DYNAMIC OPTIMIZATION FOR SWITCHED TIME-DELAY SYSTEMS WITH STATE-DEPENDENT SWITCHING CONDITIONS*

CHONGYANG LIU†, RYAN LOXTON‡, QUN LIN‡, AND KOK LAY TEO‡

Abstract. This paper considers a dynamic optimization problem for a class of switched systems characterized by two key attributes: (i) the switching mechanism is invoked automatically when the state variables satisfy certain switching conditions; and (ii) the subsystem dynamics involve time-delays in the state variables. The decision variables in the problem, which must be selected optimally to minimize system cost, consist of a set of time-invariant system parameters in the initial state functions. To solve the dynamic optimization problem, we first show that the partial derivatives of the system state with respect to the system parameters can be expressed in terms of the solution of a set of variational switched systems. Then, on the basis of this result, we develop a gradient-based optimization algorithm to determine the optimal parameter values. Finally, we validate the proposed algorithm by solving an example problem arising in the production of 1,3-propanediol.

Key words. switched system, time-delay system, dynamic optimization, state-dependent switching, nonlinear optimization

AMS subject classifications. 34K34, 49M37, 65K10

1. Introduction.

A *switched system* operates by switching among different *subsystems* or *modes* [6, 12, 17, 32, 33]. Such systems are defined by a *switching law* that governs the order in which the modes operate (the *switching sequence*), and the times at which mode changes occur (the *switching times*). Switched systems are closely related to *impulsive systems*, which experience instantaneous changes in the state (instead of the mode) at certain time points [30].

Switched systems are optimized by changing some combination of the switching sequence, switching times, and other input parameters in the mode dynamics. The switching times and input parameters are normally continuous-valued and can be determined using *gradient-based optimization techniques* [18, 19, 21, 23]. In contrast, the switching sequence is a discrete variable that poses a much greater challenge for optimization. Since this paper focuses on systems with pre-fixed switching sequence, we simply mention that methods for optimizing the switching sequence are available—albeit computationally demanding—and direct the reader to [18] for a survey.

The key idea in applying gradient-based optimization methods is to parameterize the state—and therefore the cost and constraint functions—in terms of the switching times and input parameters. This essentially converts the dynamic optimization problem into a *nonlinear programming problem* that can be solved using existing methods [18, 19]. This approach requires the gradients of the cost and constraint functions, which cannot be expressed in closed form except in the rare case when the switched system has an analytical solution. More commonly, the cost and constraint gradients must be numerically evaluated using the *variational system*, which gives the partial derivatives of the state with respect to the controllable variables. The existence of the variational system has been established for many classes of switched and impulsive

*This work was supported by the Natural Science Foundation of China (grant 11771008), the Natural Science Foundation of Shandong Province, China (grants ZR2017MA005 and ZR2015AL010), and the Australian Research Council (grant DP140100289).

†School of Mathematics and Information Science, Shandong Institute of Business and Technology, Yantai, China (liu_chongyang@yahoo.com).

‡School of Electrical Engineering, Computing, and Mathematical Sciences, Curtin University, Perth, Australia (r.loxton@curtin.edu.au, q.lin@curtin.edu.au, k.l.teo@curtin.edu.au).

systems; see, for example, references [3, 10, 14, 21, 25, 26]. For these systems, the optimization procedure is normally applied in conjunction with a *time-scaling transformation* that maps the switching times into fixed points in a new time scale, thus yielding an equivalent optimization problem that is easier to solve [18, 25, 37].

The form of the variational system for a switched system depends critically on whether the switching conditions—which dictate when the system transitions from one mode to another—are *time-dependent* or *state-dependent*. Time-dependent switching conditions define precise times (either fixed or controllable) at which mode changes occur. State-dependent switching conditions instead define certain surfaces in the state space that, when hit by the state trajectory, trigger mode changes. Switched systems with state-dependent switching conditions are the most challenging because their switching times are unknown in advance and can only be generated implicitly as the system is being solved. Dynamic optimization problems for such systems have been considered in references [8, 27], which present formulas for evaluating the gradients of cost functions in Bolza and Lagrange form. The gradient derivations are based on the differentiability of the state trajectory with respect to the controllable parameters—an issue studied in [3, 11, 14, 35] for systems with state-triggered discrete events. As an alternative to tackling the state-dependent switching conditions directly (as in [8, 27]), several methods have been proposed for converting the state-dependent switching conditions into time-dependent switching conditions under additional assumptions and constraints [7, 23, 36]. These methods, however, are not for general use: reference [7] only considers dual-mode switched systems, reference [23] is focused on a specific applied problem, and reference [36] proposes a heuristic approach that is not precisely equivalent to solving the original problem with state-based switchings.

The solution differentiability results mentioned above for systems with state-dependent discrete events (either mode changes or state jumps) do not allow time-delays. To the best of our knowledge, the only known results on solution differentiability for systems with *both* delays and state-dependent discrete events are in [15]. These results extend the work in [16], which considers delays but no discrete events.

In this paper, we derive the variational system for a class of switched delay-differential equations in which the mode switches are defined by nonlinear equations of the state variables and the initial state functions depend on controllable parameters. This variational system matches the one derived in [15] under different technical conditions and using a different proof strategy (for a more general class of systems in which the time-delays may depend on the state). Specifically, the results in [15] rely on a boundedness condition for the right-hand side dynamics, whereas our new results are based on a less restrictive linear growth condition. As a consequence, our results apply over an infinite time horizon. The proof in [15] uses the *method of steps* to convert the system with time-delays into a sequence of ordinary differential equations without delays, after which the standard theory for ODEs can be applied. We instead use a different proof strategy that is based on the equivalent integral form of the switched system, with the time-delays maintained in all stages of the proof. As a byproduct, two additional results are obtained that supplement the results in [15]:

- Although the partial derivatives of the state with respect to the decision parameters usually do not exist at the switching times, the one-sided partial derivatives always exist; and
- The state is multi-variate continuous with respect to the decision parameters—this does not follow immediately from the results in [15], because the mere existence of partial derivatives is not sufficient for full multi-variate continuity.

The variational system derived in this paper (see Sections 3 and 4) underpins a gradient-based algorithm for optimizing the decision parameters in the initial state functions. Furthermore, as we discuss in the paper, our algorithm can be easily extended to switched systems that also contain decision parameters in the mode dynamics and switching conditions. Section 5 illustrates the application of our approach to the fed-batch fermentation process for 1,3-propanediol, an organic compound widely used in the production of paints, adhesives, and laminates [28].

2. Problem formulation. Consider the following switched time-delay system consisting of N modes operating in succession:

$$\dot{x}(t) = f^i(x(t), x(t-\gamma_1), \dots, x(t-\gamma_m)), \quad t \in (\tau_{i-1}, \tau_i), \quad i = 1, \dots, N, \tag{2.1a}$$

$$x(t) = \phi(t, \zeta), \quad t \le 0, \tag{2.1b}$$

where $x(t) \in R^n$ is the state; γ_j, $j = 1, \dots, m$, are given time-delays; $\zeta \in R^p$ is a vector of time-invariant system parameters; τ_i, $i = 1, \dots, N-1$, are switching times listed in increasing order, with $\tau_0 := 0$ and $\tau_N := \infty$; and $f^i : R^{(m+1)n} \to R^n$, $i = 1, \dots, N$, and $\phi : R \times R^p \to R^n$ are given continuously differentiable functions.

We assume that each f^i, $i = 1, \dots, N$, satisfies the linear growth condition

$$|f^i(w^0, \dots, w^m)| \le L(1 + |w^0| + \cdots + |w^m|), \quad (w^0, \dots, w^m) \in R^{(m+1)n}, \tag{2.2}$$

where $L > 0$ is a real constant and $|\cdot|$ denotes the Euclidean norm.

The switching times in (2.1a) specify when the system switches from one mode to another. Unlike in time-dependent switched systems (which are widespread in the literature), the switching times here are not independent decision variables, but are instead governed by a set of state-dependent switching conditions. More specifically, a switch from mode $i-1$ to mode i occurs when the state satisfies

$$g_i(x(t)) = 0, \tag{2.3}$$

where $g_i : R^n \to R$ is a given continuously differentiable function. Thus, the switched system (2.1) evolves as follows. After beginning in state $\phi(0, \zeta)$ at time $t = 0$, the system runs smoothly according to (2.1a) with $i = 1$ until $g_1(x(t)) = 0$. Then, the system switches into mode 2 and again runs smoothly according to (2.1a) with $i = 2$ until $g_2(x(t)) = 0$, and so on. This behaviour implies that the switching sequence—that is, the order in which the different modes in (2.1a) are activated—is fixed and known, which is the case in many practical systems, such as the dual-mode fed-batch fermentation system considered in Section 5 of this paper.

Equation (2.3) is essentially a stopping condition for the ith mode. Hence, given $\tau_{i-1} \ge 0$ for some $i \ge 1$, the next switching time τ_i is defined recursively by

$$\tau_i := \inf\{t > \tau_{i-1} : g_i(x(t)) = 0\}, \tag{2.4}$$

where $\tau_0 = 0$ and $\tau_i = \infty$ if $g_i(x(t)) \ne 0$ for all $t > \tau_{i-1}$.

System (2.1) is controlled by manipulating the time-invariant system parameters. Define

$$\mathcal{Z} := \{\zeta \in R^p : a_k \le \zeta_k \le b_k,\ k = 1, \dots, p\},$$

where a_k and b_k are given constants such that $a_k < b_k$. Any vector $\zeta \in \mathcal{Z}$ is called a *feasible parameter vector*. Although in (2.1) the system parameters only appear in the

initial function ϕ, the results in this paper can be easily extended to the case where the system parameters also appear explicitly in the right-hand side dynamics (2.1a) and/or the switching functions (2.3). Indeed, ζ can be replaced where necessary by an auxiliary state vector $\bar{x}(t)$ satisfying $\dot{\bar{x}}(t) = 0$ for $t > 0$ and $\bar{x}(t) = \zeta$ for $t \leq 0$.

We show in Section 4 that, under the stated conditions on the dynamic functions f^i, $i = 1, \ldots, N$, and the initial function ϕ, system (2.1) with switching times given by (2.4) has a unique solution $x(\cdot|\zeta)$ corresponding to each feasible parameter vector $\zeta \in \mathcal{Z}$. This solution satisfies the dynamics (2.1a) almost everywhere on $[0, \infty)$, and the initial condition (2.1b) everywhere on $(-\infty, 0]$.

Now, define a cost function J as follows:

$$J(\zeta) := \Phi(x(T|\zeta)), \tag{2.5}$$

where $T > 0$ is a given terminal time and $\Phi : R^n \to R$ is a given continuously differentiable function. We state our dynamic optimization problem as follows.

PROBLEM (P). *Find a feasible parameter vector $\zeta \in \mathcal{Z}$ such that the cost function (2.5) is minimized.*

Problem (P) is a unique problem involving two main difficulties: (i) the mode dynamics in the switched system (2.1) are influenced by multiple state-delays; and (ii) the mode switches are governed by a state-dependent switching mechanism. We now develop a gradient-based computational method for solving Problem (P).

3. Main results. Let e^k denote the kth unit basis vector in R^p. Furthermore, let

$$Z_N := \{1, \ldots, N\}, \quad Z_{N-1} := \{1, \ldots, N-1\}.$$

Consider the $n \times p$ *state variation matrix* $\partial x(t|\zeta)/\partial \zeta$ whose kth column is defined by

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

assuming the limit on the right-hand side exists. This limit certainly exists for systems governed by ordinary differential equations (see, for example, the results in [4] and other well-known references), but this is not necessarily the case for system (2.1). In fact, we will show that the state variation matrix for system (2.1) does not exist in some circumstances, even though the functions defining the system are smooth. This is due to the presence of time-delays and state-dependent switching conditions.

Notwithstanding, if the state variation matrix does exist at $t = T$, then the cost function (2.5) can be differentiated using the chain rule to obtain

$$\frac{\partial J(\zeta)}{\partial \zeta} = \frac{\partial \Phi(x(T|\zeta))}{\partial x} \frac{\partial x(T|\zeta)}{\partial \zeta}, \tag{3.1}$$

where $\partial J(\zeta)/\partial \zeta$ is a $1 \times p$ row vector whose kth element is the partial derivative of J with respect to the kth system parameter. This formula, if tractable, can be incorporated into well-known computational methods such as sequential quadratic programming to solve Problem (P) as a nonlinear optimization problem.

We now introduce two conditions on the parameter vector that are fundamental to evaluating equation (3.1).

CONDITION 3.1. *For a given $\zeta \in \mathcal{Z}$,*

$$g_i(x(\tau_{i-1}|\zeta)) \neq 0, \quad i \in Z_{N-1} : \tau_{i-1} < \infty.$$

This condition guarantees that consecutive switching conditions do not occur at the same time, thus ensuring the switching mechanism in (2.1) is well-defined and the switching times are distinct. To prove this formally, suppose to the contrary that Condition 3.1 holds and $\tau_{i-1} = \tau_i$ for some $i \in Z_{N-1}$. Then there exists a sequence $\{s_j\}_{j=1}^{\infty} \subset (\tau_{i-1}, \infty)$ such that $g_i(x(s_j|\zeta)) = 0$ for each integer j and $s_j \to \tau_{i-1}$ as $j \to \infty$. Hence, since $g_i(\cdot)$ is continuous,

$$g_i(x(\tau_{i-1}|\zeta)) = \lim_{j\to\infty} g_i(x(s_j|\zeta)) = 0,$$

which contradicts Condition 3.1. This shows that, for any parameter vector satisfying Condition 3.1, the corresponding switching times are distinct—specifically, $\tau_{i-1} < \tau_i$ for each integer $i \in Z_{N-1}$ with $\tau_{i-1} < \infty$. Condition 3.1 is always satisfied in systems that change mode when the state variables reach critical thresholds, such as the hang glider in [22] and the fed-batch fermentation process in Section 5.

CONDITION 3.2. *For a given $\zeta \in \mathcal{Z}$,*

$$\frac{\partial g_i(x(\tau_i|\zeta))}{\partial x} f^i(x(\tau_i|\zeta), x(\tau_i - \gamma_1|\zeta), \ldots, x(\tau_i - \gamma_m|\zeta)) \neq 0, \quad i \in Z_{N-1} : \tau_i < \infty.$$

Condition 3.2 simply states that the dot product of $\partial g_i/\partial x$ (which is orthogonal to the switching surface g_i) and f^i (which is tangent to the state trajectory) is non-zero at the ith switching time. This condition thus ensures that the state trajectory does not approach the switching surfaces at a tangent. Similar conditions are common in the literature—see, for example, [8, 9, 15, 20, 22, 27].

We are now ready to present formulas for both the state variation matrix and the derivatives of the switching times with respect to the parameter vector. Here and elsewhere, we will use the notation $\partial\tilde{x}^j$ to denote differentiation with respect to $x(t - \gamma_j)$, with $\partial\tilde{x}^0$ denoting differentiation with respect to $x(t)$ (that is, $\gamma_0 := 0$).

THEOREM 3.1. *Suppose $\zeta \in \mathcal{Z}$ satisfies Conditions 3.1 and 3.2. Then for each $k = 1, \ldots, p$,*

$$\frac{\partial x(t|\zeta)}{\partial \zeta_k} = \Lambda_k(t), \quad t \in (\tau_{i-1}, \tau_i), \quad i \in Z_N, \tag{3.2}$$

and

$$\frac{\partial \tau_i(\zeta)}{\partial \zeta_k} = -\frac{\partial g_i(x(\tau_i))}{\partial x}\Lambda_k(\tau_i-) \div \left\{ \frac{\partial g_i(x(\tau_i))}{\partial x} f^i(x(\tau_i), x(\tau_i - \gamma_1), \ldots, x(\tau_i - \gamma_m)) \right\}, \quad i \in Z_{N-1} : \tau_i < \infty, \tag{3.3}$$

where $\Lambda_k(\cdot)$ satisfies the following variational system:

$$\dot{\Lambda}_k(t) = \sum_{j=0}^{m} \frac{\partial f^i(x(t), x(t-\gamma_1), \ldots, x(t-\gamma_m))}{\partial \tilde{x}^j} \Lambda_k(t - \gamma_j), \ t \in (\tau_{i-1}, \tau_i), \ i \in Z_N, \tag{3.4a}$$

with initial conditions

$$\Lambda_k(t) = \frac{\partial \phi(t, \zeta)}{\partial \zeta_k}, \quad t \leq 0, \tag{3.4b}$$

$$\Lambda_k(0+) = \frac{\partial \phi(0, \zeta)}{\partial \zeta_k}, \tag{3.4c}$$

and intermediate jump conditions

$$\Lambda_k(\tau_i+) = \Lambda_k(\tau_i-) + \frac{\partial \tau_i(\zeta)}{\partial \zeta_k}\big\{f^i(x(\tau_i), x(\tau_i-\gamma_1), \ldots, x(\tau_i-\gamma_m)) - f^{i+1}(x(\tau_i), x(\tau_i-\gamma_1), \ldots, x(\tau_i-\gamma_m))\big\},\ i \in Z_{N-1} : \tau_i < \infty. \tag{3.4d}$$

A detailed proof of Theorem 3.1 is given in Sections 4.2-4.8. Note that Condition 3.2 ensures the denominator in (3.3) is non-zero.

The switching times $t = \tau_i$, $i \in Z_{N-1}$, are deliberately excluded from equation (3.2) in Theorem 3.1. This is because the state variation matrix only exists at the switching times in rare circumstances, as the next result shows.

THEOREM 3.2. *Suppose $\zeta \in \mathcal{Z}$ satisfies Conditions 3.1 and 3.2 and let $\Lambda_k(\cdot)$ denote the solution of (3.4) corresponding to ζ. Then for each $k = 1, \ldots, p$ and each $i \in Z_{N-1}$ with $\tau_i < \infty$, one of the following scenarios holds.*

- *If $f^i(x(\tau_i), x(\tau_i-\gamma_1), \ldots, x(\tau_i-\gamma_m)) = f^{i+1}(x(\tau_i), x(\tau_i-\gamma_1), \ldots, x(\tau_i-\gamma_m))$ or $\partial\tau_i(\zeta)/\partial\zeta_k = 0$, then*

$$\frac{\partial x(\tau_i|\zeta)}{\partial \zeta_k} = \Lambda_k(\tau_i+) = \Lambda_k(\tau_i-).$$

- *If $f^i(x(\tau_i), x(\tau_i-\gamma_1), \ldots, x(\tau_i-\gamma_m)) \neq f^{i+1}(x(\tau_i), x(\tau_i-\gamma_1), \ldots, x(\tau_i-\gamma_m))$ and $\partial\tau_i(\zeta)/\partial\zeta_k > 0$, then*

$$\frac{\partial^{\pm} x(\tau_i|\zeta)}{\partial \zeta_k} := \lim_{\epsilon \to 0\pm} \frac{x(\tau_i|\zeta+\epsilon e^k) - x(\tau_i|\zeta)}{\epsilon} = \Lambda_k(\tau_i\mp).$$

- *If $f^i(x(\tau_i), x(\tau_i-\gamma_1), \ldots, x(\tau_i-\gamma_m)) \neq f^{i+1}(x(\tau_i), x(\tau_i-\gamma_1), \ldots, x(\tau_i-\gamma_m))$ and $\partial\tau_i(\zeta)/\partial\zeta_k < 0$, then*

$$\frac{\partial^{\pm} x(\tau_i|\zeta)}{\partial \zeta_k} := \lim_{\epsilon \to 0\pm} \frac{x(\tau_i|\zeta+\epsilon e^k) - x(\tau_i|\zeta)}{\epsilon} = \Lambda_k(\tau_i\pm).$$

Theorem 3.2 is proved in Section 4.9.

In the first scenario of Theorem 3.2, the state variation exists at $t = \tau_i$. In the last two scenarios (the more likely scenarios), the state variation does not exist at $t = \tau_i$ because the left and right partial derivatives of the state with respect to the kth system parameter are different—a consequence of $\Lambda_k(\cdot)$ being discontinuous at the ith switching time.

We now use Theorems 3.1 and 3.2 to derive the left and right partial derivatives of the cost function J. First, Taylor's theorem implies that for each $\epsilon \neq 0$ and each $k = 1, \ldots, p$, there exists a constant $\eta_{\epsilon,k} \in (0,1)$ such that

$$J(\zeta+\epsilon e^k) - J(\zeta) = \frac{\partial \Phi((1-\eta_{\epsilon,k})x(T|\zeta) + \eta_{\epsilon,k} x(T|\zeta+\epsilon e^k))}{\partial x}\big\{x(T|\zeta+\epsilon e^k) - x(T|\zeta)\big\}.$$

Collectively, Theorems 3.1 and 3.2 show that the system state is differentiable from the left and right with respect to each system parameter, assuming Conditions 3.1 and 3.2

hold. This implies $x(T|\zeta+\epsilon e^k)\to x(T|\zeta)$ as $\epsilon\to 0\pm$. Thus, for each $k=1,\dots,p$,

$$\begin{aligned}(3.5)\qquad \frac{\partial^{\pm}J(\zeta)}{\partial\zeta_k} &:= \lim_{\epsilon\to 0\pm}\frac{J(\zeta+\epsilon e^k)-J(\zeta)}{\epsilon}\\ &= \frac{\partial\Phi(x(T|\zeta))}{\partial x}\lim_{\epsilon\to 0\pm}\frac{x(T|\zeta+\epsilon e^k)-x(T|\zeta)}{\epsilon}\\ &= \frac{\partial\Phi(x(T|\zeta))}{\partial x}\begin{cases}\Lambda_k(T|\zeta), & \text{if } \tau_i(\zeta)\neq T,\ i=1,\dots,N-1,\\ \Lambda_k(T\mp|\zeta), & \text{if } \tau_i(\zeta)=T \text{ and } \partial\tau_i(\zeta)/\partial\zeta_k\geq 0,\\ \Lambda_k(T\pm|\zeta), & \text{if } \tau_i(\zeta)=T \text{ and } \partial\tau_i(\zeta)/\partial\zeta_k\leq 0,\end{cases}\end{aligned}$$

where $\Lambda_k(\cdot|\zeta)$ is the solution of (3.4) corresponding to ζ.

Equation (3.5) shows that the left and right partial derivatives of J exist at all $\zeta\in\mathcal{Z}$ satisfying Conditions 3.1 and 3.2. In practice, these conditions can be easily checked for a given $\zeta\in\mathcal{Z}$ by numerically solving the switched time-delay system (2.1). Note that if T coincides with a switching time satisfying one of the last two scenarios in Theorem 3.2, then the left and right partial derivatives of J with respect to ζ_k may differ, since in this case $\Lambda_k(T-)\neq\Lambda_k(T+)$.

Since J has well-defined left and right partial derivatives (under Conditions 3.1 and 3.2), it is clearly individually continuous with respect to each parameter. In fact, as we show in Section 4.10, Conditions 3.1 and 3.2 imply full continuity of J, and thus if these conditions hold at every point in the compact set $\mathcal{Z}$, then Problem (P) is guaranteed to admit an optimal solution. This result is summarized below.

THEOREM 3.3. *Suppose Conditions 3.1 and 3.2 hold at every point $\zeta\in\mathcal{Z}$. Then Problem (P) admits an optimal solution.*

The left and right partial derivatives of J, as defined in (3.5), can be used to identify search directions for optimizing the parameter vector. Indeed, if $\partial^+J(\zeta)/\partial\zeta_k<0$, then e^k is a descent direction of J at ζ , and if $\partial^-J(\zeta)/\partial\zeta_k>0$, then $-e^k$ is a descent direction of J at ζ. Performing a line search along a descent direction will yield an improved point with lower cost [29].

If none of the switching times coincide with the terminal time, or if the conditions for the first scenario in Theorem 3.2 are satisfied at the terminal time, then the left and right partial derivatives of J derived above become the full partial derivatives as shown in (3.1). We now present the following line search optimization algorithm for solving Problem (P) to local optimality.

1. Choose an initial point $\zeta\in\mathcal{Z}$.
2. Form an expanded switched time-delay system by combining the state system (2.1) with the variational system (3.4) for each $k=1,\dots,p$.
3. Solve the expanded system mode by mode, checking Conditions 3.1 and 3.2 at the start and end of each mode. If these conditions are violated at any stage, then stop with error.
4. Use $x(\cdot|\zeta)$ and $\Lambda_k(\cdot|\zeta)$, $k=1,\dots,p$, to determine the left and right partial derivatives of J according to equation (3.5).
5. Use $\partial^{\pm}J(\zeta)/\partial\zeta_k$, $k=1,\dots,p$, to check local optimality conditions at ζ. If the local optimality conditions hold, then stop; otherwise, continue to Step 6.
6. Use $\partial^{\pm}J(\zeta)/\partial\zeta_k$, $k=1,\dots,p$, to define a search direction.
7. Perform a line search along the direction from Step 6 to determine a new point $\zeta'\in\mathcal{Z}$.
8. Set $\zeta'\to\zeta$ and return to Step 2.

In most cases, the partial derivatives of J will exist and Steps 5-7 can be implemented using well-known methods in nonlinear optimization [29]. For example, the negative gradient can be used as the search direction. We refer readers to [19, 34] for a detailed discussion on solving optimal control and dynamic optimization problems using this approach. If any of the full partial derivatives of J do not exist—that is, a mode switch satisfying one of the last two scenarios in Theorem 3.2 occurs at the terminal time—then the signs of the left and right partial derivatives can be used to identify an appropriate descent direction along one of the coordinate axes.

4. Proofs of the main results. This section proves four key results stated earlier: existence and uniqueness of $x(\cdot|\zeta)$ (Section 4.1), Theorem 3.1 on the state variation matrix (Sections 4.2-4.8), Theorem 3.2 on the state variation matrix at the switching times (Section 4.9), and continuity of the cost function J (Section 4.10).

4.1. Solution existence and uniqueness for (2.1). For a given $\zeta \in \mathcal{Z}$, define a set of auxiliary systems recursively as follows:

$$\dot{z}^i(t) = f^i(z^i(t), z^i(t-\gamma_1), \ldots, z^i(t-\gamma_m)), \quad t > \rho_{i-1}, \tag{4.1a}$$

$$z^i(t) = z^{i-1}(t), \quad t \leq \rho_{i-1}, \tag{4.1b}$$

and

$$\rho_i = \begin{cases} \inf\{\, t > \rho_{i-1} : g_i(z^i(t)) = 0 \,\}, & \text{if } i \leq N-1, \\ \infty, & \text{if } i = N, \end{cases} \tag{4.2}$$

where $z^0(t) := \phi(t, \zeta)$ and $\rho_0 := 0$. Given ρ_{i-1} and $z^{i-1}(\cdot)$ for some $i \in Z_N$, existence of a unique solution to (4.1) can be deduced by dividing $[\rho_{i-1}, \infty)$ into consecutive subintervals of length $\min\{\gamma_1, \ldots, \gamma_m\}$, and then applying known existence and uniqueness results for non-delay systems (see [1, 2]) on each subinterval. The continuity and differentiability assumptions on ϕ and f^i, $i = 1, \ldots, N$, and the linear growth condition (2.2), ensure that the results in [1, 2] are applicable here. Using $z^0(t) = \phi(t, \zeta)$ and $\rho_0 = 0$ as the base case, it follows from induction that $z^i(\cdot)$ and ρ_i are well-defined for each $i \in Z_N$; in fact, $z^N(\cdot)$ satisfies (2.1) with $\rho_i = \tau_i$, $i = 0, \ldots, N$, and $z^N(\cdot)$ is the only solution of (2.1) because each $z^i(\cdot)$ is unique.

4.2. Preliminaries for Theorem 3.1. Let $\zeta \in \mathcal{Z}$ and $k \in \{1, \ldots, p\}$ be arbitrary and consider the perturbed parameter vector $\zeta + \epsilon e^k$, where $\epsilon \in [a_k - \zeta_k, b_k - \zeta_k]$ to ensure that $\zeta + \epsilon e^k \in \mathcal{Z}$. Let $z^{i,\epsilon}(\cdot)$, $i \in Z_N$, denote the trajectories obtained by solving (4.1) recursively for parameter vector $\zeta + \epsilon e^k$, starting with $z^{0,\epsilon}(t) = \phi(t, \zeta + \epsilon e^k)$ and $\rho_0 = 0$. Based on the arguments in Section 4.1, ρ_i defined by (4.2) for $\zeta + \epsilon e^k$ is equal to $\tau_i^\epsilon := \tau_i(\zeta + \epsilon e^k)$, and $z^{i,\epsilon}(t) = x(t|\zeta + \epsilon e^k)$ for all $t \leq \tau_i^\epsilon$.

For any function ψ of ϵ, we write $\psi(\epsilon) = O(\epsilon^d)$ if there exists a real number $M > 0$ such that $|\psi(\epsilon)| \leq M|\epsilon|^d$ for all ϵ of sufficiently small magnitude, and $\psi(\epsilon) = \theta(\epsilon)$ if $\psi(\epsilon) \to 0$ as $\epsilon \to 0$. In particular, $\psi(\epsilon) = O(1)$ means that ψ is uniformly bounded with respect to ϵ.

Let $\bar{\gamma} := \max\{\gamma_1, \ldots, \gamma_m\}$ and

$$\mu^{i,\epsilon}(t) := z^{i,\epsilon}(t) - z^{i,0}(t), \quad i = 0, \ldots, N.$$

We will show by induction that, for each integer i,

$$\max_{t\in[-\bar\gamma,T_{\max}]}\left|z^{i,\epsilon}(t)\right| = O(1) \text{ for every } T_{\max}>0, \tag{4.3}$$

$$\max_{t\in[-\bar\gamma,T_{\max}]}\left|\mu^{i,\epsilon}(t)\right| = O(\epsilon) \text{ for every } T_{\max}>0, \tag{4.4}$$

$$\lim_{\epsilon\to 0}\epsilon^{-1}\mu^{i,\epsilon}(t) = \Lambda_k(t-), \quad t\in(-\infty,\tau_i^0]\setminus\bigcup_{l=0}^{i-1}\{\tau_l^0\}, \tag{4.5}$$

$$\lim_{\epsilon\to 0}\tau_i^\epsilon = \tau_i^0, \tag{4.6}$$

and, when τ_i^0 is finite,

$$\lim_{\epsilon\to 0}\frac{\tau_i^\epsilon-\tau_i^0}{\epsilon} = \begin{cases} 0, & \text{if } i=0,\\ -\dfrac{\partial g_i(z^{i,0}(\tau_i^0))}{\partial x}\Lambda_k(\tau_i^0-) \\ \quad \div\left\{\dfrac{\partial g_i(z^{i,0}(\tau_i^0))}{\partial x}f^i(z^{i,0}(\tau_i^0),z^{i,0}(\tau_i^0-\gamma_1),\dots,z^{i,0}(\tau_i^0-\gamma_m))\right\}, & \text{if } i\ge 1. \end{cases} \tag{4.7}$$

Equation (4.3) means that the solution of (4.1) for $\zeta+\epsilon e^k$ is uniformly bounded with respect to ϵ, and equations (4.4) and (4.5) mean that the solution is continuous and differentiable at ζ with respect to the kth system parameter. Equations (4.6) and (4.7) mean that the switching times are continuous and differentiable at ζ with respect to the kth system parameter.

If equations (4.5) and (4.6) hold, then given any $t\in(\tau_{i-1},\tau_i)$, $i\in Z_N$, we have $t<\tau_i^\epsilon$ for all ϵ of sufficiently small magnitude, and hence $x(t|\zeta+\epsilon e^k)=z^{i,\epsilon}(t)$, implying

$$\frac{\partial x(t|\zeta)}{\partial\zeta_k} = \lim_{\epsilon\to 0}\frac{x(t|\zeta+\epsilon e^k)-x(t|\zeta)}{\epsilon} = \lim_{\epsilon\to 0}\epsilon^{-1}\mu^{i,\epsilon}(t) = \Lambda_k(t-) = \Lambda_k(t),$$

which proves equation (3.2). In addition, for each $i\in Z_{N-1}$ with τ_i finite, if (4.7) holds, then since $z^{i,0}(\tau_i^0-\gamma_j)=x(\tau_i-\gamma_j|\zeta)$, $j=0,\dots,m$,

$$\begin{aligned}\frac{\partial\tau_i(\zeta)}{\partial\zeta_k} &= \lim_{\epsilon\to 0}\frac{\tau_i^\epsilon-\tau_i^0}{\epsilon}\\ &= -\frac{\partial g_i(z^{i,0}(\tau_i^0))}{\partial x}\Lambda_k(\tau_i^0-)\div\left\{\frac{\partial g_i(z^{i,0}(\tau_i^0))}{\partial x}f^i(z^{i,0}(\tau_i^0),z^{i,0}(\tau_i^0-\gamma_1),\dots,z^{i,0}(\tau_i^0-\gamma_m))\right\}\\ &= -\frac{\partial g_i(x(\tau_i))}{\partial x}\Lambda_k(\tau_i-)\div\left\{\frac{\partial g_i(x(\tau_i))}{\partial x}f^i(x(\tau_i),x(\tau_i-\gamma_1),\dots,x(\tau_i-\gamma_m))\right\},\end{aligned}$$

which proves equation (3.3). Theorem 3.1 thus follows immediately from equations (4.5)-(4.7), which must be proved in conjunction with equations (4.3) and (4.4).

The basis step $i=0$ for (4.3)-(4.7) is easily established. Indeed, equations (4.3) and (4.5)-(4.7) for $i=0$ are a direct consequence of $\tau_0^\epsilon=0$ and $z^{0,\epsilon}(t)=\phi(t,\zeta+\epsilon e^k)$. For equation (4.4), since ϕ is continuously differentiable on $[-\bar\gamma,T_{\max}]\times\mathcal{Z}$,

$$\begin{aligned}\max_{t\in[-\bar\gamma,T_{\max}]}\left|\mu^{0,\epsilon}(t)\right| &= \max_{t\in[-\bar\gamma,T_{\max}]}\left|\phi(t,\zeta+\epsilon e^k)-\phi(t,\zeta)\right|\\ &\le |\epsilon|\max_{t\in[-\bar\gamma,T_{\max}]}\int_0^1\left|\frac{\partial\phi(t,\zeta+\epsilon\eta e^k)}{\partial\zeta_k}\right|d\eta = O(\epsilon).\end{aligned}$$

Now, for our inductive hypothesis, suppose that equations (4.3)-(4.7) hold for each $i = 1, \ldots, q$, where $q \leq N-1$. If $\tau_q^0 = \infty$, then clearly (4.3)-(4.6) also hold for $i = q+1$ (with (4.7) irrelevant) because $z^{q+1,\epsilon}(t) = z^{q,\epsilon}(t)$ when $|\epsilon|$ is small. Thus, it suffices to prove (4.3)-(4.7) for $i = q+1$ when $\tau_q^0 < \infty$. This is done in Sections 4.3-4.8.

4.3. Inductive step for equation (4.3). For each $i = 0, \ldots, q+1$, define

$$\hat{f}^{i,\epsilon}(s,\eta) := \begin{cases} f^i(z^{i,0}(s-\gamma_0) + \eta\mu^{i,\epsilon}(s-\gamma_0), \ldots, z^{i,0}(s-\gamma_m) + \eta\mu^{i,\epsilon}(s-\gamma_m)), & \text{if } i \geq 1, \\ \dot{\phi}(s, \zeta + \epsilon\eta e^k), & \text{if } i = 0, \end{cases}$$

and, for $i \geq 1$, let $\partial\hat{f}^{i,\epsilon}(s,\eta)/\partial\tilde{x}^j$ denote the respective partial derivatives.

From (4.1) for $i = q+1$, we have

$$z^{q+1,\epsilon}(t) = \begin{cases} z^{q,\epsilon}(\tau_q^\epsilon) + \displaystyle\int_{\tau_q^\epsilon}^t \hat{f}^{q+1,\epsilon}(s,1)ds, & \text{if } t \in [\tau_q^\epsilon, T_{\max}], \\ z^{q,\epsilon}(t), & \text{if } t \in [-\bar{\gamma}, \tau_q^\epsilon]. \end{cases}$$

Thus, for $t \in [-\bar{\gamma}, \tau_q^\epsilon]$,

$$\big|z^{q+1,\epsilon}(t)\big| = \big|z^{q,\epsilon}(t)\big| \leq \max_{s\in[-\bar{\gamma},T_{\max}]} \big|z^{q,\epsilon}(s)\big|, \tag{4.8}$$

and for $t \in [\tau_q^\epsilon, T_{\max}]$, using the linear growth condition (2.2),

$$\begin{aligned} \big|z^{q+1,\epsilon}(t)\big| &\leq \big|z^{q,\epsilon}(\tau_q^\epsilon)\big| + \int_{\tau_q^\epsilon}^t \big|\hat{f}^{q+1,\epsilon}(s,1)\big|ds \\ &\leq \max_{s\in[-\bar{\gamma},T_{\max}]} \big|z^{q,\epsilon}(s)\big| + LT_{\max} + \sum_{j=0}^m \int_{\tau_q^\epsilon}^t L\big|z^{q+1,\epsilon}(s-\gamma_j)\big|ds \\ &\leq \max_{s\in[-\bar{\gamma},T_{\max}]} \big|z^{q,\epsilon}(s)\big| + LT_{\max} + \int_{-\bar{\gamma}}^t (m+1)L\big|z^{q+1,\epsilon}(s)\big|ds. \end{aligned} \tag{4.9}$$

Combining (4.8) and (4.9), it follows from (4.3) for $i = q$ that

$$\big|z^{q+1,\epsilon}(t)\big| \leq O(1) + \int_{-\bar{\gamma}}^t (m+1)L\big|z^{q+1,\epsilon}(s)\big|ds, \quad t \in [-\bar{\gamma}, T_{\max}].$$

Finally, from the Gronwall inequality [2],

$$\big|z^{q+1,\epsilon}(t)\big| \leq O(1)\exp(L(m+1)(T_{\max}+\bar{\gamma})) = O(1), \quad t \in [-\bar{\gamma}, T_{\max}],$$

which proves equation (4.3) for $i = q+1$.

4.4. Inductive step for equation (4.4). It follows from equation (4.7) for $i = q$ that

$$\big|\tau_q^\epsilon - \tau_q^0\big| = |\epsilon| \cdot \left| \frac{\tau_q^\epsilon - \tau_q^0}{\epsilon} - \lim_{\epsilon\to 0} \frac{\tau_q^\epsilon - \tau_q^0}{\epsilon} + \lim_{\epsilon\to 0} \frac{\tau_q^\epsilon - \tau_q^0}{\epsilon} \right| = O(\epsilon). \tag{4.10}$$

There are four cases to consider for $t \in [-\bar{\gamma}, T_{\max}]$:

(i) $t < \min(\tau_q^\epsilon, \tau_q^0)$ (ii) $\tau_q^0 \leq t < \tau_q^\epsilon$ (iii) $\tau_q^\epsilon \leq t < \tau_q^0$ (iv) $t \geq \max(\tau_q^\epsilon, \tau_q^0)$.

Using (4.1) and the fundamental theorem of calculus, we can derive the following formula for $\mu^{q+1,\epsilon}(t) = z^{q+1,\epsilon}(t) - z^{q+1,0}(t)$ that covers all four cases:

$$\begin{aligned}\mu^{q+1,\epsilon}(t) = \mu^{q,\epsilon}(t) &+ \alpha_{qt}\int_{\tau_q^\epsilon}^t \big\{\hat{f}^{q+1,\epsilon}(s,1) - \hat{f}^{q,\epsilon}(s,1)\big\}ds \\ &+ \beta_{qt}\int_{\tau_q^0}^t \big\{\hat{f}^{q,0}(s,0) - \hat{f}^{q+1,0}(s,0)\big\}ds,\end{aligned} \tag{4.11}$$

where α_{qt} and β_{qt} are binary parameters indicating whether $t \geq \tau_q^\epsilon$ and $t \geq \tau_q^0$, respectively.

If at most one of $t \geq \tau_q^\epsilon$ and $t \geq \tau_q^0$ holds (cases (i)-(iii) above), then from (4.11),

$$\begin{aligned}\big|\mu^{q+1,\epsilon}(t)\big| &\leq \big|\mu^{q,\epsilon}(t)\big| + (f_{\max}^q + f_{\max}^{q+1})(\max(\tau_q^\epsilon, \tau_q^0) - \min(\tau_q^\epsilon, \tau_q^0)) \\ &= \big|\mu^{q,\epsilon}(t)\big| + (f_{\max}^q + f_{\max}^{q+1})\big|\tau_q^\epsilon - \tau_q^0\big|,\end{aligned} \tag{4.12}$$

where $f_{\max}^q$ and $f_{\max}^{q+1}$ are upper bounds for the norms of $\hat{f}^{q,\epsilon}(s,\eta)$ and $\hat{f}^{q+1,\epsilon}(s,\eta)$, respectively. These upper bounds exist because, by virtue of the inductive hypothesis and Section 4.3, both $z^{q,\epsilon}(\cdot)$ and $z^{q+1,\epsilon}(\cdot)$ are uniformly bounded on $[-\bar{\gamma}, T_{\max}]$ with respect to ϵ, and f^q and f^{q+1} are continuous functions.

On the other hand, if both $t \geq \tau_q^\epsilon$ and $t \geq \tau_q^0$ (case (iv) above), then (4.11) becomes

$$\begin{aligned}\mu^{q+1,\epsilon}(t) &= \mu^{q,\epsilon}(t) + \int_{\tau_q^\epsilon}^t \big\{\hat{f}^{q+1,\epsilon}(s,1) - \hat{f}^{q,\epsilon}(s,1)\big\}ds + \int_{\tau_q^0}^t \big\{\hat{f}^{q,0}(s,0) - \hat{f}^{q+1,0}(s,0)\big\}ds \\ &= z^{q,\epsilon}(\tau_q^\epsilon) - z^{q,0}(\tau_q^0) + \int_{\tau_q^\epsilon}^t \hat{f}^{q+1,\epsilon}(s,1)ds - \int_{\tau_q^0}^t \hat{f}^{q+1,0}(s,0)ds \\ &= \mu^{q,\epsilon}(\tau_q^0) + \int_{\tau_q^0}^{\tau_q^\epsilon} \big\{\hat{f}^{q,\epsilon}(s,1) - \hat{f}^{q+1,\epsilon}(s,1)\big\}ds + \int_{\tau_q^0}^t \big\{\hat{f}^{q+1,\epsilon}(s,1) - \hat{f}^{q+1,0}(s,0)\big\}ds,\end{aligned}$$

provided that $\tau_{q-1}^\epsilon < \tau_q^0$ when $q \geq 1$. Thus, using the mean value theorem,

$$\begin{aligned}\mu^{q+1,\epsilon}(t) = \mu^{q,\epsilon}(\tau_q^0) &+ \int_{\tau_q^0}^{\tau_q^\epsilon} \big\{\hat{f}^{q,\epsilon}(s,1) - \hat{f}^{q+1,\epsilon}(s,1)\big\}ds \\ &+ \sum_{j=0}^m \int_{\tau_q^0}^t \int_0^1 \frac{\partial \hat{f}^{q+1,\epsilon}(s,\eta)}{\partial \tilde{x}^j}\mu^{q+1,\epsilon}(s-\gamma_j)d\eta ds.\end{aligned} \tag{4.13}$$

Taking the norm of both sides gives

$$\begin{aligned}\big|\mu^{q+1,\epsilon}(t)\big| \leq \big|\mu^{q,\epsilon}(\tau_q^0)\big| &+ (f_{\max}^q + f_{\max}^{q+1})\big|\tau_q^\epsilon - \tau_q^0\big| \\ &+ \sum_{j=0}^m \int_{\tau_q^0}^t \partial f_{\max}^{q+1}\big|\mu^{q+1,\epsilon}(s-\gamma_j)\big|ds,\end{aligned} \tag{4.14}$$

where $\partial f_{\max}^{q+1}$ is an upper bound for the norm of $\partial \hat{f}^{q+1,\epsilon}(s,\eta)/\partial \tilde{x}^j$ (again, an upper bound exists because $z^{q+1,\epsilon}(\cdot)$ is uniformly bounded and f^{q+1} is continuously differentiable). Recall that this inequality requires $\tau_{q-1}^\epsilon < \tau_q^0$ when $q \geq 1$. By the inductive hypothesis and Condition 3.1,

$$\lim_{\epsilon \to 0} \tau_{q-1}^\epsilon = \tau_{q-1}^0 = \tau_{q-1}(\zeta) < \tau_q(\zeta) = \tau_q^0.$$

Hence, when ϵ is of sufficiently small magnitude, $\tau_{q-1}^{\epsilon} < \tau_q^0$ and so inequality (4.14) is valid.

Now, combining (4.12) and (4.14) and shifting the time variable in the integral gives, for all $t \in [-\bar{\gamma}, T_{\max}]$,

$$\big|\mu^{q+1,\epsilon}(t)\big| \leq \max_{s\in[-\bar{\gamma},T_{\max}]} \big|\mu^{q,\epsilon}(s)\big| + (f_{\max}^q + f_{\max}^{q+1})\big|\tau_q^{\epsilon} - \tau_q^0\big| + \int_{-\bar{\gamma}}^{t} (m+1)\partial f_{\max}^{q+1}\big|\mu^{q+1,\epsilon}(s)\big|ds.$$

Therefore, since $\mu^{q,\epsilon}(s) = O(\epsilon)$ and $\tau_q^{\epsilon} - \tau_q^0 = O(\epsilon)$ from (4.4) and (4.10), respectively,

$$\big|\mu^{q+1,\epsilon}(t)\big| \leq O(\epsilon) + \int_{-\bar{\gamma}}^{t} (m+1)\partial f_{\max}^{q+1}\big|\mu^{q+1,\epsilon}(s)\big|ds, \quad t \in [-\bar{\gamma}, T_{\max}].$$

Finally, applying the Gronwall inequality [2] gives

$$\big|\mu^{q+1,\epsilon}(t)\big| \leq O(\epsilon)\exp((m+1)(T_{\max}+\bar{\gamma})\partial f_{\max}^{q+1}) = O(\epsilon), \quad t \in [-\bar{\gamma}, T_{\max}],$$

from which we immediately infer equation (4.4) for $i = q+1$.

4.5. Some auxiliary results. We now prove some auxiliary results that are needed in the next subsection. We have already shown in Sections 4.3 and 4.4 that for any $T_{\max} > 0$, $z^{q+1,\epsilon}(\cdot)$ is uniformly bounded on $[-\bar{\gamma}, T_{\max}]$ with respect to ϵ, and $z^{q+1,\epsilon}(\cdot) \to z^{q+1,0}(\cdot)$ uniformly on $[-\bar{\gamma}, T_{\max}]$ as $\epsilon \to 0$. Thus, since f^{q+1} is a continuously differentiable function, the following limit holds uniformly with respect to $s \in [0, T_{\max}]$ and $\eta \in [0,1]$:

$$\lim_{\epsilon\to 0} \frac{\partial \hat{f}^{q+1,\epsilon}(s,\eta)}{\partial \tilde{x}^j} = \frac{\partial \hat{f}^{q+1,0}(s,0)}{\partial \tilde{x}^j}, \quad j = 0,\ldots,m.$$

This implies that, for any $T_{\max} > 0$,

$$\max_{s\in[0,T_{\max}]} \int_0^1 \left| \frac{\partial \hat{f}^{q+1,\epsilon}(s,\eta)}{\partial \tilde{x}^j} - \frac{\partial \hat{f}^{q+1,0}(s,0)}{\partial \tilde{x}^j} \right| d\eta = \theta(\epsilon), \quad j = 0,\ldots,m. \tag{4.15}$$

Now, for $s \in [0, T_{\max}]$ and $q \geq 1$, by the mean value theorem,

$$\begin{aligned} \big|\hat{f}^{q,\epsilon}(s,1) - \hat{f}^{q,0}(s,0)\big| &\leq \sum_{j=0}^{m} \int_0^1 \left|\frac{\partial \hat{f}^{q,\epsilon}(s,\eta)}{\partial \tilde{x}^j}\right| \cdot \big|\mu^{q,\epsilon}(s-\gamma_j)\big| d\eta \\ &\leq (m+1)\partial f_{\max}^q O(\epsilon) = O(\epsilon), \end{aligned} \tag{4.16}$$

where $\partial f_{\max}^q$ is an upper bound for the norm of $\partial \hat{f}^{q,\epsilon}(s,\eta)/\partial \tilde{x}^j$. Furthermore, for $s', s'' \in [0, T_{\max}]$ and $q \geq 1$,

$$\begin{aligned} \big|\hat{f}^{q,0}(s'',0) - \hat{f}^{q,0}(s',0)\big| &\leq \sum_{j=0}^{m} \int_{s'}^{s''} \left|\frac{\partial \hat{f}^{q,0}(s,0)}{\partial \tilde{x}^j}\right| \cdot \big|\dot{z}^{q,0}(s-\gamma_j)\big| ds \\ &\leq (m+1)\partial f_{\max}^q \max_{i=0,\ldots,q} f_{\max}^i \big|s'' - s'\big| \\ &= O(1)\big|s'' - s'\big|, \end{aligned} \tag{4.17}$$

where, for each $i = 0,\ldots,q$, the upper bound $f_{\max}^i$ for the norm of $\hat{f}^{i,\epsilon}(s,\eta)$ exists because $z^{i,\epsilon}(\cdot)$ is uniformly bounded and f^i and ϕ are continuous functions. Choosing

$T_{\max} > \max(\tau_q^\epsilon, \tau_q^0)$ and using (4.10), (4.16) and (4.17) gives

$$\begin{aligned}(4.18)\qquad &\int_{\min(\tau_q^\epsilon,\tau_q^0)}^{\max(\tau_q^\epsilon,\tau_q^0)} \big|\hat{f}^{q,\epsilon}(s,1) - \hat{f}^{q,0}(\tau_q^0,0)\big| ds \\ &\le \int_{\min(\tau_q^\epsilon,\tau_q^0)}^{\max(\tau_q^\epsilon,\tau_q^0)} \Big\{\big|\hat{f}^{q,\epsilon}(s,1) - \hat{f}^{q,0}(s,0)\big| + \big|\hat{f}^{q,0}(s,0) - \hat{f}^{q,0}(\tau_q^0,0)\big|\Big\} ds \\ &\le \Big\{O(\epsilon) + O(1)\big|\tau_q^\epsilon - \tau_q^0\big|\Big\}\big|\tau_q^\epsilon - \tau_q^0\big| = O(\epsilon^2).\end{aligned}$$

Similarly, by an almost identical proof,

$$(4.19)\qquad \int_{\min(\tau_q^\epsilon,\tau_q^0)}^{\max(\tau_q^\epsilon,\tau_q^0)} \big|\hat{f}^{q+1,\epsilon}(s,1) - \hat{f}^{q+1,0}(\tau_q^0,0)\big| ds = O(\epsilon^2).$$

Finally, for our last auxiliary result,

$$\begin{aligned}\int_{-\bar{\gamma}}^{\max(\tau_q^\epsilon,\tau_q^0)} \big|\epsilon^{-1}\mu^{q+1,\epsilon}(s) - \Lambda_k(s)\big| ds &\le \int_{-\bar{\gamma}}^{\tau_q^0} \big|\epsilon^{-1}\mu^{q+1,\epsilon}(s) - \Lambda_k(s)\big| ds \\ &\quad + \int_{\tau_q^0}^{\max(\tau_q^\epsilon,\tau_q^0)} \big|\epsilon^{-1}\mu^{q+1,\epsilon}(s) - \Lambda_k(s)\big| ds.\end{aligned}$$

Choosing $T_{\max} > \max(\tau_q^\epsilon, \tau_q^0)$, we know from Section 4.4 that $\epsilon^{-1}\mu^{q+1,\epsilon}(\cdot) = O(1)$ uniformly on $[-\bar{\gamma}, \max(\tau_q^\epsilon, \tau_q^0)]$. Moreover, by virtue of (4.5) and (4.6) for $i = q$, for almost every $s \in [-\bar{\gamma}, \tau_q^0)$, we have $\epsilon^{-1}\mu^{q+1,\epsilon}(s) = \epsilon^{-1}\mu^{q,\epsilon}(s) \to \Lambda_k(s)$ as $\epsilon \to 0$. Hence, by the Lebesgue dominated convergence theorem [5], the first integral on the right-hand side above converges to zero as $\epsilon \to 0$, and by (4.10), the second integral is of order $O(\epsilon)$:

$$\begin{aligned}(4.20)\qquad &\int_{-\bar{\gamma}}^{\max(\tau_q^\epsilon,\tau_q^0)} \big|\epsilon^{-1}\mu^{q+1,\epsilon}(s) - \Lambda_k(s)\big| ds \\ &\le \theta(\epsilon) + \Big\{O(1) + \max_{s\in[-\bar{\gamma},T_{\max}]} \big|\Lambda_k(s)\big|\Big\} \cdot \big|\tau_q^\epsilon - \tau_q^0\big| = \theta(\epsilon) + O(\epsilon) = \theta(\epsilon).\end{aligned}$$

Identities (4.15) and (4.18)-(4.20) are used to prove equation (4.5) for $i = q+1$ in the next subsection.

4.6. Inductive step for equation (4.5). If $t < \tau_q^0$, then (4.6) for $i = q$ implies that $t < \tau_q^\epsilon$ for all ϵ of sufficiently small magnitude and thus $z^{q+1,\epsilon}(t) = z^{q,\epsilon}(t)$. Hence, if $t \ne \tau_l^0$, $l = 0, \ldots, q-1$, then (4.5) for $i = q$ implies

$$\lim_{\epsilon\to 0} \frac{z^{q+1,\epsilon}(t) - z^{q+1,0}(t)}{\epsilon} = \lim_{\epsilon\to 0} \frac{z^{q,\epsilon}(t) - z^{q,0}(t)}{\epsilon} = \lim_{\epsilon\to 0} \epsilon^{-1}\mu^{q,\epsilon}(t) = \Lambda_k(t-),$$

which proves (4.5) for $i = q+1$ in the case $t < \tau_q^0$. We now consider the case $t > \tau_q^0$.

From (4.13), for all $t \ge \max(\tau_q^\epsilon, \tau_q^0)$,

$$\begin{aligned}\mu^{q+1,\epsilon}(t) = \mu^{q,\epsilon}(\tau_q^0) &+ \int_{\tau_q^0}^{\tau_q^\epsilon} \big\{\hat{f}^{q,\epsilon}(s,1) - \hat{f}^{q+1,\epsilon}(s,1)\big\} ds \\ &+ \sum_{j=0}^{m} \int_{\tau_q^0}^{t} \int_0^1 \frac{\partial \hat{f}^{q+1,\epsilon}(s,\eta)}{\partial \tilde{x}^j} \mu^{q+1,\epsilon}(s-\gamma_j) d\eta ds,\end{aligned}$$

provided that $\tau_q^0 > \tau_{q-1}^\epsilon$ when $q \geq 1$, which holds when ϵ is of sufficiently small magnitude (see Section 4.4). Now, choose an arbitrary $T_{\max} > \max(\tau_q^\epsilon, \tau_q^0)$. Then for all $t \leq T_{\max}$, it follows from (4.15) and the result in Section 4.4 that

$$\begin{aligned}(4.21)\qquad &\int_{\tau_q^0}^t \int_0^1 \frac{\partial \hat{f}^{q+1,\epsilon}(s,\eta)}{\partial \tilde{x}^j} \mu^{q+1,\epsilon}(s-\gamma_j) d\eta ds \\ &= \int_{\tau_q^0}^t \frac{\partial \hat{f}^{q+1,0}(s,0)}{\partial \tilde{x}^j} \mu^{q+1,\epsilon}(s-\gamma_j) ds + \theta(\epsilon) O(\epsilon).\end{aligned}$$

Furthermore, it follows from (4.18) and (4.19) that

$$\begin{aligned}(4.22)\qquad &\int_{\tau_q^0}^{\tau_q^\epsilon} \big\{ \hat{f}^{q,\epsilon}(s,1) - \hat{f}^{q+1,\epsilon}(s,1) \big\} ds \\ &= (\tau_q^\epsilon - \tau_q^0) \Big\{ \hat{f}^{q,0}(\tau_q^0,0) - \hat{f}^{q+1,0}(\tau_q^0,0) \Big\} + O(\epsilon^2).\end{aligned}$$

Using equations (4.21) and (4.22) to simplify the expression for $\mu^{q+1,\epsilon}(t)$ gives, for all times t satisfying $\max(\tau_q^\epsilon, \tau_q^0) \leq t \leq T_{\max}$,

$$\begin{aligned}(4.23)\qquad \mu^{q+1,\epsilon}(t) &= \mu^{q,\epsilon}(\tau_q^0) + (\tau_q^\epsilon - \tau_q^0) \Big\{ \hat{f}^{q,0}(\tau_q^0,0) - \hat{f}^{q+1,0}(\tau_q^0,0) \Big\} \\ &\quad + \sum_{j=0}^m \int_{\tau_q^0}^t \frac{\partial \hat{f}^{q+1,0}(s,0)}{\partial \tilde{x}^j} \mu^{q+1,\epsilon}(s-\gamma_j) ds + O(\epsilon^2) + \theta(\epsilon) O(\epsilon).\end{aligned}$$

Now, the solution of the variational system on $(\tau_q^0, \tau_{q+1}^0]$ can be expressed as

$$(4.24)\qquad \Lambda_k(t-) = \Lambda_k(\tau_q^0+) + \sum_{j=0}^m \int_{\tau_q^0}^t \frac{\partial \hat{f}^{q+1,0}(s,0)}{\partial \tilde{x}^j} \Lambda_k(s-\gamma_j) ds, \quad t \in (\tau_q^0, \tau_{q+1}^0].$$

Multiplying (4.23) by ϵ^{-1} and then subtracting (4.24) gives

$$\begin{aligned}\epsilon^{-1}\mu^{q+1,\epsilon}(t) - \Lambda_k(t-) &= \epsilon^{-1}\mu^{q,\epsilon}(\tau_q^0) - \Lambda_k(\tau_q^0+) + \epsilon^{-1}(\tau_q^\epsilon - \tau_q^0) \Big\{ \hat{f}^{q,0}(\tau_q^0,0) - \hat{f}^{q+1,0}(\tau_q^0,0) \Big\} \\ &\quad + \sum_{j=0}^m \int_{\tau_q^0}^t \frac{\partial \hat{f}^{q+1,0}(s,0)}{\partial \tilde{x}^j} \big\{ \epsilon^{-1}\mu^{q+1,\epsilon}(s-\gamma_j) - \Lambda_k(s-\gamma_j) \big\} ds + \theta(\epsilon), \\ &\qquad\qquad t \in (\max(\tau_q^\epsilon, \tau_q^0), \min(\tau_{q+1}^0, T_{\max})],\end{aligned}$$

which holds when ϵ is of sufficiently small magnitude. Hence, by taking the norm of both sides and changing the variable of integration in the last integral, we obtain

$$\begin{aligned}\big|\epsilon^{-1}\mu^{q+1,\epsilon}(t) - \Lambda_k(t-)\big| &\leq \lambda^{q,\epsilon} + \int_{-\bar{\gamma}}^t (m+1)\partial f_{\max}^{q+1} \big|\epsilon^{-1}\mu^{q+1,\epsilon}(s) - \Lambda_k(s)\big| ds + \theta(\epsilon), \\ &\qquad\qquad t \in (\max(\tau_q^\epsilon, \tau_q^0), \min(\tau_{q+1}^0, T_{\max})],\end{aligned}$$

where $\partial f_{\max}^{q+1}$ is as defined in Section 4.4 and

$$\lambda^{q,\epsilon} := \left| \epsilon^{-1}\mu^{q,\epsilon}(\tau_q^0) - \Lambda_k(\tau_q^0+) + \epsilon^{-1}(\tau_q^\epsilon - \tau_q^0) \Big\{ \hat{f}^{q,0}(\tau_q^0,0) - \hat{f}^{q+1,0}(\tau_q^0,0) \Big\} \right|.$$

Clearly, by equations (4.5) and (4.7) for $i = q$ and the jump condition in the variational system,

$$\lim_{\epsilon\to 0}\lambda^{q,\epsilon} = \lim_{\epsilon\to 0}\Big|\epsilon^{-1}\mu^{q,\epsilon}(\tau_q^0) - \Lambda_k(\tau_q^0-) + \Big\{\frac{\tau_q^\epsilon - \tau_q^0}{\epsilon} - \frac{\partial\tau_q(\zeta)}{\partial\zeta_k}\Big\}\cdot\Big\{\hat f^{q,0}(\tau_q^0,0) - \hat f^{q+1,0}(\tau_q^0,0)\Big\}\Big| = 0. \tag{4.25}$$

Using (4.20) and (4.25), we obtain

$$\big|\epsilon^{-1}\mu^{q+1,\epsilon}(t) - \Lambda_k(t-)\big| \le \theta(\epsilon) + \int_{\max(\tau_q^\epsilon,\tau_q^0)}^{t}(m+1)\partial f_{\max}^{q+1}\big|\epsilon^{-1}\mu^{q+1,\epsilon}(s) - \Lambda_k(s)\big|ds,$$
$$t\in(\max(\tau_q^\epsilon,\tau_q^0),\min(\tau_{q+1}^0,T_{\max})].$$

Now, applying the Gronwall inequality [2] gives

$$\big|\epsilon^{-1}\mu^{q+1,\epsilon}(t) - \Lambda_k(t-)\big| \le \theta(\epsilon)\exp((m+1)\partial f_{\max}^{q+1}T_{\max}), \quad t\in(\max(\tau_q^\epsilon,\tau_q^0),\min(\tau_{q+1}^0,T_{\max})], \tag{4.26}$$

which holds for all ϵ of sufficiently small magnitude. Finally, for any fixed time point $t\in(\tau_q^0,\tau_{q+1}^0]$, we can choose $T_{\max} > t$ so that $t\in(\tau_q^\epsilon,\min(\tau_{q+1}^0,T_{\max})]$ when the magnitude of ϵ is sufficiently small, and then by (4.26) we have $\epsilon^{-1}\mu^{q+1,\epsilon}(t)\to\Lambda_k(t-)$, which completes the proof of equation (4.5) for $i = q+1$.

4.7. Inductive step for equation (4.6). If $q = N-1$, then $\tau_{q+1}^\epsilon = \tau_{q+1}^0 = \infty$ and thus equation (4.6) for $i = q+1$ clearly holds. Hence, we assume that $q < N-1$.

Condition 3.1 and the definition of τ_{q+1}^0 imply that there exists $\delta > 0$ such that $\tau_q^0 - \delta < \tau_q^0 < \tau_{q+1}^0 - \delta$ and

$$g_{q+1}(z^{q+1,0}(t)) \ne 0, \quad t\in[\tau_q^0-\delta,\tau_{q+1}^0-\delta], \tag{4.27}$$

where $\tau_{q+1}^0 - \delta = \infty$ if $\tau_{q+1}^0 = \infty$. For any $T_{\max} > \tau_q^0$, we know from Section 4.4 that $z^{q+1,\epsilon}(t)\to z^{q+1,0}(t)$ uniformly on $[-\bar\gamma,T_{\max}]$ as $\epsilon\to 0$. Thus, from (4.6), when ϵ is of sufficiently small magnitude,

$$g_{q+1}(z^{q+1,\epsilon}(t)) \ne 0, \quad t\in[\tau_q^\epsilon,\min(\tau_{q+1}^0-\delta,T_{\max})]. \tag{4.28}$$

If $\tau_{q+1}^0 = \infty$, then this becomes

$$g_{q+1}(z^{q+1,\epsilon}(t)) \ne 0, \quad t\in[\tau_q^\epsilon,T_{\max}],$$

which implies $\tau_{q+1}^\epsilon \ge T_{\max}$. Since $T_{\max}$ was chosen arbitrarily, we can take $T_{\max}\to\infty$ to yield $\tau_{q+1}^\epsilon\to\infty$ as $\epsilon\to 0$, proving (4.6) for $i = q+1$.

We now consider the case $\tau_{q+1}^0 < \infty$. Clearly,

$$g_{q+1}(z^{q+1,0}(\tau_{q+1}^0)) = 0. \tag{4.29}$$

Furthermore, by Condition 3.2,

$$\frac{\partial g_{q+1}(z^{q+1,0}(\tau_{q+1}^0))}{\partial x}\hat f^{q+1,0}(\tau_{q+1}^0,0) = \frac{\partial g_{q+1}(x(\tau_{q+1}^0|\zeta))}{\partial x}f^{q+1}(x(\tau_{q+1}^0|\zeta),x(\tau_{q+1}^0-\gamma_1|\zeta),\dots,x(\tau_{q+1}^0-\gamma_m|\zeta)) \ne 0.$$

Thus, by continuity, we may assume that $\delta > 0$ in (4.27) also satisfies

$$\begin{aligned}(4.30)\quad \frac{d}{dt}\Big\{g_{q+1}(z^{q+1,0}(t))\Big\} &= \frac{\partial g_{q+1}(z^{q+1,0}(t))}{\partial x}\dot{z}^{q+1,0}(t)\\ &= \frac{\partial g_{q+1}(z^{q+1,0}(t))}{\partial x}\hat{f}^{q+1,0}(t,0) \neq 0, \quad t \in (\tau^0_{q+1}-\delta, \tau^0_{q+1}+\delta).\end{aligned}$$

This shows that $g_{q+1}(z^{q+1,0}(\cdot))$ is either strictly increasing or strictly decreasing on $(\tau^0_{q+1}-\delta, \tau^0_{q+1}+\delta)$. Therefore, in view of (4.29) and (4.30), $g_{q+1}(z^{q+1,0}(\cdot))$ has different sign at $\tau^0_{q+1}-\delta$ and $\tau^0_{q+1}+\delta$:

$$g_{q+1}(z^{q+1,0}(\tau^0_{q+1}-\delta)) \cdot g_{q+1}(z^{q+1,0}(\tau^0_{q+1}+\delta)) < 0.$$

Choosing $T_{\max} > \tau^0_{q+1}+\delta$, since $z^{q+1,\epsilon}(t) \to z^{q+1,0}(t)$ uniformly on $[-\bar{\gamma}, T_{\max}]$ as $\epsilon \to 0$, when ϵ is sufficiently small,

$$(4.31)\qquad g_{q+1}(z^{q+1,\epsilon}(\tau^0_{q+1}-\delta)) \cdot g_{q+1}(z^{q+1,\epsilon}(\tau^0_{q+1}+\delta)) < 0,$$

which implies that $g_{q+1}(z^{q+1,\epsilon}(\cdot))$, like $g_{q+1}(z^{q+1,0}(\cdot))$, has different sign at $\tau^0_{q+1}-\delta$ and $\tau^0_{q+1}+\delta$. Combining (4.28) and (4.31) gives $\tau^0_{q+1}-\delta < \tau^\epsilon_{q+1} < \tau^0_{q+1}+\delta$. The proof is completed by taking $\delta \to 0$.

4.8. Inductive step for equation (4.7). We assume that $q < N-1$, since equation (4.7) is not applicable when $i = N$. First, define

$$\hat{g}^\epsilon_{q+1}(\eta) := g_{q+1}(\eta z^{q+1,\epsilon}(\tau^\epsilon_{q+1}) + (1-\eta)z^{q+1,0}(\tau^0_{q+1}))$$

and let $\partial\hat{g}^\epsilon_{q+1}(\eta)/\partial x$ denote the respective derivative. By Taylor's theorem, there exists a constant $\eta_\epsilon \in (0,1)$ such that

$$(4.32)\qquad 0 = \hat{g}^\epsilon_{q+1}(1) - \hat{g}^0_{q+1}(0) = \frac{\partial\hat{g}^\epsilon_{q+1}(\eta_\epsilon)}{\partial x}\big\{z^{q+1,\epsilon}(\tau^\epsilon_{q+1}) - z^{q+1,0}(\tau^0_{q+1})\big\}.$$

Now, since $\tau^\epsilon_q \to \tau^0_q < \tau^0_{q+1}$ as $\epsilon \to 0$, we have $\tau^\epsilon_q < \tau^0_{q+1}$ when ϵ is of sufficiently small magnitude, and thus

$$\begin{aligned}(4.33)\quad & z^{q+1,\epsilon}(\tau^\epsilon_{q+1}) - z^{q+1,0}(\tau^0_{q+1}) = z^{q+1,\epsilon}(\tau^\epsilon_{q+1}) - z^{q+1,\epsilon}(\tau^0_{q+1}) + \mu^{q+1,\epsilon}(\tau^0_{q+1})\\ &= \int_{\tau^0_{q+1}}^{\tau^\epsilon_{q+1}} \hat{f}^{q+1,\epsilon}(s,1)ds + \mu^{q+1,\epsilon}(\tau^0_{q+1})\\ &= (\tau^\epsilon_{q+1}-\tau^0_{q+1})\int_0^1 \hat{f}^{q+1,\epsilon}(\eta\tau^\epsilon_{q+1} + (1-\eta)\tau^0_{q+1}, 1)d\eta + \mu^{q+1,\epsilon}(\tau^0_{q+1}).\end{aligned}$$

Substituting (4.33) into (4.32) and rearranging gives

$$\begin{aligned}(4.34)\quad & (\tau^\epsilon_{q+1}-\tau^0_{q+1})\frac{\partial\hat{g}^\epsilon_{q+1}(\eta_\epsilon)}{\partial x}\int_0^1 \hat{f}^{q+1,\epsilon}(\eta\tau^\epsilon_{q+1} + (1-\eta)\tau^0_{q+1}, 1)d\eta\\ &\qquad = -\frac{\partial\hat{g}^\epsilon_{q+1}(\eta_\epsilon)}{\partial x}\mu^{q+1,\epsilon}(\tau^0_{q+1}),\end{aligned}$$

which holds for all ϵ of sufficiently small magnitude.

Clearly, since $\tau^\epsilon_{q+1} \to \tau^0_{q+1}$ as $\epsilon \to 0$, we can choose $T_{\max} > \max(\tau^\epsilon_{q+1}, \tau^0_{q+1})$, and thus by Condition 3.2 and the results in Section 4.4,

$$\lim_{\epsilon\to 0}\frac{\partial\hat{g}^\epsilon_{q+1}(\eta_\epsilon)}{\partial x}\int_0^1 \hat{f}^{q+1,\epsilon}(\eta\tau^\epsilon_{q+1} + (1-\eta)\tau^0_{q+1}, 1)d\eta = \frac{\partial\hat{g}^0_{q+1}(0)}{\partial x}\hat{f}^{q+1,0}(\tau^0_{q+1}, 0) \neq 0.$$

This ensures that we can rearrange (4.34) to give

$$\tau^{\epsilon}_{q+1} - \tau^{0}_{q+1} = -\frac{\partial \hat{g}^{\epsilon}_{q+1}(\eta_\epsilon)}{\partial x}\mu^{q+1,\epsilon}(\tau^{0}_{q+1}) \div \left\{ \frac{\partial \hat{g}^{\epsilon}_{q+1}(\eta_\epsilon)}{\partial x} \int_0^1 \hat{f}^{q+1,\epsilon}(\eta\tau^{\epsilon}_{q+1} + (1-\eta)\tau^{0}_{q+1}, 1)d\eta \right\}.$$

Next, dividing both sides by ϵ yields

$$\begin{aligned}\frac{\tau^{\epsilon}_{q+1} - \tau^{0}_{q+1}}{\epsilon} &= -\frac{\partial \hat{g}^{\epsilon}_{q+1}(\eta_\epsilon)}{\partial x}\{\epsilon^{-1}\mu^{q+1,\epsilon}(\tau^{0}_{q+1})\} \\ &\quad \div \left\{ \frac{\partial \hat{g}^{\epsilon}_{q+1}(\eta_\epsilon)}{\partial x} \int_0^1 \hat{f}^{q+1,\epsilon}(\eta\tau^{\epsilon}_{q+1} + (1-\eta)\tau^{0}_{q+1}, 1)d\eta \right\}.\end{aligned}$$

Thus, using the result from Section 4.6, we obtain

$$\lim_{\epsilon\to 0} \frac{\tau^{\epsilon}_{q+1} - \tau^{0}_{q+1}}{\epsilon} = -\frac{\partial \hat{g}^{0}_{q+1}(0)}{\partial x}\Lambda_k(\tau^{0}_{q+1}-) \div \left\{ \frac{\partial \hat{g}^{0}_{q+1}(0)}{\partial x} \hat{f}^{q+1,0}(\tau^{0}_{q+1}, 0) \right\},$$

which proves equation (4.7) for $i = q+1$, as required.

4.9. Proof of Theorem 3.2. With Theorem 3.1 proved, we now turn our attention to Theorem 3.2. Consider $i \in Z_{N-1}$ with $\tau_i < \infty$ and let $k \in \{1, \dots, p\}$. Then from the auxiliary system (4.1), when $|\epsilon|$ is sufficiently small so that $\tau^{0}_{i} < \tau^{\epsilon}_{i+1}$,

$$\begin{aligned}x(\tau^{0}_{i}|\zeta + \epsilon e^k) - x(\tau^{0}_{i}|\zeta) &= z^{i,\epsilon}(\tau^{0}_{i}) - z^{i,0}(\tau^{0}_{i}) + \int_{\min(\tau^{\epsilon}_{i},\tau^{0}_{i})}^{\tau^{0}_{i}} \{\hat{f}^{i+1,\epsilon}(s,1) - \hat{f}^{i,\epsilon}(s,1)\}ds \\ &= \mu^{i,\epsilon}(\tau^{0}_{i}) + \int_{\min(\tau^{\epsilon}_{i},\tau^{0}_{i})}^{\tau^{0}_{i}} \{\hat{f}^{i+1,\epsilon}(s,1) - \hat{f}^{i,\epsilon}(s,1)\}ds.\end{aligned}$$

Therefore, from identities (4.18) and (4.19) in Section 4.5,

$$x(\tau^{0}_{i}|\zeta + \epsilon e^k) - x(\tau^{0}_{i}|\zeta) = \mu^{i,\epsilon}(\tau^{0}_{i}) + \{\tau^{0}_{i} - \min(\tau^{\epsilon}_{i}, \tau^{0}_{i})\} \cdot \{\hat{f}^{i+1,0}(\tau^{0}_{i}, 0) - \hat{f}^{i,0}(\tau^{0}_{i}, 0)\} + O(\epsilon^2).$$

Hence,

$$\frac{x(\tau^{0}_{i}|\zeta + \epsilon e^k) - x(\tau^{0}_{i}|\zeta)}{\epsilon} = \epsilon^{-1}\mu^{i,\epsilon}(\tau^{0}_{i}) + \frac{\tau^{0}_{i} - \min(\tau^{\epsilon}_{i}, \tau^{0}_{i})}{\epsilon} \cdot \{\hat{f}^{i+1,0}(\tau^{0}_{i}, 0) - \hat{f}^{i,0}(\tau^{0}_{i}, 0)\} + O(\epsilon).$$

If $\partial\tau_i/\partial\zeta_k > 0$, then $\tau^{\epsilon}_{i} \to \tau^{0}_{i}\pm$ as $\epsilon \to 0\pm$, and consequently,

$$\lim_{\epsilon\to 0+} \frac{\tau^{0}_{i} - \min(\tau^{\epsilon}_{i}, \tau^{0}_{i})}{\epsilon} = 0, \qquad \lim_{\epsilon\to 0-} \frac{\tau^{0}_{i} - \min(\tau^{\epsilon}_{i}, \tau^{0}_{i})}{\epsilon} = -\frac{\partial\tau_i}{\partial\zeta_k}.$$

Thus, since $\epsilon^{-1}\mu^{i,\epsilon}(\tau^{0}_{i}) \to \Lambda_k(\tau^{0}_{i}-)$ by (4.5),

$$\lim_{\epsilon\to 0+} \frac{x(\tau^{0}_{i}|\zeta + \epsilon e^k) - x(\tau^{0}_{i}|\zeta)}{\epsilon} = \Lambda_k(\tau^{0}_{i}-)$$

and

$$\lim_{\epsilon\to 0-} \frac{x(\tau^{0}_{i}|\zeta + \epsilon e^k) - x(\tau^{0}_{i}|\zeta)}{\epsilon} = \Lambda_k(\tau^{0}_{i}-) + \frac{\partial\tau_i(\zeta)}{\partial\zeta_k} \cdot \{\hat{f}^{i,0}(\tau^{0}_{i}, 0) - \hat{f}^{i+1,0}(\tau^{0}_{i}, 0)\} = \Lambda_k(\tau^{0}_{i}+).$$

Similarly, if $\partial\tau_i/\partial\zeta_k < 0$, then

$$\lim_{\epsilon\to 0+} \frac{x(\tau^{0}_{i}|\zeta + \epsilon e^k) - x(\tau^{0}_{i}|\zeta)}{\epsilon} = \Lambda_k(\tau^{0}_{i}-) + \frac{\partial\tau_i(\zeta)}{\partial\zeta_k} \cdot \{\hat{f}^{i,0}(\tau^{0}_{i}, 0) - \hat{f}^{i+1,0}(\tau^{0}_{i}, 0)\} = \Lambda_k(\tau^{0}_{i}+)$$

and

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

Finally, if either $\partial\tau_i/\partial\zeta_k = 0$ or $\hat{f}^{i+1,0}(\tau_i^0, 0) = \hat{f}^{i,0}(\tau_i^0, 0)$, then clearly we have $\Lambda_k(\tau_i^0+) = \Lambda_k(\tau_i^0-)$ and

$$\frac{\partial x(\tau_i^0|\zeta)}{\partial \zeta_k} = \lim_{\epsilon \to 0} \frac{x(\tau_i^0|\zeta + \epsilon e^k) - x(\tau_i^0|\zeta)}{\epsilon} = \Lambda_k(\tau_i^0-) = \Lambda_k(\tau_i^0+).$$

4.10. Continuity of the cost function. Let $\zeta \in \mathcal{Z}$ be an arbitrary parameter vector satisfying Conditions 3.1 and 3.2. Equations (4.4) and (4.6) show that $z^i(\cdot)$, $i \in Z_N$, and $\tau_i(\cdot)$, $i \in Z_{N-1}$, are individually continuous with respect to each system parameter at ζ, but this does not necessarily imply full continuity (in the space R^p) at ζ. Nevertheless, full continuity can be derived as follows. Consider a perturbed parameter vector $\zeta + \sigma \in \mathcal{Z}$ and let $z^{i,\sigma}(\cdot)$, $\tau_i^\sigma(\cdot)$, and $\mu^{i,\sigma}(\cdot)$ denote the analogues of $z^{i,\epsilon}(\cdot)$, $\tau_i^\epsilon(\cdot)$, and $\mu^{i,\epsilon}(\cdot)$ with σ instead of ϵe^k. Then the proofs in Sections 4.3, 4.4, and 4.7 can be easily modified to derive the following variations of equations (4.3), (4.4), and (4.6) for each i:

$$\max_{t \in [-\bar{\gamma}, T_{\max}]} \left| z^{i,\sigma}(t) \right| = O(1) \text{ for every } T_{\max} > 0, \tag{4.35}$$

$$\max_{t \in [-\bar{\gamma}, T_{\max}]} \left| \mu^{i,\sigma}(t) \right| = \theta(\sigma) \text{ for every } T_{\max} > 0, \tag{4.36}$$

$$\lim_{\sigma \to 0} \tau_i^\sigma = \tau_i^0, \tag{4.37}$$

where $O(1)$ in (4.35) means that the left-hand side is uniformly bounded with respect to σ, and $\theta(\sigma)$ in (4.36) means that the left-hand side converges to zero as $\sigma \to 0$. In (4.36) and (4.37), the convergence $\sigma \to 0$ can be along any path to the origin, but in the corresponding results (4.4) and (4.6), the convergence is restricted to be along one of the coordinate axes. Choosing $T_{\max} > T$, equation (4.36) for $i = N$ gives

$$\begin{aligned} \left| x(T|\zeta + \sigma) - x(T|\zeta) \right| &= \left| z^{N,\sigma}(T) - z^{N,0}(T) \right| \\ &= \left| \mu^{N,\sigma}(T) \right| \leq \max_{t \in [-\bar{\gamma}, T_{\max}]} \left| \mu^{N,\sigma}(t) \right| = \theta(\sigma), \end{aligned}$$

which shows that $x(T|\zeta + \sigma) \to x(T|\zeta)$ as $\sigma \to 0$. Hence, since Φ is continuous, we also have $J(\zeta + \sigma) = \Phi(x(T|\zeta + \sigma)) \to \Phi(x(T|\zeta)) = J(\zeta)$ as $\sigma \to 0$, as required. Note that equation (4.36) is the key continuity result here; this result must be proved simultaneously with (4.35) and (4.37) via induction.

5. Numerical example. We consider a fed-batch fermentation process for converting glycerol to 1,3-propanediol (1,3-PD). This process oscillates between two modes: batch mode (during which there is no input feed) and feeding mode (during which glycerol and alkali are added continuously to the fermentor). Mode switches occur when the concentration of glycerol reaches certain lower and upper thresholds. Moreover, the fermentation process involves a time-delay because nutrient metabolization does not immediately lead to the production of new biomass.

The model is based on the work in [24, 28]. Let $x(t) = (x_1(t), x_2(t), x_3(t), x_4(t))^\top$, where t is time (hours), $x_1(t)$ is the biomass concentration (gL^{-1}), $x_2(t)$ is the glycerol

TABLE 5.1
Model parameters in equations (5.3)-(5.5).

Δ_1	k_1	m_1	Y_1	Δ_2	k_2	m_2	Y_2	Δ_3	k_3
0.8037	0.4856	0.2977	144.9120	7.8367	9.4632	12.2577	80.8439	20.2757	38.7500

concentration (mmolL^{-1}), $x_3(t)$ is the 1,3-PD concentration (mmolL^{-1}), and $x_4(t)$ is the fluid volume (L). The process dynamics due to natural fermentation are

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{x}_3(t) \\ \dot{x}_4(t) \end{bmatrix} = \begin{bmatrix} \mu(x_2(t), x_3(t))x_1(t-\gamma_1) \\ -q_2(x_2(t), x_3(t))x_1(t-\gamma_1) \\ q_3(x_2(t), x_3(t))x_1(t-\gamma_1) \\ 0 \end{bmatrix} := f^{\text{ferm}}(x(t), x_1(t-\gamma_1)), \tag{5.1}$$

where $\gamma_1 = 0.1568$ is the time-delay; $\mu(\cdot,\cdot)$ is the cell growth rate; $q_2(\cdot,\cdot)$ is the substrate consumption rate; and $q_3(\cdot,\cdot)$ is the 1,3-PD formation rate. The process dynamics due to the input feed are

$$\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{x}_3(t) \\ \dot{x}_4(t) \end{bmatrix} = \frac{u(t)}{x_4(t)} \begin{bmatrix} -x_1(t) \\ rc_{s0} - x_2(t) \\ -x_3(t) \\ x_4(t) \end{bmatrix} := f^{\text{feed}}(x(t), u(t)), \tag{5.2}$$

where $u(t)$ is the input feeding rate (Lh^{-1}); $r = 0.5714$ is the proportion of glycerol in the input feed; and $c_{s0} = 10762$ mmolL^{-1} is the concentration of glycerol in the input feed. The functions $\mu(\cdot,\cdot)$, $q_2(\cdot,\cdot)$ and $q_3(\cdot,\cdot)$ in (5.1) are

$$\mu(x_2(t), x_3(t)) = \frac{\Delta_1 x_2(t)}{x_2(t) + k_1}\left(1 - \frac{x_2(t)}{x_2^*}\right)\left(1 - \frac{x_3(t)}{x_3^*}\right)^3, \tag{5.3}$$

$$q_2(x_2(t), x_3(t)) = m_1 + Y_1\mu(x_2(t), x_3(t)) + \frac{\Delta_2 x_2(t)}{x_2(t) + k_2}, \tag{5.4}$$

$$q_3(x_2(t), x_3(t)) = -m_2 + Y_2\mu(x_2(t), x_3(t)) + \frac{\Delta_3 x_2(t)}{x_2(t) + k_3}, \tag{5.5}$$

where $x_2^* = 2039$ mmolL^{-1} and $x_3^* = 1036$ mmolL^{-1} are, respectively, the critical concentrations of glycerol and 1,3-PD, and the values of the other parameters are given in Table 5.1.

Let N_{feed} be an upper bound for the number of feeding modes. Then the total number of potential modes is $N = 2N_{\text{feed}} + 1$ (consisting of N_{feed} feeding modes and $N_{\text{feed}} + 1$ batch modes, since the process starts and finishes in batch mode). During batch mode, there is no input and the process is governed by (5.1) only. During feeding mode, the process is governed by both (5.1) and (5.2). Thus,

$$\dot{x}(t) = \begin{cases} f^{\text{ferm}}(x(t), x_1(t-\gamma_1)), & \text{for batch mode,} \\ f^{\text{ferm}}(x(t), x_1(t-\gamma_1)) + f^{\text{feed}}(x(t), \zeta_i), & \text{for } i\text{th feeding mode,} \end{cases} \tag{5.6}$$

where ζ_i is the feeding rate during the ith feeding mode, with

$$1.0043 \le \zeta_i \le 1.9266. \tag{5.7}$$

Glycerol is consumed as the biomass grows. During batch mode, no new glycerol is added and eventually the glycerol concentration will become too low, necessitating a switch into feeding mode. The corresponding switching condition is

$$x_2(t) - \zeta_{N_{\text{feed}}+1} = 0,$$

where $\zeta_{N_{\text{feed}}+1}$ is the lower switching concentration (a parameter to be optimized). On the other hand, when the glycerol concentration becomes too high during feeding mode, cell growth is inhibited. Thus, if too much glycerol is added, then the process must switch back into batch mode. The corresponding switching condition is

$$x_2(t) - \zeta_{N_{\text{feed}}+2} = 0,$$

where $\zeta_{N_{\text{feed}}+2}$ is the upper switching concentration (another parameter to be optimized). The bound constraints on $\zeta_{N_{\text{feed}}+1}$ and $\zeta_{N_{\text{feed}}+2}$ are

$$50 \leq \zeta_{N_{\text{feed}}+1} \leq 260, \quad 300 \leq \zeta_{N_{\text{feed}}+2} \leq 600. \tag{5.8}$$

Note that the system parameters in this example appear explicitly in the dynamics and switching conditions. Thus, to apply Theorem 3.1, we replace the system parameters with auxiliary state variables $x_{4+k}(t)$, $k = 1, \ldots, N_{\text{feed}} + 2$, where $\dot{x}_{4+k}(t) = 0$, $t > 0$, and $x_{4+k}(t) = \zeta_k$, $t \leq 0$. Letting δ_{ki} denote the Kronecker delta function and ∂x and $\partial \tilde{x}_1$ denote differentiation with respect to $x(t)$ and $x_1(t - \gamma_1)$, respectively, the variational system corresponding to ζ_k is

$$\dot{\Lambda}_k(t) = \begin{cases} \dfrac{\partial f^{\text{ferm}}}{\partial x}\Lambda_k(t) + \dfrac{\partial f^{\text{ferm}}}{\partial \tilde{x}_1}\Lambda_{k1}(t-\gamma_1), & \text{batch mode,} \\ \dfrac{\partial f^{\text{ferm}}}{\partial x}\Lambda_k(t) + \dfrac{\partial f^{\text{ferm}}}{\partial \tilde{x}_1}\Lambda_{k1}(t-\gamma_1) + \dfrac{\partial f^{\text{feed}}}{\partial x}\Lambda_k(t) + \delta_{ki}\dfrac{\partial f^{\text{feed}}}{\partial u}, & i\text{th feeding mode,} \end{cases}$$

with jump conditions

$$\Lambda_k(\tau_i+) = \begin{cases} \Lambda_k(\tau_i-) - \dfrac{\partial \tau_i}{\partial \zeta_k} f^{\text{feed}}(x(\tau_i), \zeta_{(i+1)/2}), & \text{if mode } i = \text{batch,} \\ \Lambda_k(\tau_i-) + \dfrac{\partial \tau_i}{\partial \zeta_k} f^{\text{feed}}(x(\tau_i), \zeta_{i/2}), & \text{if mode } i = \text{feeding.} \end{cases}$$

Furthermore, for a switch from batch mode to feeding mode,

$$\frac{\partial \tau_i}{\partial \zeta_k} = \begin{cases} (1 - \Lambda_{k2}(\tau_i-)) \div f_2^{\text{ferm}}(x(\tau_i), x_1(\tau_i - \gamma_1)), & \text{if } k = N_{\text{feed}} + 1, \\ -\Lambda_{k2}(\tau_i-) \div f_2^{\text{ferm}}(x(\tau_i), x_1(\tau_i - \gamma_1)), & \text{otherwise.} \end{cases}$$

Similarly, for a switch from feeding mode to batch mode,

$$\frac{\partial \tau_i}{\partial \zeta_k} = \begin{cases} (1 - \Lambda_{k2}(\tau_i-)) \div \left\{ f_2^{\text{ferm}}(x(\tau_i), x_1(\tau_i - \gamma_1)) + f_2^{\text{feed}}(x(\tau_i), \zeta_{i/2}) \right\}, & \text{if } k = N_{\text{feed}} + 2, \\ -\Lambda_{k2}(\tau_i-) \div \left\{ f_2^{\text{ferm}}(x(\tau_i), x_1(\tau_i - \gamma_1)) + f_2^{\text{feed}}(x(\tau_i), \zeta_{i/2}) \right\}, & \text{otherwise.} \end{cases}$$

Since the bounds in (5.8) ensure $\zeta_{N_{\text{feed}}+1} < \zeta_{N_{\text{feed}}+2}$, Condition 3.1 is clearly satisfied at all feasible points. For Condition 3.2, we require

$$0 \neq \begin{cases} -q_2(x_2(\tau_i), x_3(\tau_i))x_1(\tau_i - \gamma_1), & \text{if mode } i = \text{batch,} \\ -q_2(x_2(\tau_i), x_3(\tau_i))x_1(\tau_i - \gamma_1) + \dfrac{\zeta_{i/2}(rc_{s0} - x_2(\tau_i))}{x_4(\tau_i)}, & \text{if mode } i = \text{feeding.} \end{cases}$$

TABLE 5.2
Optimal control parameters for the example: $\zeta_1, \ldots, \zeta_4$ are the optimal feeding rates and ζ_{49} and ζ_{50} are the optimal switching concentrations. The optimal values of $\zeta_5, \ldots, \zeta_{48}$ are irrelevant because they represent feeding rates after the terminal time.

Parameter	ζ_1	ζ_2	ζ_3	ζ_4	ζ_{49}	ζ_{50}
Optimal value	1.62662	1.36951	1.45283	1.64830	245.76512	581.35390

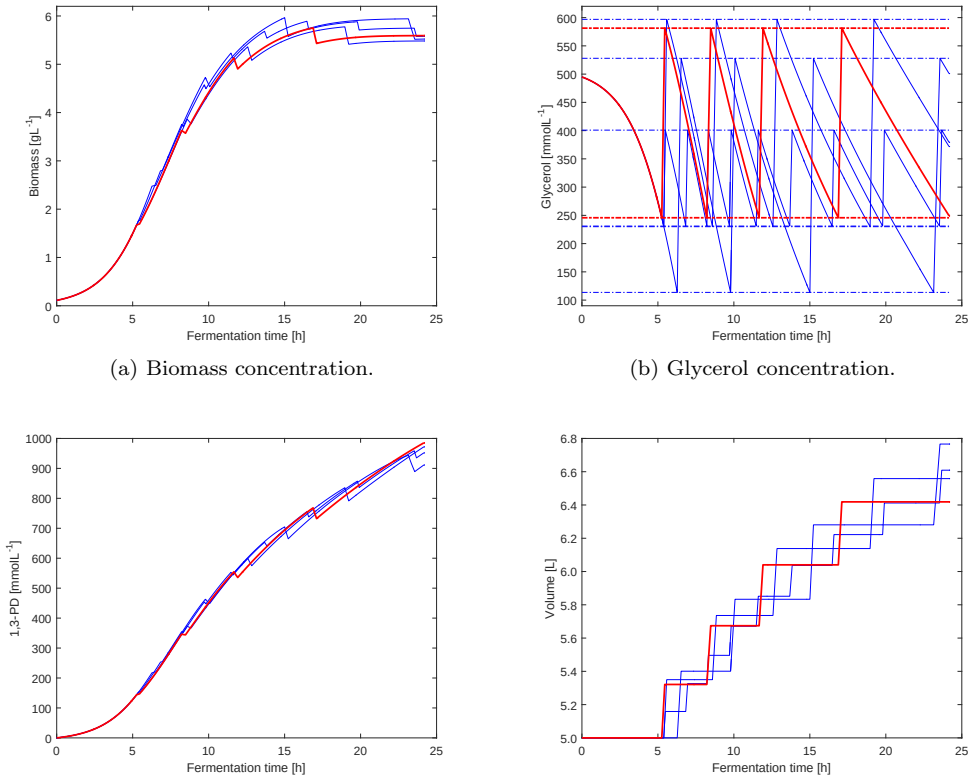
TABLE 5.3
Optimal mode durations for the example.

	Duration				
	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$
Batch mode i	5.25482	2.76659	3.18638	4.95994	7.05542
Feeding mode i	0.19745	0.25777	0.25193	0.22971	–

The expression above for batch mode is always non-zero because in practice both q_2 and x_1 are non-zero. The expression for feeding mode is also non-zero because, during feeding mode, the glycerol loss from natural fermentation (first term) is dominated by the glycerol addition from the input feed (second term). The linear growth condition is also satisfied in this example because x_4 is non-decreasing and, for biologically meaningful trajectories, $\mu(\cdot, \cdot)$ is bounded.

The initial function ϕ for the dynamics (5.6) was obtained by applying cubic spline interpolation to the experimental data in [28]. The terminal time for the fermentation process is $T = 24.16$ hours [24] and our chosen upper bound for the number of feeding modes is $N_{\text{feed}} = 48$. The aim is to maximize the concentration of 1,3-PD at the terminal time. Thus, the dynamic optimization problem is: choose the parameters ζ_k, $k = 1, \ldots, N_{\text{feed}} + 2$, to minimize the cost function $-x_3(T)$ subject to the bound constraints (5.7) and (5.8).

This problem was solved using a Fortran program that implements the gradient-based optimization procedure in Section 3. The program uses NLPQLP [31] to perform the optimization iterations (optimality check and line search), and LSODAR [13] to solve the differential equations. Our gradient-based optimization strategy generates critical points satisfying local optimality conditions, but global optimality is not guaranteed. Hence, to achieve a good estimate of the global solution, it is necessary to repeat the optimization process from different starting points. We performed 100 test runs, with each run starting from a different randomly-selected initial point. The average optimal cost over all runs was -977.12854, and the best result of -986.16815 was achieved on run 73 with the control parameters in Table 5.2. This control strategy produces 8 switches (5 batch modes and 4 feeding modes) and the corresponding mode durations are listed in Table 5.3. Fig. 5.1 shows both the optimal state trajectories (in bold red) and the intermediate state trajectories at every fifth iteration (in blue). The concentrations of biomass and 1,3-PD decrease during feeding modes 2-4 because of the dilution effect from the new input feed. The control strategy in Table 5.2 is essentially a state feedback strategy that produces more 1,3-PD compared with the time-dependent switching strategy in [24] (an increase of 5.789%), but with far fewer switches—in fact, the method in [24] requires over 1000 switches. Fig. 5.2 shows how the optimal 1,3-PD concentration varies over the 100 test runs.



(a) Biomass concentration.

(b) Glycerol concentration.

(c) 1,3-PD concentration.

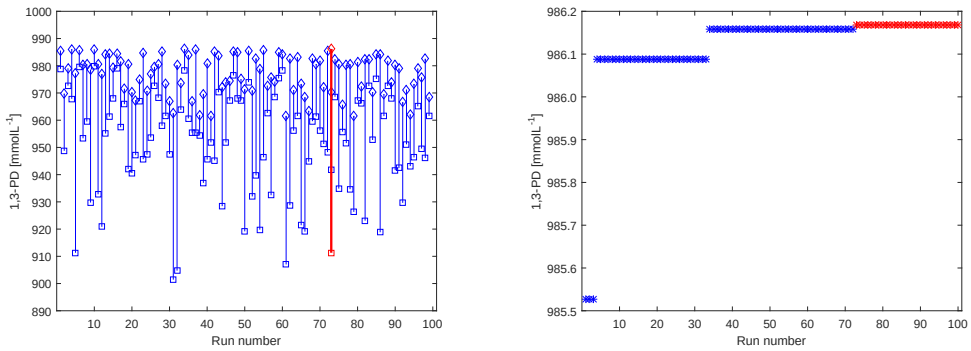
(d) Volume of culture fluid.

FIG. 5.1. *Convergence of the state trajectories; the bold red lines are the optimal trajectories, the blue lines are intermediate trajectories, and the dashed lines in (b) are the switching surfaces.*

6. Conclusion. This paper has considered a class of dynamic optimization problems for a general switched time-delay system with state-dependent switching conditions. Our main result shows that the state variation matrix in such problems can be computed by solving a certain variational system, and this forms the foundation of a gradient-based optimization method for generating locally-optimal solutions. Numerical results for an example problem arising in 1,3-propanediol production show that the method is effective. The complexity of the gradient derivation in Section 4 is due to the switching times depending implicitly on the control parameters, as well as the presence of multiple state-delays. The gradient derivation required two technical conditions (see Section 3), which are reasonable in practice and are similar to assumptions in previous related work. Importantly, our results show that the state variation matrix often does not exist at the switching points. This is different to time-dependent switched systems, where the state variation with respect to system parameters always exists, even at the switching points [18, 19].

REFERENCES

[1] N. U. AHMED, *Elements of Finite-dimensional Systems and Control Theory*, Longman Scientific and Technical, Essex, 1988.



(a) Initial and final 1,3-PD concentrations. (b) Increase in the 1,3-PD concentration.

FIG. 5.2. *Variation in the 1,3-PD concentration over 100 test runs; red denotes the best solution.*

[2] N. U. AHMED, *Dynamic Systems and Control with Applications*, World Scientific, Singapore, 2006.

[3] M. U. AKHMET, *On the smoothness of solutions of impulsive autonomous systems*, Nonlinear Anal., 60 (2005), pp. 311–324.

[4] D. K. ARROWSMITH AND C. M. PLACE, *Ordinary Differential Equations*, Chapman and Hall, London, 1982.

[5] R. G. BARTLE, *The Elements of Integration and Lebesgue Measure*, John Wiley, New York, 1966.

[6] S. C. BENGEA AND R. A. DECARLO, *Optimal control of switching systems*, Automatica J. IFAC, 41 (2005), pp. 11–27.

[7] E. BLANCHARD, R. LOXTON AND V. REHBOCK, *Dynamic optimization of dual-mode hybrid systems with state-dependent switching conditions*, Optim. Methods Softw., 33 (2018), pp. 297–310.

[8] M. BOCCADORO, Y. WARDI, M. EGERSTEDT AND E. VERRIEST, *Optimal control of switching surfaces in hybrid dynamical systems*, Discrete Event Dyn. Syst., 15 (2005), pp. 433–448.

[9] C. G. CASSANDRAS, *The event-driven paradigm for control, communication and optimization*, J. Control Decis., 1 (2014), pp. 3–17.

[10] K. G. DISHLIEVA, *Differentiability of solutions of impulsive differential equations with respect to the impulsive perturbations*, Nonlinear Anal. Real World Appl., 12 (2011), pp. 3541–3551.

[11] S. GALAN, W. F. FEEHERY AND P. I. BARTON, *Parametric sensitivity functions for hybrid discrete/continuous systems*, Appl. Numer. Math., 31 (1999), pp. 17–47.

[12] J. HESPANHA, D. LIBERZON AND A. S. MORSE, *Overcoming the limitations of adaptive control by means of logic-based switching*, Systems Control Lett., 49 (2003), pp. 49–56.

[13] A. C. HINDMARSH, *Large ordinary differential equation systems and software*, IEEE Control Syst. Mag., 2 (1982), pp. 24–30.

[14] V. LAKSHMIKANTHAM, D. D. BAINOV AND P. S. SIMEONOV, *Theory of Impulsive Differential Equations*, World Scientific, Singapore, 1989.

[15] S. M. LENZ, *Impulsive Hybrid Discrete-continuous Delay Differential Equations*, Ph.D. Dissertation, Heidelberg University, 2014.

[16] S. M. LENZ, J. P. SCHLODER AND H. G. BOCK, *Numerical computation of derivatives in systems of delay differential equations*, Math. Comput. Simulation, 96 (2014), pp. 124–156.

[17] D. LIBERZON, *Switching in Systems and Control*, Birkhäuser, Boston, 2003.

[18] Q. LIN, R. LOXTON AND K. L. TEO, *Optimal control of nonlinear switched systems: Computational methods and applications*, J. Oper. Res. Soc. China, 1 (2013), pp. 275–311.

[19] Q. LIN, R. LOXTON AND K. L. TEO, *The control parameterization method for nonlinear optimal control: A survey*, J. Ind. Manag. Optim., 10 (2014), pp. 275–309.

[20] Q. LIN, R. LOXTON, K. L. TEO AND Y. H. WU, *A new computational method for a class of free terminal time optimal control problems*, Pac. J. Optim., 7 (2011), pp. 63–81.

[21] Q. LIN, R. LOXTON, K. L. TEO AND Y. H. WU, *A new computational method for optimizing nonlinear impulsive systems*, Dyn. Contin. Discrete Impuls. Syst. Ser. B Appl. Algorithms, 18 (2011), pp. 59–76.

[22] Q. Lin, R. Loxton, K. L. Teo and Y. H. Wu, *Optimal control computation for nonlinear systems with state-dependent stopping criteria*, Automatica J. IFAC, 48 (2012), pp. 2116–2129.

[23] C. Y. Liu, Z. Gong, E. Feng and H. Yin, *Optimal switching control of a fed-batch fermentation process*, J. Global Optim., 52 (2012), pp. 265–280.

[24] C. Y. Liu, R. Loxton and K. L. Teo, *Switching time and parameter optimization in nonlinear switched systems with multiple time-delays*, J. Optim. Theory Appl., 163 (2014), pp. 957–988.

[25] R. Loxton, Q. Lin and K. L Teo, *Switching time optimization for nonlinear switched systems: Direct optimization and the time-scaling transformation*, Pac. J. Optim., 10 (2014), pp. 537–560.

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

[27] T. R. Mehta and M. Egerstedt, *Multi-modal control using adaptive motion description languages*, Automatica J. IFAC, 44 (2008), pp. 1912–1917.

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

[29] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd Edition, Springer-Verlag, New York, 2006.

[30] A. M. Samoilenko and N. A. Perestyuk, *Impulsive Differential Equations*, World Scientific, Singapore, 1995.

[31] K. Schittkowski, *NLPQLP: A Fortran implementation of a sequential quadratic programming algorithm with distributed and non-monotone line search – User's guide*, University of Bayreuth, Bayreuth, 2007.

[32] C. Seatzu, D. Corona, A. Giua and A. Bemporad, *Optimal control of continuous-time switched affine systems*, IEEE Trans. Automat. Control, 51 (2006), pp. 726–741.

[33] H. J. Sussmann, *A maximum principle for hybrid optimal control problems*, in Proceedings of the 38th IEEE Conference on Decision and Control, Phoenix, 1999, pp. 3972–3977.

[34] K. L. Teo, C. J. Goh and K. H. Wong, *A Unified Computational Approach to Optimal Control Problems*, Longman Scientific and Technical, Essex, 1991.

[35] R. Von Schwerin, M. Winckler and V. Schulz, *Parameter estimation in discontinuous descriptor models*, In: D. Bestle and W. Schiehlen (eds.), IUTAM Symposium on Optimization of Mechanical Systems, 43 (1996), pp. 269–276.

[36] X. Xu and P. J. Antsaklis, *Optimal control of switched systems based on parameterization of the switching instants*, IEEE Trans. Automat. Control, 49 (2004), pp. 2–15.

[37] C. Yu, Q. Lin, R. Loxton, K. L. Teo and G. Wang, *A hybrid time-scaling transformation for time-delay optimal control problems*, J. Optim. Theory Appl., 169 (2016), pp. 876–901.