Optimal Control of Nonlinear Switched Systems with Multiple Time-delays

Chongyang Liu · Ryan Loxton · Kok Lay Teo

Abstract In this paper, we consider an optimal control problem involving a general switched system that evolves by switching between several subsystems of nonlinear delay-differential equations. The control variables in this system consist of: (i) the times at which the subsystem switches occur; and (ii) 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 control strategy. 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.

C. Liu
School of Mathematics and Information Science, Shandong Institute of Business and Technology, Yantai, China
E-mail: liu_chongyang@yahoo.com
R. Loxton
Department of Mathematics and Statistics, Curtin University, Perth, Australia
E-mail: r.loxton@curtin.edu.au
K.L. Teo
Department of Mathematics and Statistics, Curtin University, Perth, Australia
E-mail: k.l.teo@curtin.edu.au

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 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 optimal control of switched systems is an important and challenging research topic for control theorists [9–14]. In the most general case, determining an optimal control 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 [21]. Indeed, switched systems with time-delays have various applications in areas such as power systems [22] and network control systems [23]. The presence of delays in a switched system complicates the search for an optimal control policy. In particular, the time-scaling transformation mentioned above, a powerful tool for solving switched system optimal control problems [7,13,18,19], is not applicable to switched systems with time-delays [20]. In fact, optimal control techniques for switched sys-

tems 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 separable systems with a single delay. Reference [27] presents an effective optimal control algorithm for switched systems with time-delays that 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: (i) it is only applicable to switched systems with a single delay and no system parameters; (ii) it involves integrating a large number of auxiliary differential equations (one auxiliary system for each subsystem); and (iii) 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 control variables to be selected optimally. The optimal control problem involves choosing these control 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:

$$\dot{x}(t) = f^i(t, x(t), \tilde{x}(t), \zeta), \quad t \in (\tau_{i-1}, \tau_i], \quad i = 1, 2, \ldots, p, \tag{1a}$$

$$x(t) = \phi(t, \zeta), \quad t \leq 0, \tag{1b}$$

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*; $\tilde{x}(t) = (x(t - \alpha_1)^\top, x(t - \alpha_2)^\top, \ldots, x(t - \alpha_m)^\top)^\top \in \mathbb{R}^{nm}$ is the *delayed 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 ζ_1 and $x(t - \alpha_1)$, subsystem 2 may only involve ζ_2 and $x(t - \alpha_2)$, and so on.

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

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

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

Define

$$\mathcal{Z} := \{(\zeta_1, \zeta_2, \ldots, \zeta_v)^\top \in \mathbb{R}^v : a_q \leq \zeta_q \leq b_q, \ q = 1, 2, \ldots, v\}, \tag{3}$$

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

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

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

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

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

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

Assumptions 1-2 ensure that the switched system (1) has a unique solution $x(\cdot|\tau, \zeta)$ corresponding to each pair $(\tau, \zeta) \in \mathcal{T} \times \mathcal{Z}$ [31]. This solution is called the *state trajectory*.

We suppose that the cost function 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:

$$J(\tau, \zeta) := \Phi(x(T|\tau, \zeta), \zeta), \tag{4}$$

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:

$$\sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_i} \mathcal{L}^i(t, x(t), \tilde{x}(t), \zeta) dt,$$

where $\mathcal{L}^i : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^{nm} \times \mathbb{R}^v \to \mathbb{R}, i = 1, 2, \ldots, p$, are given continuously differentiable functions. It is clear that this term can be replaced by $x_{n+1}(T)$, where x_{n+1} satisfies the dynamics

$$\dot{x}_{n+1}(t) = \mathcal{L}^i(t, x(t), \tilde{x}(t), \zeta), \quad t \in (\tau_{i-1}, \tau_i], \quad i = 1, 2, \ldots, p,$$

$$x_{n+1}(t) = 0, \quad t \leq 0.$$

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 optimal control 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 optimal control problem involving a nonlinear switched system with multiple time-delays, where the switching times and system parameters are control variables to be optimized. It is well known that variable switching times pose a significant challenge for conventional numerical optimization techniques [18,28,29]. 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 [20], 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 τ and ζ. 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 $\tau = (\tau_1, \tau_2, \ldots, \tau_{p-1})^\top$ such that

$$0 = \tau_0 < \tau_1 < \cdots < \tau_{p-1} < \tau_p = T.$$

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

then be exploited in Section 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 ζ is called the *state variation* with respect to ζ. The following result, which can be proved in a similar manner as that given for the proof of Theorem 2 in [37], 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 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 $\Gamma^q(\cdot|\tau,\zeta)$ is the solution of the following auxiliary switched system:

$$\dot{\Gamma}^q(s) = \frac{\partial f^i(s, x(s), \tilde{x}(s), \zeta)}{\partial x}\Gamma^q(s) + \sum_{j=1}^{m}\frac{\partial f^i(s, x(s), \tilde{x}(s), \zeta)}{\partial\tilde{x}^j}\Gamma^q(s - \alpha_j)$$

$$+ \frac{\partial f^i(s, x(s), \tilde{x}(s), \zeta)}{\partial\zeta_q}, \quad s \in (\tau_{i-1}, \tau_i], \quad i = 1, 2, \ldots, p, \tag{6}$$

with the initial condition

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

Theorem 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 computational optimal control 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 $(\tau, \zeta) \in \mathcal{S} \times \mathcal{Z}$ be arbitrary but fixed. For notational simplicity, we write $x(\cdot)$ instead of $x(\cdot|\tau,\zeta)$ and $x^\epsilon(\cdot)$ instead of $x(\cdot|\tau + \epsilon e^k, \zeta)$, where e^k denotes the kth unit basis vector in $\mathbb{R}^{p-1}$.

Define

$$\Theta := (\tau_{k-1} - \tau_k, \tau_{k+1} - \tau_k).$$

Note that Θ is a non-empty open interval. Clearly,

$$\epsilon \in \Theta \quad \Longleftrightarrow \quad \tau + \epsilon e^k \in \mathcal{S}.$$

Now, for each $\epsilon \in \Theta$, define the following functions:

$$\varphi^\epsilon(t) := x^\epsilon(t) - x(t), \quad t \leq T, \tag{8}$$

$$\theta^{\epsilon,j}(t) := x^\epsilon(t - \alpha_j) - x(t - \alpha_j), \quad t \leq T, \quad j = 1, 2, \ldots, m. \tag{9}$$

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

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

Obviously, for each $\epsilon \in \Theta$,

$$\theta^{\epsilon,j}(t) = \varphi^\epsilon(t - \alpha_j), \quad t \leq T, \quad j = 1, 2, \ldots, m, \tag{10}$$

and

$$\varphi^\epsilon(t) = 0, \quad t \leq \min\{\tau_k + \epsilon, \tau_k\}. \tag{11}$$

The following result can be proved in a similar manner as that given for the proof of Lemma 6.4.2 in [32]. Note that Assumption 2 plays a key role in the proof.

Lemma 1 *There exists a positive real number $L_2 > 0$ such that for all $\epsilon \in \Theta$,*

$$|x^\epsilon(t)| \leq L_2, \quad t \in [-\alpha_{\max}, T], \tag{12}$$

where $\alpha_{\max} = \max_{j \in \{1, 2, \ldots, m\}} \{\alpha_j\}$.

The next lemma gives an important upper bound for the functions defined in (8) and (9).

Lemma 2 *There exists a positive real number $L_3 > 0$ such that for all $\epsilon \in \Theta$,*

$$|\varphi^\epsilon(t)|, \max_{j \in \{1,2,\ldots,m\}} |\theta^{\epsilon,j}(t)| \le L_3 |\epsilon|, \quad t \in [0,T]. \tag{13}$$

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 τ, and let $i^\epsilon(s)$ denote the active subsystem at time s corresponding to the switching time vector $\tau + \epsilon e^k$. Then

$$x(t) = \phi(0, \zeta) + \int_0^t f^{i(s)}(s, x(s), \tilde{x}(s), \zeta) ds \tag{14}$$

and

$$x^\epsilon(t) = \phi(0, \zeta) + \int_0^t f^{i^\epsilon(s)}(s, x^\epsilon(s), \tilde{x}^\epsilon(s), \zeta) ds. \tag{15}$$

Combining (14) and (15) gives

$$\begin{aligned}
\varphi^\epsilon(t) &= x^\epsilon(t) - x(t) \\
&= \int_0^t \left\{ f^{i^\epsilon(s)}(s, x^\epsilon(s), \tilde{x}^\epsilon(s), \zeta) - f^{i(s)}(s, x(s), \tilde{x}(s), \zeta) \right\} ds.
\end{aligned} \tag{16}$$

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

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

Thus, we can rewrite (16) as follows:

$$\begin{aligned}
\varphi^\epsilon(t) &= \int_{I_\epsilon \cap [0,t]} \left\{ f^{i^\epsilon(s)}(s, x^\epsilon(s), \tilde{x}^\epsilon(s), \zeta) - f^{i(s)}(s, x(s), \tilde{x}(s), \zeta) \right\} ds \\
&\quad + \int_{[0,t] \setminus I_\epsilon} \left\{ f^{i(s)}(s, x^\epsilon(s), \tilde{x}^\epsilon(s), \zeta) - f^{i(s)}(s, x(s), \tilde{x}(s), \zeta) \right\} ds.
\end{aligned} \tag{17}$$

Taking the norm of both sides of (17) yields

$$
|\varphi^\epsilon(t)| \leq \int_0^t \big| f^{i(s)}(s, x^\epsilon(s), \tilde{x}^\epsilon(s), \zeta) - f^{i(s)}(s, x(s), \tilde{x}(s), \zeta) \big| ds
$$
$$
+ \int_{I_\epsilon} \big| f^{i^\epsilon(s)}(s, x^\epsilon(s), \tilde{x}^\epsilon(s), \zeta) - f^{i(s)}(s, x(s), \tilde{x}(s), \zeta) \big| ds. \tag{18}
$$

Now, by Assumption 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:

$$
\int_0^t \big| f^{i(s)}(s, x^\epsilon(s), \tilde{x}^\epsilon(s), \zeta) - f^{i(s)}(s, x(s), \tilde{x}(s), \zeta) \big| ds
$$
$$
\leq \int_0^t M_1 |\varphi^\epsilon(s)| ds + \sum_{j=1}^m \int_0^t M_1 |\theta^{\epsilon,j}(s)| ds. \tag{19}
$$

Furthermore, there exists another constant $M_2 > 0$ such that for all $\epsilon \in \Theta$,

$$
\big| f^l(s, x^\epsilon(s), \tilde{x}^\epsilon(s), \zeta) \big| \leq M_2, \quad s \in [0, T], \quad l = 1, 2, \ldots, p.
$$

Thus, the last integral term on the right-hand side of (18) can be simplified as follows:

$$
\int_{I_\epsilon} \big| f^{i^\epsilon(s)}(s, x^\epsilon(s), \tilde{x}^\epsilon(s), \zeta) - f^{i(s)}(s, x(s), \tilde{x}(s), \zeta) \big| ds \leq 2M_2 |\epsilon|. \tag{20}
$$

Substituting (19) and (20) into (18), and then simplifying using (10) and (11), we obtain

$$
|\varphi^\epsilon(t)| \leq 2M_2 |\epsilon| + \int_0^t M_1 |\varphi^\epsilon(s)| ds + \sum_{j=1}^m \int_0^t M_1 |\varphi^\epsilon(s - \alpha_j)| ds
$$
$$
\leq 2M_2 |\epsilon| + (m+1) M_1 \int_0^t |\varphi^\epsilon(s)| ds.
$$

Thus, by the Gronwall-Bellman Lemma [32], it follows that

$$
|\varphi^\epsilon(t)| \leq L_3 |\epsilon|, \tag{21}
$$

where $L_3 = 2M_2 \exp(mM_1 T + M_1 T)$. Therefore, by (10) and (11),

$$
\max_{j \in \{1,2,\ldots,m\}} |\theta^{\epsilon,j}(t)| = \max_{j \in \{1,2,\ldots,m\}} |\varphi^\epsilon(t - \alpha_j)| \leq L_3 |\epsilon|. \tag{22}
$$

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

Now, consider the following auxiliary switched system:

$$\dot{\Lambda}^k(s) = \frac{\partial f^i(s, x(s), \tilde{x}(s), \zeta)}{\partial x} \Lambda^k(s) + \sum_{j=1}^{m} \frac{\partial f^i(s, x(s), \tilde{x}(s), \zeta)}{\partial \tilde{x}^j} \Lambda^k(s - \alpha_j),$$

$$s \in (\tau_{i-1}, \tau_i], \quad i = k+1, \ldots, p, \tag{23}$$

with the jump condition

$$\Lambda^k(\tau_k^+) = f^k(\tau_k, x(\tau_k), \tilde{x}(\tau_k), \zeta) - f^{k+1}(\tau_k, x(\tau_k), \tilde{x}(\tau_k), \zeta) \tag{24}$$

and the initial condition

$$\Lambda^k(s) = 0, \quad s < \tau_k. \tag{25}$$

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 τ_k is equal to Λ^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 τ_k is defined by

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

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

For each $\epsilon \in \Theta \cap [0, \infty)$, define

$$\rho_k^\epsilon := \int_{\tau_k}^{\tau_k + \epsilon} \left\{ f^k(s, x^\epsilon(s), \tilde{x}^\epsilon(s), \zeta) - f^k(\tau_k, x(\tau_k), \tilde{x}(\tau_k), \zeta) \right\} ds \tag{26}$$

and

$$\rho_{k+1}^\epsilon := \int_{\tau_k}^{\tau_k + \epsilon} \left\{ f^{k+1}(s, x^\epsilon(s), \tilde{x}^\epsilon(s), \zeta) - f^{k+1}(\tau_k, x(\tau_k), \tilde{x}(\tau_k), \zeta) \right\} ds. \tag{27}$$

The next lemma gives an upper bound for the functions defined in (26) and (27).

Lemma 3 *There exists a constant $L_4 > 0$ such that for all $\epsilon \in \Theta \cap [0, \infty)$,*

$$|\rho_k^\epsilon|, \ |\rho_{k+1}^\epsilon| \le L_4 \epsilon^2.$$

Proof From the definition of ρ_k^ϵ in (26), we have

$$|\rho_k^\epsilon| \le \int_{\tau_k}^{\tau_k+\epsilon} \big| f^k(s, x^\epsilon(s), \tilde{x}^\epsilon(s), \zeta) - f^k(s, x(s), \tilde{x}(s), \zeta) \big| ds$$
$$+ \int_{\tau_k}^{\tau_k+\epsilon} \big| f^k(s, x(s), \tilde{x}(s), \zeta) - f^k(\tau_k, x(\tau_k), \tilde{x}(\tau_k), \zeta) \big| ds.$$

Thus, since f^k is Lipschitz continuous,

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

where $M_1 > 0$ is the Lipschitz constant defined in the proof of Lemma 2. Applying Lemma 2, we obtain

$$|\rho_k^\epsilon| \le M_1 L_3 \epsilon^2 + m M_1 L_3 \epsilon^2 + M_1 \epsilon^2 + \int_{\tau_k}^{\tau_k+\epsilon} M_1 |x(s) - x(\tau_k)| ds$$
$$+ \sum_{j=1}^{m} \int_{\tau_k}^{\tau_k+\epsilon} M_1 |x(s - \alpha_j) - x(\tau_k - \alpha_j)| ds. \tag{28}$$

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

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

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

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

and

$$|x(s - \alpha_j) - x(\tau_k - \alpha_j)| \le \int_{\tau_k - \alpha_j}^{s - \alpha_j} |\dot{x}(\eta)| d\eta \le M_2 \epsilon. \tag{30}$$

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

$$|\rho_k^\epsilon| \le M_1 L_3 \epsilon^2 + m M_1 L_3 \epsilon^2 + M_1 \epsilon^2 + M_1 M_2 \epsilon^2 + m M_1 M_2 \epsilon^2. \tag{31}$$

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

$$|\rho_{k+1}^\epsilon| \le M_1 L_3 \epsilon^2 + m M_1 L_3 \epsilon^2 + M_1 \epsilon^2 + M_1 M_2 \epsilon^2 + m M_1 M_2 \epsilon^2. \tag{32}$$

Choosing $L_4 = M_1 L_3 + m M_1 L_3 + M_1 + M_1 M_2 + m M_1 M_2$ in (31) and (32) completes the proof.

On the basis of Lemmas 1-3, we now establish the existence of the right state variation with respect to the switching times.

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

$$\lim_{\epsilon \to 0+} \epsilon^{-1} \varphi^\epsilon(t) = \Lambda^k(t|\tau, \zeta), \tag{33}$$

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

$$\lim_{\epsilon \to 0+} \epsilon^{-1} \varphi^\epsilon(\tau_k) = 0. \tag{34}$$

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

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

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

and

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

By Assumption 1 and Lemma 1, there exists a positive constant $N_1 > 0$ such that for all $\epsilon \in \Theta$,

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

and

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

where $|\cdot|$ denotes the natural matrix norm on $\mathbb{R}^{n\times n}$. Moreover, by Lemma 2, the following limits exist uniformly with respect to $\eta \in [0,1]$ and $s \in [0,T]$:

$$\lim_{\epsilon \to 0+} \{x(s) + \eta\varphi^\epsilon(s)\} = x(s),$$

$$\lim_{\epsilon \to 0+} \{\tilde{x}(s) + \eta\theta^\epsilon(s)\} = \tilde{x}(s).$$

Thus, it follows from Assumption 1 that for each $\delta > 0$, there exists a corresponding $\epsilon' > 0$ such that for all $\epsilon \in \Theta$ satisfying $|\epsilon| < \epsilon'$,

$$\left| \frac{\partial \bar{f}^{i,\epsilon}(s,\eta)}{\partial x} - \frac{\partial \bar{f}^{i,\epsilon}(s,0)}{\partial x} \right| < \delta, \ (s,\eta) \in [\tau_{i-1},\tau_i] \times [0,1], \ i = 1,2,\ldots,p, \tag{35}$$

and

$$\left| \frac{\partial \bar{f}^{i,\epsilon}(s,\eta)}{\partial \tilde{x}^j} - \frac{\partial \bar{f}^{i,\epsilon}(s,0)}{\partial \tilde{x}^j} \right| < \delta, \ (s,\eta) \in [\tau_{i-1},\tau_i] \times [0,1], \ i = 1,2,\ldots,p,$$

$$j = 1,2,\ldots,m. \tag{36}$$

Inequalities (35) and (36), together with Lemma 2, imply that for all $\epsilon \in \Theta$ with $|\epsilon| < \epsilon'$,

$$|\Delta_1^{i,\epsilon}(s,\eta)| \le \delta L_3 |\epsilon|, \quad |\Delta_2^{i,\epsilon}(s,\eta)| \le \delta m L_3 |\epsilon|, \quad i = 1,2,\ldots,p. \tag{37}$$

By the chain rule, we obtain

$$\frac{\partial \bar{f}^{i,\epsilon}(s,\eta)}{\partial \eta} = \frac{\partial \bar{f}^{i,\epsilon}(s,\eta)}{\partial x}\varphi^\epsilon(s) + \sum_{j=1}^{m} \frac{\partial \bar{f}^{i,\epsilon}(s,\eta)}{\partial \tilde{x}^j}\theta^{\epsilon,j}(s). \tag{38}$$

We can rewrite (38) as follows:

$$\frac{\partial \bar{f}^{i,\epsilon}(s,\eta)}{\partial \eta} = \Delta_1^{i,\epsilon}(s,\eta) + \Delta_2^{i,\epsilon}(s,\eta) + \frac{\partial \bar{f}^{i,\epsilon}(s,0)}{\partial x}\varphi^\epsilon(s)$$

$$+ \sum_{j=1}^{m} \frac{\partial \bar{f}^{i,\epsilon}(s,0)}{\partial \tilde{x}^j}\theta^{\epsilon,j}(s). \tag{39}$$

Now, let $t \in [0,T]$ be a fixed time point. We consider two cases: (i) $t \le \tau_k$; and (ii) $t > \tau_k$.

Consider case (i). In this case, $t \leq \tau_k$, and thus it follows from (11) that

$$\lim_{\epsilon \to 0+} \epsilon^{-1} \varphi^\epsilon(t) = 0. \tag{40}$$

In particular, (40) holds when $t = \tau_k$, thus proving equation (34). If $t < \tau_k$, then in view of (25),

$$\Lambda^k(t) = 0 = \lim_{\epsilon \to 0+} \epsilon^{-1} \varphi^\epsilon(t).$$

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

Now, consider case (ii). 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 ϵ' corresponds to $\delta = \gamma$ in (35) and (36). Then $\tau_k + \epsilon < \tau_{k+1}$ and $\tau_k + \epsilon < t$. Consequently,

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

Using (26) and (27), we obtain

$$\varphi^\epsilon(t) = \rho_k^\epsilon - \rho_{k+1}^\epsilon + \epsilon \left\{ \bar{f}^{k,\epsilon}(\tau_k, 0) - \bar{f}^{k+1,\epsilon}(\tau_k, 0) \right\}$$
$$+ \sum_{l=k+1}^{\varsigma} \int_{\tau_{l-1}}^{\min\{\tau_l, t\}} \left\{ \bar{f}^{l,\epsilon}(s,1) - \bar{f}^{l,\epsilon}(s,0) \right\} ds. \tag{42}$$

By using the fundamental theorem of calculus, (42) can be written as

$$\varphi^\epsilon(t) = \rho_k^\epsilon - \rho_{k+1}^\epsilon + \epsilon \left\{ \bar{f}^{k,\epsilon}(\tau_k, 0) - \bar{f}^{k+1,\epsilon}(\tau_k, 0) \right\}$$
$$+ \sum_{l=k+1}^{\varsigma} \int_{\tau_{l-1}}^{\min\{\tau_l, t\}} \left(\int_0^1 \frac{\partial \bar{f}^{l,\epsilon}(s,\eta)}{\partial \eta} d\eta \right) ds. \tag{43}$$

Substituting (39) into (43) gives

$$
\begin{aligned}
\varphi^\epsilon(t) = {} & \rho_k^\epsilon - \rho_{k+1}^\epsilon + \epsilon\{\bar{f}^{k,\epsilon}(\tau_k,0) - \bar{f}^{k+1,\epsilon}(\tau_k,0)\} \\
& + \sum_{l=k+1}^{\varsigma} \int_{\tau_{l-1}}^{\min\{\tau_l,t\}} \left(\int_0^1 \Delta_1^{l,\epsilon}(s,\eta)d\eta + \int_0^1 \Delta_2^{l,\epsilon}(s,\eta)d\eta \right. \\
& \left. + \frac{\partial \bar{f}^{l,\epsilon}(s,0)}{\partial x}\varphi^\epsilon(s) + \sum_{j=1}^m \frac{\partial \bar{f}^{l,\epsilon}(s,0)}{\partial \tilde{x}^j}\theta^{\epsilon,j}(s) \right) ds.
\end{aligned} \tag{44}
$$

Next, integrating the auxiliary system (23)-(25) yields

$$
\begin{aligned}
\Lambda^k(t) = {} & \bar{f}^{k,\epsilon}(\tau_k,0) - \bar{f}^{k+1,\epsilon}(\tau_k,0) + \sum_{l=k+1}^{\varsigma} \int_{\tau_{l-1}}^{\min\{\tau_l,t\}} \left(\frac{\partial \bar{f}^{l,\epsilon}(s,0)}{\partial x}\Lambda^k(s) \right. \\
& \left. + \sum_{j=1}^m \frac{\partial \bar{f}^{l,\epsilon}(s,0)}{\partial \tilde{x}^j}\Lambda^k(s-\alpha_j) \right) ds.
\end{aligned} \tag{45}
$$

Multiplying (44) by ϵ^{-1}, subtracting (45), taking the norm of both sides and then applying (37) with $\delta = \gamma$, we obtain

$$
\begin{aligned}
|\epsilon^{-1}\varphi^\epsilon(t) - \Lambda^k(t)| \leq {} & \epsilon^{-1}|\rho_k^\epsilon| + \epsilon^{-1}|\rho_{k+1}^\epsilon| + (L_3 T + mL_3 T)\gamma \\
& + \int_{\tau_k}^t N_1|\epsilon^{-1}\varphi^\epsilon(s) - \Lambda^k(s)|ds \\
& + \sum_{j=1}^m \int_{\tau_k}^t N_1|\epsilon^{-1}\varphi^\epsilon(s-\alpha_j) - \Lambda^k(s-\alpha_j)|ds.
\end{aligned} \tag{46}
$$

In view of (11) and (25), the last integral term on the right-hand side of (46) can be simplified as follows:

$$
\begin{aligned}
\sum_{j=1}^m \int_{\tau_k}^t N_1|\epsilon^{-1}\varphi^\epsilon(s-\alpha_j) - \Lambda^k(s-\alpha_j)|ds &= \sum_{j=1}^m \int_{\tau_k-\alpha_j}^{t-\alpha_j} N_1|\epsilon^{-1}\varphi^\epsilon(s) - \Lambda^k(s)|ds \\
&\leq \int_{\tau_k}^t mN_1|\epsilon^{-1}\varphi^\epsilon(s) - \Lambda^k(s)|ds.
\end{aligned} \tag{47}
$$

Thus, using Lemma 3 and (47) to simplify (46), we obtain

$$
\begin{aligned}
|\epsilon^{-1}\varphi^\epsilon(t) - \Lambda^k(t)| \leq {} & 2L_4\epsilon + (L_3 T + mL_3 T)\gamma \\
& + (m+1)N_1 \int_{\tau_k}^t |\epsilon^{-1}\varphi^\epsilon(s) - \Lambda^k(s)|ds.
\end{aligned}
$$

Since $\epsilon < \gamma$ and $\tau_k + \gamma < t$, this inequality becomes

$$|\epsilon^{-1}\varphi^\epsilon(t) - \Lambda^k(t)| \leq 2L_4\gamma + (L_3T + mL_3T)\gamma + (m+1)N_1 \int_{\tau_k}^{\tau_k+\gamma} |\epsilon^{-1}\varphi^\epsilon(s) - \Lambda^k(s)|ds$$

$$+ (m+1)N_1 \int_{\tau_k+\gamma}^{t} |\epsilon^{-1}\varphi^\epsilon(s) - \Lambda^k(s)|ds. \tag{48}$$

Using (13), we have

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

where $N_2 > 0$ is an upper bound for $|\Lambda^k|$ (recall that Λ^k is piecewise continuous and therefore bounded on $[0, T]$). Substituting (49) into (48) gives

$$|\epsilon^{-1}\varphi^\epsilon(t) - \Lambda^k(t)| \leq N_3\gamma + (m+1)N_1 \int_{\tau_k+\gamma}^{t} |\epsilon^{-1}\varphi^\epsilon(s) - \Lambda^k(s)|ds,$$

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

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

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

3.3 Left State Variation with Respect to the Switching Times

The left state variation at time t with respect to τ_k is defined by

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

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

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

$$\rho_k^\epsilon := \int_{\tau_k+\epsilon}^{\tau_k} \{f^k(s, x(s), \tilde{x}(s), \zeta) - f^k(\tau_k, x(\tau_k), \tilde{x}(\tau_k), \zeta)\}ds, \tag{50}$$

$$\rho_{k+1}^\epsilon := \int_{\tau_k+\epsilon}^{\tau_k} \{f^{k+1}(s, x^\epsilon(s), \tilde{x}^\epsilon(s), \zeta) - f^{k+1}(\tau_k, x(\tau_k), \tilde{x}(\tau_k), \zeta)\}ds. \tag{51}$$

By using similar arguments as in the proof of Lemma 3, we obtain the following result.

Lemma 4 *There exists a constant $L_5 > 0$ such that for all $\epsilon \in \Theta \cap (-\infty, 0]$,*

$$|\rho_k^\epsilon|,\ |\rho_{k+1}^\epsilon| \le L_5 \epsilon^2.$$

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

Theorem 3 *Let $k \in \{1, 2, \ldots, p-1\}$ and $(\tau, \zeta) \in \mathcal{S} \times \mathcal{Z}$. Then for all time points $t \in [0, T]$,*

$$\lim_{\epsilon \to 0-} \epsilon^{-1} \varphi^\epsilon(t) = \Lambda^k(t|\tau, \zeta), \tag{52}$$

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

Proof Let $\bar{f}^{i,\epsilon}, \Delta_1^{i,\epsilon}$, and $\Delta_2^{i,\epsilon}$ be as defined in the proof of Theorem 1. Recall that for each $\delta > 0$, there exists a corresponding $\epsilon' > 0$ such that for all $\epsilon \in \Theta$ satisfying $|\epsilon| < \epsilon'$,

$$|\Delta_1^{i,\epsilon}(s, \eta)| \le \delta L_3 |\epsilon|, \quad |\Delta_2^{i,\epsilon}(s, \eta)| \le \delta m L_3 |\epsilon|, \quad i = 1, 2, \ldots, p, \tag{53}$$

where L_3 is as defined in Lemma 2.

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

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

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

Thus, in view of (25),

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

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

For case (ii), there exists an integer $\varsigma \in \{k+1, \ldots, p\}$ such that $t \in [\tau_{\varsigma-1}, \tau_\varsigma)$ if $\varsigma \le p-1$, and $t \in [\tau_{p-1}, \tau_p]$ if $\varsigma = p$. Let $\delta > 0$ be arbitrary and choose $\epsilon \in \Theta$ such that

$$-\min\{\epsilon', \delta\} < \epsilon < 0. \tag{54}$$

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

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

Therefore, from (50) and (51),

$$\varphi^\epsilon(t) = \rho_{k+1}^\epsilon - \rho_k^\epsilon - \epsilon \left\{ \bar{f}^{k+1,\epsilon}(\tau_k,0) - \bar{f}^{k,\epsilon}(\tau_k,0) \right\}$$
$$+ \sum_{l=k+1}^{\varsigma} \int_{\tau_{l-1}}^{\min\{\tau_l,t\}} \left\{ \bar{f}^{l,\epsilon}(s,1) - \bar{f}^{l,\epsilon}(s,0) \right\} ds. \tag{55}$$

Using the fundamental theorem of calculus, (55) can be written as

$$\varphi^\epsilon(t) = \rho_{k+1}^\epsilon - \rho_k^\epsilon - \epsilon \left\{ \bar{f}^{k+1,\epsilon}(\tau_k,0) - \bar{f}^{k,\epsilon}(\tau_k,0) \right\}$$
$$+ \sum_{l=k+1}^{\varsigma} \int_{\tau_{l-1}}^{\min\{\tau_l,t\}} \frac{\partial \bar{f}^{l,\epsilon}(s,0)}{\partial x} \varphi^\epsilon(s) ds + \sum_{l=k+1}^{\varsigma} \int_{\tau_{l-1}}^{\min\{\tau_l,t\}} \sum_{j=1}^m \frac{\partial \bar{f}^{l,\epsilon}(s,0)}{\partial \tilde{x}^j} \theta^{\epsilon,j}(s) ds$$
$$+ \sum_{l=k+1}^{\varsigma} \int_{\tau_{l-1}}^{\min\{\tau_l,t\}} \left(\int_0^1 \Delta_1^{l,\epsilon}(s,\eta) d\eta + \int_0^1 \Delta_2^{l,\epsilon}(s,\eta) d\eta \right) ds. \tag{56}$$

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

$$\Lambda^k(t) = \bar{f}^{k,\epsilon}(\tau_k,0) - \bar{f}^{k+1,\epsilon}(\tau_k,0) + \sum_{l=k+1}^{\varsigma} \int_{\tau_{l-1}}^{\min\{\tau_l,t\}} \left(\frac{\partial \bar{f}^{l,\epsilon}(s,0)}{\partial x} \Lambda^k(s) \right.$$
$$\left. + \sum_{j=1}^m \frac{\partial \bar{f}^{l,\epsilon}(s,0)}{\partial \tilde{x}^j} \Lambda^k(s-\alpha_j) \right) ds. \tag{57}$$

From (56) and (57), we obtain

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

where N_1 is the upper bound for $|\partial \bar{f}^{i,\epsilon}/\partial x|$ and $|\partial \bar{f}^{i,\epsilon}/\partial \tilde{x}^j|$ defined in the proof of Theorem 2. By Lemma 4, (10) and (53), we see that for all $\epsilon \in \Theta$ satisfying inequality (54),

$$|\epsilon^{-1}\varphi^\epsilon(t) - \Lambda^k(t)| \leq 2L_5|\epsilon| + (m+1)\delta L_3 T + \int_{\tau_k}^t N_1|\epsilon^{-1}\varphi^\epsilon(s) - \Lambda^k(s)|ds$$
$$+ \sum_{j=1}^m \int_{\tau_k}^t N_1|\epsilon^{-1}\varphi^\epsilon(s-\alpha_j) - \Lambda^k(s-\alpha_j)|ds. \tag{58}$$

The last integral term in (58) can be simplified as

$$\sum_{j=1}^m \int_{\tau_k}^t N_1|\epsilon^{-1}\varphi^\epsilon(s-\alpha_j) - \Lambda^k(s-\alpha_j)|ds = \sum_{j=1}^m \int_{\tau_k-\alpha_j}^{t-\alpha_j} N_1|\epsilon^{-1}\varphi^\epsilon(s) - \Lambda^k(s)|ds$$
$$\leq \int_{-\alpha_{\max}}^t mN_1|\epsilon^{-1}\varphi^\epsilon(s) - \Lambda^k(s)|ds. \tag{59}$$

Since $\epsilon \in \Theta \cap (-\infty, 0]$, we have $-\alpha_{\max} < \tau_k + \epsilon < \tau_k$, and thus

$$\int_{-\alpha_{\max}}^t mN_1|\epsilon^{-1}\varphi^\epsilon(s) - \Lambda^k(s)|ds = \int_{-\alpha_{\max}}^{\tau_k+\epsilon} mN_1|\epsilon^{-1}\varphi^\epsilon(s) - \Lambda^k(s)|ds$$
$$+ \int_{\tau_k+\epsilon}^{\tau_k} mN_1|\epsilon^{-1}\varphi^\epsilon(s) - \Lambda^k(s)|ds$$
$$+ \int_{\tau_k}^t mN_1|\epsilon^{-1}\varphi^\epsilon(s) - \Lambda^k(s)|ds. \tag{60}$$

Furthermore, using (13), we have

$$|\epsilon^{-1}\varphi^\epsilon(s) - \Lambda^k(s)| \leq |\epsilon|^{-1}|\varphi^\epsilon(s)| + |\Lambda^k(s)| \leq L_3 + N_2, \quad s \in [\tau_k + \epsilon, \tau_k], \tag{61}$$

where N_2 is as defined in the proof of Theorem 2. Substituting (61) into (60) and then using (11) gives

$$\int_{-\alpha_{\max}}^t mN_1|\epsilon^{-1}\varphi^\epsilon(s) - \Lambda^k(s)|ds \leq (mL_3N_1 + mN_2N_1)|\epsilon|$$
$$+ \int_{\tau_k}^t mN_1|\epsilon^{-1}\varphi^\epsilon(s) - \Lambda^k(s)|ds. \tag{62}$$

Combining (58), (59) and (62), we obtain

$$|\epsilon^{-1}\varphi^\epsilon(t) - \Lambda^k(t)| \leq (2L_5 + mL_3N_1 + mN_2N_1)|\epsilon| + (m+1)L_3T\delta$$
$$+ \int_{\tau_k}^t (m+1)N_1|\epsilon^{-1}\varphi^\epsilon(s) - \Lambda^k(s)|ds.$$

Recall from (54) that $|\epsilon| < \delta$. Thus, by the Gronwall-Bellman Lemma [32],

$$|\epsilon^{-1}\varphi^\epsilon(t) - \Lambda^k(t)| \leq (2L_5 + mL_3N_1 + mN_2N_1 + (m+1)L_3T)\delta\exp((m+1)N_1T),$$

which holds whenever $\epsilon \in \Theta$ satisfies (54). Since $\delta > 0$ is arbitrary, this shows that (52) holds when $t \geq \tau_k$.

3.4 State Variation with Respect to the Switching Times

Combining Theorems 2 and 3 yields the following fundamental result, which gives the full state variation with respect to each switching time.

Theorem 4 *Let $k \in \{1, 2, \ldots, p-1\}$ and $(\tau, \zeta) \in \mathcal{S} \times \mathcal{Z}$. Then for all time points $t \neq \tau_k$,*

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

From Theorems 2 and 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 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 1).

4 Gradient Formulae

By using Theorems 1 and 4, it is possible to express the gradient of the cost function in Problem (P) in terms of the solution 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

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

and

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

For each $(\tau,\zeta) \in \mathcal{T} \times \mathcal{Z}$, consider the following *costate system*:

$$
\dot{\lambda}(t) = -\sum_{i=1}^{p} \left(\frac{\partial \tilde{f}^i(t|\tau,\zeta)}{\partial x} \right)^{\top} \lambda(t)\chi_{[\tau_{i-1},\tau_i)}(t)
$$
$$
- \sum_{i=1}^{p}\sum_{j=1}^{m} \left(\frac{\partial \tilde{f}^i(t+\alpha_j|\tau,\zeta)}{\partial \tilde{x}^j} \right)^{\top} \lambda(t+\alpha_j)\chi_{[\tau_{i-1}-\alpha_j,\tau_i-\alpha_j)}(t), \quad t \in [0,T], \tag{63}
$$

with the terminal conditions

$$
\lambda(T) = \left(\frac{\partial \Phi(x(T|\tau,\zeta),\zeta)}{\partial x} \right)^{\top}, \tag{64}
$$

$$
\lambda(t) = 0, \quad t > T, \tag{65}
$$

where, for a given interval I, χ_I denotes the indicator function of I defined by

$$\chi_I(t) := \begin{cases} 1, & \text{if } t \in I, \\ 0, & \text{otherwise.} \end{cases} \tag{66}$$

Note that if $t > T - \alpha_j$, then $t + \alpha_j > T$, and thus the value of $\dfrac{\partial \tilde{f}^i(t + \alpha_j | \tau, \zeta)}{\partial \tilde{x}^j}$ 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 equation (65)).

Let $\lambda(\cdot | \tau, \zeta)$ denote the solution of system (63)-(65) corresponding to the given pair $(\tau, \zeta) \in \mathcal{T} \times \mathcal{Z}$. We now express the gradients of the cost function with respect to the system parameters in terms of $\lambda(\cdot | \tau, \zeta)$.

Theorem 5 *Let $q \in \{1, 2, \ldots, v\}$ and $(\tau, \zeta) \in \mathcal{T} \times \mathcal{Z}$. Then*

$$\frac{\partial J(\tau, \zeta)}{\partial \zeta_q} = \frac{\partial \Phi(x(T), \zeta)}{\partial \zeta_q} + \sum_{i=1}^p \int_{\tau_{i-1}}^{\tau_i} \lambda(s)^\top \frac{\partial \tilde{f}^i(s | \tau, \zeta)}{\partial \zeta_q} ds + \lambda(0^+)^\top \frac{\partial \phi(0, \zeta)}{\partial \zeta_q}$$

$$+ \sum_{i=1}^p \sum_{j=1}^m \int_{\tau_{i-1} - \alpha_j}^{\tau_i - \alpha_j} \lambda(s + \alpha_j)^\top \frac{\partial \tilde{f}^i(s + \alpha_j | \tau, \zeta)}{\partial \tilde{x}^j} \frac{\partial \phi(s, \zeta)}{\partial \zeta_q} \chi_{(-\infty, 0)}(s) ds,$$

where $x(\cdot) = x(\cdot | \tau, \zeta)$, $\lambda(\cdot) = \lambda(\cdot | \tau, \zeta)$ and $\chi_{(-\infty, 0)}$ is as defined in (66).

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:

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

$$= \Phi(x(T), \zeta) + \sum_{i=1}^p \int_{\tau_{i-1}}^{\tau_i} w(s)^\top f^i(s, x(s), \tilde{x}(s), \zeta) ds - \int_0^T w(s)^\top \dot{x}(s) ds,$$

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

Applying integration by parts to the last integral term gives

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

$$+ w(0^+)^\top \phi(0, \zeta) + \int_0^T \dot{w}(s)^\top x(s) ds. \tag{67}$$

Differentiating (67) with respect to ζ_q gives

$$
\begin{aligned}
\frac{\partial J(\tau,\zeta)}{\partial \zeta_q} = {} & \left(\frac{\partial \Phi(x(T),\zeta)}{\partial x} - w(T)^\top \right) \frac{\partial x(T)}{\partial \zeta_q} + \frac{\partial \Phi(x(T),\zeta)}{\partial \zeta_q} + w(0^+)^\top \frac{\partial \phi(0,\zeta)}{\partial \zeta_q} \\
& + \sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_i} w(s)^\top \frac{\partial \tilde{f}^i(s)}{\partial \zeta_q} ds + \sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_i} \left(w(s)^\top \frac{\partial \tilde{f}^i(s)}{\partial x} + \dot{w}(s)^\top \right) \frac{\partial x(s)}{\partial \zeta_q} ds \\
& + \sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_i} \sum_{j=1}^{m} w(s)^\top \frac{\partial \tilde{f}^i(s)}{\partial \tilde{x}^j} \frac{\partial x(s-\alpha_j)}{\partial \zeta_q} ds,
\end{aligned}
\tag{68}
$$

where we have omitted the arguments τ and ζ in $\dfrac{\partial \tilde{f}^i(s|\tau,\zeta)}{\partial x}$ and $\dfrac{\partial \tilde{f}^i(s|\tau,\zeta)}{\partial \tilde{x}^j}$. Performing a change of variable in the last term on the right-hand side of (68) yields

$$
\begin{aligned}
& \sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_i} \sum_{j=1}^{m} w(s)^\top \frac{\partial \tilde{f}^i(s)}{\partial \tilde{x}^j} \frac{\partial x(s-\alpha_j)}{\partial \zeta_q} ds \\
& = \sum_{i=1}^{p} \sum_{j=1}^{m} \int_{\tau_{i-1}-\alpha_j}^{\tau_i-\alpha_j} w(s+\alpha_j)^\top \frac{\partial \tilde{f}^i(s+\alpha_j)}{\partial \tilde{x}^j} \frac{\partial x(s)}{\partial \zeta_q} ds.
\end{aligned}
\tag{69}
$$

Since $x(s) = \phi(s,\zeta)$ for all $s \leq 0$, equation (69) can be rewritten as follows:

$$
\begin{aligned}
& \sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_i} \sum_{j=1}^{m} w(s)^\top \frac{\partial \tilde{f}^i(s)}{\partial \tilde{x}^j} \frac{\partial x(s-\alpha_j)}{\partial \zeta_q} ds \\
& = \sum_{i=1}^{p} \sum_{j=1}^{m} \int_{\tau_{i-1}-\alpha_j}^{\tau_i-\alpha_j} w(s+\alpha_j)^\top \frac{\partial \tilde{f}^i(s+\alpha_j)}{\partial \tilde{x}^j} \frac{\partial x(s)}{\partial \zeta_q} \chi_{[0,+\infty)}(s) ds \\
& \quad + \sum_{i=1}^{p} \sum_{j=1}^{m} \int_{\tau_{i-1}-\alpha_j}^{\tau_i-\alpha_j} w(s+\alpha_j)^\top \frac{\partial \tilde{f}^i(s+\alpha_j)}{\partial \tilde{x}^j} \frac{\partial \phi(s,\zeta)}{\partial \zeta_q} \chi_{(-\infty,0)}(s) ds,
\end{aligned}
\tag{70}
$$

where $\chi_{[0,+\infty)}$ and $\chi_{(-\infty,0)}$ are as defined in (66). Substituting (70) into (68) yields

$$
\begin{aligned}
\frac{\partial J(\tau,\zeta)}{\partial \zeta_q} = {} & \left(\frac{\partial \Phi(x(T),\zeta)}{\partial x} - w(T)^\top \right) \frac{\partial x(T)}{\partial \zeta_q} + \frac{\partial \Phi(x(T),\zeta)}{\partial \zeta_q} + w(0^+)^\top \frac{\partial \phi(0,\zeta)}{\partial \zeta_q} \\
& + \sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_i} w(s)^\top \frac{\partial \tilde{f}^i(s)}{\partial \zeta_q} ds + \sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_i} \left(w(s)^\top \frac{\partial \tilde{f}^i(s)}{\partial x} + \dot{w}(s)^\top \right) \frac{\partial x(s)}{\partial \zeta_q} ds \\
& + \sum_{i=1}^{p} \sum_{j=1}^{m} \int_{\tau_{i-1}-\alpha_j}^{\tau_i-\alpha_j} w(s+\alpha_j)^\top \frac{\partial \tilde{f}^i(s+\alpha_j)}{\partial \tilde{x}^j} \frac{\partial x(s)}{\partial \zeta_q} \chi_{[0,+\infty)}(s) ds \\
& + \sum_{i=1}^{p} \sum_{j=1}^{m} \int_{\tau_{i-1}-\alpha_j}^{\tau_i-\alpha_j} w(s+\alpha_j)^\top \frac{\partial \tilde{f}^i(s+\alpha_j)}{\partial \tilde{x}^j} \frac{\partial \phi(s,\zeta)}{\partial \zeta_q} \chi_{(-\infty,0)}(s) ds.
\end{aligned}
$$

This equation can be rewritten as follows:

$$
\begin{aligned}
\frac{\partial J(\tau,\zeta)}{\partial \zeta_q} =& \left(\frac{\partial \Phi(x(T),\zeta)}{\partial x} - w(T)^\top\right)\frac{\partial x(T)}{\partial \zeta_q} + \frac{\partial \Phi(x(T),\zeta)}{\partial \zeta_q} + w(0^+)^\top\frac{\partial \phi(0,\zeta)}{\partial \zeta_q} \\
&+ \sum_{i=1}^{p}\int_{\tau_{i-1}}^{\tau_i} w(s)^\top\frac{\partial \tilde{f}^i(s)}{\partial \zeta_q}ds + \int_0^T\left(\dot{w}(s)^\top + \sum_{i=1}^{p} w(s)^\top\frac{\partial \tilde{f}^i(s)}{\partial x}\chi_{[\tau_{i-1},\tau_i)}(s)\right. \\
&+ \sum_{i=1}^{p}\sum_{j=1}^{m} w(s+\alpha_j)^\top\frac{\partial \tilde{f}^i(s+\alpha_j)}{\partial \tilde{x}^j}\chi_{[\tau_{i-1}-\alpha_j,\tau_i-\alpha_j)}(s)\left.\right)\frac{\partial x(s)}{\partial \zeta_q}ds \\
&+ \sum_{i=1}^{p}\sum_{j=1}^{m}\int_{\tau_{i-1}-\alpha_j}^{\tau_i-\alpha_j} w(s+\alpha_j)^\top\frac{\partial \tilde{f}^i(s+\alpha_j)}{\partial \tilde{x}^j}\frac{\partial \phi(s,\zeta)}{\partial \zeta_q}\chi_{(-\infty,0)}(s)ds.
\end{aligned}
$$

Choosing $w(\cdot) = \lambda(\cdot|\tau,\zeta)$ and substituting (63)-(65) into the above equation completes the proof.

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

Theorem 6 *Let $k \in \{1,2,\ldots,p-1\}$ and $(\tau,\zeta) \in \mathcal{T} \times \mathcal{Z}$. 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 5. Recall from equation (67) that

$$
\begin{aligned}
J(\tau,\zeta) =& \Phi(x(T),\zeta) + \sum_{i=1}^{p}\int_{\tau_{i-1}}^{\tau_i} w(s)^\top f^i(s,x(s),\tilde{x}(s),\zeta)ds - w(T)^\top x(T) \\
&+ w(0^+)^\top\phi(0,\zeta) + \int_0^T \dot{w}(s)^\top x(s)ds,
\end{aligned}
\tag{71}
$$

where, as in the proof of Theorem 5, we have omitted the arguments τ and ζ for clarity.

Differentiating (71) with respect to τ_k yields

$$
\begin{aligned}
\frac{\partial J(\tau,\zeta)}{\partial \tau_k} =& \left(\frac{\partial \Phi(x(T),\zeta)}{\partial x} - w(T)^\top\right)\frac{\partial x(T)}{\partial \tau_k} + w(\tau_k)^\top f^k(\tau_k, x(\tau_k), \tilde{x}(\tau_k), \zeta) \\
&- w(\tau_k)^\top f^{k+1}(\tau_k, x(\tau_k), \tilde{x}(\tau_k), \zeta) + \sum_{i=1}^{p}\int_{\tau_{i-1}}^{\tau_i}\left(w(s)^\top\frac{\partial \tilde{f}^i(s)}{\partial x}\right. \\
&+ \dot{w}(s)^\top\left.\right)\frac{\partial x(s)}{\partial \tau_k}ds + \sum_{i=1}^{p}\int_{\tau_{i-1}}^{\tau_i}\sum_{j=1}^{m} w(s)^\top\frac{\partial \tilde{f}^i(s)}{\partial \tilde{x}^j}\frac{\partial x(s-\alpha_j)}{\partial \tau_k}ds,
\end{aligned}
\tag{72}
$$

where, as in the proof of Theorem 5, we omit the arguments τ and ζ in $\dfrac{\partial \tilde{f}^i(s|\tau,\zeta)}{\partial x}$ and $\dfrac{\partial \tilde{f}^i(s|\tau,\zeta)}{\partial \tilde{x}^j}$.

Performing a change of variable in the last term on the right-hand side of (72) gives

$$
\sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_i} \sum_{j=1}^{m} w(s)^\top \frac{\partial \tilde{f}^i(s)}{\partial \tilde{x}^j} \frac{\partial x(s-\alpha_j)}{\partial \tau_k} ds
$$
$$
= \sum_{i=1}^{p} \sum_{j=1}^{m} \int_{\tau_{i-1}-\alpha_j}^{\tau_i-\alpha_j} w(s+\alpha_j)^\top \frac{\partial \tilde{f}^i(s+\alpha_j)}{\partial \tilde{x}^j} \frac{\partial x(s)}{\partial \tau_k} ds. \tag{73}
$$

Recall that $x(s) = \phi(s,\zeta)$ for all $s \leq 0$. Then equation (73) can be rewritten as follows:

$$
\sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_i} \sum_{j=1}^{m} w(s)^\top \frac{\partial \tilde{f}^i(s)}{\partial \tilde{x}^j} \frac{\partial x(s-\alpha_j)}{\partial \tau_k} ds
$$
$$
= \sum_{i=1}^{p} \sum_{j=1}^{m} \int_{\tau_{i-1}-\alpha_j}^{\tau_i-\alpha_j} w(s+\alpha_j)^\top \frac{\partial \tilde{f}^i(s+\alpha_j)}{\partial \tilde{x}^j} \frac{\partial x(s)}{\partial \tau_k} \chi_{[0,+\infty)}(s) ds, \tag{74}
$$

where $\chi_{[0,+\infty)}$ is as defined in (66). Substituting (74) into (72) gives

$$
\frac{\partial J(\tau,\zeta)}{\partial \tau_k} = \left(\frac{\partial \Phi(x(T),\zeta)}{\partial x} - w(T)^\top \right) \frac{\partial x(T)}{\partial \tau_k} + w(\tau_k)^\top f^k(\tau_k, x(\tau_k), \tilde{x}(\tau_k), \zeta)
$$
$$
- w(\tau_k)^\top f^{k+1}(\tau_k, x(\tau_k), \tilde{x}(\tau_k), \zeta) + \sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_i} \left(w(s)^\top \frac{\partial \tilde{f}^i(s)}{\partial x} + \dot{w}(s)^\top \right) \frac{\partial x(s)}{\partial \tau_k} ds
$$
$$
+ \sum_{i=1}^{p} \sum_{j=1}^{m} \int_{\tau_{i-1}-\alpha_j}^{\tau_i-\alpha_j} w(s+\alpha_j)^\top \frac{\partial \tilde{f}^i(s+\alpha_j)}{\partial \tilde{x}^j} \frac{\partial x(s)}{\partial \tau_k} \chi_{[0,+\infty)}(s) ds.
$$

This equation can be rearranged as follows:

$$
\frac{\partial J(\tau,\zeta)}{\partial \tau_k} = \left(\frac{\partial \Phi(x(T),\zeta)}{\partial x} - w(T)^\top \right) \frac{\partial x(T)}{\partial \tau_k} + w(\tau_k)^\top f^k(\tau_k, x(\tau_k), \tilde{x}(\tau_k), \zeta)
$$
$$
- w(\tau_k)^\top f^{k+1}(\tau_k, x(\tau_k), \tilde{x}(\tau_k), \zeta) + \int_0^T \left(\sum_{i=1}^{p} w(s)^\top \frac{\partial \tilde{f}^i(s)}{\partial x} \chi_{[\tau_{i-1},\tau_i)}(s) \right.
$$
$$
\left. + \dot{w}(s)^\top + \sum_{i=1}^{p} \sum_{j=1}^{m} w(s+\alpha_j)^\top \frac{\partial \tilde{f}^i(s+\alpha_j)}{\partial \tilde{x}^j} \chi_{[\tau_{i-1}-\alpha_j,\tau_i-\alpha_j)}(s) \right) \frac{\partial x(s)}{\partial \tau_k} ds. \tag{75}
$$

Choosing $w(\cdot) = \lambda(\cdot|\tau,\zeta)$ and substituting (63)-(65) into (75) completes the proof.

On the basis of Theorems 5 and 6, we can compute the cost function $J(\tau,\zeta)$ and its gradient at a given pair $(\tau,\zeta) \in \mathcal{T} \times \mathcal{Z}$ using the following computational procedure.

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

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

Step 3. Using $x(\cdot|\tau, \zeta)$, solve the costate system (63)-(65) from $t = T$ to $t = 0$ to obtain $\lambda(\cdot|\tau, \zeta)$.

Step 4. Using $x(\cdot|\tau, \zeta)$ and $\lambda(\cdot|\tau, \zeta)$, compute $\dfrac{\partial J(\tau, \zeta)}{\partial \zeta_q}, q = 1, 2, \ldots, v$, and $\dfrac{\partial J(\tau, \zeta)}{\partial \tau_k}, k = 1, 2, \ldots, p - 1,$

via the formulae in Theorems 5 and 6.

This computational procedure can be integrated with a standard gradient-based optimization method (e.g., sequential quadratic programming) to solve Problem (P) as a nonlinear programming problem. In the 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 Section 4 with the optimization software NLPQLP [33]. This program invokes the differential equation software LSODA [34] to solve the state and costate systems. Lagrange interpolation [35] is used when LSODA requires the value of the state or costate at an intermediate time between two adjacent knot points.

5.1 Example 1

Consider the following switched system as formulated in reference [27]:

$$\text{subsystem 1:} \begin{cases} \dot{x}_1(t) = 2x_1(t)x_2(t) + x_2(t - 0.1), \\ \dot{x}_2(t) = 3x_1(t) + 4x_2(t - 0.1), \end{cases} \tag{76}$$

$$\text{subsystem 2:} \begin{cases} \dot{x}_1(t) = -2x_1(t)x_2(t) + \sin(x_2(t - 0.1)), \\ \dot{x}_2(t) = x_1(t)x_2(t) + x_1(t - 0.1)x_2(t - 0.1), \end{cases} \tag{77}$$

$$\text{subsystem 3:} \begin{cases} \dot{x}_1(t) = t^2 - 2x_1(t) + 3x_2(t - 0.1), \\ \dot{x}_2(t) = -x_2(t) + x_1(t - 0.1)x_2(t - 0.1), \end{cases} \tag{78}$$

with the initial conditions

$$x_1(t) = t - 1, \quad x_2(t) = t^2 + 1, \quad t \le 0. \tag{79}$$

As in [27], we assume that the switching sequence is $1 \to 2 \to 3$. Let τ_1 denote the time at which the system switches from subsystem 1 to subsystem 2, and let τ_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 τ_1 and τ_2 such that the cost function

$$J(\tau_1, \tau_2) = (x_1(1) - 0.5)^2 + (x_2(1) - 0.25)^2 \tag{80}$$

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 7 optimization iterations. The optimal cost obtained by our program is $J = 4.537 \times 10^{-3}$, which is much better than the optimal cost of $J = 0.0128$ obtained in [27]. The corresponding state trajectory is shown in Fig. 1.
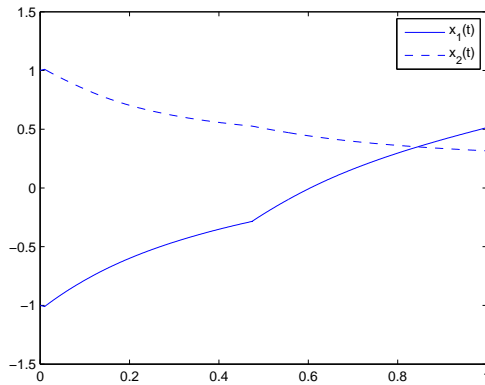

Fig. 1 The optimal state trajectories for Example 1.

5.2 Example 2

Consider the following nonlinear switched system with 2 time-delays:

$$\text{subsystem 1:} \begin{cases} \dot{x}_1(t) = -5x_1(t) - 4x_2(t) - 3x_1(t-0.5) + 2x_2(t-0.1) \\ \qquad +u_1(t) + 0.1\tanh(x_1(t)), \\ \dot{x}_2(t) = 0.1x_1(t) - 7x_2(t) + 0.5u_1(t) - \sin(x_2(t-0.1)), \end{cases} \tag{81}$$

$$\text{subsystem 2:} \begin{cases} \dot{x}_1(t) = -4x_1(t) + 0.5x_2(t) + 0.2\sin(x_2(t)) + t^2 + 8, \\ \dot{x}_2(t) = 5x_1(t) - 5x_2(t) + 0.5\sin(x_1(t-0.1)) - u_2(t), \end{cases} \tag{82}$$

with the initial conditions

$$x_1(t) = 6, \quad x_2(t) = t^2 + 2, \quad t \le 0. \tag{83}$$

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 τ_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 $0.1 \le \zeta_1 \le 1.5$ and $0.1 \le \zeta_2, \zeta_3, \zeta_4 \le 1.0$. Our goal is to find an optimal switching time τ_1 and an optimal feedback gain parameter vector $\zeta = (\zeta_1, \zeta_2, \zeta_3, \zeta_4)^\top$ such that the cost function

$$J(\tau_1, \zeta) = (x_1(1.5) - 2)^2 + (x_2(1.5) - 1)^2 \tag{84}$$

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 (84) with respect to ζ is calculated using the formulae in Section 4, with the extended Simpson's rule [35] used to compute the integrals. Starting from the initial guesses $\tau_1 = 0.1$ and $\zeta = (0.7, 0.5, 0.5, 0.8)^\top$, our program computes the optimal switching time as $\tau_1 = 1.3304$ and the optimal parameter vector as $\zeta = (1.3149, 0.5635, 0.3256, 0.6672)^\top$ after 22 optimization iterations. Moreover, the corresponding optimal cost is 3.9943×10^{-30}, which can be regarded as zero in numerical computation. The optimal feedback control and the optimal state trajectory are shown in Fig. 2.
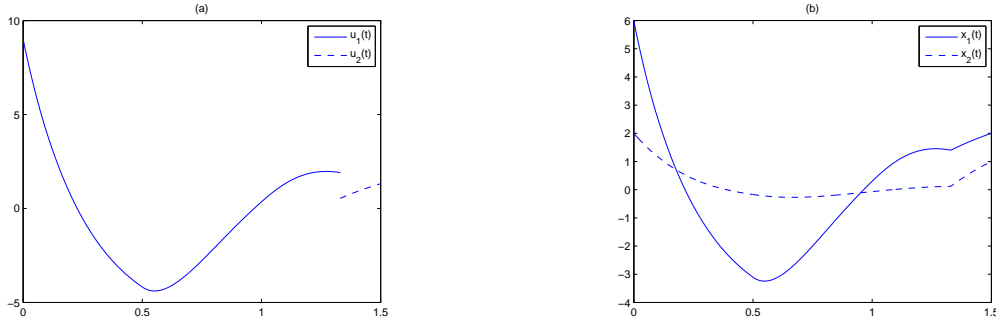
Fig. 2 (a). Feedback controls for Example 2. (b). State trajectories for Example 2.

5.3 Example 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 [36], time-delays exist in the fermentation process.

Based on our previous work [37], the dynamic model for batch mode is given by

$$
\begin{cases}
\dot{x}_1(t) = \mu(x_2(t))x_1(t - 0.16), \\
\dot{x}_2(t) = -q_2(x_2(t))x_1(t - 0.16), \\
\dot{x}_3(t) = q_3(x_2(t))x_1(t - 0.16), \\
\dot{x}_4(t) = 0,
\end{cases}
\tag{85}
$$

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

$$\mu(x_2(t)) = \frac{\Delta_1 x_2(t)}{x_2(t) + k_1},$$

$$q_2(x_2(t)) = m_2 + Y_2\mu(x_2(t)) + \frac{\Delta_2 x_2(t)}{x_2(t) + k_2},$$

$$q_3(x_2(t)) = -m_3 + Y_3\mu(x_2(t)) + \frac{\Delta_3 x_2(t)}{x_2(t) + k_3},$$

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 [37].

The dynamic model for feeding mode is given by

$$\begin{cases} \dot{x}_1(t) = \mu(x_2(t))x_1(t - 0.16) - D(x_4(t), \zeta)x_1(t), \\ \dot{x}_2(t) = D(x_4(t), \zeta)((1 + r)^{-1}c_{s0} - x_2(t)) - q_2(x_2(t))x_1(t - 0.16), \\ \dot{x}_3(t) = q_3(x_2(t))x_1(t - 0.16) - D(x_4(t), \zeta)x_3(t), \\ \dot{x}_4(t) = (1 + r)\zeta, \end{cases} \quad (86)$$

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

$$D(x_4(t), \zeta) := \frac{(1 + r)\zeta}{x_4(t)}. \quad (87)$$

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

The initial function ϕ for the system (85) and (86) is obtained by applying cubic spline interpolation to the experimental data before the zero time point. The total number of switching times is 1355 and the terminal time is $T = 24.16$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 seconds 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 τ_1) and the end 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 ζ also needs to be optimally chosen. Let $\tau = (\tau_1, \ldots, \tau_{10})^\top$. The aim of the fermentation control is to

Table 1 Optimal values, initial guesses, and lower and upper bounds for the switching times in Example 3.

Switching times	τ_1	τ_2	τ_3	τ_4	τ_5
Optimal values	5.4096	5.4112	6.1894	7.2174	8.9117
Initial guesses	5.3300	5.3314	6.1097	7.1378	8.8319
Lower bounds	5.3000	5.3014	6.0797	7.1078	8.8019
Upper bounds	5.5000	5.5022	6.2800	7.3078	9.0022
Switching times	τ_6	τ_7	τ_8	τ_9	τ_{10}
Optimal values	12.2172	15.9110	18.1602	19.9101	23.9101
Initial guesses	12.1372	15.8311	18.0808	19.8306	23.8303
Lower bounds	12.1072	15.8011	18.0503	19.8003	23.8003
Upper bounds	12.3078	16.0022	18.2511	20.0011	24.0011

obtain high concentration of 1,3-PD at the terminal time. Thus, the cost function in this example is

$$J(\tau, \zeta) = -x_3(T|\tau, \zeta). \tag{88}$$

The optimal control problem is: choose the switching time vector τ and the system parameter ζ such that the cost function (88) is minimized subject to $\zeta \in [0.7954, 1.1009]$ and the switching time bounds given in Table 1.

The gradient of the cost function (88) with respect to ζ is calculated using the formulae in Section 4, with the extended Simpson's rule [35] used to compute the integrals. By choosing the initial guess of the feeding rate $\zeta = 8.1211 \text{Lh}^{-1}$ and the initial guesses of switching times as in Table 1, our program obtains the optimal feeding rate $\zeta^* = 0.8867 \text{Lh}^{-1}$ and the optimal switching times listed in Table 1. 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, $5.75\text{s}, 7.28\text{s}, 8.00\text{s}, 7.76\text{s}, 7.51\text{s}, 5.13\text{s}, 2.22\text{s}, 1.98\text{s}$ and 1.73s. Moreover, the optimal 1,3-PD concentration at the terminal time is 932.2mmolL^{-1}. This is a 16.96% improvement over the experimental results in [38], which are obtained using a non-optimal control scheme. The optimal states of 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 optimal control of switched systems with multiple time-delays and multiple system parameters. We first established the existence of the partial derivatives of the system
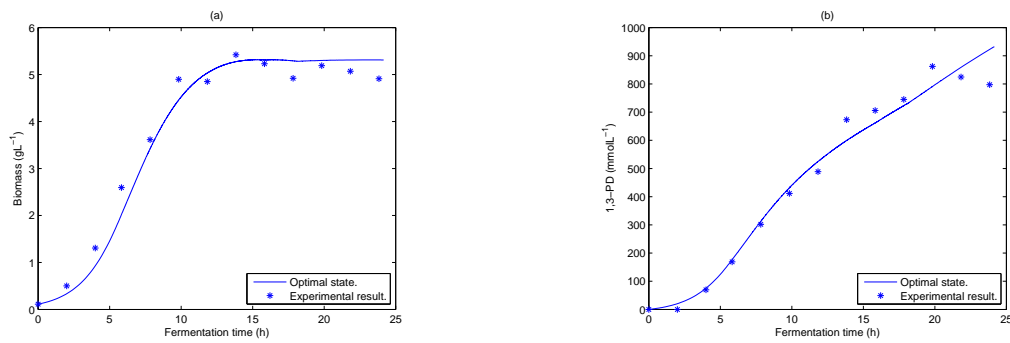
Fig. 3 Optimal state trajectories for Example 3.

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 optimal control problem. The effectiveness and applicability of this algorithm was verified using three numerical examples.

Acknowledgments

References

1. van der Schaft, A., Schumacher, H.: An Introduction to Hybrid Dynamical Systems. Springer, Berlin (2000)

2. Ramadge, P., Wonham, W.: The control of discrete event systems. Proc. IEEE **77**(1), 81–98 (1989)

3. Astrom, K.J., Puruta, K.: Swing up a pendulum by energy control. Automatica **36**, 287–295 (2000)

4. Hespanha, J., Liberzon, D., Morse, A.S.: Overcoming the limitations of adaptive control by means of logic-based switching. Syst. Control Lett. **49**, 49–56 (2003)

5. Howlett, P.: Optimal strategies for the control of a train. Automatica **32**, 519–532 (1996)

6. Liu, C.Y., Feng, E.M., Yin, H.C.: Optimal switching control for microbial fed-batch culture. Nonlinear Anal. Hybrid Syst. **2**, 1168–1174 (2008)

7. Loxton, R., Teo, K.L., Rehbock, V., Ling, W.K.: Optimal switching instants for a switched-capacitor DC/DC power converter. Automatica **45**, 973–980 (2009)

8. Woon, S.F., Rehbock, V., Loxton, R.: Towards global solutions of optimal discrete-valued control problems. Optim. Control Appl. Meth. **33**, 576–594 (2012)

9. Seidman, T.I.: Optimal control for switching systems. In: Proceedings of the 21st Annual Conference on Information Science and Systems, pp. 485–489 (1987).

10. Branicky, M.S., Borkar, V.S., Mitter, S.K.: A unified framework for hybrid control: Model and optimal control theory. IEEE Trans. Automat. Contr. **43**, 31–45 (1998)

11. Sussmann, H.: A maximum principle for hybrid optimal control problems. In: Proceedings of the 38th IEEE Conference on Decision and Control, pp. 3972–3977 (1999)

12. Luus, R., Chen, Y.Q.: Optimal switching control via direct search optimization. Asian J. Control **6**, 302–306 (2004)

13. Xu, X.P., Antsaklis, P.J.: Optimal control of switched systems based on parameterization of the switching times. IEEE Trans. Automat. Contr. **49**, 2–15 (2004)

14. Bengea, S.C., Raymond, A.D.: Optimal control of switching systems. Automatica **41**, 11–27 (2005)

15. Giua, A., Seatze, C., Van Der Mee, C.: Optimal control of autonomous linear systems switched with a preassigned finite sequence. In: Proceedings of the 2001 IEEE International Symposium on Intelligent Control, pp. 144–146 (2001)

16. Xu, X.P., Antsaklis, P.J.: Optimal control of switched autonomous systems. In: Proceedings of the 41st IEEE Conference on Decision and Control, pp. 4401–4406 (2002)

17. Egerstedt, M., Wardi, Y., Delmotte, F.: Optimal control of switching times in switched dynamical systems. In: Proceedings of the 42nd IEEE Conference on Decision and Control, pp. 2138–2143 (2003)

18. Lin, Q., Loxton, R., Teo, K.L., Wu, Y.H.: A new computational method for optimizing nonlinear impulsive systems. Dyn. Contin. Discrete Impuls. Syst. Ser. B **18**, 59–76 (2011)

19. Loxton, R., Teo, K.L., Rehbock, V.: Computational method for a class of switched system optimal control problems. IEEE Trans. Automat. Contr. **54**, 2455–2460 (2009)

20. Lin, Q., Loxton, R., Teo, K.L.: The control parameterization method for nonlinear optimal control: A survey. J. Ind. Manag. Optim. Preprint.

21. Richard, J.P.: Time-delay systems: An overview of some recent advances and open problems. Automatica **29**, 1667–1694 (2003)

22. Meyer, C., Schroder, S., De Doncker, R.W.: Solid-state circuit breakers and current limiters for medium-voltage systems having distributed power systems. IEEE Trans. Power Electr. **19**, 1333–1340 (2004).

23. Kim, D.K., Park, P.G., Ko, J.W.: Output-feedback H_∞ control of systems over communication networks using a deterministic switching system approach. Automatica, **40**, 1205–1212 (2004)

24. Verriest, E.I., Delmotte, F., Egerstedt, M.: Optimal impulsive control of point delay systems with refractory period. In: Proceedings of the 5th IFAC Workshop on Time Delay Systems, Leuven, Belgium, September (2004)

25. Verriest, E.I.: Optimal control for switched point delay systems with refractory period. In: Proceedings of the 16th IFAC World Congress, Prague, Czech Republic, July (2005)

26. Delmotte, F., Verriest, E.I., Egestedt, M.: Optimal impulsive control of delay systems. ESAIM Control Optim. CA. **14**, 767–779 (2008)

27. Wu, C.Z., Teo, K.L., Li, R., Zhao, Y.: Optimal control of switched systems with time delay. Appl. Math. Lett. **19**, 1062–1067 (2006)

28. Li, R., Teo, K.L., Wong, K.H., Duan, G.R.: Control parameterization enhancing transform for optimal control of switched systems. Math. Comput. Model. **43**, 1393–1403 (2006)

29. Farhadinia, B., Teo, K.L., Loxton, R.: A computational method for a class of non-standard time optimal control problems involving multiple time horizons. Math. Comput. Model. **49**, 1682–1691 (2009)

30. Nocedal, J., Wright, S.J.: Numerical Optimization. Springer-Verlag, New York (1999)

31. Ahmed, N.U.: Elements of Finite-dimensional Systems and Control Theory. Longman Scientific and Technical, Essex (1988)

32. Teo, K.L., Goh, C.J., Wong, K.H.: A Unified Computational Approach to Optimal Control Problems. Longman Scientific and Technical, Essex (1991)

33. Schittkowski, K.: NLPQLP: A fortran implementation of a sequential quadratic programming algorithm with distributed and non-monotone line search-user's guid. University of Bayreuth, Bayreuth (2007)

34. Hindmarsh, A.C.: Large ordinary differential equation systems and software. IEEE Contr. Syst. Mag. **2**, 24–30 (1982)

35. Stoer, J., Bulirsch R.: Introduction to Numerical Analysis. Springer-Verlag, New York (1980)

36. Xiu, Z.L., Song, B.H., Sun, L.H., Zeng, A.P.: Theoretical analysis of effects of metabolic overflow and time delay on the performance and dynamic behavior of a two-stage fermentation process. Biochem. Eng. J. **11**, 101–109 (2002)

37. Liu, C.Y., Loxton, R., Teo, K.L.: Optimal parameter selection for nonlinear multistage systems with time-delays. submitted to Computational Optimization and Applications.

38. Mu, Y., Zhang, D.J., Teng, H., Wang, W., Xiu, Z.L.: Microbial production of 1,3-Propanediol by *Klebsiella pneumoniae* using crude glycerol from biodiesel preparation. Biotechnol. Lett. **28**, 1755–1759 (2006)