Yokohama Publishers
ISSN 1349-8169 ONLINE JOURNAL

Pacific Journal of Optimization
Vol. 10, No. 3, pp. 537–560, July 2014

SWITCHING TIME OPTIMIZATION FOR NONLINEAR SWITCHED SYSTEMS: DIRECT OPTIMIZATION AND THE TIME-SCALING TRANSFORMATION

Ryan Loxton*, Qun Lin†, and Kok Lay Teo

This paper is dedicated to Professor Shu-Cherng Fang on the occasion of his 60th birthday.

Abstract: Given a switched system with multiple operating modes, a fundamental problem is to determine the optimal times at which the system should switch from one mode to another. This paper investigates two computational approaches for solving this problem—the direct optimization approach and the time-scaling approach. The direct optimization approach involves optimizing the mode switching times directly using gradient-based optimization methods such as sequential quadratic programming. The time-scaling approach involves transforming the switched system with variable switching times into an equivalent switched system with fixed switching times, where the decision parameters in the new switched system represent the mode durations in the original system. The optimal values for these new decision parameters—which can be obtained using conventional dynamic optimization techniques—then yield the optimal switching times for the original system. It is widely claimed in the literature that the time-scaling approach is superior to the direct optimization approach. However, the reasons given for its superiority are often vague, and sometimes incorrect. In this paper, we rigorously explicate the major advantages of the time-scaling transformation. We also compare the time-scaling and direct optimization approaches by solving a trajectory optimization problem involving the classical Dubins vehicle model.

Key words: *switched system, switching times, time-scaling transformation, gradient-based optimization*

Mathematics Subject Classification: *49M37, 90C30, 93C65*

1 Introduction

Switched systems are dynamic systems that evolve by switching between different subsystems or modes. Such systems arise frequently in critical applications such as robotics [4], sensor scheduling [22], electric power converters [13], and fed-batch fermentation processes [11].

In general, a switched system is controlled by varying the *switching times* (the times at which the system switches from one mode to another) and the *switching sequence* (the sequence in which the system assumes the various available modes). In this paper, we focus on the special case in which only the switching times can be varied; the switching sequence is assumed to be fixed and known. We investigate the problem of choosing optimal values

*The first author acknowledges support from the National Natural Science Foundation of China through the International Young Scientists Research Fellowship program (NSFC grant number 11350110208).

†The first and second authors acknowledge support from the State Key Laboratory for Industrial Control Technology at Zhejiang University, China (Open Research Project ICT1301).

for the switching times so that the switched system under consideration operates in the best possible manner. This problem is called the *switching time optimization problem* (STO problem).

The STO problem is a type of nonlinear optimization problem in which a finite number of decision variables (the switching times) need to be chosen to minimize a given cost function (typically a function of the final state reached by the system). In principle, this problem can be solved using standard nonlinear optimization methods—for example, sequential quadratic programming. However, such methods rely on the gradient of the cost function, which is difficult to compute in the STO problem because of the presence of dynamic constraints. In fact, it is often claimed in the literature that the gradient of the cost function in the STO problem is "discontinuous" [5, 8, 17] and "not effective for numerical computation" [9, 12, 14]. For this reason, a novel *time-scaling transformation* was developed in [8] for converting the STO problem into an equivalent problem that is easier to solve. The time-scaling transformation is now widely used to optimize switching times in a variety of different problem settings [6, 13, 22, 23].

Despite its popularity, the rationale for applying the time-scaling transformation is poorly understood in the literature. The common perception is that the time-scaling transformation is necessary because "the gradient of the cost function [in the STO problem] is discontinuous" [8], yet there is no proof of this claim in the literature. In fact, as we will show later, this claim is actually false—a surprising result that raises the issue of whether the time-scaling transformation really is superior to the direct optimization approach. The purpose of this paper is to investigate this important issue. We will show that the time-scaling transformation is indeed superior to direct optimization, but for different reasons than those commonly given in the literature.

2 The Switching Time Optimization Problem

A typical nonlinear switched system can be described mathematically as follows:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}^i(t, \boldsymbol{x}(t)), \quad t \in (\tau_{i-1}, \tau_i), \quad i = 1, \ldots, p, \tag{2.1}$$

$$\boldsymbol{x}(\tau_i^+) = \boldsymbol{x}(\tau_i^-), \quad i = 1, \ldots, p-1, \tag{2.2}$$

$$\boldsymbol{x}(0^+) = \boldsymbol{x}^0, \tag{2.3}$$

where $p \geq 2$ is the number of system modes; $\boldsymbol{x}(t) \in \mathbb{R}^n$ is the *state vector*; $\tau_0 = 0$ is the *initial time*; τ_i, $i = 1, \ldots, p-1$, are the *switching times* or *switching instants*; $\tau_p = T > 0$ is a given *terminal time*; $\boldsymbol{x}^0 \in \mathbb{R}^n$ is a given *initial state*; and $\boldsymbol{f}^i : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$, $i = 1, \ldots, p$, are given functions.

System (2.1)-(2.3) is controlled by manipulating the switching times τ_i, $i = 1, \ldots, p-1$. These switching times must satisfy the following ordering constraints:

$$0 = \tau_0 \leq \tau_1 \leq \tau_2 \leq \cdots \leq \tau_{p-1} \leq \tau_p = T. \tag{2.4}$$

Note that it is possible to remove "non-optimal" subsystems in (2.1)-(2.3) by merging two or more adjacent switching times (this may be necessary when the optimal number of switches is unknown). For example, subsystem i can be removed by setting $\tau_{i-1} = \tau_i$.

Let $\mathcal{T}$ denote the set of all vectors $\boldsymbol{\tau} = [\tau_1, \ldots, \tau_{p-1}]^\top \in \mathbb{R}^{p-1}$ satisfying inequality (2.4). Such vectors are called *admissible switching time vectors*.

Throughout this paper, we assume that system (2.1)-(2.3) satisfies the following conditions.

Assumption 2.1. The given functions $\boldsymbol{f}^i : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$, $i = 1, \ldots, p$, are continuously differentiable.

Assumption 2.2. There exists a positive constant $L_1 > 0$ such that the following linear growth inequality is satisfied:

$$\|\boldsymbol{f}^i(\eta, \boldsymbol{\xi})\| \leq L_1(1 + |\eta| + \|\boldsymbol{\xi}\|), \quad (\eta, \boldsymbol{\xi}) \in \mathbb{R} \times \mathbb{R}^n, \quad i = 1, \ldots, p,$$

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

Assumptions 2.1 and 2.2 ensure that for each admissible switching time vector $\boldsymbol{\tau} \in \mathcal{T}$, there exists a unique absolutely continuous function $\boldsymbol{x}(\cdot|\boldsymbol{\tau}) : [0, T] \to \mathbb{R}^n$ satisfying the dynamics (2.1) almost everywhere, the intermediate conditions (2.2), and the initial condition (2.3) (see references [1,2] for the proof). The function $\boldsymbol{x}(\cdot|\boldsymbol{\tau})$ is called the *solution* or *state trajectory* of (2.1)-(2.3).

It follows from Lemma 6.4.2 in [20] that $\boldsymbol{x}(\cdot|\boldsymbol{\tau})$ is uniformly bounded on $[0, T]$ with respect to the switching time vector $\boldsymbol{\tau} \in \mathcal{T}$. Hence, there exists a positive constant $L_2 > 0$ such that

$$\|\boldsymbol{x}(t|\boldsymbol{\tau})\| \leq L_2, \quad t \in [0, T], \quad \boldsymbol{\tau} \in \mathcal{T}. \tag{2.5}$$

Therefore, by Assumption 2.1, there exists another positive constant $L_3 > 0$ such that

$$\left\|\boldsymbol{f}^i(t, \boldsymbol{x}(t|\boldsymbol{\tau}))\right\| \leq L_3, \quad t \in [0, T], \quad \boldsymbol{\tau} \in \mathcal{T}, \quad i = 1, \ldots, p, \tag{2.6}$$

and

$$\left\|\frac{\partial \boldsymbol{f}^i(t, \boldsymbol{x}(t|\boldsymbol{\tau}))}{\partial t}\right\| \leq L_3, \quad \left\|\frac{\partial \boldsymbol{f}^i(t, \boldsymbol{x}(t|\boldsymbol{\tau}))}{\partial \boldsymbol{x}}\right\| \leq L_3, \quad t \in [0, T], \quad \boldsymbol{\tau} \in \mathcal{T}, \quad i = 1, \ldots, p, \tag{2.7}$$

where $\|\cdot\|$ in the second inequality in (2.7) denotes the natural matrix norm on $\mathbb{R}^{n \times n}$. The bounds in (2.7) ensure that the functions $\boldsymbol{f}^i$, $i = 1, \ldots, p$, satisfy the following Lipschitz condition: for any $t_1, t_2 \in [0, T]$ and $\boldsymbol{\tau}_1, \boldsymbol{\tau}_2 \in \mathcal{T}$,

$$\left\|\boldsymbol{f}^i(t_1, \boldsymbol{x}(t_1|\boldsymbol{\tau}_1)) - \boldsymbol{f}^i(t_2, \boldsymbol{x}(t_2|\boldsymbol{\tau}_2))\right\| \leq L_3|t_1 - t_2| + L_3\left\|\boldsymbol{x}(t_1|\boldsymbol{\tau}_1) - \boldsymbol{x}(t_2|\boldsymbol{\tau}_2)\right\|, \quad i = 1, \ldots, p. \tag{2.8}$$

Our goal is to choose the switching time vector $\boldsymbol{\tau} \in \mathcal{T}$ to minimize the following cost function:

$$J(\boldsymbol{\tau}) = \Phi(\boldsymbol{x}(T|\boldsymbol{\tau})), \tag{2.9}$$

where $\Phi : \mathbb{R}^n \to \mathbb{R}$ is a given function. Thus, our STO problem can be stated formally as follows: *Find an admissible switching time vector $\boldsymbol{\tau}^* \in \mathcal{T}$ such that*

$$J(\boldsymbol{\tau}^*) = \inf_{\boldsymbol{\tau} \in \mathcal{T}} J(\boldsymbol{\tau}).$$

This STO problem can be easily generalized to include state jumps, nonlinear constraints, and additional control parameters in the functions $\boldsymbol{f}^i$, $i = 1, \ldots, p$. However, since we focus solely on the optimization of switching times in this paper, the simple formulation given here is sufficient for our purposes.

Before concluding this section, we make the following assumption regarding the cost function (2.9).

Assumption 2.3. The given function $\Phi : \mathbb{R}^n \to \mathbb{R}$ is continuously differentiable.

3 Direct Optimization Approach

The STO problem is a nonlinear optimization problem in which the cost function J depends *implicitly* on the decision vector $\boldsymbol{\tau}$. To compute the value of $J(\boldsymbol{\tau})$, we must first solve (2.1)-(2.3) for the given switching time vector $\boldsymbol{\tau}$, and then substitute the resulting final state $\boldsymbol{x}(T|\boldsymbol{\tau})$ into equation (2.9). Since $\boldsymbol{\tau}$ influences J implicitly through the governing switched system (2.1)-(2.3), computing the gradient of J is a difficult task. However, if an explicit formula for computing the gradient of J is known, then the STO problem can be solved via the following gradient-based optimization algorithm:

1. Choose an initial switching time vector $\boldsymbol{\tau} \in \mathcal{T}$.
2. Compute $J(\boldsymbol{\tau})$.
3. Compute $\partial J(\boldsymbol{\tau})/\partial \boldsymbol{\tau}$.
4. Use $J(\boldsymbol{\tau})$ (obtained in Step 2) and $\partial J(\boldsymbol{\tau})/\partial \boldsymbol{\tau}$ (obtained in Step 3) to determine an appropriate search direction.
5. Perform a line search along the search direction determined in Step 4 to obtain a new switching time vector $\boldsymbol{\tau}' \in \mathcal{T}$.
6. Set $\boldsymbol{\tau}' \to \boldsymbol{\tau}$ and return to Step 2.

There are many well-known methods for computing the search direction in Step 4 and performing the line search in Step 5 (see [15,16]). Thus, the main challenge with implementing the above algorithm is to determine the gradient of J in Step 3. As we mentioned in the introduction, it is often claimed in the literature that the gradient of J with respect to the switching times is not suitable for numerical computation. We will show in this section that these claims are incorrect: it is possible to derive implementable formulae for the gradient of J with respect to the switching times, and these formulae can be readily incorporated into the above algorithm to solve the STO problem effectively.

3.1 Gradient Computation: Variational Method

For each $j = 1, \ldots, p-1$, consider the following *variational system*:

$$\dot{\boldsymbol{\phi}}^j(t) = \frac{\partial \boldsymbol{f}^i(t, \boldsymbol{x}(t|\boldsymbol{\tau}))}{\partial \boldsymbol{x}} \boldsymbol{\phi}^j(t), \quad t \in (\tau_{i-1}, \tau_i), \quad i = j+1, \ldots, p, \tag{3.1}$$

$$\boldsymbol{\phi}^j(\tau_i^+) = \begin{cases} \boldsymbol{f}^j(\tau_j, \boldsymbol{x}(\tau_j|\boldsymbol{\tau})) - \boldsymbol{f}^{j+1}(\tau_j, \boldsymbol{x}(\tau_j|\boldsymbol{\tau})), & \text{if } i = j, \\ \boldsymbol{\phi}^j(\tau_i^-), & \text{if } i \in \{j+1, \ldots, p\} \text{ and } \tau_i > \tau_j, \end{cases} \tag{3.2}$$

$$\boldsymbol{\phi}^j(t) = \mathbf{0}, \quad t \in [0, \tau_j), \tag{3.3}$$

where $\tau_i^+ = T$ in (3.2) if $\tau_i = T$. By virtue of Assumptions 2.1 and 2.2, for each $\boldsymbol{\tau} \in \mathcal{T}$, there exists a unique right-continuous function $\boldsymbol{\phi}^j(\cdot|\boldsymbol{\tau}) : [0, T] \to \mathbb{R}^n$ satisfying the dynamics (3.1) almost everywhere, the intermediate conditions (3.2), and the initial condition (3.3) (see [1,2]).

The variational system (3.1)-(3.3) includes an instantaneous jump at $t = \tau_j$. Since $\boldsymbol{\phi}^j(\cdot|\boldsymbol{\tau})$ is right-continuous,

$$\boldsymbol{\phi}^j(\tau_j|\boldsymbol{\tau}) = \boldsymbol{\phi}^j(\tau_j^+|\boldsymbol{\tau}) = \boldsymbol{f}^j(\tau_j, \boldsymbol{x}(\tau_j|\boldsymbol{\tau})) - \boldsymbol{f}^{j+1}(\tau_j, \boldsymbol{x}(\tau_j|\boldsymbol{\tau})).$$

Thus, $\boldsymbol{\phi}^j(\cdot|\boldsymbol{\tau})$ assumes its right limit at $t = \tau_j$.

Recall that $\boldsymbol{x}(t|\boldsymbol{\tau})$, the state vector in the STO problem, is a function of both t and $\boldsymbol{\tau}$. Our goal in this section is to investigate the partial derivatives of the state vector with respect to the switching times. The following fundamental result shows that $\boldsymbol{\phi}^j(\cdot|\boldsymbol{\tau})$ coincides with the left partial derivative of the state vector with respect to τ_j.

Theorem 3.1. *Let $j \in \{1, \ldots, p-1\}$ and $\boldsymbol{\tau} = [\tau_1, \ldots, \tau_{p-1}]^\top \in \mathcal{T}$. Furthermore, suppose that $\tau_{j-1} < \tau_j$. Then*

$$\lim_{\epsilon \to 0-} \frac{\boldsymbol{x}(t|\boldsymbol{\tau} + \epsilon \boldsymbol{e}^j) - \boldsymbol{x}(t|\boldsymbol{\tau})}{\epsilon} = \boldsymbol{\phi}^j(t|\boldsymbol{\tau}), \quad t \in [0, T], \tag{3.4}$$

where $\boldsymbol{e}^j$ is the jth unit basis vector in $\mathbb{R}^{p-1}$.

Proof. We prove the theorem in three steps.

Step 1: Preliminaries

For notational simplicity, let $\boldsymbol{x}(\cdot)$ denote $\boldsymbol{x}(\cdot|\boldsymbol{\tau})$, and let $\boldsymbol{x}^\epsilon(\cdot)$ denote $\boldsymbol{x}(\cdot|\boldsymbol{\tau} + \epsilon \boldsymbol{e}^j)$. Furthermore, let Ω denote the set of all ϵ such that $\boldsymbol{\tau} + \epsilon \boldsymbol{e}^j \in \mathcal{T}$.

For each $\epsilon \in \Omega$, define

$$\boldsymbol{\varphi}^\epsilon(\eta) = \boldsymbol{x}^\epsilon(\eta) - \boldsymbol{x}(\eta), \quad \eta \in [0, T],$$

and

$$\boldsymbol{\omega}_i^\epsilon(\eta, \alpha) = \left\{ \frac{\partial \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta) + \alpha \boldsymbol{\varphi}^\epsilon(\eta))}{\partial \boldsymbol{x}} - \frac{\partial \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta))}{\partial \boldsymbol{x}} \right\} \boldsymbol{\varphi}^\epsilon(\eta), \quad (\eta, \alpha) \in [0, T] \times [0, 1], \quad i = 1, \ldots, p.$$

In Appendix A, we show that there exists a positive constant $M_1 > 0$ such that for all $\epsilon \in \Omega$,

$$\|\boldsymbol{\varphi}^\epsilon(\eta)\| \le M_1 |\epsilon|, \quad \eta \in [0, T]. \tag{3.5}$$

Note that $\partial \boldsymbol{f}^i / \partial \boldsymbol{x}$ is continuous (recall Assumption 2.1) and $\boldsymbol{x}^\epsilon$ is uniformly bounded with respect to ϵ (recall inequality (2.5)). Hence, there exists a compact convex set $\mathcal{X} \subset \mathbb{R}^n$ such that $\boldsymbol{x}^\epsilon(\eta) \in \mathcal{X}$ for all $\eta \in [0, T]$ and $\epsilon \in \Omega$. It follows that $\partial \boldsymbol{f}^i / \partial \boldsymbol{x}$ is uniformly continuous on $[0, T] \times \mathcal{X}$.

Now, let $\gamma > 0$ be arbitrary. Since $\partial \boldsymbol{f}^i / \partial \boldsymbol{x}$ is uniformly continuous on $[0, T] \times \mathcal{X}$ and $\boldsymbol{\varphi}^\epsilon \to \boldsymbol{0}$ uniformly on $[0, T]$ as $\epsilon \to 0$ (recall inequality (3.5)), there exists a positive constant $\epsilon' > 0$ such that for all $\epsilon \in \Omega$ with $|\epsilon| < \epsilon'$,

$$\left\| \frac{\partial \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta) + \alpha \boldsymbol{\varphi}^\epsilon(\eta))}{\partial \boldsymbol{x}} - \frac{\partial \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta))}{\partial \boldsymbol{x}} \right\| < \gamma, \quad (\eta, \alpha) \in [0, T] \times [0, 1], \quad i = 1, \ldots, p.$$

Thus, from (3.5),

$$\begin{aligned} \|\boldsymbol{\omega}_i^\epsilon(\eta, \alpha)\| \le \left\| \frac{\partial \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta) + \alpha \boldsymbol{\varphi}^\epsilon(\eta))}{\partial \boldsymbol{x}} - \frac{\partial \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta))}{\partial \boldsymbol{x}} \right\| \cdot \|\boldsymbol{\varphi}^\epsilon(\eta)\| \le M_1 \gamma |\epsilon|, \\ (\eta, \alpha) \in [0, T] \times [0, 1], \quad i = 1, \ldots, p, \end{aligned} \tag{3.6}$$

which holds whenever $\epsilon \in \Omega$ satisfies $|\epsilon| < \epsilon'$. Now, for each $\epsilon \in \Omega \cap (-\infty, 0]$, define

$$\begin{aligned}\boldsymbol{\rho}_j^\epsilon &= \int_{\tau_j+\epsilon}^{\tau_j} \left\{\boldsymbol{f}^j(\eta, \boldsymbol{x}(\eta)) - \boldsymbol{f}^j(\tau_j, \boldsymbol{x}(\tau_j))\right\} d\eta, \\ \boldsymbol{\rho}_{j+1}^\epsilon &= \int_{\tau_j+\epsilon}^{\tau_j} \left\{\boldsymbol{f}^{j+1}(\eta, \boldsymbol{x}^\epsilon(\eta)) - \boldsymbol{f}^{j+1}(\tau_j, \boldsymbol{x}(\tau_j))\right\} d\eta.\end{aligned}$$

In Appendix B, we show that there exists a positive constant $M_2 > 0$ such that

$$\|\boldsymbol{\rho}_j^\epsilon\| \le M_2 \epsilon^2, \quad \|\boldsymbol{\rho}_{j+1}^\epsilon\| \le M_2 \epsilon^2. \tag{3.7}$$

We now complete the proof of (3.4) by considering two cases: (i) $t < \tau_j$; and (ii) $t \ge \tau_j$.

Step 2: Case $t < \tau_j$

If $t < \tau_j$, then for all $\epsilon < 0$ of sufficiently small magnitude,

$$\boldsymbol{x}^\epsilon(t) = \boldsymbol{x}(t).$$

Thus, by (3.3),

$$\lim_{\epsilon \to 0-} \frac{\boldsymbol{x}^\epsilon(t) - \boldsymbol{x}(t)}{\epsilon} = \mathbf{0} = \boldsymbol{\phi}^j(t|\boldsymbol{\tau}).$$

This proves equation (3.4) for $t < \tau_j$.

Step 3: Case $t \ge \tau_j$

Let $\varsigma \in \{j+1, \ldots, p\}$ be such that $t \in [\tau_{\varsigma-1}, \tau_\varsigma]$. Furthermore, choose ϵ so that

$$\max\{-\gamma, -\epsilon', \tau_{j-1} - \tau_j\} < \epsilon < 0, \tag{3.8}$$

where γ and ϵ' are as defined in Step 1.

It is clear that $\epsilon \in \Omega$. Thus, since $\tau_{j-1} < \tau_j + \epsilon < \tau_j$,

$$\boldsymbol{x}(t) = \boldsymbol{x}(\tau_j + \epsilon) + \int_{\tau_j+\epsilon}^{\tau_j} \boldsymbol{f}^j(\eta, \boldsymbol{x}(\eta)) d\eta + \sum_{i=j+1}^{\varsigma} \int_{\tau_{i-1}}^{\min\{t, \tau_i\}} \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta)) d\eta \tag{3.9}$$

and

$$\boldsymbol{x}^\epsilon(t) = \boldsymbol{x}^\epsilon(\tau_j + \epsilon) + \sum_{i=j+1}^{\varsigma} \int_{\tau_{i-1}+\epsilon\delta_{i,j+1}}^{\min\{t, \tau_i\}} \boldsymbol{f}^i(\eta, \boldsymbol{x}^\epsilon(\eta)) d\eta, \tag{3.10}$$

where $\delta_{i,j+1}$ denotes the Kronecker delta function defined by

$$\delta_{i,j+1} = \begin{cases} 1, & \text{if } i = j+1, \\ 0, & \text{otherwise.} \end{cases}$$

Equation (3.10) can be written as

$$\boldsymbol{x}^\epsilon(t) = \boldsymbol{x}^\epsilon(\tau_j + \epsilon) + \int_{\tau_j+\epsilon}^{\tau_j} \boldsymbol{f}^{j+1}(\eta, \boldsymbol{x}^\epsilon(\eta)) d\eta + \sum_{i=j+1}^{\varsigma} \int_{\tau_{i-1}}^{\min\{t, \tau_i\}} \boldsymbol{f}^i(\eta, \boldsymbol{x}^\epsilon(\eta)) d\eta. \tag{3.11}$$

Since $\epsilon < 0$, it is easy to see that $\boldsymbol{x}^{\epsilon}(\tau_j+\epsilon) = \boldsymbol{x}(\tau_j+\epsilon)$. Thus, combining equations (3.9) and (3.11) gives

$$\begin{aligned}\boldsymbol{\varphi}^{\epsilon}(t) &= \boldsymbol{x}^{\epsilon}(t) - \boldsymbol{x}(t)\\ &= \int_{\tau_j+\epsilon}^{\tau_j} \big\{\boldsymbol{f}^{j+1}(\eta, \boldsymbol{x}^{\epsilon}(\eta)) - \boldsymbol{f}^{j}(\eta, \boldsymbol{x}(\eta))\big\}d\eta\\ &\quad + \sum_{i=j+1}^{\varsigma} \int_{\tau_{i-1}}^{\min\{t,\tau_i\}} \big\{\boldsymbol{f}^{i}(\eta, \boldsymbol{x}^{\epsilon}(\eta)) - \boldsymbol{f}^{i}(\eta, \boldsymbol{x}(\eta))\big\}d\eta.\end{aligned}$$

Therefore,

$$\begin{aligned}\boldsymbol{\varphi}^{\epsilon}(t) = \boldsymbol{\rho}_{j+1}^{\epsilon} - \boldsymbol{\rho}_{j}^{\epsilon} - \epsilon \boldsymbol{f}^{j+1}(\tau_j, \boldsymbol{x}(\tau_j)) + \epsilon \boldsymbol{f}^{j}(\tau_j, \boldsymbol{x}(\tau_j))&\\ + \sum_{i=j+1}^{\varsigma} \int_{\tau_{i-1}}^{\min\{t,\tau_i\}} \big\{\boldsymbol{f}^{i}(\eta, \boldsymbol{x}^{\epsilon}(\eta)) - \boldsymbol{f}^{i}(\eta, \boldsymbol{x}(\eta))\big\}d\eta,&\end{aligned} \tag{3.12}$$

where $\boldsymbol{\rho}_{j}^{\epsilon}$ and $\boldsymbol{\rho}_{j+1}^{\epsilon}$ are as defined in Step 1.

Using the fundamental theorem of calculus, equation (3.12) can be written as

$$\begin{aligned}\boldsymbol{\varphi}^{\epsilon}(t) = \boldsymbol{\rho}_{j+1}^{\epsilon} - \boldsymbol{\rho}_{j}^{\epsilon} - \epsilon \boldsymbol{f}^{j+1}(\tau_j, \boldsymbol{x}(\tau_j)) + \epsilon \boldsymbol{f}^{j}(\tau_j, \boldsymbol{x}(\tau_j))&\\ + \sum_{i=j+1}^{\varsigma} \int_{\tau_{i-1}}^{\min\{t,\tau_i\}} \int_0^1 \frac{\partial}{\partial \alpha}\big\{\boldsymbol{f}^{i}(\eta, \boldsymbol{x}(\eta) + \alpha \boldsymbol{\varphi}^{\epsilon}(\eta))\big\}d\alpha d\eta.&\end{aligned}$$

Hence,

$$\begin{aligned}\boldsymbol{\varphi}^{\epsilon}(t) = \boldsymbol{\rho}_{j+1}^{\epsilon} - \boldsymbol{\rho}_{j}^{\epsilon} - \epsilon \boldsymbol{f}^{j+1}(\tau_j, \boldsymbol{x}(\tau_j)) + \epsilon \boldsymbol{f}^{j}(\tau_j, \boldsymbol{x}(\tau_j))&\\ + \sum_{i=j+1}^{\varsigma} \int_{\tau_{i-1}}^{\min\{t,\tau_i\}} \int_0^1 \frac{\partial \boldsymbol{f}^{i}(\eta, \boldsymbol{x}(\eta) + \alpha \boldsymbol{\varphi}^{\epsilon}(\eta))}{\partial \boldsymbol{x}} \boldsymbol{\varphi}^{\epsilon}(\eta) d\alpha d\eta.&\end{aligned}$$

Rearranging this equation gives

$$\begin{aligned}\boldsymbol{\varphi}^{\epsilon}(t) = \boldsymbol{\rho}_{j+1}^{\epsilon} - \boldsymbol{\rho}_{j}^{\epsilon} - \epsilon \boldsymbol{f}^{j+1}(\tau_j, \boldsymbol{x}(\tau_j)) + \epsilon \boldsymbol{f}^{j}(\tau_j, \boldsymbol{x}(\tau_j))&\\ + \sum_{i=j+1}^{\varsigma} \int_{\tau_{i-1}}^{\min\{t,\tau_i\}} \frac{\partial \boldsymbol{f}^{i}(\eta, \boldsymbol{x}(\eta))}{\partial \boldsymbol{x}} \boldsymbol{\varphi}^{\epsilon}(\eta) d\eta&\\ + \sum_{i=j+1}^{\varsigma} \int_{\tau_{i-1}}^{\min\{t,\tau_i\}} \int_0^1 \boldsymbol{\omega}_i^{\epsilon}(\eta, \alpha) d\alpha d\eta,&\end{aligned} \tag{3.13}$$

where $\boldsymbol{\omega}_i^{\epsilon}$ is as defined in Step 1. Now, the right-continuous solution of the variational system (3.1)-(3.3) is

$$\boldsymbol{\phi}^{j}(t) = \boldsymbol{f}^{j}(\tau_j, \boldsymbol{x}(\tau_j)) - \boldsymbol{f}^{j+1}(\tau_j, \boldsymbol{x}(\tau_j)) + \sum_{i=j+1}^{\varsigma} \int_{\tau_{i-1}}^{\min\{t,\tau_i\}} \frac{\partial \boldsymbol{f}^{i}(\eta, \boldsymbol{x}(\eta))}{\partial \boldsymbol{x}} \boldsymbol{\phi}^{j}(\eta) d\eta, \tag{3.14}$$

where $\boldsymbol{\phi}^{j}(\cdot) = \boldsymbol{\phi}^{j}(\cdot|\boldsymbol{\tau})$. Combining (3.13) and (3.14) gives

$$\begin{aligned}\big\|\epsilon^{-1}\boldsymbol{\varphi}^{\epsilon}(t) - \boldsymbol{\phi}^{j}(t)\big\| \le |\epsilon|^{-1}\|\boldsymbol{\rho}_{j}^{\epsilon}\| + |\epsilon|^{-1}\|\boldsymbol{\rho}_{j+1}^{\epsilon}\| + \int_{\tau_j}^{t} L_3 \big\|\epsilon^{-1}\boldsymbol{\varphi}^{\epsilon}(\eta) - \boldsymbol{\phi}^{j}(\eta)\big\| d\eta&\\ + \sum_{i=j+1}^{\varsigma} \int_{\tau_{i-1}}^{\min\{t,\tau_i\}} \int_0^1 |\epsilon|^{-1}\big\|\boldsymbol{\omega}_i^{\epsilon}(\eta, \alpha)\big\| d\alpha d\eta,&\end{aligned}$$

where L_3 is as defined in (2.7). Since $|\epsilon| < \epsilon'$ and $|\epsilon| < \gamma$ (see (3.8)), we can apply inequalities (3.6) and (3.7) to obtain

$$\begin{aligned}\big\|\epsilon^{-1}\boldsymbol{\varphi}^{\epsilon}(t) - \boldsymbol{\phi}^{j}(t)\big\| &\le 2M_2|\epsilon| + M_1T\gamma + \int_{\tau_j}^{t} L_3\big\|\epsilon^{-1}\boldsymbol{\varphi}^{\epsilon}(\eta) - \boldsymbol{\phi}^{j}(\eta)\big\| d\eta \\ &\le 2M_2\gamma + M_1T\gamma + \int_{\tau_j}^{t} L_3\big\|\epsilon^{-1}\boldsymbol{\varphi}^{\epsilon}(\eta) - \boldsymbol{\phi}^{j}(\eta)\big\| d\eta.\end{aligned}$$

This inequality holds for any $t \ge \tau_j$, uniformly with respect to ϵ. Thus, from the Gronwall-Bellman inequality,

$$\big\|\epsilon^{-1}\boldsymbol{\varphi}^{\epsilon}(t) - \boldsymbol{\phi}^{j}(t)\big\| \le (2M_2 + M_1T)\exp(L_3T)\gamma,$$

which holds for all ϵ satisfying (3.8). Since $\gamma > 0$ was chosen arbitrarily, equation (3.4) follows. □

According to Theorem 3.1, the left partial derivative of the state vector with respect to τ_j is equal to the solution of the variational system (3.1)-(3.3). Note that the condition $\tau_{j-1} < \tau_j$ in Theorem 3.1 is essential: it ensures that the perturbed switching time vector $\boldsymbol{\tau} + \epsilon\boldsymbol{e}^j$ satisfies the ordering constraints (2.4), and is therefore admissible for the STO problem.

Our next result gives the right partial derivative of the state vector with respect to τ_j.

Theorem 3.2. *Let $j \in \{1,\dots,p-1\}$ and $\boldsymbol{\tau} = [\tau_1,\dots,\tau_{p-1}]^\top \in \mathcal{T}$. Furthermore, suppose that $\tau_j < \tau_{j+1}$. Then*

$$\lim_{\epsilon\to 0+} \frac{\boldsymbol{x}(t|\boldsymbol{\tau} + \epsilon\boldsymbol{e}^j) - \boldsymbol{x}(t|\boldsymbol{\tau})}{\epsilon} = \boldsymbol{\phi}^j(t|\boldsymbol{\tau}), \quad t \in [0,T] \setminus \{\tau_j\}, \tag{3.15}$$

and

$$\lim_{\epsilon\to 0+} \frac{\boldsymbol{x}(\tau_j|\boldsymbol{\tau} + \epsilon\boldsymbol{e}^j) - \boldsymbol{x}(\tau_j|\boldsymbol{\tau})}{\epsilon} = \mathbf{0}, \tag{3.16}$$

where $\boldsymbol{e}^j$ is the jth unit basis vector in $\mathbb{R}^{p-1}$.

Proof. As with the proof of Theorem 3.1, we prove Theorem 3.2 in three steps.

Step 1: Preliminaries

Let Ω, $\boldsymbol{x}(\cdot)$, $\boldsymbol{x}^{\epsilon}(\cdot)$, $\boldsymbol{\varphi}^{\epsilon}(\cdot)$, and $\boldsymbol{\omega}_i^{\epsilon}(\cdot,\cdot)$ be as defined in the proof of Theorem 3.1. Recall that for all $\epsilon \in \Omega$,

$$\|\boldsymbol{\varphi}^{\epsilon}(\eta)\| \le M_1|\epsilon|, \quad \eta \in [0,T], \tag{3.17}$$

where $M_1 > 0$ is a constant independent of ϵ.

Also recall that for each $\gamma > 0$, there exists a corresponding real number $\epsilon' > 0$ such that for all $\epsilon \in \Omega$ satisfying $|\epsilon| < \epsilon'$,

$$\|\boldsymbol{\omega}_i^{\epsilon}(\eta,\alpha)\| \le M_1\gamma|\epsilon|, \quad (\eta,\alpha) \in [0,T]\times[0,1], \quad i = 1,\dots,p. \tag{3.18}$$

Now, for each $\epsilon \in \Omega \cap [0,\infty)$, define

$$\begin{aligned}\boldsymbol{\rho}_j^{\epsilon} &= \int_{\tau_j}^{\tau_j+\epsilon} \big\{\boldsymbol{f}^j(\eta,\boldsymbol{x}^{\epsilon}(\eta)) - \boldsymbol{f}^j(\tau_j,\boldsymbol{x}(\tau_j))\big\} d\eta, \\ \boldsymbol{\rho}_{j+1}^{\epsilon} &= \int_{\tau_j}^{\tau_j+\epsilon} \big\{\boldsymbol{f}^{j+1}(\eta,\boldsymbol{x}^{\epsilon}(\eta)) - \boldsymbol{f}^{j+1}(\tau_j,\boldsymbol{x}(\tau_j))\big\} d\eta.\end{aligned}$$

Notice that these definitions for $\boldsymbol{\rho}_j^\epsilon$ and $\boldsymbol{\rho}_{j+1}^\epsilon$ are slightly different from the definitions given in the proof of Theorem 3.1. Nevertheless, inequality (3.7) still holds:

$$\|\boldsymbol{\rho}_j^\epsilon\| \leq M_2\epsilon^2, \quad \|\boldsymbol{\rho}_{j+1}^\epsilon\| \leq M_2\epsilon^2, \tag{3.19}$$

where $M_2 > 0$ is a constant independent of ϵ.

We now complete the proof by considering two cases: (i) $t \leq \tau_j$; and (ii) $t > \tau_j$.

Step 2: Case $t \leq \tau_j$

If $t \leq \tau_j$, then for all $\epsilon \in (0, \tau_{j+1} - \tau_j)$,

$$\boldsymbol{x}^\epsilon(t) = \boldsymbol{x}(t).$$

Therefore,

$$\lim_{\epsilon \to 0+} \frac{\boldsymbol{x}^\epsilon(t) - \boldsymbol{x}(t)}{\epsilon} = \mathbf{0}.$$

Equations (3.15) and (3.16) then follow immediately.

Step 3: Case $t > \tau_j$

For each $s > \tau_j$, let $\varsigma(s) \in \{j+1, \ldots, p\}$ denote the unique integer such that $s \in (\tau_{\varsigma(s)-1}, \tau_{\varsigma(s)}]$. Furthermore, let $\gamma \in (0, t - \tau_j)$ be arbitrary but fixed and choose ϵ so that

$$0 < \epsilon < \min\{\gamma, \epsilon', \tau_{j+1} - \tau_j\}, \tag{3.20}$$

where ϵ' is as defined in Step 1. Then clearly $\epsilon \in \Omega$ and $\tau_j + \epsilon < \tau_j + \gamma < t$. Thus, for all $s \in [\tau_j + \gamma, t]$,

$$\boldsymbol{x}(s) = \boldsymbol{x}(\tau_j) + \sum_{i=j+1}^{\varsigma(s)} \int_{\tau_{i-1}}^{\min\{s,\tau_i\}} \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta))d\eta \tag{3.21}$$

and

$$\boldsymbol{x}^\epsilon(s) = \boldsymbol{x}^\epsilon(\tau_j) + \int_{\tau_j}^{\tau_j+\epsilon} \boldsymbol{f}^j(\eta, \boldsymbol{x}^\epsilon(\eta))d\eta + \sum_{i=j+1}^{\varsigma(s)} \int_{\tau_{i-1}+\epsilon\delta_{i,j+1}}^{\min\{s,\tau_i\}} \boldsymbol{f}^i(\eta, \boldsymbol{x}^\epsilon(\eta))d\eta, \tag{3.22}$$

where $\delta_{i,j+1}$ is the Kronecker delta function defined in the proof of Theorem 3.1. Equation (3.22) can be written as

$$\boldsymbol{x}^\epsilon(s) = \boldsymbol{x}^\epsilon(\tau_j) + \int_{\tau_j}^{\tau_j+\epsilon} \big\{\boldsymbol{f}^j(\eta, \boldsymbol{x}^\epsilon(\eta)) - \boldsymbol{f}^{j+1}(\eta, \boldsymbol{x}^\epsilon(\eta))\big\}d\eta + \sum_{i=j+1}^{\varsigma(s)} \int_{\tau_{i-1}}^{\min\{s,\tau_i\}} \boldsymbol{f}^i(\eta, \boldsymbol{x}^\epsilon(\eta))d\eta. \tag{3.23}$$

Since $\boldsymbol{x}^\epsilon(\tau_j) = \boldsymbol{x}(\tau_j)$, subtracting (3.21) from (3.23) gives

$$\begin{aligned}\boldsymbol{\varphi}^\epsilon(s) \quad = \quad & \int_{\tau_j}^{\tau_j+\epsilon} \big\{\boldsymbol{f}^j(\eta, \boldsymbol{x}^\epsilon(\eta)) - \boldsymbol{f}^{j+1}(\eta, \boldsymbol{x}^\epsilon(\eta))\big\}d\eta \\ & + \sum_{i=j+1}^{\varsigma(s)} \int_{\tau_{i-1}}^{\min\{s,\tau_i\}} \big\{\boldsymbol{f}^i(\eta, \boldsymbol{x}^\epsilon(\eta)) - \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta))\big\}d\eta.\end{aligned}$$

Therefore,

$$\begin{aligned}\boldsymbol{\varphi}^{\epsilon}(s) &= \boldsymbol{\rho}_j^{\epsilon} - \boldsymbol{\rho}_{j+1}^{\epsilon} + \epsilon \boldsymbol{f}^j(\tau_j, \boldsymbol{x}(\tau_j)) - \epsilon \boldsymbol{f}^{j+1}(\tau_j, \boldsymbol{x}(\tau_j)) \\ &\quad + \sum_{i=j+1}^{\varsigma(s)} \int_{\tau_{i-1}}^{\min\{s,\tau_i\}} \big\{ \boldsymbol{f}^i(\eta, \boldsymbol{x}^{\epsilon}(\eta)) - \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta)) \big\} d\eta,\end{aligned} \tag{3.24}$$

where $\boldsymbol{\rho}_j^{\epsilon}$ and $\boldsymbol{\rho}_{j+1}^{\epsilon}$ are as defined in Step 1. Using the fundamental theorem of calculus, equation (3.24) can be written as

$$\begin{aligned}\boldsymbol{\varphi}^{\epsilon}(s) &= \boldsymbol{\rho}_j^{\epsilon} - \boldsymbol{\rho}_{j+1}^{\epsilon} + \epsilon \boldsymbol{f}^j(\tau_j, \boldsymbol{x}(\tau_j)) - \epsilon \boldsymbol{f}^{j+1}(\tau_j, \boldsymbol{x}(\tau_j)) \\ &\quad + \sum_{i=j+1}^{\varsigma(s)} \int_{\tau_{i-1}}^{\min\{s,\tau_i\}} \int_0^1 \frac{\partial}{\partial \alpha} \Big\{ \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta) + \alpha \boldsymbol{\varphi}^{\epsilon}(\eta)) \Big\} d\alpha d\eta.\end{aligned}$$

Hence,

$$\begin{aligned}\boldsymbol{\varphi}^{\epsilon}(s) &= \boldsymbol{\rho}_j^{\epsilon} - \boldsymbol{\rho}_{j+1}^{\epsilon} + \epsilon \boldsymbol{f}^j(\tau_j, \boldsymbol{x}(\tau_j)) - \epsilon \boldsymbol{f}^{j+1}(\tau_j, \boldsymbol{x}(\tau_j)) \\ &\quad + \sum_{i=j+1}^{\varsigma(s)} \int_{\tau_{i-1}}^{\min\{s,\tau_i\}} \int_0^1 \frac{\partial \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta) + \alpha \boldsymbol{\varphi}^{\epsilon}(\eta))}{\partial \boldsymbol{x}} \boldsymbol{\varphi}^{\epsilon}(\eta) d\alpha d\eta.\end{aligned}$$

Rearranging gives

$$\begin{aligned}\boldsymbol{\varphi}^{\epsilon}(s) &= \boldsymbol{\rho}_j^{\epsilon} - \boldsymbol{\rho}_{j+1}^{\epsilon} + \epsilon \boldsymbol{f}^j(\tau_j, \boldsymbol{x}(\tau_j)) - \epsilon \boldsymbol{f}^{j+1}(\tau_j, \boldsymbol{x}(\tau_j)) \\ &\quad + \sum_{i=j+1}^{\varsigma(s)} \int_{\tau_{i-1}}^{\min\{s,\tau_i\}} \frac{\partial \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta))}{\partial \boldsymbol{x}} \boldsymbol{\varphi}^{\epsilon}(\eta) d\eta \\ &\quad + \sum_{i=j+1}^{\varsigma(s)} \int_{\tau_{i-1}}^{\min\{s,\tau_i\}} \int_0^1 \boldsymbol{\omega}_i^{\epsilon}(\eta, \alpha) d\alpha d\eta,\end{aligned} \tag{3.25}$$

where $\boldsymbol{\omega}_i^{\epsilon}$ is as defined in the proof of Theorem 3.1. Now, for any $s \in [\tau_j + \gamma, t]$,

$$\boldsymbol{\phi}^j(s) = \boldsymbol{\phi}^j(\tau_j^+) + \sum_{i=j+1}^{\varsigma(s)} \int_{\tau_{i-1}}^{\min\{s,\tau_i\}} \frac{\partial \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta))}{\partial \boldsymbol{x}} \boldsymbol{\phi}^j(\eta) d\eta,$$

where $\boldsymbol{\phi}^j(\cdot) = \boldsymbol{\phi}^j(\cdot|\boldsymbol{\tau})$. Thus, from the intermediate conditions (3.2),

$$\boldsymbol{\phi}^j(s) = \boldsymbol{f}^j(\tau_j, \boldsymbol{x}(\tau_j)) - \boldsymbol{f}^{j+1}(\tau_j, \boldsymbol{x}(\tau_j)) + \sum_{i=j+1}^{\varsigma(s)} \int_{\tau_{i-1}}^{\min\{s,\tau_i\}} \frac{\partial \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta))}{\partial \boldsymbol{x}} \boldsymbol{\phi}^j(\eta) d\eta. \tag{3.26}$$

Combining (3.25) and (3.26) yields

$$\begin{aligned}\big\|\epsilon^{-1}\boldsymbol{\varphi}^{\epsilon}(s) - \boldsymbol{\phi}^j(s)\big\| &\le \epsilon^{-1}\|\boldsymbol{\rho}_j^{\epsilon}\| + \epsilon^{-1}\|\boldsymbol{\rho}_{j+1}^{\epsilon}\| + \int_{\tau_j}^{s} L_3 \big\|\epsilon^{-1}\boldsymbol{\varphi}^{\epsilon}(\eta) - \boldsymbol{\phi}^j(\eta)\big\| d\eta \\ &\quad + \sum_{i=j+1}^{\varsigma(s)} \int_{\tau_{i-1}}^{\min\{s,\tau_i\}} \int_0^1 \epsilon^{-1} \big\|\boldsymbol{\omega}_i^{\epsilon}(\eta, \alpha)\big\| d\alpha d\eta,\end{aligned}$$

where $L_3 > 0$ is as defined in (2.7). Hence, since ϵ satisfies (3.20), it follows from inequalities (3.18) and (3.19) that

$$\left\|\epsilon^{-1}\boldsymbol{\varphi}^{\epsilon}(s) - \boldsymbol{\phi}^{j}(s)\right\| \leq 2M_2\epsilon + M_1T\gamma + \int_{\tau_j}^{s} L_3\left\|\epsilon^{-1}\boldsymbol{\varphi}^{\epsilon}(\eta) - \boldsymbol{\phi}^{j}(\eta)\right\|d\eta,$$

which holds for all $s \in [\tau_j + \gamma, t]$. Therefore, since $\epsilon < \gamma$,

$$\begin{aligned}\left\|\epsilon^{-1}\boldsymbol{\varphi}^{\epsilon}(s) - \boldsymbol{\phi}^{j}(s)\right\| \leq 2M_2\gamma + M_1T\gamma &+ \int_{\tau_j}^{\tau_j+\gamma} L_3\left\|\epsilon^{-1}\boldsymbol{\varphi}^{\epsilon}(\eta) - \boldsymbol{\phi}^{j}(\eta)\right\|d\eta \\ &+ \int_{\tau_j+\gamma}^{s} L_3\left\|\epsilon^{-1}\boldsymbol{\varphi}^{\epsilon}(\eta) - \boldsymbol{\phi}^{j}(\eta)\right\|d\eta.\end{aligned} \tag{3.27}$$

Now, using (3.17), we obtain

$$\left\|\epsilon^{-1}\boldsymbol{\varphi}^{\epsilon}(\eta) - \boldsymbol{\phi}^{j}(\eta)\right\| \leq \epsilon^{-1}\|\boldsymbol{\varphi}^{\epsilon}(\eta)\| + \|\boldsymbol{\phi}^{j}(\eta)\| \leq M_1 + M_3, \quad \eta \in [\tau_j, \tau_j + \gamma], \tag{3.28}$$

where $M_3 > 0$ is an upper bound for $\|\boldsymbol{\phi}^j\|$ (recall that $\boldsymbol{\phi}^j$ is piecewise continuous). Substituting (3.28) into (3.27) gives

$$\left\|\epsilon^{-1}\boldsymbol{\varphi}^{\epsilon}(s) - \boldsymbol{\phi}^{j}(s)\right\| \leq M_4\gamma + \int_{\tau_j+\gamma}^{s} L_3\left\|\epsilon^{-1}\boldsymbol{\varphi}^{\epsilon}(\eta) - \boldsymbol{\phi}^{j}(\eta)\right\|d\eta,$$

where $M_4 = 2M_2 + M_1T + L_3(M_1 + M_3)$. This inequality holds for all $s \in [\tau_j + \gamma, t]$. Thus, by the Gronwall-Bellman inequality,

$$\left\|\epsilon^{-1}\boldsymbol{\varphi}^{\epsilon}(s) - \boldsymbol{\phi}^{j}(s)\right\| \leq M_4 \exp(L_3T)\gamma, \quad s \in [\tau_j + \gamma, t].$$

Choosing $s = t$ gives

$$\left\|\epsilon^{-1}\boldsymbol{\varphi}^{\epsilon}(t) - \boldsymbol{\phi}^{j}(t)\right\| \leq M_4 \exp(L_3T)\gamma,$$

which holds for all ϵ satisfying (3.20). Since $\gamma > 0$ was chosen arbitrarily, this shows that (3.15) holds when $t > \tau_j$. ☐

By combining Theorems 3.1 and 3.2, we obtain the full partial derivative of the state vector with respect to τ_j. This partial derivative is called the *state variation*.

Theorem 3.3. *Let $j \in \{1, \dots, p-1\}$ and $\boldsymbol{\tau} = [\tau_1, \dots, \tau_{p-1}]^\top \in \mathcal{T}$. Suppose that $\tau_{j-1} < \tau_j < \tau_{j+1}$. Then*

$$\frac{\partial \boldsymbol{x}(t|\boldsymbol{\tau})}{\partial \tau_j} = \lim_{\epsilon \to 0} \frac{\boldsymbol{x}(t|\boldsymbol{\tau} + \epsilon \boldsymbol{e}^j) - \boldsymbol{x}(t|\boldsymbol{\tau})}{\epsilon} = \boldsymbol{\phi}^j(t|\boldsymbol{\tau}), \quad t \in [0, T] \setminus \{\tau_j\}.$$

Note that the state variation in Theorem 3.3 does not exist at $t = \tau_j$, as the left and right partial derivatives differ at this time point. Note also that the condition $\tau_{j-1} < \tau_j < \tau_{j+1}$ in Theorem 3.3 is necessary to ensure that the perturbed switching time vector $\boldsymbol{\tau} + \epsilon \boldsymbol{e}^j$ is admissible for the STO problem.

On the basis of Theorem 3.3, we are now able to derive the gradient the cost function in the STO problem.

Theorem 3.4. *Let* $\boldsymbol{\tau} = [\tau_1, \ldots, \tau_{p-1}]^\top \in \mathcal{T}$ *and suppose that* $\tau_{j-1} < \tau_j$, $j = 1, \ldots, p$. *Then*

$$\frac{\partial J(\boldsymbol{\tau})}{\partial \tau_j} = \frac{\partial \Phi(\boldsymbol{x}(T|\boldsymbol{\tau}))}{\partial \boldsymbol{x}} \boldsymbol{\phi}^j(T|\boldsymbol{\tau}), \quad j = 1, \ldots, p-1.$$

Proof. This result follows immediately from Theorem 3.3 and the chain rule of differentiation:

$$\frac{\partial J(\boldsymbol{\tau})}{\partial \tau_j} = \frac{\partial}{\partial \tau_j}\big\{\Phi(\boldsymbol{x}(T|\boldsymbol{\tau}))\big\} = \frac{\partial \Phi(\boldsymbol{x}(T|\boldsymbol{\tau}))}{\partial \boldsymbol{x}} \frac{\partial \boldsymbol{x}(T|\boldsymbol{\tau})}{\partial \tau_j} = \frac{\partial \Phi(\boldsymbol{x}(T|\boldsymbol{\tau}))}{\partial \boldsymbol{x}} \boldsymbol{\phi}^j(T|\boldsymbol{\tau}),$$

as required. □

In view of Theorem 3.4, we can compute the gradient of J at each $\boldsymbol{\tau} \in \mathcal{T}$ satisfying $\tau_{j-1} < \tau_j$, $j = 1, \ldots, p$, by invoking the following procedure:

1. Combine the original state system (2.1)-(2.3) with the variational systems (3.1)-(3.3) to form an enlarged switched system.

2. Solve the enlarged system constructed in Step 1 to obtain $\boldsymbol{x}(\cdot|\boldsymbol{\tau})$ and $\boldsymbol{\phi}^j(\cdot|\boldsymbol{\tau})$, $j = 1, \ldots, p-1$.

3. Use $\boldsymbol{x}(T|\boldsymbol{\tau})$ and $\boldsymbol{\phi}^j(T|\boldsymbol{\tau})$, $j = 1, \ldots, p-1$, to compute $\partial J/\partial \boldsymbol{\tau}$ via the formulae in Theorem 3.4.

This gradient computation method—called the *variational method*—can be used in conjunction with the gradient-based optimization algorithm given at the beginning of this section to determine an optimal switching time vector for the STO problem.

$\boxed{3.2}$ Gradient Computation: Costate Method

We now develop an alternative gradient computation procedure called the *costate method*. With the costate method, we define an auxiliary system called the *costate system* as follows:

$$\dot{\boldsymbol{\lambda}}(t) = -\left[\frac{\partial \boldsymbol{f}^i(t, \boldsymbol{x}(t|\boldsymbol{\tau}))}{\partial \boldsymbol{x}}\right]^\top \boldsymbol{\lambda}(t), \quad t \in (\tau_{i-1}, \tau_i), \quad i = 1, \ldots, p, \tag{3.29}$$

$$\boldsymbol{\lambda}(\tau_i^+) = \boldsymbol{\lambda}(\tau_i^-), \quad i = 1, \ldots, p-1, \tag{3.30}$$

$$\boldsymbol{\lambda}(T) = \left[\frac{\partial \Phi(\boldsymbol{x}(T|\boldsymbol{\tau}))}{\partial \boldsymbol{x}}\right]^\top. \tag{3.31}$$

Since the costate system involves a terminal condition rather than an initial condition, it must be solved backward in time starting from $t = T$. Let $\boldsymbol{\lambda}(\cdot|\boldsymbol{\tau}) : [0, T] \to \mathbb{R}^n$ denote the unique continuous solution of (3.29)-(3.31) corresponding to $\boldsymbol{\tau} \in \mathcal{T}$. We now derive the gradient of J in terms of $\boldsymbol{\lambda}(\cdot|\boldsymbol{\tau})$.

Theorem 3.5. *Let* $\boldsymbol{\tau} = [\tau_1, \ldots, \tau_{p-1}]^\top \in \mathcal{T}$ *and suppose that* $\tau_{j-1} < \tau_j$, $j = 1, \ldots, p$. *Then*

$$\frac{\partial J(\boldsymbol{\tau})}{\partial \tau_j} = \boldsymbol{\lambda}(\tau_j|\boldsymbol{\tau})^\top \boldsymbol{f}^j(\tau_j, \boldsymbol{x}(\tau_j|\boldsymbol{\tau})) - \boldsymbol{\lambda}(\tau_j|\boldsymbol{\tau})^\top \boldsymbol{f}^{j+1}(\tau_j, \boldsymbol{x}(\tau_j|\boldsymbol{\tau})), \quad j = 1, \ldots, p-1.$$

Proof. Let $\boldsymbol{v} : [0,T] \to \mathbb{R}^n$ denote an arbitrary absolutely continuous function. Then we can express the cost function J as follows:

$$J(\boldsymbol{\tau}) = \Phi(\boldsymbol{x}(T|\boldsymbol{\tau})) = \Phi(\boldsymbol{x}(T|\boldsymbol{\tau})) + \sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_i} \boldsymbol{v}(t)^\top \boldsymbol{f}^i(t, \boldsymbol{x}(t|\boldsymbol{\tau}))dt - \sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_i} \boldsymbol{v}(t)^\top \dot{\boldsymbol{x}}(t|\boldsymbol{\tau})dt.$$

Using integration by parts,

$$\begin{aligned} J(\boldsymbol{\tau}) = \Phi(\boldsymbol{x}(T|\boldsymbol{\tau})) &+ \sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_i} \boldsymbol{v}(t)^\top \boldsymbol{f}^i(t, \boldsymbol{x}(t|\boldsymbol{\tau}))dt \\ &- \sum_{i=1}^{p} \left\{ \boldsymbol{v}(\tau_i)^\top \boldsymbol{x}(\tau_i|\boldsymbol{\tau}) - \boldsymbol{v}(\tau_{i-1})^\top \boldsymbol{x}(\tau_{i-1}|\boldsymbol{\tau}) - \int_{\tau_{i-1}}^{\tau_i} \dot{\boldsymbol{v}}(t)^\top \boldsymbol{x}(t|\boldsymbol{\tau})dt \right\} \\ = \Phi(\boldsymbol{x}(T|\boldsymbol{\tau})) &- \boldsymbol{v}(T)^\top \boldsymbol{x}(T|\boldsymbol{\tau}) + \boldsymbol{v}(0)^\top \boldsymbol{x}^0 \\ &+ \sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_i} \left\{ \boldsymbol{v}(t)^\top \boldsymbol{f}^i(t, \boldsymbol{x}(t|\boldsymbol{\tau})) + \dot{\boldsymbol{v}}(t)^\top \boldsymbol{x}(t|\boldsymbol{\tau}) \right\} dt. \end{aligned}$$

Using the Leibniz rule, this equation can be differentiated with respect to τ_j to give

$$\begin{aligned} \frac{\partial J(\boldsymbol{\tau})}{\partial \tau_j} = &\frac{\partial \Phi(\boldsymbol{x}(T|\boldsymbol{\tau}))}{\partial \boldsymbol{x}} \frac{\partial \boldsymbol{x}(T|\boldsymbol{\tau})}{\partial \tau_j} - \boldsymbol{v}(T)^\top \frac{\partial \boldsymbol{x}(T|\boldsymbol{\tau})}{\partial \tau_j} \\ &+ \boldsymbol{v}(\tau_j)^\top \boldsymbol{f}^j(\tau_j, \boldsymbol{x}(\tau_j|\boldsymbol{\tau})) - \boldsymbol{v}(\tau_j)^\top \boldsymbol{f}^{j+1}(\tau_j, \boldsymbol{x}(\tau_j|\boldsymbol{\tau})) \\ &+ \sum_{i=1}^{p} \int_{\tau_{i-1}}^{\tau_i} \left\{ \boldsymbol{v}(t)^\top \frac{\partial \boldsymbol{f}^i(t, \boldsymbol{x}(t|\boldsymbol{\tau}))}{\partial \boldsymbol{x}} \frac{\partial \boldsymbol{x}(t|\boldsymbol{\tau})}{\partial \tau_j} + \dot{\boldsymbol{v}}(t)^\top \frac{\partial \boldsymbol{x}(t|\boldsymbol{\tau})}{\partial \tau_j} \right\} dt. \end{aligned}$$

Recall that $\boldsymbol{v} : [0,T] \to \mathbb{R}^n$ was chosen arbitrarily. Setting $\boldsymbol{v}(\cdot) = \boldsymbol{\lambda}(\cdot|\boldsymbol{\tau})$ and applying the costate equations (3.29)-(3.31), we obtain

$$\frac{\partial J(\boldsymbol{\tau})}{\partial \tau_j} = \boldsymbol{\lambda}(\tau_j|\boldsymbol{\tau})^\top \boldsymbol{f}^j(\tau_j, \boldsymbol{x}(\tau_j|\boldsymbol{\tau})) - \boldsymbol{\lambda}(\tau_j|\boldsymbol{\tau})^\top \boldsymbol{f}^{j+1}(\tau_j, \boldsymbol{x}(\tau_j|\boldsymbol{\tau})),$$

as required. □

Note that Theorem 3.5 gives similar gradient formulae to those derived in [20] (see Chapter 5) and [7] (see Proposition 1). However, the derivation of Theorem 3.5 is completely different to the derivations given in these references.

On the basis of Theorem 3.5, we now present the following *costate method* for computing the gradient of J at each $\boldsymbol{\tau} \in \mathcal{T}$ with $\tau_{j-1} < \tau_j$, $j = 1, \ldots, p$:

1. Solve the state system (2.1)-(2.3) forward in time to obtain $\boldsymbol{x}(\cdot|\boldsymbol{\tau})$.
2. Solve the costate system (3.29)-(3.31) backward in time to obtain $\boldsymbol{\lambda}(\cdot|\boldsymbol{\tau})$.
3. Use $\boldsymbol{x}(\tau_j|\boldsymbol{\tau})$ and $\boldsymbol{\lambda}(\tau_j|\boldsymbol{\tau})$, $j = 1, \ldots, p-1$, to compute $\partial J/\partial \boldsymbol{\tau}$ via the formulae in Theorem 3.5.

Notice that the costate method involves backward integration, whereas the variational method only involves forward integration. Since the state and costate systems are solved in different directions, numerical interpolation is generally needed when implementing the costate method. However, one advantage of the costate method is that it typically requires solving fewer differential equations—there is only one costate system compared with $p-1$ variational systems.

$\boxed{3.3}$ Difficulties with the Direct Optimization Approach

The direct optimization approach to solving the STO problem involves combining one of the two gradient computation procedures (either the variational method or the costate method) with the gradient-based optimization algorithm described at the beginning of this section. As mentioned earlier, it is often claimed in the literature that the direct optimization approach is difficult to implement numerically because the gradient of the cost function in the STO problem is discontinuous (see [5, 8, 19, 21]). However, there is no proof of this claim in the literature. In fact, as we will soon show, the reason why no such proof exists is because the claim itself is incorrect. To show this, we will need the following lemma.

Lemma 3.6. *Let* $\{\boldsymbol{\tau}^k\}_{k=1}^{\infty} \subset \mathcal{T}$ *be a sequence of switching time vectors such that* $\boldsymbol{\tau}^k \to \boldsymbol{\tau} \in \mathcal{T}$ *as* $k \to \infty$. *Then* $\boldsymbol{x}(\cdot|\boldsymbol{\tau}^k) \to \boldsymbol{x}(\cdot|\boldsymbol{\tau})$ *uniformly on* $[0,T]$ *as* $k \to \infty$.

Proof. Consider the indicator functions $\chi_{(\tau_{i-1},\tau_i)} : \mathbb{R} \to \mathbb{R}$ and $\chi_{(\tau_{i-1}^k,\tau_i^k)} : \mathbb{R} \to \mathbb{R}$ defined as follows:

$$\chi_{(\tau_{i-1},\tau_i)}(t) = \begin{cases} 1, & \text{if } t \in (\tau_{i-1},\tau_i), \\ 0, & \text{if } t \notin (\tau_{i-1},\tau_i), \end{cases} \qquad \chi_{(\tau_{i-1}^k,\tau_i^k)}(t) = \begin{cases} 1, & \text{if } t \in (\tau_{i-1}^k,\tau_i^k), \\ 0, & \text{if } t \notin (\tau_{i-1}^k,\tau_i^k). \end{cases}$$

Let $t' \in [0,T] \setminus \{\tau_0, \tau_1, \ldots, \tau_p\}$ be a fixed time point. Then there exists an integer $\varsigma \in \{1, \ldots, p\}$ such that $t' \in (\tau_{\varsigma-1}, \tau_\varsigma)$ and $t' \in (\tau_{\varsigma-1}^k, \tau_\varsigma^k)$ for all sufficiently large k. Thus, when k is sufficiently large,

$$\chi_{(\tau_{i-1}^k,\tau_i^k)}(t') = \chi_{(\tau_{i-1},\tau_i)}(t') = \begin{cases} 1, & \text{if } i = \varsigma, \\ 0, & \text{if } i \neq \varsigma. \end{cases}$$

Since $t' \in [0,T] \setminus \{\tau_0, \tau_1, \ldots, \tau_p\}$ was selected arbitrarily, this shows that $\chi_{(\tau_{i-1}^k,\tau_i^k)} \to \chi_{(\tau_{i-1},\tau_i)}$ almost everywhere as $k \to \infty$.

Now, using (2.1)-(2.3), the state trajectories corresponding to $\boldsymbol{\tau}$ and $\boldsymbol{\tau}^k$ can be written as follows:

$$\boldsymbol{x}(t|\boldsymbol{\tau}) = \boldsymbol{x}^0 + \sum_{i=1}^{p} \int_0^t \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta|\boldsymbol{\tau}))\chi_{(\tau_{i-1},\tau_i)}(\eta)d\eta$$

and

$$\boldsymbol{x}(t|\boldsymbol{\tau}^k) = \boldsymbol{x}^0 + \sum_{i=1}^{p} \int_0^t \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta|\boldsymbol{\tau}^k))\chi_{(\tau_{i-1}^k,\tau_i^k)}(\eta)d\eta. \tag{3.32}$$

Thus,

$$\left\|\boldsymbol{x}(t|\boldsymbol{\tau}^k) - \boldsymbol{x}(t|\boldsymbol{\tau})\right\| \leq \sum_{i=1}^{p} \int_0^t \left\|\boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta|\boldsymbol{\tau}^k))\chi_{(\tau_{i-1}^k,\tau_i^k)}(\eta) - \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta|\boldsymbol{\tau}))\chi_{(\tau_{i-1},\tau_i)}(\eta)\right\|d\eta.$$

Using the triangle inequality, we obtain

$$\begin{aligned}\left\|\boldsymbol{x}(t|\boldsymbol{\tau}^k) - \boldsymbol{x}(t|\boldsymbol{\tau})\right\| \leq &\sum_{i=1}^{p} \int_0^t \left\|\boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta|\boldsymbol{\tau}^k))\chi_{(\tau_{i-1}^k,\tau_i^k)}(\eta) - \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta|\boldsymbol{\tau}))\chi_{(\tau_{i-1}^k,\tau_i^k)}(\eta)\right\|d\eta \\ &+ \sum_{i=1}^{p} \int_0^t \left\|\boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta|\boldsymbol{\tau}))\chi_{(\tau_{i-1}^k,\tau_i^k)}(\eta) - \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta|\boldsymbol{\tau}))\chi_{(\tau_{i-1},\tau_i)}(\eta)\right\|d\eta.\end{aligned}$$

Since $\chi_{(\tau_{i-1}^k,\tau_i^k)}(\eta) \in \{0,1\}$, this inequality can be manipulated to yield

$$\left\|\boldsymbol{x}(t|\boldsymbol{\tau}^k) - \boldsymbol{x}(t|\boldsymbol{\tau})\right\| \leq \Delta_k + \sum_{i=1}^{p} \int_0^t \left\|\boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta|\boldsymbol{\tau}^k)) - \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta|\boldsymbol{\tau}))\right\| d\eta, \tag{3.33}$$

where

$$\Delta_k = \sum_{i=1}^{p} \int_0^T \left\|\boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta|\boldsymbol{\tau}))\chi_{(\tau_{i-1}^k,\tau_i^k)}(\eta) - \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta|\boldsymbol{\tau}))\chi_{(\tau_{i-1},\tau_i)}(\eta)\right\| d\eta.$$

Since $\chi_{(\tau_{i-1}^k,\tau_i^k)} \to \chi_{(\tau_{i-1},\tau_i)}$ almost everywhere as $k \to \infty$, it follows from Lebesgue Dominated Convergence Theorem that $\Delta_k \to 0$ as $k \to \infty$. Furthermore, by (2.8),

$$\left\|\boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta|\boldsymbol{\tau}^k)) - \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta|\boldsymbol{\tau}))\right\| \leq L_3\left\|\boldsymbol{x}(\eta|\boldsymbol{\tau}^k) - \boldsymbol{x}(\eta|\boldsymbol{\tau})\right\|, \quad \eta \in [0,T], \quad i = 1,\ldots,p.$$

Substituting this inequality into (3.33) yields

$$\left\|\boldsymbol{x}(t|\boldsymbol{\tau}^k) - \boldsymbol{x}(t|\boldsymbol{\tau})\right\| \leq \Delta_k + \int_0^t pL_3\left\|\boldsymbol{x}(\eta|\boldsymbol{\tau}^k) - \boldsymbol{x}(\eta|\boldsymbol{\tau})\right\| d\eta.$$

Thus, by the Gronwall-Bellman inequality,

$$\left\|\boldsymbol{x}(t|\boldsymbol{\tau}^k) - \boldsymbol{x}(t|\boldsymbol{\tau})\right\| \leq \Delta_k \exp(pL_3T).$$

Since $\Delta_k \to 0$ as $k \to \infty$, it follows that $\boldsymbol{x}(\cdot|\boldsymbol{\tau}^k) \to \boldsymbol{x}(\cdot|\boldsymbol{\tau})$ uniformly. □

We now prove the analogue of Lemma 3.6 for the costate system (3.29)-(3.31).

Lemma 3.7. *Let $\{\boldsymbol{\tau}^k\}_{k=1}^{\infty} \subset \mathcal{T}$ be a sequence of switching time vectors such that $\boldsymbol{\tau}^k \to \boldsymbol{\tau} \in \mathcal{T}$ as $k \to \infty$. Then $\boldsymbol{\lambda}(\cdot|\boldsymbol{\tau}^k) \to \boldsymbol{\lambda}(\cdot|\boldsymbol{\tau})$ uniformly on $[0,T]$ as $k \to \infty$.*

Proof. It follows from (3.29)-(3.31) that for all $t \in [0,T]$,

$$\boldsymbol{\lambda}(t|\boldsymbol{\tau}) = \boldsymbol{\lambda}(T|\boldsymbol{\tau}) + \sum_{i=1}^{p} \int_t^T \left[\frac{\partial \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta|\boldsymbol{\tau}))}{\partial \boldsymbol{x}}\right]^{\top} \boldsymbol{\lambda}(\eta|\boldsymbol{\tau})\chi_{(\tau_{i-1},\tau_i)}(\eta) d\eta$$

and

$$\boldsymbol{\lambda}(t|\boldsymbol{\tau}^k) = \boldsymbol{\lambda}(T|\boldsymbol{\tau}^k) + \sum_{i=1}^{p} \int_t^T \left[\frac{\partial \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta|\boldsymbol{\tau}^k))}{\partial \boldsymbol{x}}\right]^{\top} \boldsymbol{\lambda}(\eta|\boldsymbol{\tau}^k)\chi_{(\tau_{i-1}^k,\tau_i^k)}(\eta) d\eta,$$

where, as in the proof of Lemma 3.6, $\chi_{(\tau_{i-1},\tau_i)} : \mathbb{R} \to \mathbb{R}$ and $\chi_{(\tau_{i-1}^k,\tau_i^k)} : \mathbb{R} \to \mathbb{R}$ are defined by

$$\chi_{(\tau_{i-1},\tau_i)}(t) = \begin{cases} 1, & \text{if } t \in (\tau_{i-1},\tau_i), \\ 0, & \text{if } t \notin (\tau_{i-1},\tau_i), \end{cases} \qquad \chi_{(\tau_{i-1}^k,\tau_i^k)}(t) = \begin{cases} 1, & \text{if } t \in (\tau_{i-1}^k,\tau_i^k), \\ 0, & \text{if } t \notin (\tau_{i-1}^k,\tau_i^k). \end{cases}$$

Thus,

$$\begin{aligned} &\left\|\boldsymbol{\lambda}(t|\boldsymbol{\tau}^k) - \boldsymbol{\lambda}(t|\boldsymbol{\tau})\right\| \leq \left\|\boldsymbol{\lambda}(T|\boldsymbol{\tau}^k) - \boldsymbol{\lambda}(T|\boldsymbol{\tau})\right\| \\ &\quad + \sum_{i=1}^{p} \int_t^T \left\|\boldsymbol{\lambda}(\eta|\boldsymbol{\tau}^k)^{\top} \frac{\partial \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta|\boldsymbol{\tau}^k))}{\partial \boldsymbol{x}} \chi_{(\tau_{i-1}^k,\tau_i^k)}(\eta) - \boldsymbol{\lambda}(\eta|\boldsymbol{\tau})^{\top} \frac{\partial \boldsymbol{f}^i(\eta, \boldsymbol{x}(\eta|\boldsymbol{\tau}))}{\partial \boldsymbol{x}} \chi_{(\tau_{i-1},\tau_i)}(\eta)\right\| d\eta. \end{aligned}$$

Therefore, by the triangle inequality,

$$\begin{aligned}\big\|\boldsymbol{\lambda}(t|\boldsymbol{\tau}^k)-\boldsymbol{\lambda}(t|\boldsymbol{\tau})\big\| &\le \upsilon_k+\big\|\boldsymbol{\lambda}(T|\boldsymbol{\tau}^k)-\boldsymbol{\lambda}(T|\boldsymbol{\tau})\big\| \\ &+\sum_{i=1}^{p}\int_t^T\left\|\boldsymbol{\lambda}(\eta|\boldsymbol{\tau}^k)^\top\frac{\partial \boldsymbol{f}^i(\eta,\boldsymbol{x}(\eta|\boldsymbol{\tau}^k))}{\partial \boldsymbol{x}}-\boldsymbol{\lambda}(\eta|\boldsymbol{\tau})^\top\frac{\partial \boldsymbol{f}^i(\eta,\boldsymbol{x}(\eta|\boldsymbol{\tau}^k))}{\partial \boldsymbol{x}}\right\|d\eta \\ &+\sum_{i=1}^{p}\int_t^T\left\|\boldsymbol{\lambda}(\eta|\boldsymbol{\tau})^\top\frac{\partial \boldsymbol{f}^i(\eta,\boldsymbol{x}(\eta|\boldsymbol{\tau}^k))}{\partial \boldsymbol{x}}-\boldsymbol{\lambda}(\eta|\boldsymbol{\tau})^\top\frac{\partial \boldsymbol{f}^i(\eta,\boldsymbol{x}(\eta|\boldsymbol{\tau}))}{\partial \boldsymbol{x}}\right\|d\eta,\end{aligned} \tag{3.34}$$

where

$$\upsilon_k=\sum_{i=1}^{p}\int_0^T\left\|\boldsymbol{\lambda}(\eta|\boldsymbol{\tau})^\top\frac{\partial \boldsymbol{f}^i(\eta,\boldsymbol{x}(\eta|\boldsymbol{\tau}))}{\partial \boldsymbol{x}}\chi_{(\tau_{i-1}^k,\tau_i^k)}(\eta)-\boldsymbol{\lambda}(\eta|\boldsymbol{\tau})^\top\frac{\partial \boldsymbol{f}^i(\eta,\boldsymbol{x}(\eta|\boldsymbol{\tau}))}{\partial \boldsymbol{x}}\chi_{(\tau_{i-1},\tau_i)}(\eta)\right\|d\eta.$$

Recall from the proof of Lemma 3.6 that $\chi_{(\tau_{i-1}^k,\tau_i^k)}\to\chi_{(\tau_{i-1},\tau_i)}$ almost everywhere as $k\to\infty$. Thus, it follows from Lebesgue Dominated Convergence Theorem that $\upsilon_k\to 0$ as $k\to\infty$.

Let $\gamma>0$ be arbitrary. Then it follows from inequality (2.5), Assumptions 2.1 and 2.3, and Lemma 3.6 that when k is sufficiently large,

$$\big\|\boldsymbol{\lambda}(T|\boldsymbol{\tau}^k)-\boldsymbol{\lambda}(T|\boldsymbol{\tau})\big\|=\left\|\frac{\partial\Phi(\boldsymbol{x}(T|\boldsymbol{\tau}^k))}{\partial \boldsymbol{x}}-\frac{\partial\Phi(\boldsymbol{x}(T|\boldsymbol{\tau}))}{\partial \boldsymbol{x}}\right\|<\gamma$$

and

$$\left\|\frac{\partial \boldsymbol{f}^i(\eta,\boldsymbol{x}(\eta|\boldsymbol{\tau}^k))}{\partial \boldsymbol{x}}-\frac{\partial \boldsymbol{f}^i(\eta,\boldsymbol{x}(\eta|\boldsymbol{\tau}))}{\partial \boldsymbol{x}}\right\|<\gamma,\quad \eta\in[0,T],\quad i=1,\dots,p.$$

Thus, when k is sufficiently large, (3.34) can be simplified as follows:

$$\begin{aligned}\big\|\boldsymbol{\lambda}(t|\boldsymbol{\tau}^k)-\boldsymbol{\lambda}(t|\boldsymbol{\tau})\big\| &\le \upsilon_k+(pN_1T+1)\gamma \\ &+\sum_{i=1}^{p}\int_t^T\left\|\left[\frac{\partial \boldsymbol{f}^i(\eta,\boldsymbol{x}(\eta|\boldsymbol{\tau}^k))}{\partial \boldsymbol{x}}\right]^\top\boldsymbol{\lambda}(\eta|\boldsymbol{\tau}^k)-\left[\frac{\partial \boldsymbol{f}^i(\eta,\boldsymbol{x}(\eta|\boldsymbol{\tau}^k))}{\partial \boldsymbol{x}}\right]^\top\boldsymbol{\lambda}(\eta|\boldsymbol{\tau})\right\|d\eta,\end{aligned}$$

where N_1 is an upper bound for the Euclidean norm of $\boldsymbol{\lambda}(\eta|\boldsymbol{\tau})$ (recall that $\boldsymbol{\lambda}(\cdot|\boldsymbol{\tau})$ is a continuous function defined on the compact interval $[0,T]$). By Assumption 2.1 and inequality (2.5), there exists a positive constant $N_2>0$ such that

$$\left\|\left[\frac{\partial \boldsymbol{f}^i(t,\boldsymbol{x}(t|\boldsymbol{\tau}^k))}{\partial \boldsymbol{x}}\right]^\top\right\|\le N_2,\quad t\in[0,T],\quad \boldsymbol{\tau}^k\in\mathcal{T},\quad i=1,\dots,p.$$

Hence,

$$\big\|\boldsymbol{\lambda}(t|\boldsymbol{\tau}^k)-\boldsymbol{\lambda}(t|\boldsymbol{\tau})\big\|\le\upsilon_k+(pN_1T+1)\gamma+\int_t^T pN_2\big\|\boldsymbol{\lambda}(\eta|\boldsymbol{\tau}^k)-\boldsymbol{\lambda}(\eta|\boldsymbol{\tau})\big\|d\eta.$$

Therefore, by the Gronwall-Bellman inequality,

$$\big\|\boldsymbol{\lambda}(t|\boldsymbol{\tau}^k)-\boldsymbol{\lambda}(t|\boldsymbol{\tau})\big\|\le(\upsilon_k+(pN_1T+1)\gamma)\exp(pN_2T),$$

which holds whenever k is sufficiently large. Since γ was selected arbitrarily, and $\upsilon_k\to 0$ as $k\to\infty$, it follows that $\boldsymbol{\lambda}(t|\boldsymbol{\tau}^k)\to\boldsymbol{\lambda}(t|\boldsymbol{\tau})$ as $k\to\infty$, uniformly with respect to t. □

Recall from Theorem 3.5 that for each admissible switching time vector $\boldsymbol{\tau} \in \mathcal{T}$ with $\tau_{j-1} < \tau_j$, $j = 1, \ldots, p$, the partial derivatives of the cost function J are given by

$$\frac{\partial J(\boldsymbol{\tau})}{\partial \tau_j} = \boldsymbol{\lambda}(\tau_j|\boldsymbol{\tau})^\top \boldsymbol{f}^j(\tau_j, \boldsymbol{x}(\tau_j|\boldsymbol{\tau})) - \boldsymbol{\lambda}(\tau_j|\boldsymbol{\tau})^\top \boldsymbol{f}^{j+1}(\tau_j, \boldsymbol{x}(\tau_j|\boldsymbol{\tau})), \quad j = 1, \ldots, p-1.$$

It follows immediately from Assumption 2.1 and Lemmas 3.6 and 3.7 that these partial derivatives are continuous with respect to $\boldsymbol{\tau}$. Thus, contrary to what is commonly claimed in the literature, the gradient of the cost function in the STO problem is continuous with respect to the switching times. The confusion around this issue seems to stem from the presence of an instantaneous jump in the variational system (3.1)-(3.3), which causes the state variation to be discontinuous (as a function of time). However, although the state variation is discontinuous, the cost function's gradient is not.

So what are the difficulties associated with implementing the direct optimization approach? There are two difficulties that should be mentioned:

(i) Piecewise integration of the variational and/or costate systems is cumbersome to implement numerically because the switching times change at each optimization iteration; and

(ii) The gradient formulae in Theorems 3.4 and 3.5 only exist when the switching times are *distinct* (this causes problems if the optimization algorithm tries to merge two adjacent switching times to delete a "non-optimal" subsystem).

For these reasons, the direct optimization approach is rarely used in practice. Instead, the *time-scaling transformation* [8,9,19] is usually invoked to transform the STO problem into an equivalent problem with fixed switching times. This allows one to circumvent the difficulties mentioned above. The time-scaling transformation is discussed in the next section.

4 Indirect Optimization: The Time-Scaling Approach

The time-scaling approach involves solving the STO problem *indirectly* by transforming it into an equivalent problem that is easier to solve. This is done by introducing a new time variable $s \in [0, p]$, along with a new set of decision parameters $\theta_i = \tau_i - \tau_{i-1}$, $i = 1, \ldots, p$. The new time variable s is related to the old time variable t through the following equations:

$$\frac{dt(s)}{ds} = \theta_i, \quad s \in (i-1, i), \quad i = 1, \ldots, p, \tag{4.1}$$

$$t(i^+) = t(i^-), \quad i = 1, \ldots, p-1, \tag{4.2}$$

$$t(0) = 0, \tag{4.3}$$

$$t(p) = T, \tag{4.4}$$

where $T > 0$ is the terminal time. Note that equations (4.1)-(4.4) define $t = t(s)$ as a non-negative and non-decreasing piecewise linear function of s.

Integrating (4.1)-(4.3) gives

$$t(i) = t(0) + \sum_{j=1}^{i} \int_{j-1}^{j} \frac{dt(s)}{ds} ds = \sum_{j=1}^{i} \int_{j-1}^{j} \theta_j ds = \sum_{j=1}^{i} (\tau_j - \tau_{j-1}) = \tau_i, \quad i = 1, \ldots, p.$$

Thus, under the time-scaling transformation defined by equations (4.1)-(4.4), $s = i$ in the new time horizon is mapped to $t = \tau_i$ in the original time horizon.

Now, let $\tilde{\boldsymbol{x}}(s) = \boldsymbol{x}(t(s))$. Then under the time-scaling transformation, the dynamics (2.1) become

$$\frac{d}{ds}\{\tilde{\boldsymbol{x}}(s)\} = \frac{d}{ds}\{\boldsymbol{x}(t(s))\} = \theta_i \boldsymbol{f}^i(t(s), \boldsymbol{x}(t(s))) = \theta_i \boldsymbol{f}^i(t(s), \tilde{\boldsymbol{x}}(s)), \quad s \in (i-1, i), \quad i = 1, \ldots, p. \tag{4.5}$$

Furthermore, the intermediate conditions (2.2) and the initial condition (2.3) become

$$\tilde{\boldsymbol{x}}(i^+) = \lim_{s \to i+} \boldsymbol{x}(t(s)) = \boldsymbol{x}(\tau_i^+) = \boldsymbol{x}(\tau_i^-) = \lim_{s \to i-} \boldsymbol{x}(t(s)) = \boldsymbol{x}(i^-), \quad i = 1, \ldots, p-1, \tag{4.6}$$

and

$$\tilde{\boldsymbol{x}}(0^+) = \lim_{s \to 0+} \boldsymbol{x}(t(s)) = \boldsymbol{x}(0^+) = \boldsymbol{x}^0. \tag{4.7}$$

Equations (4.1)-(4.7) constitute a new switched system (with fixed switching times) in which t is considered as a state variable.

The subsystem durations θ_i, $i = 1, \ldots, p$, must satisfy the following constraints:

$$\theta_i \geq 0, \quad i = 1, \ldots, p, \tag{4.8}$$

and

$$\theta_1 + \theta_2 + \cdots + \theta_p = T. \tag{4.9}$$

Notice that constraint (4.9) is equivalent to (4.4).

Any $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_p]^\top \in \mathbb{R}^p$ satisfying (4.8) and (4.9) is called an *admissible duration vector*. Let $\mathcal{S}$ denote the set of all such admissible duration vectors.

Under the time-scaling transformation, the cost function (2.9) becomes

$$\tilde{J}(\boldsymbol{\theta}) = \Phi(\tilde{\boldsymbol{x}}(p|\boldsymbol{\theta})),$$

where $\tilde{\boldsymbol{x}}(\cdot|\boldsymbol{\theta})$ denotes the solution of (4.5)-(4.7) corresponding to a given duration vector $\boldsymbol{\theta} \in \mathcal{S}$. Thus, our new problem can be stated formally as follows: *Find an admissible duration vector* $\boldsymbol{\theta}^* \in \mathcal{S}$ *such that*

$$\tilde{J}(\boldsymbol{\theta}^*) = \inf_{\boldsymbol{\theta} \in \mathcal{S}} \tilde{J}(\boldsymbol{\theta}).$$

This new problem is equivalent to the original STO problem [10]. In fact, if $\boldsymbol{\theta}^* = [\theta_1^*, \ldots, \theta_p^*]^\top$ is an optimal solution for the new problem, then the optimal switching times for the original STO problem are

$$\tau_i^* = \theta_1^* + \cdots + \theta_i^*, \quad i = 1, \ldots, p-1.$$

Notice that the switching times in (4.5)-(4.7) occur at the fixed locations $s = i$, $i = 1, \ldots, p-1$. Thus, the partial derivatives of the cost function with respect to the switching times are not required to solve the equivalent problem. Instead, we require the partial derivatives of the cost function with respect to the new decision parameters θ_i, $i = 1, \ldots, p$. These partial derivatives can be obtained using conventional methods; see [9,10,12,20] for details. Moreover, unlike in the original STO problem, the partial derivatives of the cost function in the equivalent problem still exist when some of the subsystem durations are zero. This is the primary advantage of the time-scaling approach.

5 Numerical Example

Consider the Dubins vehicle model for the motion of a charged particle in a magnetic field [3]:

$$\dot{x}(t) = R \sin \alpha(t), \quad \dot{y}(t) = -R \cos \alpha(t), \quad \dot{\alpha}(t) = u(t), \tag{5.1}$$

where $(x(t), y(t))$ is the particle's position, $\alpha(t)$ is the particle's heading, $u(t) \in \{-1, 1\}$ is a discrete-valued control variable representing the direction of the magnetic field, and $R > 0$ is a given constant. The initial conditions for (5.1) are

$$x(0) = 0, \quad y(0) = 0, \quad \alpha(0) = 0. \tag{5.2}$$

Since $u(t) \in \{-1, 1\}$ for all t, we can re-write system (5.1) as follows:

$$\dot{x}(t) = R \sin \alpha(t), \quad \dot{y}(t) = -R \cos \alpha(t), \quad \dot{\alpha}(t) = (-1)^{i+1}, \quad t \in (\tau_{i-1}, \tau_i), \quad i = 1, \ldots, p, \tag{5.3}$$

where $p - 1$ is the maximum number of control switches, $\tau_0 = 0$ is the initial time, $\tau_p = T$ is the terminal time, and τ_i, $i = 1, \ldots, p - 1$, are switching times to be optimized.

Our STO problem is thus stated as follows: *Subject to the initial conditions* (5.2) *and the dynamics* (5.3)*, choose the switching times τ_i, $i = 1, \ldots, p - 1$, to minimize*

$$J = (x(T) - x^T)^2 + (y(T) - y^T)^2, \tag{5.4}$$

where (x^T, y^T) denotes the desired final state.

To solve this problem, we wrote a Fortran program based on the direct optimization approach. This program uses the optimization code NLPQLP [18] to implement Step 4 (compute search direction) and Step 5 (line search) of the gradient-based optimization algorithm presented in Section 3. For the cost function gradients, the program allows the user to select either the variational or costate method.

For $p = 4$, $R = 1$, $T = 8$, and $(x^T, y^T) = (4, 4)$, applying the variational method gives an optimal cost of $J = 9.6814 \times 10^{-3}$ and the optimal switching times

$$\tau_1^* = 3.4081, \quad \tau_2^* = 4.9388, \quad \tau_3^* = 6.4694.$$

Applying the costate method gives an optimal cost of $J = 9.6856 \times 10^{-3}$ and the optimal switching times

$$\tau_1^* = 3.4051, \quad \tau_2^* = 4.9307, \quad \tau_3^* = 6.4623.$$

As expected, both the variational and costate methods yield almost identical results. The optimal state trajectory for the particle is shown in Figure 1.

We now apply the time-scaling transformation to this problem. Under the time-scaling transformation, the state equations (5.3) become

$$\dot{\tilde{x}}(s) = R\theta_i \sin \tilde{\alpha}(s), \quad \dot{\tilde{y}}(s) = -R\theta_i \cos \tilde{\alpha}(s), \quad \dot{\tilde{\alpha}}(s) = (-1)^{i+1}\theta_i, \;\; s \in (i-1, i), \;\; i = 1, \ldots, p, \tag{5.5}$$

where $\tilde{x}(s) = x(t(s))$, $\tilde{y}(s) = y(t(s))$, $\tilde{\alpha}(s) = \alpha(t(s))$, and the overhead dot now denotes differentiation with respect to the new time variable s.

The initial conditions (5.2) become

$$\tilde{x}(0) = 0, \quad \tilde{y}(0) = 0, \quad \tilde{\alpha}(0) = 0. \tag{5.6}$$

Furthermore, the cost function (5.4) becomes

$$\tilde{J} = (\tilde{x}(p) - x^T)^2 + (\tilde{y}(p) - y^T)^2.$$

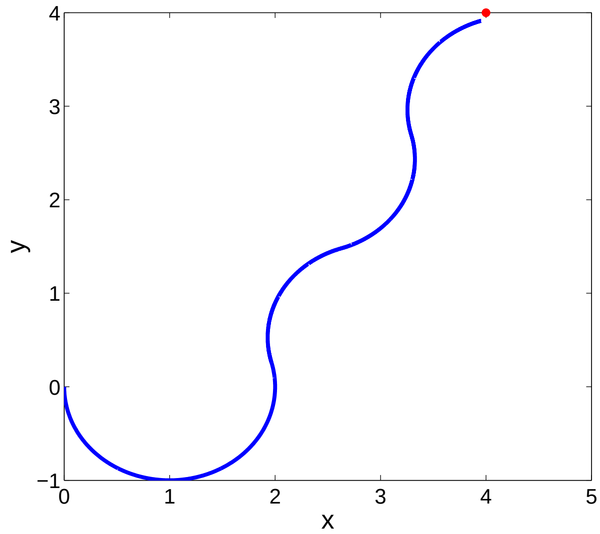

Figure 1: The optimal state trajectory obtained using the direct optimization approach in conjunction with the variational method. The red dot indicates the desired terminal position.

Thus, the transformed problem can be stated as follows: *Subject to the new dynamics* (5.5) *and the new initial conditions* (5.6)*, choose the mode durations* θ_i*,* $i = 1, \ldots, p$*, to minimize* $\tilde{J}$*.*

We wrote another Fortran program to solve this transformed problem. This new program also uses NLPQLP to implement the optimization process, and allows the user to select between the variational and costate methods for computing the cost function gradients. Applying the variational method gives an optimal cost of $\tilde{J} = 9.6814 \times 10^{-3}$ and the optimal durations

$$\theta_1^* = 3.4081, \quad \theta_2^* = 1.5306, \quad \theta_3^* = 1.5306, \quad \theta_4^* = 1.5306.$$

These optimal durations correspond to the following optimal switching times:

$$\tau_1^* = \theta_1^* = 3.4081, \quad \tau_2^* = \theta_1^* + \theta_2^* = 4.9387, \quad \tau_3^* = \theta_1^* + \theta_2^* + \theta_3^* = 6.4693.$$

Applying the costate method gives an optimal cost of $\tilde{J} = 9.6818 \times 10^{-3}$ and the optimal durations

$$\theta_1^* = 3.4092, \quad \theta_2^* = 1.5312, \quad \theta_3^* = 1.5291, \quad \theta_4^* = 1.5306.$$

The corresponding optimal switching times are

$$\tau_1^* = \theta_1^* = 3.4092, \quad \tau_2^* = \theta_1^* + \theta_2^* = 4.9404, \quad \tau_3^* = \theta_1^* + \theta_2^* + \theta_3^* = 6.4695.$$

As expected, the time-scaling transformation produces almost identical results to the direct optimization approach.

Note that under the time-scaling transformation, the variable switching times τ_i, $i = 1, \ldots, p-1$, are replaced by the conventional decision parameters θ_i, $i = 1, \ldots, p$. Hence, since the switching times in the new system (5.5) are fixed, integrating (5.5) numerically is much easier than integrating (5.3). The other advantage of the time-scaling approach is that the cost function's gradient exists even when some of the subsystem durations are zero. Recall that the gradient formulae for the direct optimization approach do not hold when two or more switching times coincide.

6 Conclusion

In this paper, we have investigated two computational approaches for solving the STO problem: the direct optimization approach and the time-scaling transformation. The direct optimization approach involves optimizing the switching times directly by combining special procedures for computing the cost function's gradient with standard nonlinear programming algorithms. The time-scaling approach, on the other hand, involves optimizing the switching times indirectly by transforming the STO problem into an equivalent problem that is easier to solve. Although the time-scaling transformation is much more popular than the direct optimization approach, the basis for its popularity is poorly described in the literature. Papers on this topic typically justify the time-scaling transformation with vague statements such as "the cost function's gradient in the STO problem is not effective for numerical implementation". In this paper, we demonstrated rigorously why the time-scaling transformation is superior to the direct optimization approach, thereby closing a significant gap in the literature. In particular, we showed that the common perception that "the cost function's gradient in the STO problem is discontinuous" is actually incorrect. We therefore hope that these new results will stimulate further research into the STO problem.

A Proof that $\boldsymbol{\varphi}^\epsilon$ is of order ϵ

Let $\epsilon \in \Omega$ be arbitrary but fixed. Furthermore, let $i(\eta)$ denote the active mode at time η corresponding to the switching time vector $\boldsymbol{\tau}$, and let $i_\epsilon(\eta)$ denote the active mode at time η corresponding to the switching time vector $\boldsymbol{\tau} + \epsilon \boldsymbol{e}^j$. Then there exists an interval $\mathcal{I}_\epsilon$ of length $|\mathcal{I}_\epsilon| = \epsilon$ such that

$$i_\epsilon(\eta) = i(\eta), \quad \eta \notin \mathcal{I}_\epsilon. \tag{A.1}$$

Now, for each $t \in [0, T]$, we have

$$\boldsymbol{\varphi}^\epsilon(t) = \boldsymbol{x}^\epsilon(t) - \boldsymbol{x}(t) = \int_0^t \left\{ \boldsymbol{f}^{i_\epsilon(\eta)}(\eta, \boldsymbol{x}^\epsilon(\eta)) - \boldsymbol{f}^{i(\eta)}(\eta, \boldsymbol{x}(\eta)) \right\} d\eta.$$

Taking the norm of both sides and then applying equation (A.1) yields

$$\begin{aligned}
\|\boldsymbol{\varphi}^\epsilon(t)\| &\le \int_{[0,t]\setminus\mathcal{I}_\epsilon} \left\| \boldsymbol{f}^{i(\eta)}(\eta, \boldsymbol{x}^\epsilon(\eta)) - \boldsymbol{f}^{i(\eta)}(\eta, \boldsymbol{x}(\eta)) \right\| d\eta \\
&\quad + \int_{\mathcal{I}_\epsilon} \left\| \boldsymbol{f}^{i_\epsilon(\eta)}(\eta, \boldsymbol{x}^\epsilon(\eta)) - \boldsymbol{f}^{i(\eta)}(\eta, \boldsymbol{x}(\eta)) \right\| d\eta \\
&\le \int_0^t \left\| \boldsymbol{f}^{i(\eta)}(\eta, \boldsymbol{x}^\epsilon(\eta)) - \boldsymbol{f}^{i(\eta)}(\eta, \boldsymbol{x}(\eta)) \right\| d\eta \\
&\quad + \int_{\mathcal{I}_\epsilon} \left\| \boldsymbol{f}^{i_\epsilon(\eta)}(\eta, \boldsymbol{x}^\epsilon(\eta)) - \boldsymbol{f}^{i(\eta)}(\eta, \boldsymbol{x}(\eta)) \right\| d\eta.
\end{aligned}$$

Therefore, from (2.6) and (2.8),

$$\|\boldsymbol{\varphi}^\epsilon(t)\| \le 2L_3|\mathcal{I}_\epsilon| + \int_0^t L_3\|\boldsymbol{\varphi}^\epsilon(\eta)\| d\eta = 2L_3|\epsilon| + \int_0^t L_3\|\boldsymbol{\varphi}^\epsilon(\eta)\| d\eta.$$

Finally, applying the Gronwall-Bellman inequality gives

$$\|\boldsymbol{\varphi}^{\epsilon}(t)\| \leq 2L_3|\epsilon|\exp(L_3T). \tag{A.2}$$

This completes the proof.

B Proof that $\boldsymbol{\rho}_j^{\epsilon}$ and $\boldsymbol{\rho}_{j+1}^{\epsilon}$ are of order ϵ^2

Let $\boldsymbol{\rho}_j^{\epsilon}$ and $\boldsymbol{\rho}_{j+1}^{\epsilon}$ be as defined in the proof of Theorem 3.1 and assume that $\epsilon \in \Omega \cap (-\infty, 0]$. Then

$$\|\boldsymbol{\rho}_j^{\epsilon}\| \leq \int_{\tau_j+\epsilon}^{\tau_j} \big\|\boldsymbol{f}^j(\eta, \boldsymbol{x}(\eta)) - \boldsymbol{f}^j(\tau_j, \boldsymbol{x}(\tau_j))\big\| d\eta \leq \int_{\tau_j+\epsilon}^{\tau_j} L_3|\eta - \tau_j| d\eta + \int_{\tau_j+\epsilon}^{\tau_j} L_3\big\|\boldsymbol{x}(\eta) - \boldsymbol{x}(\tau_j)\big\| d\eta,$$

where L_3 is the Lipschitz constant defined in (2.8). Thus,

$$\|\boldsymbol{\rho}_j^{\epsilon}\| \leq \int_{\tau_j+\epsilon}^{\tau_j} L_3|\eta - \tau_j| d\eta + \int_{\tau_j+\epsilon}^{\tau_j} L_3\big\|\boldsymbol{x}(\eta) - \boldsymbol{x}(\tau_j)\big\| d\eta \leq L_3\epsilon^2 + \int_{\tau_j+\epsilon}^{\tau_j} L_3\big\|\boldsymbol{x}(\eta) - \boldsymbol{x}(\tau_j)\big\| d\eta. \tag{B.1}$$

Since $\epsilon \in \Omega \cap (-\infty, 0]$, we must have $\tau_{j-1} \leq \tau_j + \epsilon \leq \tau_j$. Hence, for each $\eta \in [\tau_j + \epsilon, \tau_j]$,

$$\big\|\boldsymbol{x}(\eta) - \boldsymbol{x}(\tau_j)\big\| \leq \int_{\eta}^{\tau_j} \big\|\boldsymbol{f}^j(s, \boldsymbol{x}(s))\big\| ds \leq \int_{\tau_j+\epsilon}^{\tau_j} \big\|\boldsymbol{f}^j(s, \boldsymbol{x}(s))\big\| ds \leq \int_{\tau_j+\epsilon}^{\tau_j} L_3 ds = L_3|\epsilon|, \tag{B.2}$$

where the last inequality follows from (2.6). Substituting (B.2) into (B.1) gives

$$\|\boldsymbol{\rho}_j^{\epsilon}\| \leq L_3\epsilon^2 + L_3^2\epsilon^2. \tag{B.3}$$

Now, for $\boldsymbol{\rho}_{j+1}^{\epsilon}$, we have

$$\begin{aligned} \|\boldsymbol{\rho}_{j+1}^{\epsilon}\| &\leq \int_{\tau_j+\epsilon}^{\tau_j} \big\|\boldsymbol{f}^{j+1}(\eta, \boldsymbol{x}^{\epsilon}(\eta)) - \boldsymbol{f}^{j+1}(\tau_j, \boldsymbol{x}(\tau_j))\big\| d\eta \\ &\leq \int_{\tau_j+\epsilon}^{\tau_j} \big\|\boldsymbol{f}^{j+1}(\eta, \boldsymbol{x}^{\epsilon}(\eta)) - \boldsymbol{f}^{j+1}(\eta, \boldsymbol{x}(\eta))\big\| d\eta \\ &\quad + \int_{\tau_j+\epsilon}^{\tau_j} \big\|\boldsymbol{f}^{j+1}(\eta, \boldsymbol{x}(\eta)) - \boldsymbol{f}^{j+1}(\tau_j, \boldsymbol{x}(\tau_j))\big\| d\eta. \end{aligned} \tag{B.4}$$

Thus, using inequalities (2.8) and (A.2),

$$\begin{aligned} \|\boldsymbol{\rho}_{j+1}^{\epsilon}\| &\leq \int_{\tau_j+\epsilon}^{\tau_j} L_3\|\boldsymbol{\varphi}^{\epsilon}(\eta)\| d\eta + \int_{\tau_j+\epsilon}^{\tau_j} \big\|\boldsymbol{f}^{j+1}(\eta, \boldsymbol{x}(\eta)) - \boldsymbol{f}^{j+1}(\tau_j, \boldsymbol{x}(\tau_j))\big\| d\eta \\ &\leq 2L_3^2\exp(L_3T)\epsilon^2 + \int_{\tau_j+\epsilon}^{\tau_j} \big\|\boldsymbol{f}^{j+1}(\eta, \boldsymbol{x}(\eta)) - \boldsymbol{f}^{j+1}(\tau_j, \boldsymbol{x}(\tau_j))\big\| d\eta. \end{aligned}$$

The integral term on the right-hand side can be handled using the same arguments as in the derivation of (B.3). Thus,

$$\|\boldsymbol{\rho}_{j+1}^{\epsilon}\| \leq 2L_3^2\exp(L_3T)\epsilon^2 + L_3\epsilon^2 + L_3^2\epsilon^2.$$

This completes the proof.

References

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

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

[3] Y. An, C. Xu, Q. Lin, R. Loxton and K.L. Teo, Path planning for underactuated Dubins micro-robots using switching control, in *Proceedings of the 10th IEEE International Conference on Control and Automation*, Hangzhou, China, June 2013.

[4] M. Boccadoro, Y. Wardi, M. Egerstedt and E. Verriest, Optimal control of switching surfaces in hybrid dynamical systems, *Discrete Event Dyn. Syst.* 15 (2005) 433–448.

[5] L. Caccetta, I. Loosen and V. Rehbock, Computational aspects of the optimal transit path problem, *J. Ind. Manag. Optim.* 4 (2008) 95–105.

[6] M. Gerdts, A variable time transformation method for mixed-integer optimal control problems, *Optimal Control Appl. Methods* 27 (2006) 169–182.

[7] M. Kamgarpour and C. Tomlin, On optimal control of non-autonomous switched systems with a fixed mode sequence, *Automatica J. IFAC* 48 (2012) 1177–1181.

[8] H.W.J. Lee, K.L. Teo, V. Rehbock and L.S. Jennings, Control parametrization enhancing technique for optimal discrete-valued control problems, *Automatica J. IFAC* 35 (1999) 1401–1407.

[9] R. Li, K.L. Teo, K.H. Wong and G.R. Duan, Control parameterization enhancing transform for optimal control of switched systems, *Math. Comput. Modelling* 43 (2006) 1393–1403.

[10] 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) 59–76.

[11] C. Liu, Z. Gong, E. Feng and H. Yin, Modelling and optimal control for nonlinear multistage dynamical system of microbial fed-batch culture, *J. Ind. Manag. Optim.* 5 (2009) 835–850.

[12] R. Loxton, K.L. Teo and V. Rehbock, Optimal control problems with multiple characteristic time points in the objective and constraints, *Automatica J. IFAC* 44 (2008) 2923–2929.

[13] R. Loxton, K.L. Teo and V. Rehbock, Computational method for a class of switched system optimal control problems, *IEEE Trans. Automat. Control* 54 (2009) 2455–2460.

[14] R. Loxton, K.L. Teo, V. Rehbock and K.F.C. Yiu, Optimal control problems with a continuous inequality constraint on the state and the control, *Automatica J. IFAC* 45 (2009) 2250–2257.

[15] D.G. Luenberger and Y. Ye, *Linear and Nonlinear Programming*, 3rd Edition, Springer, New York, 2008.

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

[17] T. Ruby and V. Rehbock, Numerical solutions of optimal switching control problems, in *Optimization and Control with Applications*, L. Qi, K.L. Teo and X. Yang (eds.), Springer, New York, 2005, pp.447–459.

[18] 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.

[19] K.L. Teo, Control parametrization enhancing transform to optimal control problems, *Nonlinear Anal.* 63 (2005) e2223–e2236.

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

[21] K.L. Teo, L.S. Jennings, H.W.J. Lee and V. Rehbock, The control parameterization enhancing transform for constrained optimal control problems, *J. Aust. Math. Soc. Ser. B* 40 (1999) 314–335.

[22] S.F. Woon, V. Rehbock and R. Loxton, Global optimization method for continuous-time sensor scheduling, *Nonlinear Dyn. Syst. Theory* 10 (2010) 175-188.

[23] X. Xu and P.J. Antsaklis, Optimal control of switched systems based on parameterization of the switching instants, *IEEE Trans. Automat. Control* 49 (2004) 2–16.

Manuscript received 12 August 2013
revised 7 January 2014
accepted for publication 26 February 2014

Ryan Loxton
Mathematics and Statistics, Curtin University, Australia
E-mail address: `r.loxton@curtin.edu.au`

Qun Lin
Mathematics and Statistics, Curtin University, Australia
E-mail address: `q.lin@curtin.edu.au`

Kok Lay Teo
Mathematics and Statistics, Curtin University, Australia
E-mail address: `k.l.teo@curtin.edu.au`