwall–Bellman Lemma [32], it follows that

\[ |e^{-1}\varphi^\epsilon(t) - \Lambda^k(t)| \leq N_3 \gamma \exp((m + 1)N_1T), \]

whenever \( 0 < \epsilon < \min\{\gamma, \epsilon'\} \). Since \( \gamma > 0 \) was selected arbitrarily, this shows that (33) holds when \( t > \tau_k \). \( \Box \)

### 3.3 Left State Variation with Respect to the Switching Times

The left state variation at time \( t \) with respect to \( \tau_k \) is defined by

\[
\lim_{\epsilon \uparrow 0} x(t|\tau + \epsilon \epsilon^k, \zeta) - x(t|\tau, \zeta) = \lim_{\epsilon \uparrow 0} x^\epsilon(t) - x(t). 
\]

We will now establish the existence of the left state variation.

For each \( \epsilon \in \Theta \cap ] - \infty, 0] \), define

\[
\rho^\epsilon_k := \int_{\tau_k + \epsilon}^{\tau_k} \{ f^k(s, x(s), \tilde{x}(s), \zeta) - f^k(\tau_k, x(\tau_k), \tilde{x}(\tau_k), \zeta) \} \, ds, \quad (50)
\]

\[
\rho^\epsilon_{k+1} := \int_{\tau_k + \epsilon}^{\tau_k} \{ f^{k+1}(s, x^\epsilon(s), \tilde{x}^\epsilon(s), \zeta) - f^{k+1}(\tau_k, x(\tau_k), \tilde{x}(\tau_k), \zeta) \} \, ds. \quad (51)
\]

By using similar arguments as in the proof of Lemma 3.3, we obtain the following result.

**Lemma 3.4** There exists a real number \( L_5 > 0 \) such that, for all \( \epsilon \in \Theta \cap ] - \infty, 0] \),

\[ |\rho^\epsilon_k|, |\rho^\epsilon_{k+1}| \leq L_5 \epsilon^2. \]

The following theorem shows that the left state variation coincides exactly with the solution of the auxiliary switched system (23)–(25).

**Theorem 3.3** Let \( k \in \{1, 2, \ldots, p - 1\} \) and \((\tau, \zeta) \in S \times \mathcal{Z}\). Then, for all time points \( t \in [0, T] \),

\[ \lim_{\epsilon \uparrow 0} \epsilon^{-1}\varphi^\epsilon(t) = \Lambda^k(t|\tau, \zeta), \]

(52)
where $\Lambda^k(\cdot|\tau, \zeta)$ is the unique right-continuous solution of the auxiliary switched system (23)–(25).

Proof Let $\bar{f}_{1,\epsilon}, \Delta_{1,\epsilon}^i$, and $\Delta_{2,\epsilon}^i$ be as defined in the proof of Theorem 3.2. Recall that, for each $\delta > 0$, there exists a corresponding $\epsilon' > 0$ such that, for all $\epsilon \in \Theta$ satisfying $|\epsilon| < \epsilon'$,

$$|\Delta_{1,\epsilon}^i(s, \eta)| \leq \delta L_3|\epsilon|, \quad |\Delta_{2,\epsilon}^i(s, \eta)| \leq \delta mL_3|\epsilon|, \quad i = 1, 2, \ldots, p,$$

where $L_3$ is as defined in Lemma 3.2.

Let $t \in [0, T]$ be a fixed time point. We consider two cases: (1) $t < \tau_k$; and (2) $t \geq \tau_k$.

For case (1), it follows from (11) that

$$\lim_{\epsilon \uparrow 0} \epsilon^{-1} \varphi^\epsilon(t) = 0.$$ 

Thus, in view of (25),

$$\Lambda^k(t) = 0 = \lim_{\epsilon \uparrow 0} \epsilon^{-1} \varphi^\epsilon(t),$$

which proves (52) for $t < \tau_k$.

For case (2), there exists an integer $\varsigma \in \{k + 1, \ldots, p\}$ such that $t \in [\tau_{\varsigma-1}, \tau_{\varsigma}]$ if $\varsigma 

$$- \min\{\epsilon', \delta\} < \epsilon < 0.$$ 

(54)

Then, $\tau_{k-1} < \tau_k + \epsilon < \tau_k$. As a result,

$$\varphi^\epsilon(t) = \int_{\tau_k + \epsilon}^{\tau_k} \{\bar{f}_{k+1,\epsilon}(s, 1) - \bar{f}_{k,\epsilon}(s, 0)\} ds$$

$$+ \sum_{l=k+1}^{\varsigma} \int_{\tau_{l-1}}^{\min\{\tau_l, t\}} \{\bar{f}_{l,\epsilon}(s, 1) - \bar{f}_{l,\epsilon}(s, 0)\} ds.$$

Therefore, from (50) and (51),

$$\varphi^\epsilon(t) = \rho_{k+1}^\epsilon - \rho_k^\epsilon - \epsilon \left\{\bar{f}_{k+1,\epsilon}(\tau_k, 0) - \bar{f}_{k,\epsilon}(\tau_k, 0)\right\}$$

$$+ \sum_{l=k+1}^{\varsigma} \int_{\tau_{l-1}}^{\min\{\tau_l, t\}} \{f_{l,\epsilon}(s, 1) - f_{l,\epsilon}(s, 0)\} ds.$$ 

Using the fundamental theorem of calculus, (55) can be written as
\[
\varphi^\varepsilon(t) = \rho_{k+1}^\varepsilon - \rho_k^\varepsilon - \varepsilon \left\{ \tilde{f}^{k+1,\varepsilon}(\tau_k, 0) - \tilde{f}^{k,\varepsilon}(\tau_k, 0) \right\} \\
+ \sum_{l=k+1}^{\zeta} \int_{\tau_{l-1}}^{\tau_l} \frac{\partial \tilde{f}^l,\varepsilon(s, 0)}{\partial x} \varphi^\varepsilon(s) \, ds + \sum_{l=k+1}^{\zeta} \int_{\tau_{l-1}}^{\tau_l} \sum_{j=1}^{m} \frac{\partial \tilde{f}^l,\varepsilon(s, 0)}{\partial \tilde{x}^j} \theta^{\varepsilon, j}(s) \, ds \\
+ \sum_{l=k+1}^{\zeta} \int_{\tau_{l-1}}^{\tau_l} \left( \int_0^1 \Delta_1^{l,\varepsilon}(s, \eta) \, d\eta + \int_0^1 \Delta_2^{l,\varepsilon}(s, \eta) \, d\eta \right) \, ds.
\]

Now, the right-continuous solution of the auxiliary switched system (23)–(25) is

\[
\Lambda^k(t) = \tilde{f}^{k,\varepsilon}(\tau_k, 0) - \tilde{f}^{k+1,\varepsilon}(\tau_k, 0) + \sum_{l=k+1}^{\zeta} \int_{\tau_{l-1}}^{\tau_l} \left( \frac{\partial \tilde{f}^l,\varepsilon(s, 0)}{\partial x} \Lambda^k(s) \right) \, ds \\
+ \sum_{j=1}^{m} \frac{\partial \tilde{f}^l,\varepsilon(s, 0)}{\partial \tilde{x}^j} \Lambda^k(s - \alpha_j) \, ds.
\]

From (56) and (57), we obtain

\[
|\varepsilon^{-1} \varphi^\varepsilon(t) - \Lambda^k(t)| \leq |\varepsilon|^{-1} |\rho_k^\varepsilon| + |\varepsilon|^{-1} |\rho_{k+1}^\varepsilon| + \int_{\tau_k}^{t} N_1 |\varepsilon^{-1} \varphi^\varepsilon(s) - \Lambda^k(s)| \, ds \\
+ \sum_{j=1}^{m} \int_{\tau_k}^{t} N_1 |\varepsilon^{-1} \theta^{\varepsilon, j}(s) - \Lambda^k(s - \alpha_j)| \, ds \\
+ \sum_{l=k+1}^{\zeta} \int_{\tau_{l-1}}^{\tau_l} \left( \int_0^1 |\varepsilon|^{-1} |\Delta_1^{l,\varepsilon}(s, \eta)| \, d\eta + \int_0^1 |\varepsilon|^{-1} |\Delta_2^{l,\varepsilon}(s, \eta)| \, d\eta \right) \, ds,
\]

where \( N_1 \) is the upper bound for \( |\partial \tilde{f}^l,\varepsilon / \partial x| \) and \( |\partial \tilde{f}^l,\varepsilon / \partial \tilde{x}^j| \) defined in the proof of Theorem 3.2. By Lemma 3.4, (10) and (53), we see that, for all \( \varepsilon \in \Theta \) satisfying (54),

\[
|\varepsilon^{-1} \varphi^\varepsilon(t) - \Lambda^k(t)| \leq 2L_5|\varepsilon| + (m + 1)\delta L_3 T + \int_{\tau_k}^{t} N_1 |\varepsilon^{-1} \varphi^\varepsilon(s) - \Lambda^k(s)| \, ds \\
+ \sum_{j=1}^{m} \int_{\tau_k}^{t} N_1 |\varepsilon^{-1} \varphi^\varepsilon(s - \alpha_j) - \Lambda^k(s - \alpha_j)| \, ds.
\]

The last integral term in (58) can be simplified as...
\[
\sum_{j=1}^{m} \int_{\tau_k}^{t} N_1 |\epsilon^{-1} \varphi^\epsilon(s - \alpha_j) - \Lambda^k(s - \alpha_j)| \, ds = \sum_{j=1}^{m} \int_{t - \alpha_j}^{t} N_1 |\epsilon^{-1} \varphi^\epsilon(s) - \Lambda^k(s)| \, ds \\
\leq \int_{-\alpha_{\text{max}}}^{t} m N_1 |\epsilon^{-1} \varphi^\epsilon(s) - \Lambda^k(s)| \, ds.
\]  
(59)

Since \(\epsilon \in \Theta \cap ] - \infty, 0]\), we have \(-\alpha_{\text{max}} < \tau_k + \epsilon < \tau_k\), and thus
\[
\int_{-\alpha_{\text{max}}}^{t} m N_1 |\epsilon^{-1} \varphi^\epsilon(s) - \Lambda^k(s)| \, ds = \int_{-\alpha_{\text{max}}}^{\tau_k + \epsilon} m N_1 |\epsilon^{-1} \varphi^\epsilon(s) - \Lambda^k(s)| \, ds \\
+ \int_{\tau_k}^{\tau_k + \epsilon} m N_1 |\epsilon^{-1} \varphi^\epsilon(s) - \Lambda^k(s)| \, ds \\
+ \int_{\tau_k}^{t} m N_1 |\epsilon^{-1} \varphi^\epsilon(s) - \Lambda^k(s)| \, ds.
\]  
(60)

Furthermore, using (13), we have
\[
|\epsilon^{-1} \varphi^\epsilon(s) - \Lambda^k(s)| \leq |\epsilon|^{-1} |\varphi^\epsilon(s)| + |\Lambda^k(s)| \leq L_3 + N_2, \quad s \in [\tau_k + \epsilon, \tau_k],
\]  
(61)

where \(N_2\) is as defined in the proof of Theorem 3.2. Substituting (61) into (60) and then using (11) gives
\[
\int_{-\alpha_{\text{max}}}^{t} m N_1 |\epsilon^{-1} \varphi^\epsilon(s) - \Lambda^k(s)| \, ds \leq (m L_3 N_1 + m N_2 N_1)|\epsilon| \\
+ \int_{\tau_k}^{t} m N_1 |\epsilon^{-1} \varphi^\epsilon(s) - \Lambda^k(s)| \, ds.
\]  
(62)

Combining (58), (59), and (62), we obtain
\[
|\epsilon^{-1} \varphi^\epsilon(t) - \Lambda^k(t)| \leq (2L_5 + m L_3 N_1 + m N_2 N_1)|\epsilon| + (m + 1)L_3 T \delta \\
+ \int_{\tau_k}^{t} (m + 1) N_1 |\epsilon^{-1} \varphi^\epsilon(s) - \Lambda^k(s)| \, ds.
\]

Recall from (54) that \(|\epsilon| < \delta\). Thus, by the Gronwall–Bellman Lemma [32],
\[
|\epsilon^{-1} \varphi^\epsilon(t) - \Lambda^k(t)| \leq (2L_5 + m L_3 N_1 + m N_2 N_1 + (m + 1)L_3 T) \delta \exp((m + 1)N_1 T),
\]
which holds whenever $\epsilon \in \Theta$ satisfies (54). Since $\delta > 0$ was selected arbitrarily, this shows that (52) holds when $t \geq \tau_k$. □

3.4 State Variation with Respect to the Switching Times

Combining Theorems 3.2 and 3.3 yields the following fundamental result, which gives the full state variation with respect to each switching time.

**Theorem 3.4** Let $k \in \{1, 2, \ldots, p-1\}$ and $(\tau, \zeta) \in \mathcal{S} \times \mathcal{Z}$. Then for all time points $t \neq \tau_k$,

$$\frac{\partial x(t|\tau, \zeta)}{\partial \tau_k} := \lim_{\epsilon \to 0} \frac{x(t|\tau + \epsilon e^k, \zeta) - x(t|\tau, \zeta)}{\epsilon} = \lim_{\epsilon \to 0} \epsilon^{-1} \varphi^k(t) = \Lambda^k(t|\tau, \zeta).$$

From Theorems 3.2 and 3.3, we see that the left and right state variations differ at $t = \tau_k$, and thus the full state variation does not exist at this point. This is why $t = \tau_k$ is excluded from Theorem 3.4. Thus, the state variation with respect to a given switching time exists at all time points except for the switching time in question. This is different to the state variations with respect to the system parameters, which exist at every point in the time horizon (see Theorem 3.1).

4 Gradient Formulae

By using Theorems 3.1 and 3.4, it is possible to express the gradient of the cost function in Problem (P) in terms of the solutions of the auxiliary switched systems as follows:

$$\frac{\partial J(\tau, \zeta)}{\partial \zeta_q} = \frac{\partial \Phi(x(T|\tau, \zeta), \zeta)}{\partial \zeta_q} + \frac{\partial \Phi(x(T|\tau, \zeta), \zeta)}{\partial x} \Gamma^q(T|\tau, \zeta), \quad q = 1, 2, \ldots, v,$$

$$\frac{\partial J(\tau, \zeta)}{\partial \tau_k} = \frac{\partial \Phi(x(T|\tau, \zeta), \zeta)}{\partial x} \Lambda^k(T|\tau, \zeta), \quad k = 1, 2, \ldots, p - 1.$$

However, to compute these gradient formulae, a large number of auxiliary systems need to be solved (there is one auxiliary system for each system parameter and each switching time). To overcome this challenge, we now derive new gradient formulae using the so-called costate method, which is a commonly used technique in the optimal control domain [32].

Define

$$\frac{\partial f^i(t|\tau, \zeta)}{\partial x} := \frac{\partial f^i(t, x(t|\tau, \zeta), \tilde{x}(t|\tau, \zeta), \zeta)}{\partial x}, \quad t \in [0, T], \quad i = 1, 2, \ldots, p,$$

$$\frac{\partial f^i(t|\tau, \zeta)}{\partial \tilde{x}^j} := \frac{\partial f^i(t, x(t|\tau, \zeta), \tilde{x}(t|\tau, \zeta), \zeta)}{\partial \tilde{x}^j}, \quad t \in [0, T], \quad i = 1, 2, \ldots, p, \quad j = 1, 2, \ldots, m,$$
and
\[
\frac{\partial \tilde{f}_i(t|\tau, \zeta)}{\partial \zeta_q} := \frac{\partial f_i(t, x(t|\tau, \zeta), \tilde{x}(t|\tau, \zeta), \zeta)}{\partial \zeta_q}, \quad t \in [0, T],
\]
i = 1, 2, \ldots, p, \quad q = 1, 2, \ldots, v.

For each \((\tau, \zeta) \in T \times Z\), consider the following costate system:
\[
\dot{\lambda}(t) = -p \sum_{i=1}^{p} \left( \frac{\partial \tilde{f}_i(t|\tau, \zeta)}{\partial x} \right)^\top \lambda(t) \chi_{[\tau_{i-1}, \tau_i]}(t)
\]
\[-p \sum_{i=1}^{p} \sum_{j=1}^{m} \left( \frac{\partial \tilde{f}_i(t + \alpha_j|\tau, \zeta)}{\partial \tilde{x}_j} \right)^\top \lambda(t + \alpha_j) \chi_{[\tau_{i-1} - \alpha_j, \tau_i - \alpha_j]}(t), \quad t \in [0, T],
\]
(63)

with the terminal conditions
\[
\lambda(t) = 0, \quad t > T, \quad \lambda(T) = \left( \frac{\partial \Phi(x(T|\tau, \zeta), \zeta)}{\partial x} \right)^\top,
\]
(64)

where, for a given interval \(I\), \(\chi_I\) denotes the indicator function of \(I\) defined by
\[
\chi_I(t) := \begin{cases} 
1, & \text{if } t \in I, \\
0, & \text{otherwise.} 
\end{cases}
\]
(65)

Note that, if \(t > T - \alpha_j\), then \(t + \alpha_j > T\), and thus the value of \(\frac{\partial \tilde{f}_i(t + \alpha_j|\tau, \zeta)}{\partial \tilde{x}_j}\) is undefined. However, this value has no effect on the costate dynamics (63) because \(\lambda(t + \alpha_j) = 0\) when \(t > T - \alpha_j\) (see (64)).

Let \(\lambda(\cdot|\tau, \zeta)\) denote the solution of system (63) and (64) corresponding to the given pair \((\tau, \zeta) \in T \times Z\). We now express the gradients of the cost function with respect to the system parameters in terms of \(\lambda(\cdot|\tau, \zeta)\).

**Theorem 4.1** Let \(q \in \{1, 2, \ldots, v\}\) and \((\tau, \zeta) \in T \times Z\). Then,
\[
\frac{\partial J(\tau, \zeta)}{\partial \zeta_q} = \frac{\partial \Phi(x(T), \zeta)}{\partial \zeta_q} + \sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_i} \lambda(s) \left( \frac{\partial \tilde{f}_i(s|\tau, \zeta)}{\partial \zeta_q} + \left( \frac{\partial \Phi(s|\tau, \zeta)}{\partial x} \right)^\top \lambda(s) \chi_{[s-\infty, 0]}(s) \right) ds,
\]
\[
+ \sum_{i=1}^{p} \sum_{j=1}^{m} \int_{\tau_{i-1} - \alpha_j}^{\tau_i - \alpha_j} \lambda(s + \alpha_j) \left( \frac{\partial \tilde{f}_i(s + \alpha_j|\tau, \zeta)}{\partial \tilde{x}_j} \right)^\top \lambda(s + \alpha_j) \chi_{[s-\infty, 0]}(s) ds,
\]
where \(x(\cdot) = x(\cdot|\tau, \zeta)\), \(\lambda(\cdot) = \lambda(\cdot|\tau, \zeta)\) and \(\chi_{[s-\infty, 0]}\) is as defined in (65).
Proof Let \( w : [0, \infty[ \to \mathbb{R}^n \) be an arbitrary function that is continuous and differentiable almost everywhere. Then, we may express the cost function \( J \) as follows:

\[
J(\tau, \zeta) = \Phi(\dot{x}(T), \zeta) + \sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_{i}} \Big( w(s)^T \dot{f}^i(s, x(s), \tilde{x}(s), \zeta) - w(s)^T \dot{x}(s) \Big) ds
\]

where we have omitted the arguments \( \tau \) and \( \zeta \) in \( x(\cdot|\tau, \zeta) \) for simplicity.

Applying integration by parts to the last integral term gives

\[
J(\tau, \zeta) = \Phi(\dot{x}(T), \zeta) + \sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_{i}} w(s)^T \dot{f}^i(s, x(s), \tilde{x}(s), \zeta) ds - \int_{0}^{T} w(s)^T \dot{x}(s) ds,
\]

(66)

Differentiating (66) with respect to \( \zeta \) gives

\[
\frac{\partial J(\tau, \zeta)}{\partial \zeta} = \left( \frac{\partial \Phi(\dot{x}(T), \zeta)}{\partial x} - w(T)^T \right) \frac{\partial x(T)}{\partial \zeta} + \frac{\partial \Phi(\dot{x}(T), \zeta)}{\partial \zeta} + w(0^+)^T \frac{\partial \phi(0, \zeta)}{\partial \zeta} + \sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_{i}} \left( w(s)^T \frac{\partial \tilde{f}^i(s)}{\partial x} + \dot{w}(s)^T \right) \left( \frac{\partial x(s)}{\partial \zeta} \right) ds
\]

(67)

where we have omitted the arguments \( \tau \) and \( \zeta \) in \( \frac{\partial \tilde{f}^i(s)}{\partial x} \) and \( \frac{\partial \tilde{f}^i(s)}{\partial \tilde{x}} \). Performing a change of variable in the last term on the right-hand side of (67) yields

\[
\frac{\partial J(\tau, \zeta)}{\partial \zeta} = \sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_{i}} \sum_{j=1}^{m} w(s)^T \left( \frac{\partial \tilde{f}^i(s)}{\partial \tilde{x}^j} \frac{\partial x(s - \alpha_j)}{\partial \zeta} \right) ds
\]

(68)
Since \( x(s) = \phi(s, \zeta) \) for all \( s \leq 0 \), (68) can be rewritten as follows:

\[
\sum_{i=1}^{p} \sum_{j=1}^{m} \int_{\tau_{i-1}}^{\tau_{i}} w(s) \, \partial f_i(s) \, \partial x(s - \alpha_j) \, dx \\
= \sum_{i=1}^{p} \sum_{j=1}^{m} \int_{\tau_{i-1} - \alpha_j}^{\tau_{i} - \alpha_j} w(s + \alpha_j) \, \partial f_i(s + \alpha_j) \, \partial x(s) \, dx \\
+ \sum_{i=1}^{p} \sum_{j=1}^{m} \int_{\tau_{i-1} - \alpha_j}^{\tau_{i} - \alpha_j} w(s + \alpha_j) \, \partial f_i(s + \alpha_j) \, \partial \phi(s, \zeta) \, \partial x \, dx
\]

where \( \chi_{[0, +\infty]} \) and \( \chi_{[-\infty, 0]} \) are as defined in (65). Substituting (69) into (67) yields

\[
\frac{\partial J(\tau, \zeta)}{\partial \zeta_q} = \left( \frac{\partial \Phi(x(T), \zeta)}{\partial x} - w(T)^\top \right) \frac{\partial x(T)}{\partial \zeta_q} + \left( \frac{\partial \Phi(x(T), \zeta)}{\partial \zeta_q} + w(0^+)^\top \right) \frac{\partial \phi(0, \zeta)}{\partial \zeta_q} \\
+ \sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_i} w(s) \, \partial f_i(s) \, \partial \zeta_q \, dx + \sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_i} \left( \frac{\partial f_i(s)}{\partial x} + w(s) \right) \partial x(s) \, dx \\
+ \sum_{i=1}^{p} \sum_{j=1}^{m} \int_{\tau_{i-1} - \alpha_j}^{\tau_{i} - \alpha_j} w(s + \alpha_j) \, \partial f_i(s + \alpha_j) \, \partial \phi(s, \zeta) \, \partial x \, dx
\]

This equation can be rewritten as follows:

\[
\frac{\partial J(\tau, \zeta)}{\partial \zeta_q} = \left( \frac{\partial \Phi(x(T), \zeta)}{\partial x} - w(T)^\top \right) \frac{\partial x(T)}{\partial \zeta_q} + \left( \frac{\partial \Phi(x(T), \zeta)}{\partial \zeta_q} + w(0^+)^\top \right) \frac{\partial \phi(0, \zeta)}{\partial \zeta_q} \\
+ \sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_i} w(s) \, \partial f_i(s) \, \partial \zeta_q \, dx + \int_{0}^{T} \left( \frac{\partial f_i(s)}{\partial x} + \sum_{i=1}^{p} w(s) \right) \partial x(s) \, ds \\
+ \sum_{i=1}^{p} \sum_{j=1}^{m} w(s + \alpha_j) \, \partial f_i(s + \alpha_j) \, \partial \phi(s, \zeta) \, \partial x(s) \, ds \\
+ \sum_{i=1}^{p} \sum_{j=1}^{m} \int_{\tau_{i-1} - \alpha_j}^{\tau_{i} - \alpha_j} w(s + \alpha_j) \, \partial f_i(s + \alpha_j) \, \partial \phi(s, \zeta) \, \partial x \, dx
\]
Choosing $w(\cdot) = \lambda(\cdot|\tau, \zeta)$ and substituting (63) and (64) into the above equation completes the proof. \qed

The following theorem gives the gradients of the cost function with respect to the switching times.

**Theorem 4.2** Let $k \in \{1, 2, \ldots, p - 1\}$ and $(\tau, \zeta) \in T \times Z$. Then,

$$
\frac{\partial J(\tau, \zeta)}{\partial \tau_k} = \lambda(\tau_k) \top f^k(\tau_k, x(\tau_k), \tilde{x}(\tau_k), \zeta) - \lambda(\tau_k) \top f^{k+1}(\tau_k, x(\tau_k), \tilde{x}(\tau_k), \zeta),
$$

where $x(\tau_k) = x(\tau_k|\tau, \zeta)$, $\tilde{x}(\tau_k) = \tilde{x}(\tau_k|\tau, \zeta)$, and $\lambda(\tau_k) = \lambda(\tau_k|\tau, \zeta)$.

**Proof** Let $w(\cdot)$ be as defined in the proof of Theorem 4.1. Recall from (66) that

$$
J(\tau, \zeta) = \Phi(x(T), \zeta) + \sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_{i}} w(s) \top f^i(s, x(s), \tilde{x}(s), \zeta) ds - w(T) \top x(T)
$$

$$
+ w(0^+)\top \phi(0, \zeta) + \int_{0}^{T} \dot{w}(s) \top x(s) ds,
$$

(70)

where, as in the proof of Theorem 4.1, we have omitted the arguments $\tau$ and $\zeta$ for clarity.

Differentiating (70) with respect to $\tau_k$ yields

$$
\frac{\partial J(\tau, \zeta)}{\partial \tau_k} = \left( \frac{\partial \Phi(x(T), \zeta)}{\partial x} - w(T) \top \right) \frac{\partial x(T)}{\partial \tau_k} + \lambda(\tau_k) \top f^k(\tau_k, x(\tau_k), \tilde{x}(\tau_k), \zeta)
$$

$$
- w(\tau_k) \top f^{k+1}(\tau_k, x(\tau_k), \tilde{x}(\tau_k), \zeta) + \sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_{i}} \left( w(s) \top \frac{\partial f^i(s)}{\partial x} + \dot{w}(s) \top \right) \frac{\partial x(s)}{\partial \tau_k} ds
$$

$$
+ \sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_{i}} \sum_{j=1}^{m} w(s) \top \frac{\partial f^i(s)}{\partial \tilde{x}^j} \frac{\partial x(s - \alpha_j)}{\partial \tau_k} ds,
$$

(71)

where, as in the proof of Theorem 4.1, we omit the arguments $\tau$ and $\zeta$ in $\frac{\partial f^i(s|\tau, \zeta)}{\partial x}$ and $\frac{\partial f^i(s|\tau, \zeta)}{\partial \tilde{x}^j}$. Performing a change of variable in the last term on the right-hand side of (71) gives

$$
\sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_{i}} \sum_{j=1}^{m} w(s) \top \frac{\partial f^i(s)}{\partial \tilde{x}^j} \frac{\partial x(s - \alpha_j)}{\partial \tau_k} ds
$$

$$
= \sum_{i=1}^{p} \sum_{j=1}^{m} \int_{\tau_{i-1} - \alpha_j}^{\tau_i - \alpha_j} w(s + \alpha_j) \top \frac{\partial f^i(s + \alpha_j)}{\partial \tilde{x}^j} \frac{\partial x(s)}{\partial \tau_k} ds.
$$

(72)
Recall that \( x(s) = \phi(s, \xi) \) for all \( s \leq 0 \). Then, (72) can be rewritten as follows:

\[
\sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_{i}} \sum_{j=1}^{m} w(s) \top \frac{\partial f^i(s)}{\partial x^j} \frac{\partial x(s - \alpha_j)}{\partial \tau_k} ds
\]

\[
= \sum_{i=1}^{p} \sum_{j=1}^{m} \int_{\tau_{i-1}}^{\tau_{i} - \alpha_j} w(s) \tau_{i} + \alpha_j \frac{\partial f^i(s + \alpha_j)}{\partial x} \frac{\partial x(s)}{\partial \tau_k} \chi_{[0, +\infty]}(s) ds,
\]

(73)

where \( \chi_{[0, +\infty]} \) is as defined in (65). Substituting (73) into (71) gives

\[
\frac{\partial J(\tau, \xi)}{\partial \tau_k} = \left( \frac{\partial \Phi(x(T), \xi)}{\partial x} - w(T) \top \right) \frac{\partial x(T)}{\partial \tau_k} + \frac{\partial \Phi(x(T), \xi)}{\partial \tau_k} f^k(\tau_k, x(\tau_k), \tilde{x}(\tau_k), \xi)
\]

\[
- w(\tau_k) \top f^{k+1}(\tau_k, x(\tau_k), \tilde{x}(\tau_k), \xi) + \sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_{i}} \left( w(s) \top \frac{\partial f^i(s)}{\partial x} + \dot{w}(s) \top \right) \frac{\partial x(s)}{\partial \tau_k} ds
\]

\[
+ \sum_{i=1}^{p} \sum_{j=1}^{m} \int_{\tau_{i-1} - \alpha_j}^{\tau_{i} - \alpha_j} w(s + \alpha_j) \top \frac{\partial f^i(s + \alpha_j)}{\partial \tilde{x}^j} \frac{\partial x(s)}{\partial \tau_k} \chi_{[0, +\infty]}(s) ds.
\]

This equation can be rearranged as follows:

\[
\frac{\partial J(\tau, \xi)}{\partial \tau_k} = \left( \frac{\partial \Phi(x(T), \xi)}{\partial x} - w(T) \top \right) \frac{\partial x(T)}{\partial \tau_k} + \frac{\partial \Phi(x(T), \xi)}{\partial \tau_k} f^k(\tau_k, x(\tau_k), \tilde{x}(\tau_k), \xi)
\]

\[
- w(\tau_k) \top f^{k+1}(\tau_k, x(\tau_k), \tilde{x}(\tau_k), \xi) + \int_{0}^{T} \left( \sum_{i=1}^{p} w(s) \top \frac{\partial f^i(s)}{\partial x} \chi_{[\tau_{i-1}, \tau_i]}(s) \right) \frac{\partial x(s)}{\partial \tau_k} ds
\]

\[
+ \dot{w}(s) \top + \sum_{i=1}^{p} \sum_{j=1}^{m} w(s + \alpha_j) \top \frac{\partial f^i(s + \alpha_j)}{\partial \tilde{x}^j} \chi_{[\tau_{i-1} - \alpha_j, \tau_i - \alpha_j]}(s) \frac{\partial x(s)}{\partial \tau_k} ds.
\]

(74)

Choosing \( w(\cdot) = \lambda(\cdot | \tau, \xi) \) and substituting (63) and (64) into (74) completes the proof.

On the basis of Theorems 4.1 and 4.2, we can compute the cost function \( J(\tau, \xi) \) and its gradient at a given pair \( (\tau, \xi) \in T \times Z \) using the following computational procedure.

Step 1. Solve the switched system (1) from \( t = 0 \) to \( t = T \) to obtain \( x(\cdot | \tau, \xi) \).

Step 2. Using \( x(T | \tau, \xi) \), compute \( J(\tau, \xi) \).

Step 3. Using \( x(\cdot | \tau, \xi) \), solve the costate system (63)–(64) from \( t = T \) to \( t = 0 \) to obtain \( \lambda(\cdot | \tau, \xi) \).
Step 4. Using \( x(\cdot | \tau, \xi) \) and \( \lambda(\cdot | \tau, \xi) \), compute \( \frac{\partial J(\tau, \xi)}{\partial q}, q = 1, 2, \ldots, v \), and \( \frac{\partial J(\tau, \xi)}{\partial \tau_k}, k = 1, 2, \ldots, p - 1 \), via the formulae in Theorems 4.1 and 4.2.

This computational procedure can be integrated with a standard gradient-based optimization method (e.g., sequential quadratic programming [33]) to solve Problem (P) as a nonlinear programming problem. In next section, we use this approach to solve three numerical examples.

5 Numerical Examples

We consider three example problems. To solve these problems, we wrote a Fortran program that combines the gradient computation procedure in Sect. 4 with the optimization software NLPQLP [34]. This program invokes the differential equation software LSODA [35] to solve the state and costate systems. Lagrange interpolation [36] is used when LSODA requires the value of the state or costate at an intermediate time between two adjacent knot points.

Example 5.1 Consider the following switched system as formulated in Reference [27]:

for subsystem 1:
\[
\begin{align*}
\dot{x}_1(t) &= 2x_1(t)x_2(t) + x_2(t - 0.1), \\
\dot{x}_2(t) &= 3x_1(t) + 4x_2(t - 0.1),
\end{align*}
\]

for subsystem 2:
\[
\begin{align*}
\dot{x}_1(t) &= -2x_1(t)x_2(t) + \sin(x_2(t - 0.1)), \\
\dot{x}_2(t) &= x_1(t)x_2(t) + x_1(t - 0.1)x_2(t - 0.1),
\end{align*}
\]

for subsystem 3:
\[
\begin{align*}
\dot{x}_1(t) &= t^2 - 2x_1(t) + 3x_2(t - 0.1), \\
\dot{x}_2(t) &= -x_2(t) + x_1(t - 0.1)x_2(t - 0.1),
\end{align*}
\]

with the initial conditions
\[
x_1(t) = t - 1, \quad x_2(t) = t^2 + 1, \quad t \leq 0.
\]

As in [27], we assume that the switching sequence is 1 → 2 → 3. Let \( \tau_1 \) denote the time at which the system switches from subsystem 1 to subsystem 2, and let \( \tau_2 \) denote the time at which the system switches from subsystem 2 to subsystem 3. The time horizon for this system is \([\tau_0, \tau_3] = [0, 1]\).

We suppose that 0.01 \( \leq \tau_1 \leq 0.2 \) and 0.21 \( \leq \tau_2 \leq 0.9 \). Our goal is to choose the switching times \( \tau_1 \) and \( \tau_2 \) such that the cost function
\[
J(\tau_1, \tau_2) = (x_1(1) - 0.5)^2 + (x_2(1) - 0.25)^2
\]

is minimized subject to the constraints \( \tau_1 \in [0.01, 0.2] \) and \( \tau_2 \in [0.21, 0.9] \).

Starting from the initial guesses \( \tau_1 = 0.2 \) and \( \tau_2 = 0.9 \), our program computes the optimal switching times \( \tau_1 = 0.01 \) and \( \tau_2 = 0.4739 \) after seven optimization iterations. The optimal cost obtained by our program is \( J = 4.537 \times 10^{-3} \), which is
Example 5.2 Consider the following nonlinear switched system with two time-delays:

subsystem 1: \[
\begin{align*}
\dot{x}_1(t) &= -5x_1(t) - 4x_2(t) - 3x_1(t - 0.5) + 2x_2(t - 0.1) \\
+u_1(t) + 0.1 \tanh(x_1(t)), \\
\dot{x}_2(t) &= 0.1x_1(t) - 7x_2(t) + 0.5u_1(t) - \sin(x_2(t - 0.1)),
\end{align*}
\]

subsystem 2: \[
\begin{align*}
\dot{x}_1(t) &= -4x_1(t) + 0.5x_2(t) + 0.2 \sin(x_2(t)) + t^2 + 8, \\
\dot{x}_2(t) &= 5x_1(t) - 5x_2(t) + 0.5 \sin(x_1(t - 0.1)) - u_2(t),
\end{align*}
\]

with the initial conditions
\[
x_1(t) = 6, \quad x_2(t) = t^2 + 2, \quad t \leq 0.
\]

The time horizon here is [0, 1.5]. We assume that the system switches once during the time horizon from subsystem 1 to subsystem 2. Let \(\tau_1\) denote the switching point. The functions \(u_1\) and \(u_2\) are feedback controllers of the form \(u_1(t) = \zeta_1 x_1(t) + \zeta_2 x_2(t)\) and \(u_2(t) = \zeta_3 x_1(t) + \zeta_4 x_2(t)\), where \(0.1 \leq \zeta_1 \leq 1.5\) and \(0.1 \leq \zeta_2, \zeta_3, \zeta_4 \leq 1.0\). Our goal is to find an optimal switching time \(\tau_1\) and an optimal feedback gain parameter vector \(\zeta = (\zeta_1, \zeta_2, \zeta_3, \zeta_4)^T\) such that the cost function
\[
J(\tau_1, \zeta) = (x_1(1.5) - 2)^2 + (x_2(1.5) - 1)^2
\]
is minimized subject to the constraints \(\zeta_1 \in [0.1, 1.5]\) and \(\zeta_2, \zeta_3, \zeta_4 \in [0.1, 1.0]\).

The gradient of the cost function (83) with respect to \(\zeta\) is calculated using the formulae in Sect. 4, with the extended Simpson’s rule [36] used to compute the integrals. Starting from the initial guesses \(\tau_1 = 0.1\) and \(\zeta = (0.7, 0.5, 0.5, 0.8)^T\), our program computes the optimal switching time as \(\tau_1 = 1.3304\) and the optimal parameter vector as \(\zeta = (1.3149, 0.5635, 0.3256, 0.6672)^T\) after 22 optimization iterations. Moreover, the corresponding optimal cost is \(3.9943 \times 10^{-30}\), which can be regarded as zero in
numerical computation. The optimal feedback control and the optimal state trajectory are shown in Fig. 2.

**Example 5.3** Our final example involves the fed-batch fermentation process for converting glycerol to 1,3-propanediol (1,3-PD) using the microorganism *Klebsiella pneumoniae*. This process oscillates between two modes: batch mode and feeding mode. In batch mode, no substrate is added to the fermentor. In feeding mode, substrates are added with constant feeding rates to provide nutrition and maintain a suitable environment for cell growth. Since nutrient metabolization does not immediately lead to the production of new biomass [37], time-delays exist in the fermentation process.

Based on our previous work [31], the dynamic model for batch mode is given by

\[
\begin{align*}
\dot{x}_1(t) &= \mu(x_2(t))x_1(t - 0.16), \\
\dot{x}_2(t) &= -q_2(x_2(t))x_1(t - 0.16), \\
\dot{x}_3(t) &= q_3(x_2(t))x_1(t - 0.16), \\
\dot{x}_4(t) &= 0,
\end{align*}
\]

(84)

where \(x_1(t), x_2(t), x_3(t), \) and \(x_4(t)\) are, respectively, the extracellular concentrations of biomass, glycerol, 1,3-PD, and the volume of culture fluid at time \(t\) in the fermentor; \(\mu(x_2(t))\) is the specific growth rate of cells; \(q_2(x_2(t))\) is the specific consumption rate of substrate; and \(q_3(x_2(t))\) is the specific formation rate of 1,3-PD. The functions \(\mu(x_2(t)), q_2(x_2(t)), \) and \(q_3(x_2(t))\) are given by

\[
\begin{align*}
\mu(x_2(t)) &= \frac{\Delta_1 x_2(t)}{x_2(t) + k_1}, \\
q_2(x_2(t)) &= m_2 + Y_2 \mu(x_2(t)) + \frac{\Delta_2 x_2(t)}{x_2(t) + k_2}, \\
q_3(x_2(t)) &= -m_3 + Y_3 \mu(x_2(t)) + \frac{\Delta_3 x_2(t)}{x_2(t) + k_3},
\end{align*}
\]

where \(\Delta_1, k_1, m_2, Y_2, \Delta_2, k_2, m_3, Y_3, \Delta_3, \) and \(k_3\) are model parameters whose values are given in Table 1.
The dynamic model for feeding mode is given by

\[
\begin{align*}
\dot{x}_1(t) &= \mu(x_2(t))x_1(t - 0.16) - D(x_4(t), \zeta)x_1(t), \\
\dot{x}_2(t) &= D(x_4(t), \zeta)((1 + r)^{-1}c_{s0} - x_2(t)) - q_2(x_2(t))x_1(t - 0.16), \\
\dot{x}_3(t) &= q_3(x_2(t))x_1(t - 0.16) - D(x_4(t), \zeta)x_3(t), \\
\dot{x}_4(t) &= (1 + r)\zeta,
\end{align*}
\]

where \( r = 0.75 \) is the velocity ratio of adding alkali to glycerol; \( c_{s0} = 10762 \text{ mmol L}^{-1} \) is the concentration of the initial feed of glycerol; \( \zeta \) is the feeding rate of the glycerol; and \( D(x_4(t), \zeta) \) is the dilution rate at time \( t \) defined as

\[
D(x_4(t), \zeta) = \frac{(1 + r)\zeta}{x_4(t)}. \tag{86}
\]

The fed-batch fermentation process switches between the batch and feeding modes during the time horizon.

The initial function \( \phi \) for the system (84) and (85) is obtained by applying cubic spline interpolation to the experimental data before the zero time point. The total number of switching times is 1,355 and the terminal time is \( T = 24.16 \text{ h} \). The entire fermentation process consists of the first batch mode followed by nine phases, each of which consists of multiple fed-batch pairs (feeding mode followed by batch mode) of 100 s duration. In each phase, the durations of the feeding modes are the same. As a result, the end moment of the first batch mode (denoted by \( \tau_1 \)) and the end

### Table 1

The model parameters in Example 5.3 [31]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>( \Delta_1 )</th>
<th>( k_1 )</th>
<th>( m_2 )</th>
<th>( Y_2 )</th>
<th>( \Delta_2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>0.8037</td>
<td>0.4856</td>
<td>0.2977</td>
<td>144.912</td>
<td>7.8367</td>
</tr>
<tr>
<td>Parameter</td>
<td>( k_2 )</td>
<td>( m_3 )</td>
<td>( Y_3 )</td>
<td>( \Delta_3 )</td>
<td>( k_3 )</td>
</tr>
<tr>
<td>Value</td>
<td>9.4632</td>
<td>12.2557</td>
<td>80.8439</td>
<td>20.2757</td>
<td>38.75</td>
</tr>
</tbody>
</table>

### Table 2

Optimal values, initial guesses, and lower and upper bounds for the switching times in Example 5.3

<table>
<thead>
<tr>
<th>Switching time</th>
<th>( \tau_1 )</th>
<th>( \tau_2 )</th>
<th>( \tau_3 )</th>
<th>( \tau_4 )</th>
<th>( \tau_5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal value</td>
<td>5.4096</td>
<td>5.4112</td>
<td>6.1894</td>
<td>7.2174</td>
<td>8.9117</td>
</tr>
<tr>
<td>Initial guess</td>
<td>5.3300</td>
<td>5.3314</td>
<td>6.1097</td>
<td>7.1378</td>
<td>8.8319</td>
</tr>
<tr>
<td>Lower bound</td>
<td>5.3000</td>
<td>5.3014</td>
<td>6.0797</td>
<td>7.1078</td>
<td>8.8019</td>
</tr>
<tr>
<td>Upper bound</td>
<td>5.5000</td>
<td>5.5022</td>
<td>6.2800</td>
<td>7.3078</td>
<td>9.0022</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Switching time</th>
<th>( \tau_6 )</th>
<th>( \tau_7 )</th>
<th>( \tau_8 )</th>
<th>( \tau_9 )</th>
<th>( \tau_{10} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Optimal value</td>
<td>12.2172</td>
<td>15.9110</td>
<td>18.1602</td>
<td>19.9101</td>
<td>23.9101</td>
</tr>
<tr>
<td>Initial guess</td>
<td>12.1372</td>
<td>15.8311</td>
<td>18.0808</td>
<td>19.8306</td>
<td>23.8303</td>
</tr>
<tr>
<td>Lower bound</td>
<td>12.1072</td>
<td>15.8011</td>
<td>18.0503</td>
<td>19.8003</td>
<td>23.8003</td>
</tr>
</tbody>
</table>
moment of the first feeding mode in each phase (denoted by $\tau_i$, $i = 2, \ldots, 10$) are to be optimized. Furthermore, the feeding rate of glycerol $\zeta$ also needs to be optimally chosen. Let $\tau = (\tau_1, \tau_2, \ldots, \tau_{10})^T$. The aim of the fermentation optimization is to obtain high concentration of 1,3-PD at the terminal time. Thus, the cost function in this example is

$$J(\tau, \zeta) = -x_3(T|\tau, \zeta). \quad (87)$$

The optimization problem is: choose the switching time vector $\tau$ and the system parameter $\zeta$ such that the cost function (87) is minimized subject to $\zeta \in [0.7954, 1.1009]$ and the switching time bounds given in Table 2.

The gradient of the cost function (87) with respect to $\zeta$ is calculated using the formulae in Sect. 4, with the extended Simpson’s rule [36] used to compute the integrals. By choosing the initial guess $\zeta = 8.1211 \text{ L h}^{-1}$ for the feeding rate, as well as the initial guesses for the switching times in Table 2, our program obtains the optimal feeding rate $\zeta^* = 0.8867 \text{ L h}^{-1}$ and the optimal switching times listed in Table 2. Our program took 15 optimization iterations to obtain this optimal solution. The corresponding optimal durations of the feeding modes in the nine phases are, respectively, 5.75, 7.28, 8.00, 7.76, 7.51, 5.13, 2.22, 1.98, and 1.73 s. Moreover, the optimal 1,3-PD concentration at the terminal time is 932.2 mmol L$^{-1}$. This is a 16.96 % improvement over the experimental results in [38], which are obtained using a nonoptimal switching control scheme. The optimal trajectories for the biomass and 1,3-PD are shown in Fig. 3. The curves in Fig. 3 also confirm that 1,3-PD concentration at the terminal time is increased compared with the results in [38].

### 6 Conclusions

In this paper, we investigated the optimization of switched systems with multiple time-delays and multiple system parameters. We first established the existence of the partial derivatives of the system state with respect to the system parameters and switching times. We then used this result to derive the gradient of the cost function in
terms of the solutions of the state system and an auxiliary system called the costate system. On this basis, a new gradient-based optimization algorithm was developed to solve the optimization problem. The effectiveness and applicability of this algorithm were verified using three numerical examples. Based on the work in this paper, we have identified several further research topics worthy of investigation. For example, more complex switched systems would involve time-varying input signals (recall that the input parameter vector $\xi$ is time-invariant). Moreover, the results in this paper could potentially prove useful in the optimization of systems in which the delays are functions of the state variables.

Acknowledgments The first author is supported by the Natural Science Foundation for the Youth of China (Grant Nos. 11201267 and 11001153), the Tian Yuan Special Funds of the National Natural Science Foundation of China (Grant No. 11126077), and the Shandong Province Natural Science Foundation of China (Grant Nos. ZR2010AQ016 and ZR2011AL003). The second author is supported by the National Science Foundation of China (Grant No. 11350110208) and the Australian Research Council (Discovery Grant DP110100083). The third author is supported by the Australian Research Council (Discovery Grant DP110100083).

References