Department of Mathematics and Statistics

Optimal control problems involving constrained, switched, and delay systems

Ryan Christopher Loxton

This thesis is presented for the Degree of
Doctor of Philosophy
of
Curtin University of Technology

February 2010

Declaration

I affirm that the material in this thesis is the result of my own original research and has not been submitted for any other degree, diploma, or award.

.....
Ryan Christopher Loxton
February 2010

Abstract

In this thesis, we develop numerical methods for solving five nonstandard optimal control problems. The main idea of each method is to reformulate the optimal control problem as, or approximate it by, a nonlinear programming problem. The decision variables in this nonlinear programming problem influence its cost function (and constraints, if it has any) implicitly through the dynamic system. Hence, deriving the gradient of the cost and the constraint functions is a difficult task. A major focus of this thesis is on developing methods for computing these gradients. These methods can then be used in conjunction with a gradient-based optimization technique to solve the optimal control problem efficiently.

The first optimal control problem that we consider has nonlinear inequality constraints that depend on the state at two or more discrete time points. These time points are decision variables that, together with a control function, should be chosen in an optimal manner. To tackle this problem, we first approximate the control by a piecewise constant function whose values and switching times (the times at which it changes value) are decision variables. We then apply a novel time-scaling transformation that maps the switching times to fixed points in a new time horizon. This yields an approximate dynamic optimization problem with a finite number of decision variables. We develop a new algorithm, which involves integrating an auxiliary dynamic system forward in time, for computing the gradient of the cost and constraints in this approximate problem.

The second optimal control problem that we consider has nonlinear continuous inequality constraints. These constraints restrict both the state and the control at every point in the time horizon. As with the first problem, we approximate the control by a piecewise constant function and then transform the time variable. This yields an approximate semi-infinite programming problem, which can be solved using a penalty function algorithm. A solution of this problem immediately furnishes a suboptimal control for the original optimal control problem. By repeatedly increasing the number of parameters used in the approximation, we can generate a sequence of suboptimal controls. Our main result shows that the cost of these suboptimal controls converges to the minimum cost.

The third optimal control problem that we consider is an applied problem from electrical engineering. Its aim is to determine an optimal operating scheme for a switched-capacitor DC-DC power converter—an electronic device that transforms one DC voltage into another by periodically switching between several circuit topologies. Specifically, the

optimal control problem is to choose the times at which the topology switches occur so that the output voltage ripple is minimized and the load regulation is maximized. This problem is governed by a switched system with linear subsystems (each subsystem models one of the power converter's topologies). Moreover, its cost function is non-smooth. By introducing an auxiliary dynamic system and transforming the time variable (so that the topology switching times become fixed), we derive an equivalent semi-infinite programming problem. This semi-infinite programming problem, like the one that approximates the continuously-constrained optimal control problem, can be solved using a penalty function algorithm.

The fourth optimal control problem that we consider involves a general switched system, which includes the model of a switched-capacitor DC-DC power converter as a special case. This switched system evolves by switching between several subsystems of nonlinear ordinary differential equations. Furthermore, each subsystem switch is accompanied by an instantaneous change in the state. These instantaneous changes—so-called state jumps—are influenced by control variables that, together with the subsystem switching times, should be selected in an optimal manner. As with the previous optimal control problems, we tackle this problem by transforming the time variable to obtain an equivalent problem in which the switching times are fixed. However, the functions governing the state jumps in this new problem are discontinuous. To overcome this difficulty, we introduce an approximate problem whose state jumps are governed by smooth functions. This approximate problem can be solved using a nonlinear programming algorithm. We prove an important convergence result that links the approximate problem's solution with the original problem's solution.

The final optimal control problem that we consider is a parameter identification problem. The aim of this problem is to use given experimental data to identify unknown state-delays in a nonlinear delay-differential system. More precisely, the optimal control problem involves choosing the state-delays to minimize a cost function measuring the discrepancy between predicted and observed system output. We show that the gradient of this cost function can be computed by solving an auxiliary delay-differential system. On the basis of this result, the optimal control problem can be formulated—and hence solved—as a standard nonlinear programming problem.

List of publications

The following papers (which have been published or accepted for publication) were completed during PhD candidature:

- R. C. Loxton, K. L. Teo, and V. Rehbock, "On a class of optimal control problems with variable time points in the objective and constraint functionals," in *Proceedings of The 7th International Conference on Optimization Techniques and Applications*, December 2007.

- R. C. Loxton, K. L. Teo, and V. Rehbock, "Optimal control problems with multiple characteristic time points in the objective and constraints," *Automatica*, vol. 44, no. 11, pp. 2923-2929, 2008.

- B. Farhadinia, K. L. Teo, and R. C. Loxton, "A computational method for a class of non-standard time optimal control problems involving multiple time horizons," *Mathematical and Computer Modelling*, vol. 49, no. 7-8, pp. 1682-1691, 2009.

- Q. Lin, Y. H. Wu, and R. C. Loxton, "A generalized expansion method for nonlinear wave equations," *Journal of Physics A: Mathematical and Theoretical*, vol. 42, 2009.

- Q. Lin, Y. H. Wu, and R. C. Loxton, "On the Cauchy problem for a generalized Boussinesq equation," *Journal of Mathematical Analysis and Applications*, vol. 353, no. 1, pp. 186-195, 2009.

- Q. Lin, Y. H. Wu, R. C. Loxton, and S. Lai, "Linear B-spline finite element method for the improved Boussinesq equation," *Journal of Computational and Applied Mathematics*, vol. 224, no. 2, pp. 658-667, 2009.

- R. C. Loxton, K. L. Teo, and V. Rehbock, "Computational method for a class of switched system optimal control problems," *IEEE Transactions on Automatic Control*, vol. 54, no. 10, pp. 2455-2460, 2009.

- R. C. Loxton, K. L. Teo, V. Rehbock, and W. K. Ling, "Optimal switching instants for a switched-capacitor DC/DC power converter," *Automatica*, vol. 45, no. 4, pp. 973-980, 2009.

- R. C. 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*, vol. 45, no. 10, pp. 2250-2257, 2009.

- L. Y. Wang, W. H. Gui, K. L. Teo, R. C. Loxton, and C. H. Yang, "Time delayed optimal control problems with multiple characteristic time points: Computation and industrial applications," *Journal of Industrial and Management Optimization*, vol. 5, no. 4, pp. 705-718, 2009.

- H. Xu, Y. Chen, K. L. Teo, and R. C. Loxton, "An impulsive stabilizing control of a new chaotic system," *Dynamic Systems and Applications*, vol. 18, no. 2, pp. 241-250, 2009.

- Q. Lin, R. C. Loxton, K. L. Teo, and Y. H. Wu, "A new computational method for a class of free terminal time optimal control problems," *Pacific Journal of Optimization*, to appear.

- L. Y. Wang, W. H. Gui, K. L. Teo, and R. C. Loxton, "Optimal control problems arising in the zinc sulphate electrolyte purification process," *Journal of Global Optimization*, to appear.

- S. F. Woon, V. Rehbock, and R. C. Loxton, "Global optimization method for continuous time sensor scheduling," *Nonlinear Dynamics and Systems Theory*, to appear.

The following papers were completed during PhD candidature and are currently under review:

- R. C. Loxton, K. L. Teo, and V. Rehbock, "An optimization approach to state-delay identification," *IEEE Transactions on Automatic Control*, conditionally accepted.

- S. F. Woon, V. Rehbock, and R. C. Loxton, "Optimal operation for a hybrid power system."

Acknowledgements

The research reported in this thesis was carried out between February 2007 and August 2009, while I was a PhD student in the Department of Mathematics and Statistics, Curtin University of Technology. I am very grateful for the help I received from friends, family, teachers, and colleagues during this time. Without it, this thesis could not be completed.

I am especially indebted to my supervisors, Prof. Kok Lay Teo and A/Prof. Volker Rehbock. They have guided my research during the past three years with remarkable patience and enthusiasm. A/Prof. Rehbock, in particular, should be commended for persevering with me for so long: he was also the supervisor of my third-year and honours projects! In fact, he first introduced me to control theory in an undergraduate course called *Systems Theory and Control 302*. Prof. Teo, meanwhile, deserves special thanks for sponsoring me to attend conferences in Japan, Taiwan, and Adelaide.

I also thank Prof. Yong Hong Wu, the postgraduate coordinator in the Department of Mathematics and Statistics. Prof. Wu was the chairman of my thesis committee and he helped me on many occasions. He also taught me while I was an undergraduate student; I took several of his numerical analysis courses and they provided an excellent foundation for my PhD research. Furthermore, at the end of my honours year, Prof. Wu helped me to apply for an Australian Postgraduate Award (a PhD scholarship), which I was fortunate enough to receive.

Between September 2008 and January 2009, I visited the Department of Applied Mathematics at The Hong Kong Polytechnic University. I am grateful to Dr. Cedric Yiu for inviting me to Hong Kong and for supervising my research while I was there. Working in Hong Kong was an excellent opportunity to collaborate with Dr. Yiu and his PhD student Liu Jingzhen. I thank Jingzhen for helping me to settle in when I first arrived in Hong Kong and for being an excellent colleague. I also thank my other friends in Hong Kong—in particular, Hu Wenhao, An Congpei, Zhang Xingfa, Zhou Hongjun, Chen Zhangyou, Xing Na, Jia Yujie, Stacey Birkett, and Celeste Hao—for making my visit so enjoyable.

I thank all of the staff in the Department of Mathematics and Statistics for contributing to a friendly work environment. The administrative staff past and present—Joyce Yang, Shuie Liu, Florence Wong, Carey Ryken-Rapp, Lisa Holling, and Shally Wang—deserve special thanks for providing kind and professional help on numerous occasions. I

am also particularly grateful to Prof. Lou Caccetta, Dr. Ventsi Rumchev, and Dr. Greg Gamble. Prof. Caccetta gave me the opportunity to teach many talented school students in the department's mathematics enrichment program. This was a very rewarding (and also challenging!) experience. Prof. Caccetta also gave me valuable advice concerning mathematics research. Dr. Rumchev, meanwhile, was the lecturer-in-charge of most of the courses that I taught as a sessional tutor. I really appreciate his advice and encouragement. Last but not least, Dr. Gamble was my LaTeX saviour on many occasions!

I have really enjoyed working with the other students in the Department of Mathematics and Statistics, especially Lin Qun, Xu Honglei, Wang Lingyun, Siew Fang Woon, Tiffany Jones, Minou Rabiei, Li Bin, Jimmy Liu, Li Rui, Kyle Chow, Agustinus Ribal, Jonathan Blanchard, Waseem Alshanti, Zhou Jingyang, Ian Loosen, Renato Costa, and Sarah Peursum. I thank each of them for being a wonderful friend and colleague. Additionally, I thank each of the department's academic visitors that I had the opportunity to meet. Dr. Feng Zhiguo, in particular, has been an excellent friend and mentor.

Finally, on a more personal note, I sincerely thank everyone in my family and three good friends—Yongju Rue, Li Binghui, and Winnie Wen—for their support and encouragement during my PhD candidature. I should also acknowledge assistance given by the National Natural Science Foundation of China, which supported my joint research with Xu Honglei and Prof. Teo under Grant 60704003.

Contents

1 Introduction **1**

1.1 Motivation and background 1
1.2 Overview of this thesis 9
1.3 Notation 13

2 Optimal control problems with characteristic-time constraints **17**

2.1 Introduction 17
2.2 Problem formulation 18
2.3 Problem approximation 20
2.4 Gradient computation 25
2.5 Numerical examples 30
2.5.1 Optimal observation times 30
2.5.2 Optimal chemotherapy administration 32
2.6 Conclusion 36

3 Optimal control problems with continuous inequality constraints **37**

3.1 Introduction 37
3.2 Problem formulation 38
3.3 Problem approximation 40
3.4 Time-scaling transformation 42
3.5 Solving Problem $\tilde{\mathrm{P}}_p$ 47
3.6 Convergence results 51
3.7 Numerical examples 57
3.7.1 Rayleigh's optimal control problem 58
3.7.2 Optimal control of a container crane 59
3.8 Conclusion 60

4 Optimal control of a switched-capacitor DC-DC power converter **63**

4.1 Introduction 63
4.2 Problem formulation 64
4.3 Problem transformation 67
4.4 Solving Problem $\tilde{\mathrm{Q}}$ 75

4.5 Existence of an optimal solution . . . 78
4.6 A numerical example . . . 80
4.7 Conclusion . . . 84
4.A System matrices in Section 4.6 . . . 85

5 Optimal control of a switched system 87
5.1 Introduction . . . 87
5.2 Problem formulation . . . 88
5.3 Time-scaling transformation . . . 90
5.4 Problem approximation . . . 92
5.5 Convergence results and algorithm . . . 99
5.6 A numerical example . . . 108
5.7 Conclusion . . . 109

6 State-delay identification via optimal control techniques 111
6.1 Introduction . . . 111
6.2 Problem formulation . . . 112
6.3 Preliminary results . . . 114
6.4 The main result . . . 121
6.5 Gradient computation . . . 126
6.6 Numerical examples . . . 127
6.6.1 A predator-prey model . . . 128
6.6.2 A continuously stirred tank reactor model . . . 129
6.7 Conclusion . . . 131

7 Summary and future research directions 133
7.1 Main contributions of this thesis . . . 133
7.2 Future research directions . . . 135

Bibliography 137

CHAPTER 1

Introduction

1.1 Motivation and background

A *system* is a set of interrelated components that interact according to some mechanism. The size of a system can range from minuscule—an atom is a system consisting of protons, neutrons, and electrons—to gigantic—the solar system itself is a system consisting of stars, planets, comets, astroids, and moons. Machines, computers, vehicles, the Internet, transportation networks, and buildings are examples of man-made systems that are ubiquitous in daily life.

Many systems can be influenced, to varying degrees, by human action. These include practically all man-made systems, which are designed for a specific purpose and are useful insofar as they can be manipulated to behave in a desired manner. Natural systems, on the other hand, are not constructed by humans and may be impervious to human activity. The earth's orbit of the sun, for example, is not affected by what we do on its surface. Many natural systems, however, can be influenced by human action. For example, a marine ecosystem is affected by fishing and littering, which are human activities. Furthermore, the behavior of the human body can be manipulated through medicines and devices such as pacemakers and hemodialysis machines.

For any system, a natural question to ask is: how can we influence this system to our advantage? In other words, what is the best strategy for manipulating or controlling this system? This fundamental question has occupied human thought since the dawn of mankind. Several thousand years ago, it mainly pertained to systems that provided food, such as farms, and it was probably answered through trial and error. For example, over many years of experimentation, humans learnt cultivation techniques that improved crop quality. Today, the question applies to any of a diverse range of man-made systems—from submarines to hovercrafts, from iPods to PlayStations, and from skyscrapers to open-pit mines. Sophisticated techniques have been developed for determining the most efficient ways of controlling these modern systems.

Since the advent of the digital computer, mathematical modeling has emerged as a powerful technique for investigating systems. The main idea of mathematical modeling is

to encapsulate a system's behavior in a model that can be implemented on a computer. The model can then be used to simulate the system under different control strategies, which enables the system designer to identify the best strategies. Furthermore, the model can be analyzed using mathematical techniques to give insight into the system. The major virtue of mathematical modeling is that the system's response to any control strategy can be predicted immediately by the model. Therefore, the system designer can determine a satisfactory control strategy *without* operating the system. This eliminates the need for expensive (and time-consuming) experiments that test numerous control strategies—some of which may yield very poor results—on the real system.

Static systems—those that do not change with time—are typically modeled by algebraic equations and/or inequalities. For example, many systems can be modeled algebraically as follows:

$$G_i(\boldsymbol{\sigma}) = 0, \qquad i = 1, \ldots, l, \tag{1.1}$$

and

$$G_i(\boldsymbol{\sigma}) \geq 0, \qquad i = l+1, \ldots, q, \tag{1.2}$$

where $\boldsymbol{\sigma} \in \mathbb{R}^r$ is a *control vector* and $G_i : \mathbb{R}^r \to \mathbb{R}$, $i = 1, \ldots, q$, are given functions. The components of the control vector are called *control variables* or *decision variables*; they represent quantities in the system that can be chosen by the system designer. The equations (1.1) are called *equality constraints*, and the inequalities (1.2) are called *inequality constraints*. These constraints represent the natural limitations of the system; only control vectors that satisfy them constitute valid control strategies. Accordingly, we say that a control vector is *feasible* if it satisfies the constraints (1.1)-(1.2). The set of all feasible control vectors is called the *feasible region*.

The feasible region usually comprises many control vectors. The key question is: which is the best—or optimal—control vector? To answer this question, we must express some indicator of the system's performance, such as system cost or system efficiency, as a function of the control vector. This function is called an *objective function*. If the objective function measures a quantity that the system designer wishes to maximize (for example, efficiency, revenue, or profit), then it is called a *performance index*. On the other hand, if the objective function measures a quantity that the system designer wishes to minimize (for example, cost or wastage), then it is called a *cost function*.

We say that a control vector is optimal if it is feasible and, in addition, it maximizes/minimizes the appropriate objective function over the entire feasible region. The problem of determining an optimal control vector is stated formally below.

Problem $\mathbf{P_1}$. *Find a feasible control vector—that is, a control vector satisfying the constraints* (1.1)-(1.2)*—that maximizes/minimizes a given objective function* $G_0 : \mathbb{R}^r \to \mathbb{R}$.

When the objective and constraint functions (G_i, $i = 0, \ldots, q$) are linear, Problem P_1 is called a *linear programming problem* or *linear program*. Otherwise, it is called a *nonlinear*

programming problem. Nonlinear programming problems are also called nonlinear optimization problems, mathematical programming problems, static optimization problems, or parameter optimization problems. The linear and nonlinear programming problems are classical and have been studied extensively over the past five decades. Many excellent books are devoted entirely to these topics—see, for example, [7, 23, 76, 79, 113].

Because of their special structure, linear programming problems are easier to solve than nonlinear programming problems. In fact, linear programming problems can be solved *globally* using the well-known simplex algorithm. This means that applying the simplex algorithm always yields a feasible control vector that optimizes the objective function over the entire feasible region—a so-called *global solution*. In contrast, it is difficult in general to solve nonlinear programming problems globally. The *Karush-Kuhn-Tucker conditions* (KKT conditions), which are the most important theoretical tools in nonlinear programming, furnish necessary conditions for a *local solution*—a feasible control vector that is superior to its neighbouring control vectors, but is not necessarily the best in the feasible region. Consequently, most nonlinear programming algorithms search for a local solution, which may or may not be globally optimal. Nevertheless, nonlinear programming is widely used in practice to solve complicated optimization problems that arise in industry. Moreover, new algorithms have recently been proposed for solving nonlinear programming problems globally—see, for example, [67, 128–130, 134] and the references cited therein.

The KKT conditions provide the theoretical foundation for *gradient-based nonlinear programming algorithms* (gradient-based NLP algorithms). These algorithms use the gradient of the objective and constraint functions to compute a sequence of control vectors that converges to a control vector satisfying the KKT conditions. Naturally, gradient-based NLP algorithms are only applicable if the objective and constraint functions are differentiable.

Most gradient-based NLP algorithms involve the following main steps. First, an initial candidate control vector is chosen, and the objective and constraint functions and their gradients are evaluated at this vector. Second, the information computed in the first step is used to test whether the current control vector satisfies the KKT conditions. If it does, then the algorithm stops; otherwise, the information computed in the first step is used to construct a *search direction*, which points towards an improved control vector. If the initial control vector is infeasible, then an improved control vector is one that is closer to the feasible region. On the other hand, if the initial control vector is feasible, then an improved control vector is one that is feasible and has an improved objective function value. The final step in the algorithm involves calculating the *step length*—the distance that must be travelled along the search direction to obtain an improved control vector. These steps are then repeated with the initial control vector replaced by the new one.

The procedure described above generates a sequence of candidate control vectors re-

cursively as follows:

$$\boldsymbol{\sigma}^k = \boldsymbol{\sigma}^{k-1} + \alpha_k \boldsymbol{d}^k, \qquad k \geq 1,$$

where $\boldsymbol{d}^k \in \mathbb{R}^r$ is the search direction at the kth iteration; $\alpha_k \in \mathbb{R}$ is the step length at the kth iteration; and $\boldsymbol{\sigma}^0 \in \mathbb{R}^r$ is the initial candidate control vector. Different algorithms compute the search direction and step length in different ways; this must be done appropriately to ensure that the sequence $\{\boldsymbol{\sigma}^k\}_{k=0}^{\infty}$ converges to a control vector satisfying the KKT conditions. Many high-quality implementations of gradient-based NLP algorithms are available, including NLPQLP [92, 93], FFSQP [135], and NPSOL [33].

In the previous two paragraphs, we briefly described an algorithmic framework for solving Problem P_1. This framework is also applicable to optimization problems involving *dynamic systems*. In fact, in this thesis, we will solve several complicated dynamic optimization problems by either transforming them into, or approximating them by, nonlinear programming problems.

Dynamic systems change with time and are therefore more complicated than the static system (1.1)-(1.2). They are represented mathematically by *dynamic models*, which consist of difference equations, ordinary differential equations, differential-algebraic equations, partial differential equations, delay-differential equations, stochastic differential equations, or integro-differential equations. In addition to control variables, dynamic models also have *state variables*, which describe the current state of the system. The state variables are influenced by the control variables through the equations comprising the dynamic model.

One of the most common dynamic models* is

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}\big(t, \boldsymbol{x}(t), \boldsymbol{u}(t)\big), \qquad t \in [0, T], \tag{1.3}$$

and

$$\boldsymbol{x}(0) = \boldsymbol{x}^0, \tag{1.4}$$

where $T > 0$ is a given time; $\boldsymbol{x}(t) \in \mathbb{R}^n$ is the *state vector* (whose components are the state variables) at time t; $\boldsymbol{x}^0 \in \mathbb{R}^n$ is a given initial state vector; $\boldsymbol{u}(t) \in \mathbb{R}^r$ is the *control vector* (whose components are the control variables) at time t; and $\boldsymbol{f} : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^n$ is a given function. In this model, the system starts in state $\boldsymbol{x}^0$ at time $t = 0$ and evolves smoothly in accordance with (1.3) until time $t = T$. Thus, at each time, the state's instantaneous rate of change is a function of the time, the current state, and the current value of the control variables. Many practical systems can be accurately modeled in this way; examples include penicillin production systems [80], container cranes [91], the F-8 aircraft [28, 51], and batch crystallization systems [43, 87]. The time T in (1.3) is called the *terminal time*, and the interval $[0, T]$ is called the *time horizon*.

*It is customary to refer to both the system and the system's model as "the system". Thus, in the sequel, we will often say "system" instead of the more precise term "dynamic model".

A control strategy for system (1.3)-(1.4) is a function $\boldsymbol{u} : [0, T] \to \mathbb{R}^r$ that returns the value of the control vector at each point in the time horizon. Such a control strategy is also called a *control function*. In practice, there are often constraints on the control function. For example, in the batch crystallization system described in [87], the value of the control function cannot be increased arbitrarily because it represents the temperature of a sodium aluminate solution. Accordingly, we normally assume that the range of the control function is contained within some proper subset $\mathcal{W}$ of $\mathbb{R}^r$. This set is called the *control restraint set*, and a function $\boldsymbol{u} : [0, T] \to \mathcal{W}$ is called an *admissible control function*.

The control function influences the evolution of the dynamic system through equation (1.3). The key question is: how should we choose this control function so that the system evolves in the most optimal way? To answer this question, we must formulate an appropriate objective function in terms of the system state and/or control function. For example, consider a cruise missile whose mission is to hit a target with position vector $\hat{\boldsymbol{x}} \in \mathbb{R}^3$ at time $t = T$. If the state vector $\boldsymbol{x}(t) \in \mathbb{R}^3$ is the position vector of the missile at time t, then an appropriate objective function is

$$G_0 = (x_1(T) - \hat{x}_1)^2 + (x_2(T) - \hat{x}_2)^2 + (x_3(T) - \hat{x}_3)^2. \tag{1.5}$$

This non-negative function measures the distance from the missile to the target at the terminal time; it is equal to zero if and only if the missile strikes the target. Accordingly, a suitable control strategy for the missile is one that minimizes (1.5).

Equation (1.5) is a special case of the following more general objective function:

$$G_0 = \Phi(\boldsymbol{x}(T)) + \int_0^T \mathcal{L}\big(t, \boldsymbol{x}(t), \boldsymbol{u}(t)\big)dt, \tag{1.6}$$

where $\Phi : \mathbb{R}^n \to \mathbb{R}$ and $\mathcal{L} : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$ are given functions.

We now state a classical *optimal control problem*.

Problem $\mathbf{P_2}$. *Find an admissible control function that causes the system* (1.3)-(1.4) *to evolve in such a way that the objective function* (1.6) *is minimized.*

Problem P_2 involves finding a vector-valued function (as opposed to just a vector) and is therefore more complicated than Problem P_1. Furthermore, Problem P_2 is governed by a dynamic system that changes with time (recall that Problem P_1 is governed by a static system comprised of equations and inequalities). If $\mathcal{L} = 0$ in (1.6), then Problem P_2 is called a *Mayer problem*; if $\Phi = 0$, then Problem P_2 is called a *Lagrange problem*; and if both $\mathcal{L}$ and Φ are non-zero, then Problem P_2 is called a *Bolza problem*. The Mayer, Lagrange, and Bolza problems are all equivalent: each can easily be transformed into any of the others. There are two classical results that can be used to solve Problem P_2: Pontryagin's minimum principle and Bellman's principle of optimality. We discuss each briefly, starting with the minimum principle.

The Pontryagin minimum principle was developed by Pontryagin and his colleagues [83]. It is probably the most famous result in optimal control theory, and is discussed in most books on the subject—see, for example, [1,2,53,96,133]. To state the minimum principle, we must define the so-called *Hamiltonian function* $H : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$. This function is defined as follows:

$$H(t, \boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{u}) \triangleq \mathcal{L}(t, \boldsymbol{x}, \boldsymbol{u}) + \boldsymbol{\lambda}^T \boldsymbol{f}(t, \boldsymbol{x}, \boldsymbol{u}), \qquad (t, \boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{u}) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^r.$$

The essence of the minimum principle is that an optimal control for Problem P_2 minimizes the Hamiltonian at every point in the time horizon. More precisely, if $\boldsymbol{u}^* : [0, T] \to \mathcal{W}$ is an optimal control and $\boldsymbol{x}^* : [0, T] \to \mathbb{R}^n$ is its corresponding state (defined via (1.3)-(1.4)), then

$$H\big(t, \boldsymbol{x}^*(t), \boldsymbol{\lambda}^*(t), \boldsymbol{u}^*(t)\big) = \min_{\boldsymbol{w} \in \mathcal{W}} H\big(t, \boldsymbol{x}^*(t), \boldsymbol{\lambda}^*(t), \boldsymbol{w}\big), \qquad t \in [0, T], \tag{1.7}$$

where $\boldsymbol{\lambda}^* : [0, T] \to \mathbb{R}^n$ is a function—called the *costate*—that satisfies

$$\dot{\boldsymbol{\lambda}}^*(t) = -\left[\frac{\partial H\big(t, \boldsymbol{x}^*(t), \boldsymbol{\lambda}^*(t), \boldsymbol{u}^*(t)\big)}{\partial \boldsymbol{x}}\right]^T, \qquad t \in [0, T], \tag{1.8}$$

and

$$\boldsymbol{\lambda}^*(T) = \left[\frac{\partial \Phi(\boldsymbol{x}^*(T))}{\partial \boldsymbol{x}}\right]^T. \tag{1.9}$$

How do these equations help determine an optimal control? The main idea is to rearrange equation (1.7) so that the optimal control is expressed in terms of the time, the state, and the costate. In other words, one attempts to construct a function $\bar{\boldsymbol{u}} : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^r$ such that

$$\boldsymbol{u}^*(t) = \bar{\boldsymbol{u}}(t, \boldsymbol{x}^*(t), \boldsymbol{\lambda}^*(t)), \qquad t \in [0, T]. \tag{1.10}$$

If such a function exists, then in principle it may be substituted into equations (1.3)-(1.4) and (1.8)-(1.9) to yield a two-point boundary-value problem in terms of the optimal state and costate. Solving this boundary-value problem yields a candidate optimal control that satisfies equations (1.7)-(1.9). There is no guarantee, however, that this control is optimal: the minimum principle is a *necessary* condition for optimality, but it is not sufficient in general.

A sufficient condition for optimality in Problem P_2 can be derived using Bellman's principle of dynamic programming [8]. To state the major implication of Bellman's principle, we need to introduce the following dynamic system:

$$\dot{\boldsymbol{y}}(s) = \boldsymbol{f}\big(s, \boldsymbol{y}(s), \boldsymbol{u}(s)\big), \qquad s \in [t, T], \tag{1.11}$$

and

$$\boldsymbol{y}(t) = \boldsymbol{x}, \tag{1.12}$$

where $t \geq 0$ is an initial time and $\boldsymbol{x} \in \mathbb{R}^n$ is an initial state. Notice that this system is almost the same as (1.3)-(1.4). Here, however, the initial time and state are not fixed. Let $\boldsymbol{y}(\cdot|t,\boldsymbol{x})$ denote the solution of (1.11)-(1.12) corresponding to $(t,\boldsymbol{x}) \in [0,T] \times \mathbb{R}^n$.

We define the *value function* $V : [0,T] \times \mathbb{R}^n \to \mathbb{R}$ as follows:

$$V(t,\boldsymbol{x}) \triangleq \inf_{\boldsymbol{u}\in\mathcal{U}} \left\{ \Phi(\boldsymbol{y}(T|t,\boldsymbol{x})) + \int_t^T \mathcal{L}\big(s,\boldsymbol{y}(s|t,\boldsymbol{x}),\boldsymbol{u}(s)\big)ds \right\}, \qquad (t,\boldsymbol{x}) \in [0,T] \times \mathbb{R}^n,$$

where $\mathcal{U}$ is the set consisting of all admissible control functions. It turns out that the value function satisfies the partial differential equation

$$\frac{\partial V(t,\boldsymbol{x})}{\partial t} + \inf_{\boldsymbol{w}\in\mathcal{W}} \left\{ \frac{\partial V(t,\boldsymbol{x})}{\partial \boldsymbol{x}} \boldsymbol{f}(t,\boldsymbol{x},\boldsymbol{w}) + \mathcal{L}(t,\boldsymbol{x},\boldsymbol{w}) \right\} = 0, \quad (t,\boldsymbol{x}) \in [0,T] \times \mathbb{R}^n, \quad (1.13)$$

and the boundary condition

$$V(T,\boldsymbol{x}) = \Phi(\boldsymbol{x}), \qquad \boldsymbol{x} \in \mathbb{R}^n. \quad (1.14)$$

Equation (1.13) is known as the *Hamilton-Jacobi-Bellman equation*; it is established by applying Bellman's principle of dynamic programming to Problem P_2 (see, for example, [4]). If a solution of equations (1.13)-(1.14) exists, then it may be used to construct an optimal control for Problem P_2. Hence, equations (1.13)-(1.14) constitute a sufficient condition for optimality in Problem P_2.

Pontryagin's minimum principle and the Hamilton-Jacobi-Bellman equation are powerful tools. However, it is often difficult (if not impossible) to solve (1.7)-(1.9) or (1.13)-(1.14) analytically. Consequently, numerical methods for solving optimal control problems are indispensable. Many such methods have been proposed in the literature. Some of these are based on finite-difference or finite-volume approximation schemes for the Hamilton-Jacobi-Bellman equation—see, for example, [35,118,119] and the references cited therein. Other methods are based on the minimum principle. For example, a numerical shooting method can be used to solve the two-point boundary-value problem derived from the minimum principle [1, 53]. Shooting methods, however, require an accurate initial estimate of the costate, and it is usually difficult to obtain such an estimate. Nevertheless, a new method has recently been proposed for determining the value of the costate at the initial time [19, 20]. If the initial costate can be determined in advance, then the state and costate systems can be solved simultaneously, which allows the optimal control to be constructed forward in time using equation (1.10).

Control parameterization is another approach to solving optimal control problems numerically [100]. The main idea of control parameterization is to approximate the control function by a linear combination of basis functions, the coefficients of which are decision variables to be chosen optimally. In other words, the control is approximated by a

function that is completely determined by a finite number of parameters. Applying this approximation scheme yields an approximate optimization problem with a finite number of decision variables. This approximate problem is much easier to solve than the original optimal control problem.

The control parameterization technique is the basis of several optimal control softwares, including MISER3 [47], and has been applied to a wide variety of practical optimal control problems (see, for example, [24, 58, 66, 78, 87]). It is normally implemented with piecewise constant basis functions. In this case, the time horizon is partitioned into fixed subintervals, and the control is approximated by a constant function on each subinterval. The problem is then to choose a value for the control on each subinterval so that the objective function is maximized/minimized. This is a nonlinear programming problem whose decision variables influence the objective function *implicitly* through the dynamic system. Thus, it is very difficult to derive the gradient of the objective with respect to the decision variables. Nevertheless, gradient formulae for an objective function in canonical form are derived in [39,100]. These formulae can be used in conjunction with a gradient-based NLP algorithm to solve the approximate problem, and thereby obtain an approximate solution for the original optimal control problem.

The control parameterization technique described above involves partitioning the time horizon in advance. As such, the times at which the approximate control changes its value—the so-called *switching times*—are fixed. A better approach is to consider the switching times, in addition to the control values, as decision variables. In other words, the control is approximated by a piecewise constant function whose values *and* switching times are decision variables to be selected optimally. Unfortunately, although the gradient of the objective function with respect to the switching times does exist, the formulae for computing it are very difficult to implement numerically [39, 100]. Furthermore, the governing dynamic system is very difficult to integrate if the switching times are variable. In particular, major problems can arise if consecutive switching times combine to form a single switch.

To overcome these difficulties, a novel time-scaling transformation, originally called the *control parameterization enhancing technique* (commonly abbreviated to CPET), was developed in [60]. The main idea of this technique is to transform the time horizon of an optimal control problem in such a way that the switching times become fixed. More precisely, this time-scaling transformation introduces a new time variable $s \in [0, p]$, where $p - 1$ is the number of allowed control switches, and relates s to the original time variable $t \in [0, T]$ through the differential equation

$$\dot{t}(s) = v(s), \qquad s \in [0, p], \tag{1.15}$$

and the boundary conditions

$$t(0) = 0 \tag{1.16}$$

and

$$t(p) = T, \tag{1.17}$$

where $v : [0, p] \to [0, T]$ is a non-negative piecewise constant function with switching times at the fixed points $s = 1, \ldots, p-1$. The values of v represent the duration between consecutive switching times in the original time horizon. Equations (1.15)-(1.17) can be used to transform the optimal control problem into an equivalent problem that has switching times at the fixed points $s = 1, \ldots, p-1$. Note that the time-scaling transformation can easily be modified to transform the time horizon from $[0, T]$ into $[0, 1]$ instead of $[0, p]$. In this case, the switching times are mapped to $s = 1/p, \ldots, (p-1)/p$. The key point, however, is that the time-scaling transformation always maps the switching times to *fixed* points, regardless of whether the new time horizon is $[0, p]$ or $[0, 1]$.

The time-scaling transformation was originally introduced in [60] to solve time-optimal control problems. It was subsequently applied to many other classes of optimal control problems, including constrained optimal control problems [97, 103], optimal discrete-valued control problems [61], switched system optimal control problems [64, 89, 132], impulsive optimal control problems [69, 125], optimal control problems involving multiple coupled subsystems [22], and singular optimal control problems [95]. The time-scaling transformation has also been used to solve practical problems involving sensors [59, 123], hybrid power systems [90, 122], and submarines [12], and in areas as diverse as system identification [127], differential equations [56], and mixed integer programming [57].

It is worth emphasizing that the control parameterization technique only discretizes the control. Another numerical method—the so-called *state discretization method*—proceeds by discretizing *both* the control and the state [36, 50, 114]. Applying this method yields a set of difference equations—essentially equality constraints—that approximate the governing dynamic system. As with control parameterization, state-discretization yields an approximate problem that can be solved using a gradient-based NLP algorithm. But the approximate problem is usually much larger and contains more nonlinear equality constraints than the one derived using control parameterization.

1.2 Overview of this thesis

In the previous section, we gave a brief introduction to nonlinear programming and optimal control. In particular, we introduced two problems: Problem P_1 (a nonlinear programming problem) and Problem P_2 (an optimal control problem). These are standard problems that have been studied extensively since the 1950s. The purpose of this thesis is to present new algorithms for solving five *nonstandard* optimal control problems. We

briefly describe these five problems below.

In Chapter 2, we consider an optimal control problem in which the governing dynamic system is subject to constraints of the following type:

$$\Phi\big(\boldsymbol{x}(\tau_1),\ldots,\boldsymbol{x}(\tau_m)\big)+\int_0^T \boldsymbol{\mathcal{L}}\big(\boldsymbol{x}(t),\boldsymbol{u}(t)\big)\geq 0,$$

where $\Phi:\mathbb{R}^{mn}\to\mathbb{R}$ and $\boldsymbol{\mathcal{L}}:\mathbb{R}^n\times\mathbb{R}^r\to\mathbb{R}$ are given functions and τ_j, $j=1,\ldots,m$, are time points called *characteristic times*. Such constraints—which are called *characteristic-time inequality constraints*—first appeared in optimal control problems concerned with finding effective cancer treatment strategies [77, 78]. The problems in [77, 78] have fixed characteristic times; we consider a more difficult problem in which the characteristic times are actually control variables to be determined optimally. Thus, the characteristic-time inequality constraints in Chapter 2 depend on the state at times that are initially unknown. Such constraints are much more complicated than conventional inequality constraints, which usually depend only on final state reached by the system.

In Chapter 3, we consider another type of constrained optimal control problem. The constraints in this problem are of the form

$$h\big(\boldsymbol{x}(t),\boldsymbol{u}(t)\big)\geq 0,\qquad t\in[0,T],$$

where $h:\mathbb{R}^n\times\mathbb{R}^r\to\mathbb{R}$ is a given function. Such constraints are called *continuous inequality constraints*; they are imposed at *every* point in the time horizon. A single continuous inequality constraint can actually be viewed as an infinite number of inequality constraints—one for each point in the time horizon. It is therefore not surprising that continuous inequality constraints are very difficult to handle, both theoretically and computationally. However, they are also one of the most common constraints in practice. This is because many systems have requirements that must be satisfied at all times in the time horizon, not just at several isolated times.

In Chapter 4, we consider the problem of controlling a switched-capacitor DC-DC power converter. A switched-capacitor DC-DC power converter is an electronic device, consisting primarily of capacitors and switches, that transforms one DC voltage (the input) into another (the output) by repeatedly switching between different circuit topologies. The output voltage of the power converter must be as steady as possible to ensure that the attached appliance runs correctly. It should also be robust with respect to changes and uncertainties in the load resistance and input voltage. In other words, changing the attached appliance or the voltage source should not cause drastic changes in the output. Hence, the times at which the circuit topologies are switched should be chosen so that both the *output voltage ripple* (the difference between the maximum and minimum output voltages) and the *output voltage sensitivity* (the derivative of the output voltage) are as

small as possible. We formulate this problem as an optimal control problem in which the switching times are chosen to minimize the following cost function:

$$G_0 = \alpha\left\{ \sup_{t\in[0,T]} y(t) - \inf_{t\in[0,T]} y(t) \right\} + \beta \sup_{t\in[0,T]} \left| \frac{\partial y(t)}{\partial R_L} \right| + \gamma \sup_{t\in[0,T]} \left| \frac{\partial y(t)}{\partial \boldsymbol{\sigma}} \right|_\infty, \tag{1.18}$$

where $T > 0$ is a given terminal time; $y(t) \in \mathbb{R}$ is the output voltage at time t; $R_L \in \mathbb{R}$ is the resistance of the attached appliance; $\boldsymbol{\sigma} \in \mathbb{R}^r$ is the DC input voltage; and α, β, and γ are non-negative weights. Note that the first term in (1.18) penalizes the output voltage ripple, while the second and third terms penalize the output voltage sensitivity. This optimal control problem is difficult to solve for two reasons. First, the cost function (1.18) is non-smooth and thus cannot be minimized directly using a gradient-based NLP algorithm. Second, the dynamic model of a switched-capacitor DC-DC power converter is much more complicated than the systems considered in Chapters 2 and 3. In fact, the model's state variables—and the dynamics that govern them—change instantaneously at each switching time. The dynamic model of a switched-capacitor DC-DC power converter is actually a *switched system* of the following type:

$$\dot{\boldsymbol{x}}(t) = A_k \boldsymbol{x}(t) + B_k \boldsymbol{\sigma}, \qquad t \in (t_{k-1}, t_k), \qquad k = 1, \ldots, m, \tag{1.19}$$

$$y(t) = C_k \boldsymbol{x}(t) + D_k \boldsymbol{\sigma}, \qquad t \in [t_{k-1}, t_k), \qquad k = 1, \ldots, m, \tag{1.20}$$

and

$$\boldsymbol{x}(t_k^+) = \begin{cases} \boldsymbol{x}^0, & \text{if } k = 0, \qquad (1.21a) \\ \boldsymbol{x}(t_k^-) + \boldsymbol{z}^k\big(\boldsymbol{x}(t_k^-)\big), & \text{if } k \in \{1, \ldots, m-1\}, \qquad (1.21b) \end{cases}$$

where $t_0 = 0$, $t_m = T$, and t_k, $k = 1, \ldots, m-1$, are the switching times; $\boldsymbol{x}(t) \in \mathbb{R}^n$ is the state voltage vector (whose components represent the voltages of the different capacitors) at time t; A_k, B_k, C_k, and D_k, $k = 1, \ldots, m$, are given matrices that depend on the load resistance; and $\boldsymbol{z}^k : \mathbb{R}^n \to \mathbb{R}^n$, $k = 1, \ldots, m-1$, are given functions. In this switched system, the state voltage begins at $\boldsymbol{x}^0$ at time $t = 0$ and evolves smoothly according to equation (1.19) with $k = 1$ until time $t = t_1$. The circuit topology is then switched, and this causes the state voltage to change instantaneously from $\boldsymbol{x}(t_1^-)$ to $\boldsymbol{x}(t_1^+)$; see equation (1.21b). This instantaneous change is called a *state jump*; it models the energy loss that occurs when the circuit topology is switched. Restarting from $\boldsymbol{x}(t_1^+)$, the state voltage again evolves smoothly according to (1.19) with $k = 2$ until time $t = t_2$, at which time the circuit topology switches again, and the state voltage experiences another jump from $\boldsymbol{x}(t_2^-)$ to $\boldsymbol{x}(t_2^+)$. The system continues in this way for the remainder of the time horizon.

In Chapter 5, we consider an optimal control problem involving a more general switched

system. This general switched system is of the following form:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}^k\big(\boldsymbol{x}(t), \boldsymbol{\sigma}\big), \qquad t \in (t_{k-1}, t_k), \qquad k = 1, \ldots, m,$$

and

$$\boldsymbol{x}(t_k^+) = \begin{cases} \boldsymbol{x}^0, & \text{if } k = 0, \\ \boldsymbol{x}(t_k^-) + \boldsymbol{z}^k\big(\boldsymbol{x}(t_k^-), \boldsymbol{\sigma}\big), & \text{if } k \in \{1, \ldots, m-1\} \text{ and } t_{k-1} < t_k < T, \end{cases}$$

where $t_0 = 0$, $t_m = T$, and $T > 0$ is a given terminal time; t_k, $k = 1, \ldots, m-1$, are switching times; $\boldsymbol{\sigma} \in \mathbb{R}^r$ is a control vector; and $\boldsymbol{f}^k : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^n$, $k = 1, \ldots, m$, and $\boldsymbol{z}^k : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^n$, $k = 1, \ldots, m-1$, are given functions. This switched system has *nonlinear* dynamics. Furthermore, its switching times are allowed to coincide if necessary. This means that it may be beneficial to choose $t_{k-1} = t_k$ for some $k \in \{1, \ldots, m\}$, which would effectively delete the kth subsystem. Note though, that one (and only one) state jump is applied at each *distinct* switching time—if two or more switching times coincide, then only the state jump corresponding to the first one is applied. The optimal control problem here is to choose the switching times and the control vector to minimize a given cost function.

In Chapter 6, we consider the following *delay-differential system*:

$$\dot{\boldsymbol{x}}(t) = \sum_{i=1}^{r} \boldsymbol{f}^i\big(\boldsymbol{x}(t), \boldsymbol{x}(t - \tau_i)\big), \qquad t \in (0, T], \tag{1.23}$$

$$\boldsymbol{x}(t) = \boldsymbol{z}(t), \qquad t \in [-\bar{\tau}, 0], \tag{1.24}$$

and

$$\boldsymbol{y}(t) = \boldsymbol{g}(\boldsymbol{x}(t)), \qquad t \in [-\bar{\tau}, T], \tag{1.25}$$

where $T > 0$ and $\bar{\tau} > 0$ are given real numbers; τ_i, $i = 1, \ldots, r$, are unknown state-delays that need to be identified; $\boldsymbol{x}(t) \in \mathbb{R}^n$ is the system state at time t; $\boldsymbol{y}(t) \in \mathbb{R}^m$ is the system output at time t; and $\boldsymbol{f}^i : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$, $i = 1, \ldots, r$, $\boldsymbol{z} : \mathbb{R} \to \mathbb{R}^n$, and $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^m$ are given functions. Many real-life systems are of this type; examples include aircraft [112], predator-prey systems [131], zinc production systems [116, 117], and continuously-stirred tank reactors [13]. We suppose that the system modeled by (1.23)-(1.25) has been observed (via experimentation) at p points in the time horizon. Accordingly, we have a data set

$$(t_1, \hat{\boldsymbol{y}}^1), (t_2, \hat{\boldsymbol{y}}^2), \ldots, (t_p, \hat{\boldsymbol{y}}^p),$$

where, for each $j = 1, \ldots, p$, $\hat{\boldsymbol{y}}^j$ is the output measured at time $t = t_j$. Our goal is to choose the state-delays so that the solution of (1.23)-(1.25) best fits the experimental data. Hence, our optimal control problem is to choose the state-delays to minimize the

cost function

$$G_0 = \sum_{j=1}^{p} \left| \boldsymbol{y}(t_j) - \hat{\boldsymbol{y}}^j \right|^2.$$

The unusual aspect of this optimal control problem is that the control variables are the state-delays themselves. Although many optimal control methods have been devised for delay systems—see, for example, [49, 54, 55, 62, 107, 108, 120, 121] and the references cited therein—none of them are applicable to this problem. This is because previous optimal control methods assume that the control in the system does *not* affect the delay (the delays are typically assumed to be given constants).

Finally, in Chapter 7, we summarize the main contributions of this thesis and discuss some interesting directions for future research.

1.3 Notation

To conclude this chapter, we introduce some notation.

The symbol "$\triangleq$" is used throughout this thesis to denote a definition or assignment. For example,

$$\alpha \triangleq 0$$

means that α is being assigned the value of 0. Furthermore, $\rho_{i,j}$ is used to denote the well-known *Kronecker delta* (δ is reserved for small quantities). That is,

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

We use $\hat{\rho}_{i,j}$ to denote the *cumulative Kronecker delta*. That is,

$$\hat{\rho}_{i,j} \triangleq \begin{cases} 1, & \text{if } i \leq j, \\ 0, & \text{otherwise.} \end{cases}$$

Clearly,

$$\hat{\rho}_{i,j} = \sum_{k=1}^{j} \rho_{i,k}.$$

For each $\mathcal{I} \subset \mathbb{R}$, the corresponding *indicator function* $\chi_{\mathcal{I}} : \mathbb{R} \to \mathbb{R}$ is defined by

$$\chi_{\mathcal{I}}(t) \triangleq \begin{cases} 1, & \text{if } t \in \mathcal{I}, \\ 0, & \text{otherwise.} \end{cases}$$

Vectors in the n-dimensional Euclidean space $\mathbb{R}^n$ are written in boldface, and their components are indexed with subscripts. Thus $x_1, \ldots, x_n$ are the components of $\boldsymbol{x} \in \mathbb{R}^n$.

There is one exception: to avoid confusion, we denote a vector of switching times by $\boldsymbol{\nu}$ instead of $\boldsymbol{t}$. Thus, if $t_1, \ldots, t_{m-1}$, are switching times, then

$$\boldsymbol{\nu} = [t_1, \ldots, t_{m-1}]^T \in \mathbb{R}^{m-1} \tag{1.26}$$

is the corresponding switching-time vector.

All vectors are considered column vectors. Hence,

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$$

and

$$[x_1, \ldots, x_n]^T \in \mathbb{R}^n.$$

For each $i = 1, \ldots, n$, we denote the ith standard unit basis vector in $\mathbb{R}^n$ by $\boldsymbol{e}^{n,i}$. We also denote a vector of zeros—the so-called *zero vector* or *null vector*—by $\mathbf{0}$ (its dimension will be obvious from the context).

We denote the Euclidean norm by "$|\cdot|$". Thus, if $\boldsymbol{x} \in \mathbb{R}^n$, then

$$|\boldsymbol{x}| = \sqrt{x_1^2 + \cdots + x_n^2}.$$

In addition, we denote the infinity norm by "$|\cdot|_\infty$". Hence, if $\boldsymbol{x} \in \mathbb{R}^n$, then

$$|\boldsymbol{x}|_\infty = \max_{1 \le i \le n} |x_i|.$$

As is customary, we let $\mathbb{R}^{n \times m}$ denote the set of all $n \times m$ real matrices. When working with such matrices, we use "$|\cdot|$" to denote the *natural*, or *induced*, matrix norm associated with the Euclidean norms in $\mathbb{R}^n$ and $\mathbb{R}^m$. More precisely, if $A \in \mathbb{R}^{n \times m}$, then

$$|A| = \sup \big\{ \, |A\boldsymbol{x}| : \ \boldsymbol{x} \in \mathbb{R}^m, \ |\boldsymbol{x}| = 1 \, \big\}.$$

Recall that

$$|A\boldsymbol{x}| \le |A| \cdot |\boldsymbol{x}|.$$

If $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable function, then

$$\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} \triangleq \left[\frac{\partial f(\boldsymbol{x})}{\partial x_1}, \ldots, \frac{\partial f(\boldsymbol{x})}{\partial x_n} \right], \qquad \boldsymbol{x} \in \mathbb{R}^n.$$

Thus,

$$\left[\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}} \right]^T \in \mathbb{R}^n, \qquad \boldsymbol{x} \in \mathbb{R}^n.$$

This vector is denoted by $\nabla f(\boldsymbol{x})$ and is called the *gradient* of the function f. Hence,

$$\nabla f(\boldsymbol{x}) \triangleq \left[\frac{\partial f(\boldsymbol{x})}{\partial \boldsymbol{x}}\right]^T = \begin{bmatrix} \dfrac{\partial f(\boldsymbol{x})}{\partial x_1} \\ \vdots \\ \dfrac{\partial f(\boldsymbol{x})}{\partial x_n} \end{bmatrix}, \qquad \boldsymbol{x} \in \mathbb{R}^n.$$

These concepts are easily generalized to vector-valued functions. Indeed, if $\boldsymbol{f} : \mathbb{R}^m \to \mathbb{R}^n$ is a differentiable function, then

$$\frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}} \triangleq \begin{bmatrix} \dfrac{\partial f_1(\boldsymbol{x})}{\partial x_1} & \cdots & \dfrac{\partial f_1(\boldsymbol{x})}{\partial x_m} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial f_n(\boldsymbol{x})}{\partial x_1} & \cdots & \dfrac{\partial f_n(\boldsymbol{x})}{\partial x_m} \end{bmatrix}, \qquad \boldsymbol{x} \in \mathbb{R}^m.$$

When dealing with a function of only one variable, we use an overhead dot to denote differentiation. Thus, if $\boldsymbol{\phi} : \mathbb{R} \to \mathbb{R}^n$ is a differentiable function, then $\dot{\boldsymbol{\phi}}$ denotes its derivative.

We will use "$\lfloor \cdot \rfloor$" to denote the well-known *floor function*. That is, $\lfloor x \rfloor$ is the greatest integer less than or equal to $x \in \mathbb{R}$. Furthermore, we will use an overhead bar to denote set closure. Hence, $\bar{\mathcal{A}}$ denotes the closure of the set $\mathcal{A}$.

We use a positive superscript to denote the right limit, and a negative superscript to denote the left limit. For example, if $\tau \in \mathbb{R}$ and $\boldsymbol{f} : \mathbb{R} \to \mathbb{R}^n$, then

$$\boldsymbol{f}(\tau^+) = \lim_{t \to \tau +} \boldsymbol{f}(t)$$

and

$$\boldsymbol{f}(\tau^-) = \lim_{t \to \tau -} \boldsymbol{f}(t).$$

Finally, let $\boldsymbol{f} : \mathbb{R} \to \mathbb{R}^n$ and $g : \mathbb{R} \to \mathbb{R}$ be two functions. We say that $\boldsymbol{f}$ is of order $\mathcal{O}(g(\epsilon))$ if there exists a real number $\alpha > 0$ such that for all ϵ of sufficiently small magnitude,

$$\big|\boldsymbol{f}(\epsilon)\big| \leq \alpha \big|g(\epsilon)\big|.$$

In this case, we write

$$\boldsymbol{f}(\epsilon) = \mathcal{O}(g(\epsilon)).$$

CHAPTER 2

Optimal control problems with characteristic-time inequality constraints*

2.1 Introduction

In this chapter, we consider an optimal control problem with inequality constraints that depend on the state at two or more discrete time points. These time points are called *characteristic times*, and the constraints themselves are called *characteristic-time inequality constraints* (CTI constraints). Optimal control problems with CTI constraints were first introduced in a study of chemotherapy administration policies for cancer treatment [77, 78]. In this study, an optimal control problem was formulated in which the chemotherapy delivery rate is the control variable to be determined optimally. This problem has conventional inequality constraints in addition to CTI constraints. The conventional constraints arise because of restrictions on the amount of chemotherapy that can be safely administered; the CTI constraints arise because of a requirement that the size of the cancer tumor decrease at least as fast as a prescribed rate.

The method proposed in [77, 78] for solving optimal control problems with CTI constraints is based on the control parameterization technique (see Chapter 1). More specifically, it involves approximating the control by a piecewise constant function, so that the optimal control problem becomes a dynamic optimization problem with a finite number of decision variables. An algorithm was developed in [77, 78] for computing the gradient of the objective and constraint functions in this approximate problem. This algorithm can be used in conjunction with a gradient-based nonlinear programming algorithm to solve the approximate problem.

The algorithm proposed in [77, 78] for computing the gradient of the CTI constraints has two main steps: first, the state system is integrated forward in time; second, an auxiliary costate system with jumps is integrated backwards in time. The jumps make the costate system difficult to integrate. Furthermore, since the state and costate systems are integrated in opposite directions, it is impossible to ensure that their knot sets coincide

*This chapter is based on [71, 72].

(unless a crude integration technique with fixed step lengths is used). This is a major problem, because the costate system actually depends on the solution of the state system. It is therefore necessary to interpolate the state when the costate system is being solved, which ultimately compromises the accuracy of the constraint gradients.

The characteristic times in the optimal control problem in [77, 78] are fixed. A more difficult problem, in which the characteristic times are actually decision variables to be determined optimally, is considered in [105]. The control parameterization technique, together with the time-scaling transformation, can be used to approximate this problem by a dynamic optimization problem with constraints that depend on the state at fixed characteristic times. This approximate problem can be readily solved using the method developed in [77, 78].

The optimal control problem that we consider in this chapter includes those formulated in [77,78,105] as special cases. Although our approach to solving this problem is also based on the control parameterization technique, we develop a new method for computing the gradient of the CTI constraints. Our new method is inspired by those in [51, 113, 126]; it involves integrating an auxiliary dynamic system, which does not have instantaneous jumps, forward in time. Accordingly, our new method has two important advantages over those in [77, 78, 105]: the difficulties involved in dealing with a discontinuous costate system are avoided; and, more importantly, no interpolation is required when solving the differential equations comprising the auxiliary system.

2.2 Problem formulation

Consider the following dynamic system:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}\big(\boldsymbol{x}(t), \boldsymbol{u}(t)\big), \qquad t \in [0, T], \tag{2.1}$$

and

$$\boldsymbol{x}(0) = \boldsymbol{x}^0, \tag{2.2}$$

where $T > 0$ is a given terminal time; $\boldsymbol{x}(t) \in \mathbb{R}^n$ is the system state at time t; $\boldsymbol{u}(t) \in \mathbb{R}^r$ is the control function at time t; $\boldsymbol{x}^0 \in \mathbb{R}^n$ is a given initial state; and $\boldsymbol{f} : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^n$ is a given function.

Define

$$\mathcal{W} \triangleq \big\{\, \boldsymbol{w} \in \mathbb{R}^r : \; a_\varsigma \le w_\varsigma \le b_\varsigma, \; \varsigma = 1, \ldots, r \,\big\},$$

where a_ς and b_ς, $\varsigma = 1, \ldots, r$, are given real numbers such that $a_\varsigma < b_\varsigma$. Any measurable function $\boldsymbol{u} : [0, T] \to \mathbb{R}^r$ such that $\boldsymbol{u}(t) \in \mathcal{W}$ for almost all $t \in [0, T]$ is called an *admissible control*. Let $\mathcal{U}$ denote the class of all such admissible controls.

We assume that the following conditions are satisfied.

Assumption 2.1. The function $\boldsymbol{f}$ is continuously differentiable.

Assumption 2.2. There exists a real number $L_1 > 0$ such that

$$\big|\boldsymbol{f}(\boldsymbol{v},\boldsymbol{w})\big| \leq L_1(1+|\boldsymbol{v}|), \qquad (\boldsymbol{v},\boldsymbol{w}) \in \mathbb{R}^n \times \mathcal{W}.$$

By Theorem 3.3.3 of [2], the system (2.1)-(2.2) has a unique solution corresponding to each admissible control $\boldsymbol{u} \in \mathcal{U}$. We denote this solution by $\boldsymbol{x}(\cdot|\boldsymbol{u})$.

We suppose that the dynamic system (2.1)-(2.2) is required to satisfy the following inequality constraints:

$$G_i(\boldsymbol{u},\boldsymbol{\tau}) \triangleq \Phi_i\big(\boldsymbol{x}(\tau_1|\boldsymbol{u}),\ldots,\boldsymbol{x}(\tau_m|\boldsymbol{u})\big) + \int_0^T \mathcal{L}_i\big(\boldsymbol{x}(t|\boldsymbol{u}),\boldsymbol{u}(t)\big)dt \geq 0, \quad i=1,\ldots,q, \tag{2.3}$$

where $\Phi_i : \mathbb{R}^{mn} \to \mathbb{R}$ and $\mathcal{L}_i : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$, $i = 1,\ldots,q$, are given functions and τ_j, $j = 1,\ldots,m$, are time points called *characteristic times*. Such constraints are called *characteristic-time inequality constraints* (CTI constraints). Since the CTI constraints depend on the state at intermediate points in the time horizon, they are more complicated than conventional inequality constraints, which usually depend only on the final state reached by the system.

The characteristic times are chosen by the system designer; we assume that they are required to satisfy

$$c_j \leq \tau_j \leq d_j, \qquad j=1,\ldots,m, \tag{2.4}$$

where c_j and d_j, $j = 1,\ldots,m$, are given real numbers such that

$$0 \leq c_j \leq d_j \leq T, \qquad j=1,\ldots,m,$$

and

$$d_{j-1} \leq c_j, \qquad j=2,\ldots,m.$$

These conditions on c_j and d_j, $j = 1,\ldots,m$, ensure that the characteristic times are indexed in chronological order. That is,

$$0 \leq \tau_1 \leq \tau_2 \leq \cdots \leq \tau_m \leq T.$$

Let $\mathcal{T}$ denote the set consisting of all vectors $\boldsymbol{\tau} \in \mathbb{R}^m$ that satisfy the constraints (2.4). A pair $(\boldsymbol{u},\boldsymbol{\tau}) \in \mathcal{U} \times \mathcal{T}$ is called a *feasible control pair*. Let $\mathcal{F}$ denote the set of all such feasible control pairs.

We now define the following optimal control problem.

Problem P. *Find a feasible control pair* $(\boldsymbol{u}, \boldsymbol{\tau}) \in \mathcal{F}$ *that minimizes the cost function*

$$G_0(\boldsymbol{u}, \boldsymbol{\tau}) \triangleq \Phi_0\big(\boldsymbol{x}(\tau_1|\boldsymbol{u}), \ldots, \boldsymbol{x}(\tau_m|\boldsymbol{u})\big) + \int_0^T \mathcal{L}_0\big(\boldsymbol{x}(t|\boldsymbol{u}), \boldsymbol{u}(t)\big)dt,$$

where $\Phi_0 : \mathbb{R}^{mn} \to \mathbb{R}$ *and* $\mathcal{L}_0 : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$ *are given functions, over* $\mathcal{F}$*.*

Recall that the optimal control problem considered in [77,78] has fixed characteristic times. Problem P is far more general: its characteristic times are actually decision variables that, together with the control function, should be determined optimally. It is clear, however, that Problem P reduces to the optimal control problem in [77, 78] when

$$c_j = d_j, \qquad j = 1, \ldots, m.$$

Furthermore, there is no loss of generality in assuming that the functions $\boldsymbol{f}$ and $\mathcal{L}_i$, $i = 0, \ldots, q$, do not depend on time explicitly. This is because time may be replaced by the state variable v whose dynamics are

$$\dot{v}(t) = 1, \qquad t \in [0, T],$$

and

$$v(0) = 0.$$

Before finishing this section, we make one further assumption.

Assumption 2.3. The functions Φ_i and $\mathcal{L}_i$, $i = 0, \ldots, q$, are continuously differentiable.

2.3 Problem approximation

In general, Problem P is too complex to solve analytically. Thus, in this section, we will approximate Problem P by a dynamic optimization problem with a finite number of decision variables.

Let $p \geq 1$ be a fixed integer and define

$$N \triangleq (m + 1)p.$$

We construct a piecewise constant approximation of the control as follows. First, let the approximate control change its value at $N - 1$ locations in the time horizon. The times at which these changes occur are called *switching times*; they are denoted by t_k, $k = 1, \ldots, N - 1$. Let $\boldsymbol{\nu} \in \mathbb{R}^{N-1}$ denote the vector whose components are the switching times. That is,

$$\boldsymbol{\nu} = [t_1, \ldots, t_{N-1}]^T.$$

This vector is called the *switching-time vector*.

The switching times are indexed in non-decreasing order. Hence,

$$0 \triangleq t_0 \leq t_1 \leq \cdots \leq t_{N-1} \leq t_N \triangleq T. \tag{2.5}$$

Furthermore, we require that

$$t_{pj} = \tau_j, \qquad j = 1, \ldots, m. \tag{2.6}$$

This constraint ensures that every pth control switch occurs at a characteristic time. In other words, there are $p - 1$ control switches between consecutive characteristic times.

Because of (2.4) and (2.6), the switching times are also subject to the following additional constraints:

$$c_j \leq t_{pj} \leq d_j, \qquad j = 1, \ldots, m. \tag{2.7}$$

Now, define subintervals $\mathcal{I}_k \subset [0, T]$, $k = 1, \ldots, N$, as follows:

$$\mathcal{I}_k \triangleq \begin{cases} [t_{k-1}, t_k), & \text{if } k \in \{1, \ldots, N-1\}, \\ [t_{k-1}, t_k], & \text{if } k = N. \end{cases}$$

On each subinterval $\mathcal{I}_k$, $k = 1, \ldots, N$, the approximate control assumes a constant value of $\boldsymbol{\sigma}^k \in \mathbb{R}^r$. Therefore, we impose the constraint

$$\boldsymbol{\sigma}^k \in \mathcal{W}, \qquad k = 1, \ldots, N. \tag{2.8}$$

This constraint ensures that the approximate control is admissible.

Let

$$\boldsymbol{\sigma} \triangleq \left[(\boldsymbol{\sigma}^1)^T, \ldots, (\boldsymbol{\sigma}^N)^T\right]^T \in \mathbb{R}^{Nr}.$$

This vector is called the *control-value vector*.

We can now express the approximate control as follows:

$$\boldsymbol{u}^p(t) = \sum_{k=1}^{N} \boldsymbol{\sigma}^k \chi_{\mathcal{I}_k}(t), \qquad t \in [0, T], \tag{2.9}$$

where, for each $\mathcal{I} \subset \mathbb{R}$, the indicator function $\chi_{\mathcal{I}} : \mathbb{R} \to \mathbb{R}$ is defined by

$$\chi_{\mathcal{I}}(t) \triangleq \begin{cases} 1, & \text{if } t \in \mathcal{I}, \\ 0, & \text{otherwise.} \end{cases}$$

Substituting (2.9) into the dynamic system (2.1)-(2.2) gives

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}\big(\boldsymbol{x}(t), \boldsymbol{\sigma}^k\big), \qquad t \in \mathcal{I}_k, \qquad k = 1, \ldots, N, \tag{2.10}$$

and

$$\boldsymbol{x}(0) = \boldsymbol{x}^0. \tag{2.11}$$

Let $\boldsymbol{x}^p(\cdot|\boldsymbol{\nu}, \boldsymbol{\sigma})$ denote the solution of this system corresponding to the switching-time vector $\boldsymbol{\nu} \in \mathbb{R}^{N-1}$ and the control-value vector $\boldsymbol{\sigma} \in \mathbb{R}^{Nr}$. Then, by virtue of (2.6), the cost and constraint functions G_i, $i = 0, \ldots, q$, become

$$G_i^p(\boldsymbol{\nu}, \boldsymbol{\sigma}) \triangleq \Phi_i\big(\boldsymbol{x}^p(t_p|\boldsymbol{\nu}, \boldsymbol{\sigma}), \ldots, \boldsymbol{x}^p(t_{mp}|\boldsymbol{\nu}, \boldsymbol{\sigma})\big) + \int_0^T \mathcal{L}\big(\boldsymbol{x}^p(t|\boldsymbol{\nu}, \boldsymbol{\sigma}), \boldsymbol{u}^p(t)\big)dt, \quad i = 0, \ldots, q.$$

Thus, when equation (2.9) is used to approximate the control, Problem P becomes the following optimization problem.

Problem $\mathbf{P}_p$. *Find a pair* $(\boldsymbol{\nu}, \boldsymbol{\sigma}) \in \mathbb{R}^{N-1} \times \mathbb{R}^{Nr}$ *that minimizes the cost function* G_0^p *subject to the linear constraints* (2.5)-(2.8) *and the CTI constraints*

$$G_i^p(\boldsymbol{\nu}, \boldsymbol{\sigma}) \geq 0, \qquad i = 1, \ldots, q.$$

Since it has a finite number of decision variables, Problem P_p can be viewed as a nonlinear programming problem. However, some of Problem P_p's decision variables are control switching times, and as we mentioned in Chapter 1, optimization problems with variable switching times are very difficult to solve. Thus, we will use the time-scaling transformation mentioned in Chapter 1 to convert Problem P_p into a new optimization problem that has fixed switching times.

First, define the set

$$\Theta \triangleq \big\{\, \boldsymbol{\theta} \in \mathbb{R}^N : \ \theta_k \geq 0, \ k = 1, \ldots, N; \ \theta_1 + \cdots + \theta_N = NT \,\big\}.$$

That is, Θ consists of all vectors of length N whose components are non-negative and sum together to give NT. For each $\boldsymbol{\theta} \in \Theta$, we can define a transformation from $[0, T]$ to a new time horizon $[0, 1]$ through the differential equation

$$\dot{t}(s) = \tilde{v}^p(s), \qquad s \in [0, 1], \tag{2.12}$$

and the initial condition

$$t(0) = 0, \tag{2.13}$$

where

$$\tilde{v}^p(s) \triangleq \sum_{k=1}^{N} \theta_k \chi_{\mathcal{J}_k}(s), \qquad s \in [0,1],$$

$$\mathcal{J}_k \triangleq \begin{cases} \big[\alpha_{k-1}, \alpha_k\big), & \text{if } k \in \{1,\dots,N-1\}, \\ \big[\alpha_{k-1}, \alpha_k\big], & \text{if } k = N, \end{cases}$$

and

$$\alpha_k \triangleq \frac{k}{N}, \qquad k = 0,\dots,N.$$

The relationship between $t \in [0,T]$ and the new time variable $s \in [0,1]$ is obtained by integrating (2.12)-(2.13):

$$t(s) = \sum_{l=1}^{\lfloor Ns \rfloor} \frac{\theta_l}{N} + \frac{\theta_{\lfloor Ns \rfloor + 1}}{N}(Ns - \lfloor Ns \rfloor), \qquad s \in [0,1]. \tag{2.14}$$

In terms of the new time variable $s \in [0,1]$, the approximate control (2.9) switches value at the fixed times $s = \alpha_k$, $k = 1,\dots,N-1$. Furthermore, the kth switching time in the new time horizon $[0,1]$ corresponds to the kth switching time in $[0,T]$. Hence, by equation (2.14),

$$t_k = t(\alpha_k) = \sum_{l=1}^{k} \frac{\theta_l}{N}, \qquad k = 1,\dots,N-1. \tag{2.15}$$

This shows that the switching times in the original time horizon are controlled by $\boldsymbol{\theta} \in \Theta$.

Equation (2.14) also shows that

$$t_0 = t(0) = 0$$

and

$$t_N = t(1) = \sum_{l=1}^{N} \frac{\theta_l}{N} = T,$$

where the last equality follows from the definition of Θ. Thus, it is clear that equations (2.12)-(2.13) define a monotonic transformation from $[0,T]$ to $[0,1]$ in which the control switching times t_k, $k = 1,\dots,N-1$, are mapped to the uniformly distributed locations $s = \alpha_k$, $k = 1,\dots,N-1$.

The approximate control in the new time horizon is

$$\tilde{\boldsymbol{u}}^p(s) \triangleq \boldsymbol{u}^p(t(s)) = \sum_{k=1}^{N} \boldsymbol{\sigma}^k \chi_{\mathcal{J}_k}(s), \qquad s \in [0,1].$$

Since $\boldsymbol{u}^p$ maps $[0,T]$ into $\mathcal{W}$, $\tilde{\boldsymbol{u}}^p$ must map $[0,1]$ into $\mathcal{W}$. Therefore, we retain the

constraints (2.8):

$$\boldsymbol{\sigma}^k \in \mathcal{W}, \qquad k = 1, \ldots, N. \tag{2.16}$$

Furthermore, recall that every pth control switch occurs at a characteristic time. Hence, for each $j = 1, \ldots, m$, the point $s = \alpha_{pj} = pj/N$ in the new time horizon corresponds to the characteristic time $t = \tau_j$ in the original time horizon. Using equation (2.15), we obtain

$$\tau_j = t_{pj} = \sum_{l=1}^{pj} \frac{\theta_l}{N}, \qquad j = 1, \ldots, m.$$

Hence, in view of (2.7), we must have the constraints

$$c_j \leq \sum_{l=1}^{pj} \frac{\theta_l}{N} \leq d_j, \qquad j = 1, \ldots, m. \tag{2.17}$$

Now, applying the time-scaling transformation defined by (2.12)-(2.13) to the original dynamic system (2.1)-(2.2) gives

$$\dot{\tilde{\boldsymbol{x}}}(s) = \tilde{v}^p(s)\boldsymbol{f}\big(\tilde{\boldsymbol{x}}(s), \tilde{\boldsymbol{u}}^p(s)\big), \qquad s \in [0, 1], \tag{2.18}$$

and

$$\tilde{\boldsymbol{x}}(0) = \boldsymbol{x}^0, \tag{2.19}$$

where

$$\tilde{\boldsymbol{x}}(s) \triangleq \boldsymbol{x}(t(s)).$$

Let $\tilde{\boldsymbol{x}}^p(\cdot|\boldsymbol{\theta}, \boldsymbol{\sigma})$ denote the solution of (2.18)-(2.19) corresponding to $(\boldsymbol{\theta}, \boldsymbol{\sigma}) \in \Theta \times \mathbb{R}^{Nr}$.

Finally, applying the transformation defined by (2.12)-(2.13) to the cost and constraint functions G_i^p, $i = 0, \ldots, q$, gives

$$\tilde{G}_i^p(\boldsymbol{\theta}, \boldsymbol{\sigma}) \triangleq \Phi_i\big(\tilde{\boldsymbol{x}}^p(\alpha_p|\boldsymbol{\theta}, \boldsymbol{\sigma}), \ldots, \tilde{\boldsymbol{x}}^p(\alpha_{pm}|\boldsymbol{\theta}, \boldsymbol{\sigma})\big) + \int_0^1 \tilde{v}^p(s)\mathcal{L}_i\big(\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta}, \boldsymbol{\sigma}), \tilde{\boldsymbol{u}}^p(s)\big)ds, \quad i = 0, \ldots, q. \tag{2.20}$$

We now state the following optimization problem, which is equivalent to Problem P_p.

Problem $\tilde{\mathrm{P}}_p$. *Find a pair $(\boldsymbol{\theta}, \boldsymbol{\sigma}) \in \Theta \times \mathbb{R}^{Nr}$ that minimizes the cost function $\tilde{G}_0^p$ subject to the linear constraints* (2.16)-(2.17) *and the CTI constraints*

$$\tilde{G}_i^p(\boldsymbol{\theta}, \boldsymbol{\sigma}) \geq 0, \qquad i = 1, \ldots, q.$$

Problem $\tilde{\mathrm{P}}_p$ is derived by replacing the switching times t_k, $k = 1, \ldots, N-1$, in Problem P_p with the new decision variables θ_k, $k = 1, \ldots, N$. Its solution $(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*) \in \Theta \times \mathbb{R}^{Nr}$

immediately furnishes the following suboptimal control for Problem P:

$$\boldsymbol{u}^{p,*}(t) = \sum_{k=1}^{N} \boldsymbol{\sigma}^{k,*} \chi_{\mathcal{I}_k^*}(t), \qquad t \in [0,T],$$

where the subintervals $\mathcal{I}_k^* \subset [0,T]$, $k = 1, \ldots, N$, have endpoints

$$t_k^* = \sum_{l=1}^{k} \frac{\theta_l^*}{N}, \qquad k = 0, \ldots, N.$$

Furthermore, the optimal characteristic times are

$$\tau_j^{p,*} = t_{pj}^* = \sum_{l=1}^{pj} \frac{\theta_l^*}{N}, \qquad j = 1, \ldots, m.$$

It can be shown (see [77]) that the cost of $(\boldsymbol{u}^{p,*}, \boldsymbol{\tau}^{p,*})$ converges to the optimal cost of Problem P as $p \to \infty$. Indeed, if $(\boldsymbol{u}^*, \boldsymbol{\tau}^*) \in \mathcal{F}$ is an optimal solution of Problem P, then

$$\lim_{p \to \infty} G_0(\boldsymbol{u}^{p,*}, \boldsymbol{\tau}^{p,*}) = G_0(\boldsymbol{u}^*, \boldsymbol{\tau}^*).$$

Furthermore, if the sequence $\{\boldsymbol{u}^{p,*}\}_{p=1}^{\infty}$ converges almost everywhere on $[0,T]$ to an admissible control $\hat{\boldsymbol{u}} \in \mathcal{U}$, and if the sequence $\{\boldsymbol{\tau}^{p,*}\}_{p=1}^{\infty}$ converges to a switching-time vector $\hat{\boldsymbol{\tau}} \in \mathcal{T}$, then the pair $(\hat{\boldsymbol{u}}, \hat{\boldsymbol{\tau}})$ is optimal for Problem P.

Problem $\tilde{\mathrm{P}}_p$ has fixed switching times and is therefore easier to solve than Problem P_p. In the next section, we will show how to compute the gradient of its cost and constraint functions with respect to the decision variables θ_k, $k = 1, \ldots, N$, and σ_ς^k, $k = 1, \ldots, N$, $\varsigma = 1, \ldots, r$.

2.4 Gradient computation

For each $k = 1, \ldots, N$, consider the following auxiliary dynamic system:

$$\dot{\boldsymbol{\psi}}^k(s) = \hat{\rho}_{k,l}\theta_l \frac{\partial \boldsymbol{f}\big(\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big)}{\partial \boldsymbol{x}} \boldsymbol{\psi}^k(s) + \rho_{k,l} \boldsymbol{f}\big(\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big), \quad s \in \mathcal{J}_l, \quad l = 1, \ldots, N, \tag{2.21}$$

and

$$\boldsymbol{\psi}^k(0) = \mathbf{0}, \tag{2.22}$$

where $(\boldsymbol{\theta},\boldsymbol{\sigma}) \in \Theta \times \mathbb{R}^{Nr}$ and

$$\rho_{k,l} \triangleq \begin{cases} 1, & \text{if } k = l, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\hat{\rho}_{k,l} \triangleq \begin{cases} 1, & \text{if } k \leq l, \\ 0, & \text{otherwise.} \end{cases}$$

Let $\boldsymbol{\psi}^k(\cdot|\boldsymbol{\theta},\boldsymbol{\sigma})$ denote the solution of (2.21)-(2.22). The following theorem shows that the partial derivatives of $\tilde{G}_i^p$, $i = 0,\dots,q$, with respect to θ_k, $k = 1,\dots,N$, can be expressed in terms of $\boldsymbol{\psi}^k(\cdot|\boldsymbol{\theta},\boldsymbol{\sigma})$.

Theorem 2.1. *Let $(\boldsymbol{\theta},\boldsymbol{\sigma}) \in \Theta \times \mathbb{R}^{Nr}$. Then for each $i = 0,\dots,q$,*

$$\begin{aligned}\frac{\partial \tilde{G}_i^p(\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \theta_k} = \sum_{j=1}^{m} \frac{\partial \Phi_i\big(\tilde{\boldsymbol{x}}^p(\alpha_p|\boldsymbol{\theta},\boldsymbol{\sigma}),\dots,\tilde{\boldsymbol{x}}^p(\alpha_{pm}|\boldsymbol{\theta},\boldsymbol{\sigma})\big)}{\partial \boldsymbol{x}(\tau_j)} \boldsymbol{\psi}^k(\alpha_{pj}|\boldsymbol{\theta},\boldsymbol{\sigma}) \\ + \sum_{l=1}^{N} \int_{\mathcal{J}_l} \theta_l \frac{\partial \mathcal{L}_i\big(\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big)}{\partial \boldsymbol{x}} \boldsymbol{\psi}^k(s|\boldsymbol{\theta},\boldsymbol{\sigma}) ds \\ + \int_{\mathcal{J}_k} \mathcal{L}_i\big(\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^k\big) ds, \qquad k = 1,\dots,N.\end{aligned}$$

Proof. Let $(\boldsymbol{\theta},\boldsymbol{\sigma}) \in \Theta \times \mathbb{R}^{Nr}$, $i \in \{0,\dots,q\}$, and $k \in \{1,\dots,N\}$ be arbitrary but fixed. It is clear from equation (2.20) that $\tilde{G}_i^p$ can be expressed as

$$\tilde{G}_i^p(\boldsymbol{\theta},\boldsymbol{\sigma}) = \Phi_i\big(\tilde{\boldsymbol{x}}^p(\alpha_p|\boldsymbol{\theta},\boldsymbol{\sigma}),\dots,\tilde{\boldsymbol{x}}^p(\alpha_{pm}|\boldsymbol{\theta},\boldsymbol{\sigma})\big) + \sum_{l=1}^{N} \int_{\mathcal{J}_l} \theta_l \mathcal{L}_i\big(\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big) ds. \quad (2.23)$$

Differentiating (2.23) with respect to θ_k yields

$$\begin{aligned}\frac{\partial \tilde{G}_i^p(\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \theta_k} = \sum_{j=1}^{m} \frac{\partial \Phi_i\big(\tilde{\boldsymbol{x}}^p(\alpha_p|\boldsymbol{\theta},\boldsymbol{\sigma}),\dots,\tilde{\boldsymbol{x}}^p(\alpha_{pm}|\boldsymbol{\theta},\boldsymbol{\sigma})\big)}{\partial \boldsymbol{x}(\tau_j)} \frac{\partial \tilde{\boldsymbol{x}}^p(\alpha_{pj}|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \theta_k} \\ + \sum_{l=1}^{N} \int_{\mathcal{J}_l} \theta_l \frac{\partial \mathcal{L}_i\big(\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big)}{\partial \boldsymbol{x}} \frac{\partial \tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \theta_k} ds \\ + \int_{\mathcal{J}_k} \mathcal{L}_i\big(\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^k\big) ds. \quad (2.24)\end{aligned}$$

Now, for each $l = 1,\dots,N$, it follows from (2.18)-(2.19) that

$$\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma}) = \tilde{\boldsymbol{x}}^p(\alpha_{l-1}|\boldsymbol{\theta},\boldsymbol{\sigma}) + \int_{\alpha_{l-1}}^{s} \theta_l \boldsymbol{f}\big(\tilde{\boldsymbol{x}}^p(\eta|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big) d\eta, \quad s \in \mathcal{J}_l. \quad (2.25)$$

If $l > k$, then differentiating (2.25) with respect to θ_k yields

$$\frac{\partial \tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \theta_k} = \frac{\partial \tilde{\boldsymbol{x}}^p(\alpha_{l-1}|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \theta_k} + \int_{\alpha_{l-1}}^{s} \theta_l \frac{\partial \boldsymbol{f}\big(\tilde{\boldsymbol{x}}^p(\eta|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big)}{\partial \boldsymbol{x}} \frac{\partial \tilde{\boldsymbol{x}}^p(\eta|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \theta_k} d\eta, \quad s \in \mathcal{J}_l. \tag{2.26}$$

On the other hand, if $l = k$, then differentiating (2.25) with respect to θ_k yields

$$\frac{\partial \tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \theta_k} = \frac{\partial \tilde{\boldsymbol{x}}^p(\alpha_{l-1}|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \theta_k} + \int_{\alpha_{l-1}}^{s} \theta_l \frac{\partial \boldsymbol{f}\big(\tilde{\boldsymbol{x}}^p(\eta|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big)}{\partial \boldsymbol{x}} \frac{\partial \tilde{\boldsymbol{x}}^p(\eta|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \theta_k} d\eta + \int_{\alpha_{l-1}}^{s} \boldsymbol{f}\big(\tilde{\boldsymbol{x}}^p(\eta|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big) d\eta, \quad s \in \mathcal{J}_l. \tag{2.27}$$

Since θ_k is the value of $\tilde{v}^p$ on the subinterval $\mathcal{J}_k$, it does not affect the state at times before $\mathcal{J}_k$. Hence, if $l < k$, then

$$\frac{\partial \tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \theta_k} = \boldsymbol{0}, \qquad s \in \mathcal{J}_l. \tag{2.28}$$

Using the Kronecker delta and the cumulative Kronecker delta, we can combine (2.26)-(2.28) into one equation as follows:

$$\frac{\partial \tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \theta_k} = \hat{\rho}_{k,l} \frac{\partial \tilde{\boldsymbol{x}}^p(\alpha_{l-1}|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \theta_k} + \int_{\alpha_{l-1}}^{s} \hat{\rho}_{k,l}\theta_l \frac{\partial \boldsymbol{f}\big(\tilde{\boldsymbol{x}}^p(\eta|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big)}{\partial \boldsymbol{x}} \frac{\partial \tilde{\boldsymbol{x}}^p(\eta|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \theta_k} d\eta + \int_{\alpha_{l-1}}^{s} \rho_{k,l} \boldsymbol{f}\big(\tilde{\boldsymbol{x}}^p(\eta|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big) d\eta, \qquad s \in \mathcal{J}_l, \qquad l = 1,\dots,N.$$

Differentiating this equation with respect to s gives

$$\frac{d}{ds}\left\{\frac{\partial \tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \theta_k}\right\} = \hat{\rho}_{k,l}\theta_l \frac{\partial \boldsymbol{f}\big(\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big)}{\partial \boldsymbol{x}} \frac{\partial \tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \theta_k} + \rho_{k,l} \boldsymbol{f}\big(\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big), \qquad s \in \mathcal{J}_l, \qquad l = 1,\dots,N. \tag{2.29}$$

Furthermore,

$$\frac{\partial \tilde{\boldsymbol{x}}^p(0|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \theta_k} = \frac{\partial}{\partial \theta_k}\big\{\boldsymbol{x}^0\big\} = \boldsymbol{0}. \tag{2.30}$$

Equations (2.29)-(2.30) show that $\partial \tilde{\boldsymbol{x}}^p(\cdot|\boldsymbol{\theta},\boldsymbol{\sigma})/\partial\theta_k$ is a solution of (2.21)-(2.22). By the theory of differential equations (see [1, 2]), such a solution is unique. Therefore,

$$\frac{\partial \tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \theta_k} = \boldsymbol{\psi}^k(s|\boldsymbol{\theta},\boldsymbol{\sigma}), \qquad s \in [0,1].$$

Substituting this equation into (2.24) completes the proof. □

We now present formulae for the partial derivatives of $\tilde{G}_i^p$, $i = 0, \ldots, q$, with respect to σ_ς^k, $k = 1, \ldots, N$, $\varsigma = 1, \ldots, r$.

For each $k = 1, \ldots, N$, and $\varsigma = 1, \ldots, r$, consider the following auxiliary dynamic system:

$$\dot{\boldsymbol{\phi}}^{k,\varsigma}(s) = \hat{\rho}_{k,l}\theta_l \frac{\partial \boldsymbol{f}\big(\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big)}{\partial \boldsymbol{x}} \boldsymbol{\phi}^{k,\varsigma}(s) + \rho_{k,l}\theta_l \frac{\partial \boldsymbol{f}\big(\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big)}{\partial u_\varsigma}, \quad s \in \mathcal{J}_l, \quad l = 1, \ldots, N, \tag{2.31}$$

and

$$\boldsymbol{\phi}^{k,\varsigma}(0) = \mathbf{0}, \tag{2.32}$$

where $(\boldsymbol{\theta}, \boldsymbol{\sigma}) \in \Theta \times \mathbb{R}^{Nr}$. Let $\boldsymbol{\phi}^{k,\varsigma}(\cdot|\boldsymbol{\theta},\boldsymbol{\sigma})$ denote the solution of (2.31)-(2.32).

We have the following important result.

Theorem 2.2. *Let $(\boldsymbol{\theta}, \boldsymbol{\sigma}) \in \Theta \times \mathbb{R}^{Nr}$. Then for each $i = 0, \ldots, q$,*

$$\begin{aligned}\frac{\partial \tilde{G}_i^p(\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \sigma_\varsigma^k} &= \sum_{j=1}^{m} \frac{\partial \Phi_i\big(\tilde{\boldsymbol{x}}^p(\alpha_p|\boldsymbol{\theta},\boldsymbol{\sigma}), \ldots, \tilde{\boldsymbol{x}}^p(\alpha_{pm}|\boldsymbol{\theta},\boldsymbol{\sigma})\big)}{\partial \boldsymbol{x}(\tau_j)} \boldsymbol{\phi}^{k,\varsigma}(\alpha_{pj}|\boldsymbol{\theta},\boldsymbol{\sigma}) \\ &\quad + \sum_{l=1}^{N} \int_{\mathcal{J}_l} \theta_l \frac{\partial \mathcal{L}_i\big(\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big)}{\partial \boldsymbol{x}} \boldsymbol{\phi}^{k,\varsigma}(s|\boldsymbol{\theta},\boldsymbol{\sigma}) ds \\ &\quad + \int_{\mathcal{J}_k} \theta_k \frac{\partial \mathcal{L}_i\big(\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^k\big)}{\partial u_\varsigma} ds, \quad k = 1, \ldots, N, \quad \varsigma = 1, \ldots, r.\end{aligned}$$

Proof. The proof is similar to the proof of Theorem 2.1. First, let $(\boldsymbol{\theta}, \boldsymbol{\sigma}) \in \Theta \times \mathbb{R}^{Nr}$, $i \in \{0, \ldots, q\}$, $k \in \{1, \ldots, N\}$, and $\varsigma \in \{1, \ldots, r\}$ be arbitrary but fixed.

Recall equation (2.23) in the proof of Theorem 2.1:

$$\tilde{G}_i^p(\boldsymbol{\theta},\boldsymbol{\sigma}) = \Phi_i\big(\tilde{\boldsymbol{x}}^p(\alpha_p|\boldsymbol{\theta},\boldsymbol{\sigma}), \ldots, \tilde{\boldsymbol{x}}^p(\alpha_{pm}|\boldsymbol{\theta},\boldsymbol{\sigma})\big) + \sum_{l=1}^{N} \int_{\mathcal{J}_l} \theta_l \mathcal{L}_i\big(\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big) ds.$$

Differentiating this equation with respect to σ_ς^k yields

$$\begin{aligned}\frac{\partial \tilde{G}_i^p(\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \sigma_\varsigma^k} &= \sum_{j=1}^{m} \frac{\partial \Phi_i\big(\tilde{\boldsymbol{x}}^p(\alpha_p|\boldsymbol{\theta},\boldsymbol{\sigma}), \ldots, \tilde{\boldsymbol{x}}^p(\alpha_{pm}|\boldsymbol{\theta},\boldsymbol{\sigma})\big)}{\partial \boldsymbol{x}(\tau_j)} \frac{\partial \tilde{\boldsymbol{x}}^p(\alpha_{pj}|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \sigma_\varsigma^k} \\ &\quad + \sum_{l=1}^{N} \int_{\mathcal{J}_l} \theta_l \frac{\partial \mathcal{L}_i\big(\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big)}{\partial \boldsymbol{x}} \frac{\partial \tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \sigma_\varsigma^k} ds \\ &\quad + \int_{\mathcal{J}_k} \theta_k \frac{\partial \mathcal{L}_i\big(\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^k\big)}{\partial u_\varsigma} ds.\end{aligned} \tag{2.33}$$

Also recall from the proof of Theorem 2.1 that for each $l = 1, \ldots, N$,

$$\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma}) = \tilde{\boldsymbol{x}}^p(\alpha_{l-1}|\boldsymbol{\theta},\boldsymbol{\sigma}) + \int_{\alpha_{l-1}}^{s} \theta_l \boldsymbol{f}\big(\tilde{\boldsymbol{x}}^p(\eta|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big)d\eta, \quad s \in \mathcal{J}_l. \tag{2.34}$$

If $l > k$, then differentiating (2.34) with respect to σ_ς^k yields

$$\frac{\partial \tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \sigma_\varsigma^k} = \frac{\partial \tilde{\boldsymbol{x}}^p(\alpha_{l-1}|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \sigma_\varsigma^k} + \int_{\alpha_{l-1}}^{s} \theta_l \frac{\partial \boldsymbol{f}\big(\tilde{\boldsymbol{x}}^p(\eta|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big)}{\partial \boldsymbol{x}} \frac{\partial \tilde{\boldsymbol{x}}^p(\eta|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \sigma_\varsigma^k} d\eta, \quad s \in \mathcal{J}_l. \tag{2.35}$$

On the other hand, if $l = k$, then differentiating (2.34) with respect to σ_ς^k yields

$$\frac{\partial \tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \sigma_\varsigma^k} = \frac{\partial \tilde{\boldsymbol{x}}^p(\alpha_{l-1}|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \sigma_\varsigma^k} + \int_{\alpha_{l-1}}^{s} \theta_l \frac{\partial \boldsymbol{f}\big(\tilde{\boldsymbol{x}}^p(\eta|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big)}{\partial \boldsymbol{x}} \frac{\partial \tilde{\boldsymbol{x}}^p(\eta|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \sigma_\varsigma^k} d\eta + \int_{\alpha_{l-1}}^{s} \theta_l \frac{\partial \boldsymbol{f}\big(\tilde{\boldsymbol{x}}^p(\eta|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big)}{\partial u_\varsigma} d\eta, \quad s \in \mathcal{J}_l. \tag{2.36}$$

It is obvious that if $l < k$, then

$$\frac{\partial \tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \sigma_\varsigma^k} = \boldsymbol{0}, \qquad s \in \mathcal{J}_l. \tag{2.37}$$

Combining equations (2.35)-(2.37) gives

$$\frac{\partial \tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \sigma_\varsigma^k} = \hat{\rho}_{k,l} \frac{\partial \tilde{\boldsymbol{x}}^p(\alpha_{l-1}|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \sigma_\varsigma^k} + \int_{\alpha_{l-1}}^{s} \hat{\rho}_{k,l}\theta_l \frac{\partial \boldsymbol{f}\big(\tilde{\boldsymbol{x}}^p(\eta|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big)}{\partial \boldsymbol{x}} \frac{\partial \tilde{\boldsymbol{x}}^p(\eta|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \sigma_\varsigma^k} d\eta + \int_{\alpha_{l-1}}^{s} \rho_{k,l}\theta_l \frac{\partial \boldsymbol{f}\big(\tilde{\boldsymbol{x}}^p(\eta|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big)}{\partial u_\varsigma} d\eta, \qquad s \in \mathcal{J}_l, \qquad l = 1, \ldots, N.$$

By differentiating this equation with respect to s, we obtain

$$\frac{d}{ds}\left\{\frac{\partial \tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \sigma_\varsigma^k}\right\} = \hat{\rho}_{k,l}\theta_l \frac{\partial \boldsymbol{f}\big(\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big)}{\partial \boldsymbol{x}} \frac{\partial \tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \sigma_\varsigma^k} + \rho_{k,l}\theta_l \frac{\partial \boldsymbol{f}\big(\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}^l\big)}{\partial u_\varsigma}, \qquad s \in \mathcal{J}_l, \qquad l = 1, \ldots, N. \tag{2.38}$$

Moreover,

$$\frac{\partial \tilde{\boldsymbol{x}}^p(0|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \sigma_\varsigma^k} = \frac{\partial}{\partial \sigma_\varsigma^k}\big\{\boldsymbol{x}^0\big\} = \boldsymbol{0}. \tag{2.39}$$

Equations (2.38)-(2.39) show that $\partial \tilde{\boldsymbol{x}}^p(\cdot|\boldsymbol{\theta},\boldsymbol{\sigma})/\partial \sigma_\varsigma^k$ is the unique solution of (2.31)-(2.32). Hence,

$$\frac{\partial \tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \sigma_\varsigma^k} = \boldsymbol{\phi}^{k,\varsigma}(s|\boldsymbol{\theta},\boldsymbol{\sigma}), \qquad s \in [0,1].$$

Substituting this equation into (2.33) completes the proof. □

The formulae in Theorems 2.1 and 2.2 express the partial derivatives of $\tilde{G}_i^p$, $i = 0, \ldots, q$, in terms of the solution of the state system (2.18)-(2.19) and the solutions of the auxiliary systems (2.21)-(2.22) and (2.31)-(2.32). The auxiliary systems cannot be solved independently, because their right-hand sides depend on the state. Nevertheless, we can combine the state system and the auxiliary systems to form an expanded initial value problem, which can be solved using any numerical integration technique. This suggests the following algorithm for computing the value and gradient of $\tilde{G}_i^p$, $i = 0, \ldots, q$, at a given $(\boldsymbol{\theta}, \boldsymbol{\sigma}) \in \Theta \times \mathbb{R}^{Nr}$.

Algorithm 2.1. Input $(\boldsymbol{\theta}, \boldsymbol{\sigma}) \in \Theta \times \mathbb{R}^{Nr}$.

(i) Obtain $\tilde{\boldsymbol{x}}^p(\cdot|\boldsymbol{\theta}, \boldsymbol{\sigma})$, $\boldsymbol{\psi}^k(\cdot|\boldsymbol{\theta}, \boldsymbol{\sigma})$, and $\boldsymbol{\phi}^{k,\varsigma}(\cdot|\boldsymbol{\theta}, \boldsymbol{\sigma})$ by solving the initial value problem consisting of (2.18)-(2.19), (2.21)-(2.22), and (2.31)-(2.32).

(ii) Use $\tilde{\boldsymbol{x}}^p(\cdot|\boldsymbol{\theta}, \boldsymbol{\sigma})$ to compute $\tilde{G}_i^p(\boldsymbol{\theta}, \boldsymbol{\sigma})$, $i = 0, \ldots, q$.

(iii) Use $\tilde{\boldsymbol{x}}^p(\cdot|\boldsymbol{\theta}, \boldsymbol{\sigma})$, $\boldsymbol{\psi}^k(\cdot|\boldsymbol{\theta}, \boldsymbol{\sigma})$, and $\boldsymbol{\phi}^{k,\varsigma}(\cdot|\boldsymbol{\theta}, \boldsymbol{\sigma})$ to compute the derivatives $\partial \tilde{G}_i^p(\boldsymbol{\theta}, \boldsymbol{\sigma})/\partial \theta_k$ and $\partial \tilde{G}_i^p(\boldsymbol{\theta}, \boldsymbol{\sigma})/\partial \sigma_\varsigma^k$, $i = 0, \ldots, q$, according to the formulae in Theorems 2.1 and 2.2.

By incorporating Algorithm 2.1 into a gradient descent algorithm, Problem $\tilde{\mathrm{P}}_p$ can be solved as a nonlinear programming problem. The solution of Problem $\tilde{\mathrm{P}}_p$ then furnishes a suboptimal control for Problem P (see the discussion at the end of Section 2.3). In the next section, we apply this approach to two examples.

2.5 Numerical examples

In this section, we solve two example problems. The first example is an optimal control problem whose characteristic times are decision variables. The second example is the optimal cancer chemotherapy problem formulated in [78] (which has fixed characteristic times).

2.5.1 Optimal observation times

Consider the following dynamic system:

$$\dot{x}_1(t) = x_2(t), \qquad t \in [0, 3], \tag{2.40a}$$

$$\dot{x}_2(t) = -u_1(t)x_2(t) - x_1(t) + u_2(t), \qquad t \in [0, 3], \tag{2.40b}$$

and

$$x_1(0) = 4, \tag{2.41a}$$

$$x_2(0) = 1, \tag{2.41b}$$

where x_1 and x_2 are state variables and u_1 and u_2 are control variables. Define a *target trajectory* as follows:

$$w(t) \triangleq 3 + \sin(2t) + \exp(t/5), \qquad t \in [0, 3].$$

We suppose that the dynamic system (2.40)-(2.41) needs to be observed five times during the interval $[0, 3]$. These observations should take place when x_1 is near the target trajectory w.

Let τ_j, $j = 1, \ldots, 5$, denote the time at which the jth observation takes place. We impose the following constraints:

$$\frac{5j-2}{10} \leq \tau_j \leq \frac{5j+2}{10}, \qquad j = 1, \ldots, 5. \tag{2.42}$$

We also impose the following constraints on the control functions:

$$-5 \leq u_1(t) \leq 5, \qquad t \in [0, 3], \tag{2.43}$$

and

$$-5 \leq u_2(t) \leq 5, \qquad t \in [0, 3]. \tag{2.44}$$

Our optimal control problem is to choose control functions u_1 and u_2 and observation times τ_j, $j = 1, \ldots, 5$, to minimize the cost function

$$G_0 = \sum_{j=1}^{5} \big(x_1(\tau_j) - w(\tau_j)\big)^2 \tag{2.45}$$

subject to the dynamic system (2.40)-(2.41) and the constraints (2.42)-(2.44). Notice that τ_j, $j = 1, \ldots, 5$, are characteristic times in the cost function (2.45).

We discretized this optimal control problem using the approximation scheme described in Section 2.3 (with $p = 2$). Problem $\tilde{\mathrm{P}}_p$ was solved using Algorithm 2.1 in conjunction with the nonlinear programming software NLPQLP (see [93]). The differential equations were solved using the LSODA solver (see [41]). Note that LSODA does not use a fixed step length; it instead varies the step length to curb local truncation error.

The results obtained are as follows. The optimal value of the cost function is

$$G_0^* = 1.2071184 \times 10^{-6}.$$

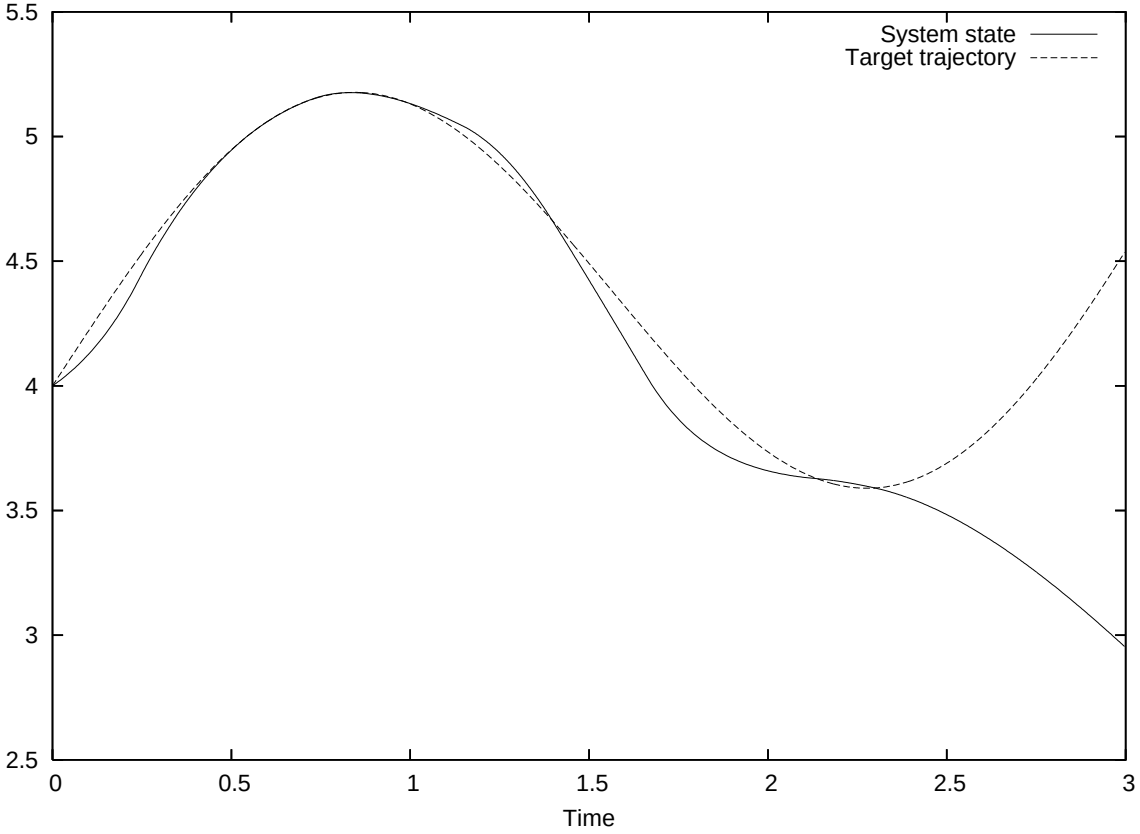

Figure 2.1: The target trajectory and optimal system state in Example 2.5.1.

The optimal observation times are

$$\tau_1^* = 0.5890,$$
$$\tau_2^* = 0.8186,$$
$$\tau_3^* = 1.3971,$$
$$\tau_4^* = 2.1349,$$
$$\tau_5^* = 2.3000.$$

The optimal state is shown in Figure 2.1 along with the target trajectory. The optimal controls are shown in Figures 2.2 and 2.3.

2.5.2 Optimal chemotherapy administration

We consider the optimal cancer chemotherapy problem formulated in [78]. This optimal control problem has the following dynamics:

$$\dot{y}(t) = -\lambda y(t) + \kappa(v(t) - v_{\text{th}})H(v(t) - v_{\text{th}}), \qquad t \in [0, T], \tag{2.46a}$$
$$\dot{v}(t) = u(t) - \gamma v(t), \qquad t \in [0, T], \tag{2.46b}$$

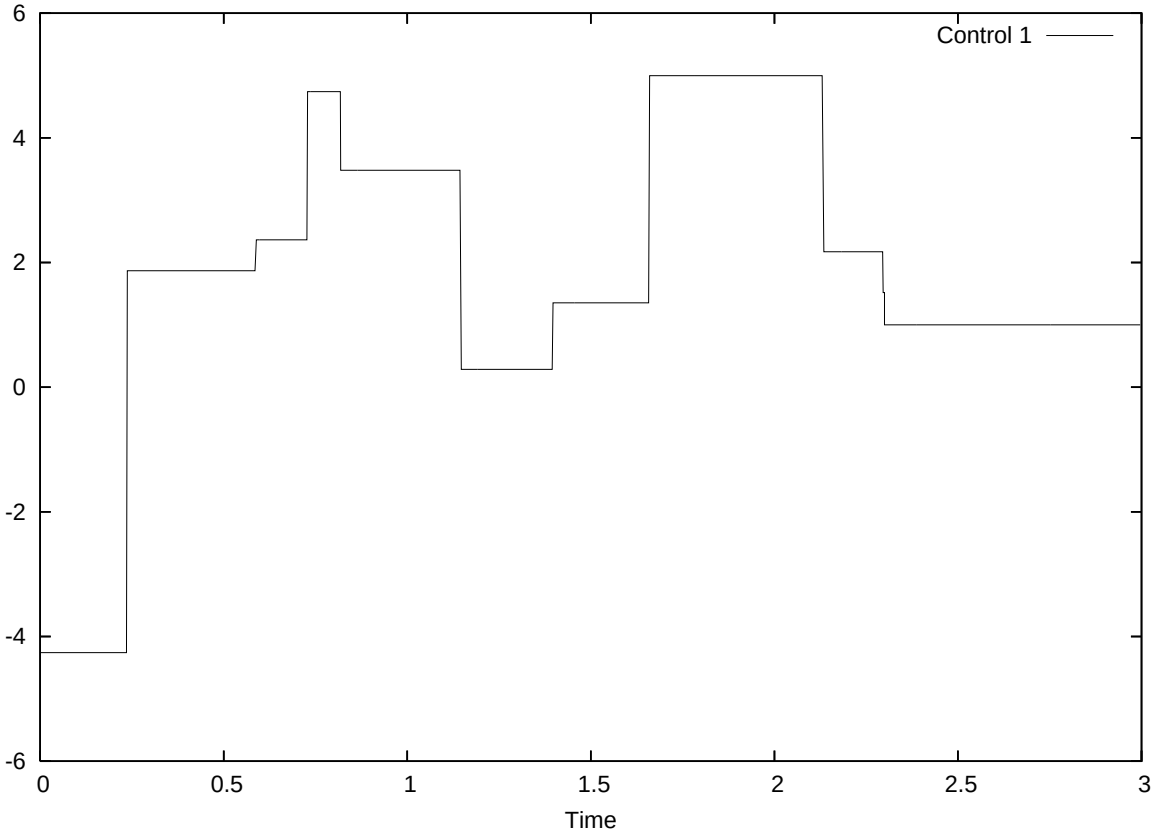

Figure 2.2: The optimal control u_1 in Example 2.5.1.

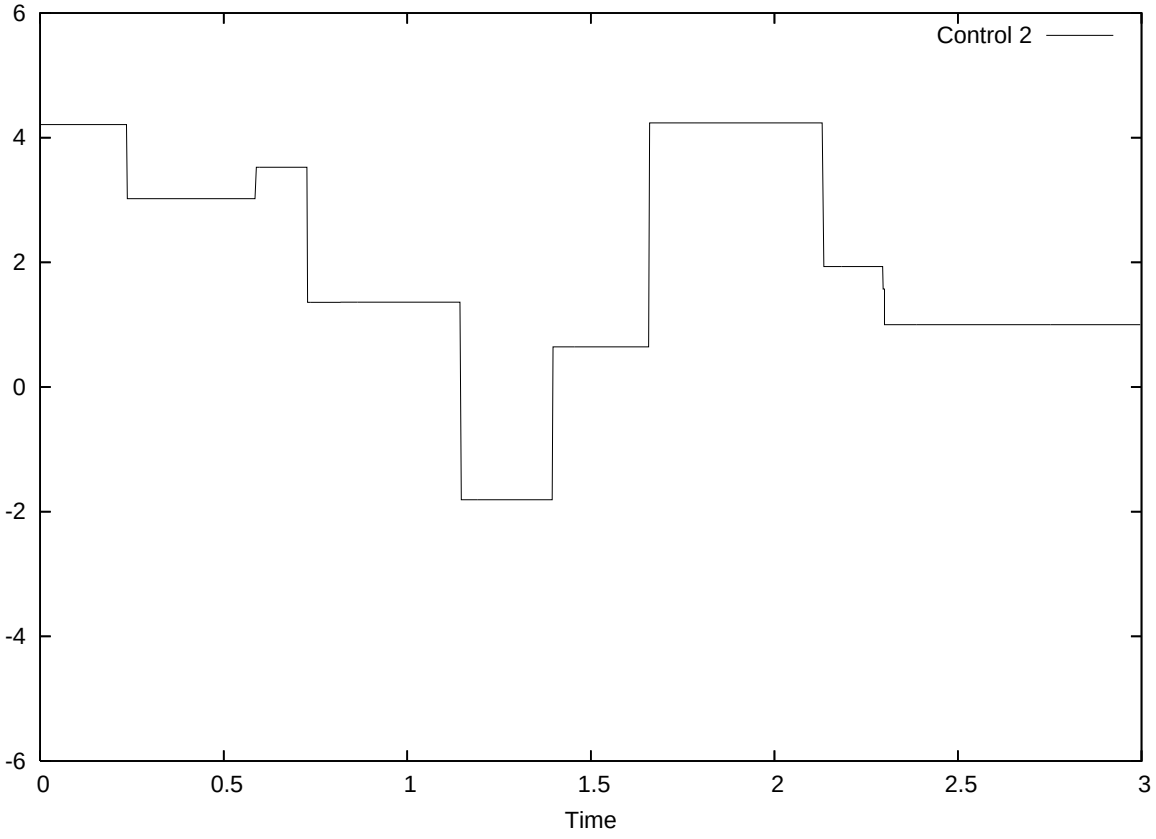

Figure 2.3: The optimal control u_2 in Example 2.5.1.

and

$$y(0) = \ln(\omega/C_0), \tag{2.47a}$$

$$v(0) = 0, \tag{2.47b}$$

where $T > 0$ is the time at which the treatment ends; λ, κ, ω, C_0, γ, and v_{th} are model parameters; $y(t) = \ln(\omega/C(t))$ and $C(t)$ is the number of tumor cells alive at time t; $v(t)$ is the concentration of the chemotherapy drug at the cancer site at time t; $u(t)$ is the rate at which the chemotherapy drug is being delivered at time t; and $H : \mathbb{R} \to \mathbb{R}$ is the Heaviside step function defined by

$$H(\eta) \triangleq \begin{cases} 1, & \text{if } \eta \geq 0, \\ 0, & \text{if } \eta < 0. \end{cases}$$

Since chemotherapy is highly toxic, there are restrictions on the amount that can be administered to the patient. These restrictions give rise to the following constraints:

$$0 \leq v(t) \leq v_{\max}, \qquad t \in [0, T], \tag{2.48}$$

and

$$\int_0^T v(t)dt \leq v_{\text{acc}}, \tag{2.49}$$

where $v_{\max} > 0$ and $v_{\text{acc}} > 0$ are given real numbers.

We choose fixed characteristic times τ_1, τ_2, and τ_3 such that

$$0 \triangleq \tau_0 < \tau_1 < \tau_2 < \tau_3 < \tau_4 \triangleq T.$$

Furthermore, we choose $p - 1$ control switching times between each characteristic time. These switching times are denoted by t_k, $k = 1, \ldots, 4p$, where

$$t_{pj} = \tau_j, \qquad j = 0, \ldots, 4,$$

and

$$0 \triangleq t_0 \leq t_1 \leq \cdots \leq t_{4p-1} \leq t_{4p} \triangleq T.$$

For each $k = 1, \ldots, 4p$, we approximate the drug delivery rate on the interval $[t_{k-1}, t_k)$ by the constant σ_k. It is shown in [78] that the constraints (2.48) then reduce to

$$v(t_k) \leq v_{\max}, \qquad k = 1, \ldots, 4p. \tag{2.50}$$

The tumor size is required to decrease sufficiently between consecutive characteristic times.

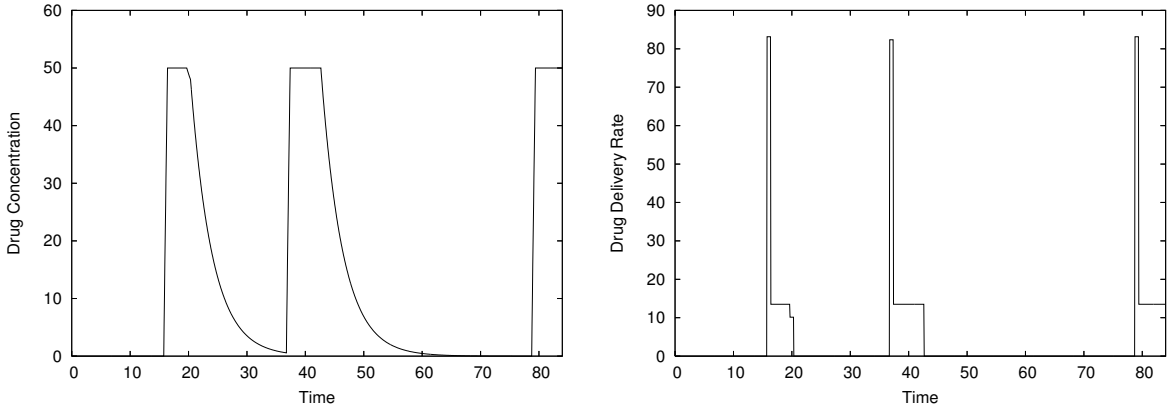

Figure 2.4: The optimal treatment regime and its corresponding drug concentration profile in Example 2.5.2.

This requirement gives rise to the following CTI constraints:

$$y(\tau_j) - y(\tau_{j-1}) + \ln(\varepsilon) \geq 0, \qquad j = 1, 2, 3, \tag{2.51}$$

where $0 < \varepsilon < 1$.

We want to choose a value for the drug delivery rate in each subinterval so that the final tumor size is minimized. Thus, our optimal control problem is as follows: find a vector $\boldsymbol{\sigma} = [\sigma_1, \ldots, \sigma_{4p}]^T \in \mathbb{R}^{4p}$ that minimizes the cost function

$$G_0 = -y(T),$$

subject to the dynamics (2.46)-(2.47) and the constraints (2.49)-(2.51). The parameter values here are: $T = 84.0$, $\lambda = 9.9 \times 10^{-4}$, $\gamma = 0.27$, $\kappa = 8.4 \times 10^{-3}$, $v_{\text{th}} = 10.0$, $v_{\max} = 50.0$, $v_{\text{acc}} = 1100.0$, $\varepsilon = 0.5$, $\omega = 10^{12}$, $C_0 = 10^{10}$, $t_k = kT/4p$, $k = 0, \ldots, 4p$, and $\tau_j = jT/4$, $j = 0, \ldots, 4$.

We solved this optimal control problem using the method proposed in Sections 2.3 and 2.4. As in Example 2.5.1, we used NLPQLP to perform the optimization and LSODA to solve the differential equations. The function values and gradients required by NLPQLP were computed using Algorithm 2.1. Initially, we choose $p = 4$; the problem was subsequently re-solved for $p = 8$, $p = 16$, and $p = 32$, using the previous solution as the initial guess at each step. The optimal solution that we obtained has a final tumor cell population of

$$C^*(T) = 3.2492 \times 10^7.$$

The optimal treatment regime and its corresponding drug concentration profile are shown in Figure 2.4.

2.6 Conclusion

In this chapter, we developed a numerical method for solving optimal control problems with nonlinear CTI constraints. The main idea of this method is to use the control parameterization technique and the time-scaling transformation to derive an approximate optimization problem that has a finite number of decision variables. We developed a novel scheme for computing the cost and constraint gradients in this approximate problem. This scheme can be used in conjunction with a gradient-based nonlinear programming algorithm to solve the approximate problem.

Our gradient computation scheme, like those in [77, 78, 105], involves integrating an auxiliary dynamic system. This auxiliary system does not have any jumps, and is thus quite different from the auxiliary systems used in [77, 78, 105]. Furthermore, our auxiliary system has an initial condition instead of a final condition, which ensures that it can be integrated forward in time in conjunction with the state system. This is a major advantage because it eliminates the need for state interpolation.

CHAPTER 3

Optimal control problems with continuous inequality constraints*

3.1 Introduction

Dynamic systems typically have requirements that must be satisfied at all times. Such requirements give rise to *continuous constraints*—constraints on the state and/or control that are imposed at each point in the time horizon. Examples of continuously-constrained systems include container cranes [91], batch crystallization systems [87], and solar-powered vehicles [29]. In this chapter, we will develop a numerical method for solving optimal control problems with continuous inequality constraints.

Finding an optimal control strategy for a continuously-constrained system is very challenging. This is because the continuous constraints restrict the system over the entire time horizon, not just at several isolated times. One continuous constraint can therefore be viewed as an infinite (actually, uncountable) number of conventional constraints. Several versions of the Pontryagin minimum principle (see Chapter 1) have been derived for continuously-constrained optimal control problems [38]. In addition, many reliable numerical methods for solving continuously-constrained optimal control problems are available. These include discretization methods [11, 15, 32], non-smooth Newton methods [30, 31], feasible direction methods [84, 85], and control parameterization methods [34, 100, 101].

Control parameterization, in particular, is a versatile method that has been used to solve a wide variety of practical optimal control problems. It was first applied to continuously-constrained optimal control problems in [34]. The method proposed in [34] uses a simple transcription, which is inspired by the one in [98], to convert the continuous inequality constraints into a conventional inequality constraint. An approximate nonlinear programming problem is obtained by first applying this transcription, and then approximating the control by a piecewise constant function. Unfortunately, this approximate problem always violates the so-called *Linear Independence Constraint Qualification* (LICQ)—a regularity condition requiring that the gradients of the active constraints be

*This chapter is based on [75].

linearly independent (see [7, 79]). Nonlinear programming algorithms usually fail to converge if the LICQ is violated. Therefore, it is very difficult to solve the approximate problem in [34] using a nonlinear programming algorithm.

Subsequently, a new transcription method—the so-called ϵ-τ method—was introduced in [100, 101]. The ϵ-τ method approximates each continuous inequality constraint by a conventional inequality constraint. A solution of the original optimal control problem is then obtained by solving a sequence of approximate optimization problems. These approximate problems will generally not violate the LICQ, and thus they are much easier to solve than those in [34]. This is a major advantage of the ϵ-τ method. Nevertheless, the ϵ-τ method is only guaranteed to converge for optimal control problems with *pure-state continuous constraints*—continuous constraints that depend explicitly on the state, but not on the control. In fact, the proofs of the convergence results in [100, 101] are not valid if one of the continuous constraints depends explicitly on the control.

Thus, the methods discussed in [34, 100, 101], which are based on control parameterization, have significant shortcomings. This motivates the work in this chapter. We will develop a new control parameterization method that is capable of handling continuous inequality constraints involving *both* the state and the control. We will also show that this new method has very strong convergence properties. More specifically, we will derive two key convergence results that hold whenever the optimal control problem satisfies standard regularity conditions (see [34, 48, 100, 101, 106]). Furthermore, our new method can be readily implemented using a gradient-based nonlinear programming algorithm. It therefore preserves one of the greatest virtues of control parameterization—ease of implementation.

3.2 Problem formulation

Consider the following dynamic system:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}\big(\boldsymbol{x}(t), \boldsymbol{u}(t)\big), \qquad t \in [0, T], \tag{3.1}$$

and

$$\boldsymbol{x}(0) = \boldsymbol{x}^0, \tag{3.2}$$

where $T > 0$ is a given terminal time; $\boldsymbol{x}(t) \in \mathbb{R}^n$ is the system state at time t; $\boldsymbol{u}(t) \in \mathbb{R}^r$ is the control input at time t; $\boldsymbol{x}^0 \in \mathbb{R}^n$ is a given initial state; and $\boldsymbol{f} : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^n$ is a given function.

Let

$$\mathcal{W} \triangleq \big\{\, \boldsymbol{w} \in \mathbb{R}^r : \; a_\varsigma \le w_\varsigma \le b_\varsigma, \; \varsigma = 1, \ldots, r \,\big\},$$

where a_ς and b_ς, $\varsigma = 1, \ldots, r$, are given real numbers such that $a_\varsigma < b_\varsigma$. Any piecewise

continuous function $\boldsymbol{u} : [0, T] \to \mathcal{W}$ that is continuous from the right is called an *admissible control*. Let $\mathcal{U}$ denote the class of all such admissible controls.

We assume that the following conditions are satisfied.

Assumption 3.1. The function $\boldsymbol{f}$ is continuously differentiable.

Assumption 3.2. There exists a real number $L_1 > 0$ such that

$$\big|\boldsymbol{f}(\boldsymbol{v}, \boldsymbol{w})\big| \leq L_1(1 + |\boldsymbol{v}|), \qquad (\boldsymbol{v}, \boldsymbol{w}) \in \mathbb{R}^n \times \mathcal{W}.$$

By Theorem 3.3.3 of [2], the system (3.1)-(3.2) has a unique solution corresponding to each admissible control $\boldsymbol{u} \in \mathcal{U}$. We denote this solution by $\boldsymbol{x}(\cdot|\boldsymbol{u})$.

We suppose that the dynamic system (3.1)-(3.2) is subject to the following continuous inequality constraints:

$$h_i\big(\boldsymbol{x}(t|\boldsymbol{u}), \boldsymbol{u}(t)\big) \geq 0, \qquad t \in [0, T), \qquad i = 1, \ldots, q, \tag{3.3}$$

where $h_i : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$, $i = 1, \ldots, q$, are given functions. An admissible control $\boldsymbol{u} \in \mathcal{U}$ that satisfies the constraints (3.3) is called a *feasible control*. Let $\mathcal{F}$ denote the class of all such feasible controls.

We now define the following optimal control problem.

Problem P. *Find a feasible control $\boldsymbol{u} \in \mathcal{F}$ that minimizes the cost function*

$$G_0(\boldsymbol{u}) \triangleq \Phi\big(\boldsymbol{x}(T|\boldsymbol{u})\big), \tag{3.4}$$

where $\Phi : \mathbb{R}^n \to \mathbb{R}$ is a given function, over $\mathcal{F}$.

Recall from Section 2.2 that there is no loss of generality in assuming that $\boldsymbol{f}$ and h_i, $i = 1, \ldots, q$, are not explicit functions of time.

The cost function (3.4) only depends on the final state reached by the system. Nevertheless, we can easily incorporate an integral cost of the form

$$\int_0^T \mathcal{L}\big(\boldsymbol{x}(t|\boldsymbol{u}), \boldsymbol{u}(t)\big)dt, \tag{3.5}$$

where $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$ is a given function, into (3.4). This is done by augmenting the state system with the auxiliary dynamics

$$\dot{v}(t) = \mathcal{L}\big(\boldsymbol{x}(t), \boldsymbol{u}(t)\big), \qquad t \in [0, T],$$

and

$$v(0) = 0.$$

Clearly, the value of v at the terminal time is equal to the integral cost (3.5).

The continuous inequality constraints considered in [100, 101] are of the form

$$g_i\big(t, \boldsymbol{x}(t|\boldsymbol{u})\big) \geq 0, \qquad t \in [0, T), \qquad i = 1, \ldots, q, \tag{3.6}$$

where $g_i : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}$, $i = 1, \ldots, q$, are given continuous functions. The continuous inequality constraints considered in this chapter are more complicated: the left-hand side of (3.3) may be an explicit function of *both* the state and the control, not just the state. Furthermore, because it depends on the control function explicitly, the left-hand side of (3.3) is usually a discontinuous function of time. In contrast, the left-hand side of (3.6), considered as a function of time, is the composition of continuous functions and is therefore continuous itself. This fact is exploited in [100,101] to prove several important convergence results.

To conclude this section, we make one further assumption.

Assumption 3.3. The functions Φ and h_i, $i = 1, \ldots, q$, are continuously differentiable.

3.3 Problem approximation

In this section, we will derive an approximation of Problem P by restricting the admissible controls to suitable piecewise constant functions. Convergence results relating a solution of this approximate problem to a solution of Problem P will be presented later in Section 3.6.

First, let $p \geq 2$ be a fixed integer. Furthermore, let Γ denote the set consisting of all vectors $\boldsymbol{\nu} = [t_1, \ldots, t_{p-1}]^T \in \mathbb{R}^{p-1}$ that satisfy the constraints

$$t_{k-1} \leq t_k, \qquad k = 1, \ldots, p, \tag{3.7}$$

where $t_0 \triangleq 0$ and $t_p \triangleq T$.

For each $\boldsymbol{\nu} = [t_1, \ldots, t_{p-1}]^T \in \Gamma$, define corresponding subintervals $\mathcal{I}_k(\boldsymbol{\nu})$, $k = 1, \ldots, p$, as follows:

$$\mathcal{I}_k(\boldsymbol{\nu}) \triangleq \begin{cases} [t_{k-1}, t_k), & \text{if } k \in \{1, \ldots, p-1\}, \\ [t_{k-1}, t_k], & \text{if } k = p. \end{cases}$$

It is clear that

$$\mathcal{I}_1(\boldsymbol{\nu}), \ldots, \mathcal{I}_p(\boldsymbol{\nu}),$$

constitutes a partition of the time horizon $[0, T]$. That is,

$$\bigcup_{k=1}^{p} \mathcal{I}_k(\boldsymbol{\nu}) = [0, T]$$

and

$$\mathcal{I}_{k_1}(\boldsymbol{\nu}) \cap \mathcal{I}_{k_2}(\boldsymbol{\nu}) = \emptyset, \qquad k_1 \neq k_2.$$

Now, define

$$\Xi \triangleq \prod_{k=1}^{p} \mathcal{W}.$$

In other words, Ξ is the set of all tuples $(\boldsymbol{\sigma}^1, \ldots, \boldsymbol{\sigma}^p)$ satisfying $\boldsymbol{\sigma}^k \in \mathcal{W}$, $k = 1, \ldots, p$.

For each $(\boldsymbol{\nu}, \boldsymbol{\sigma}) \in \Gamma \times \Xi$, we define a corresponding function $\boldsymbol{u}^p(\cdot|\boldsymbol{\nu}, \boldsymbol{\sigma}) : [0, T] \to \mathbb{R}^r$ as follows:

$$\boldsymbol{u}^p(t|\boldsymbol{\nu}, \boldsymbol{\sigma}) \triangleq \sum_{k=1}^{p} \boldsymbol{\sigma}^k \chi_{\mathcal{I}_k(\boldsymbol{\nu})}(t), \qquad t \in [0, T], \tag{3.8}$$

where, for each $\mathcal{I} \subset \mathbb{R}$, the indicator function $\chi_{\mathcal{I}} : \mathbb{R} \to \mathbb{R}$ is defined by

$$\chi_{\mathcal{I}}(t) \triangleq \begin{cases} 1, & \text{if } t \in \mathcal{I}, \\ 0, & \text{otherwise.} \end{cases}$$

Clearly,

$$\boldsymbol{u}^p(t|\boldsymbol{\nu}, \boldsymbol{\sigma}) = \boldsymbol{\sigma}^k \in \mathcal{W}, \qquad t \in \mathcal{I}_k(\boldsymbol{\nu}), \qquad k = 1, \ldots, p,$$

which shows that the range of $\boldsymbol{u}^p(\cdot|\boldsymbol{\nu}, \boldsymbol{\sigma})$ is in $\mathcal{W}$. Furthermore, it is clear that $\boldsymbol{u}^p(\cdot|\boldsymbol{\nu}, \boldsymbol{\sigma})$ is both piecewise continuous and continuous from the right. Hence, $\boldsymbol{u}^p(\cdot|\boldsymbol{\nu}, \boldsymbol{\sigma})$ is an admissible control for Problem P. Recall from Chapter 2 that the times t_k, $k = 1, \ldots, p-1$, are called *switching times*. Accordingly, each $\boldsymbol{\nu} \in \Gamma$ is called a *switching-time vector*.

We will approximate the control in Problem P by the piecewise constant function (3.8). Substituting (3.8) into the dynamic system (3.1)-(3.2) gives

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}\big(\boldsymbol{x}(t), \boldsymbol{\sigma}^k\big), \qquad t \in \mathcal{I}_k(\boldsymbol{\nu}), \qquad k = 1, \ldots, p, \tag{3.9}$$

and

$$\boldsymbol{x}(0) = \boldsymbol{x}^0. \tag{3.10}$$

Let $\boldsymbol{x}^p(\cdot|\boldsymbol{\nu}, \boldsymbol{\sigma})$ denote the solution of (3.9)-(3.10) corresponding to $(\boldsymbol{\nu}, \boldsymbol{\sigma}) \in \Gamma \times \Xi$. Clearly,

$$\boldsymbol{x}^p(\cdot|\boldsymbol{\nu}, \boldsymbol{\sigma}) = \boldsymbol{x}(\cdot|\boldsymbol{u}^p(\cdot|\boldsymbol{\nu}, \boldsymbol{\sigma})).$$

Substituting (3.8) into the continuous inequality constraints (3.3) gives

$$h_i\big(\boldsymbol{x}^p(t|\boldsymbol{\nu}, \boldsymbol{\sigma}), \boldsymbol{\sigma}^k\big) \geq 0, \qquad t \in [t_{k-1}, t_k), \qquad k = 1, \ldots, p, \qquad i = 1, \ldots, q. \tag{3.11}$$

Let Ω denote the set consisting of all pairs $(\boldsymbol{\nu}, \boldsymbol{\sigma}) \in \Gamma \times \Xi$ that satisfy the constraints (3.11).

Such pairs are called *feasible pairs*. Clearly,

$$\Omega = \big\{\, (\boldsymbol{\nu}, \boldsymbol{\sigma}) \in \Gamma \times \Xi :\ \boldsymbol{u}^p(\cdot|\boldsymbol{\nu}, \boldsymbol{\sigma}) \in \mathcal{F} \,\big\}.$$

When the controls are restricted to those of the form (3.8), the cost function (3.4) becomes

$$G_0^p(\boldsymbol{\nu}, \boldsymbol{\sigma}) \triangleq G_0(\boldsymbol{u}^p(\cdot|\boldsymbol{\nu}, \boldsymbol{\sigma})) = \Phi\big(\boldsymbol{x}^p(T|\boldsymbol{\nu}, \boldsymbol{\sigma})\big).$$

We now define the following approximate optimization problem.

Problem $\mathbf{P}_p$. *Find a feasible pair $(\boldsymbol{\nu}, \boldsymbol{\sigma}) \in \Omega$ that minimizes the cost function G_0^p over Ω.*

Remark 3.1. We emphasize that *both* the values and the switching times of the approximate control (3.8) are decision variables in Problem P_p. In contrast, the methods in [34,100,101] use a coarse partition of the time horizon to pre-assign the switching times; only the control values are determined optimally.

Remark 3.2. If $(\boldsymbol{\nu}^*, \boldsymbol{\sigma}^*) \in \Omega$ is an optimal solution of Problem P_p, then $\boldsymbol{u}^p(\cdot|\boldsymbol{\nu}^*, \boldsymbol{\sigma}^*)$ is a suboptimal control for Problem P.

3.4 Time-scaling transformation

The decision variables in Problem P_p are the values and switching times of a piecewise constant control. As we mentioned in Chapters 1 and 2, it is very difficult to solve dynamic optimization problems with variable switching times directly. Hence, in this section, we will use the time-scaling transformation to convert Problem P_p into a new problem with fixed switching times.

Let $p \geq 2$ be a fixed integer. Define

$$\Theta \triangleq \big\{\, \boldsymbol{\theta} \in \mathbb{R}^p :\ \theta_k \geq 0,\ k = 1, \ldots, p;\ \theta_1 + \cdots + \theta_p = T \,\big\}.$$

For each $\boldsymbol{\theta} \in \Theta$, define a corresponding function $\mu(\cdot|\boldsymbol{\theta}) : [0, 1] \to \mathbb{R}$ as follows:

$$\mu(s|\boldsymbol{\theta}) \triangleq \begin{cases} \displaystyle\sum_{l=1}^{\lfloor ps \rfloor} \theta_l + \theta_{\lfloor ps \rfloor + 1}(ps - \lfloor ps \rfloor), & \text{if } s \in [0, 1), \\ T, & \text{if } s = 1, \end{cases}$$

where $\lfloor \cdot \rfloor$ denotes the floor function. It is not difficult to show that $\mu(\cdot|\boldsymbol{\theta})$ is continuous, non-negative, and non-decreasing.

Now, let

$$\alpha_k \triangleq \frac{k}{p}, \qquad k = 0, \ldots, p.$$

For each $\boldsymbol{\theta} \in \Theta$, let

$$\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}) = \left[\tilde{\nu}_1(\boldsymbol{\theta}), \ldots, \tilde{\nu}_{p-1}(\boldsymbol{\theta})\right]^T \in \mathbb{R}^{p-1}$$

be the vector in $\mathbb{R}^{p-1}$ whose kth component is equal to $\mu(\alpha_k|\boldsymbol{\theta})$. Furthermore, define

$$\tilde{\nu}_0(\boldsymbol{\theta}) \triangleq \mu(0|\boldsymbol{\theta}) = 0$$

and

$$\tilde{\nu}_p(\boldsymbol{\theta}) \triangleq \mu(1|\boldsymbol{\theta}) = T.$$

Therefore,

$$\tilde{\nu}_k(\boldsymbol{\theta}) = \mu(\alpha_k|\boldsymbol{\theta}) = \sum_{l=1}^{k} \theta_l, \qquad k = 0, \ldots, p. \tag{3.12}$$

Since the components of $\boldsymbol{\theta} \in \Theta$ are non-negative, equation (3.12) implies that

$$\tilde{\nu}_{k-1}(\boldsymbol{\theta}) \leq \tilde{\nu}_k(\boldsymbol{\theta}), \qquad k = 1, \ldots, p.$$

Thus, $\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta})$ is a valid switching-time vector for Problem P_p. That is,

$$\left\{ \tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}) : \ \boldsymbol{\theta} \in \Theta \right\} \subset \Gamma. \tag{3.13}$$

It turns out that the reverse inclusion is also true. To see why, let $\boldsymbol{\nu}' = [t'_1, \ldots, t'_{p-1}]^T \in \Gamma$ and define a corresponding vector $\boldsymbol{\theta}' \in \mathbb{R}^p$ as follows:

$$\theta'_k \triangleq t'_k - t'_{k-1}, \qquad k = 1, \ldots, p,$$

where $t'_0 \triangleq 0$ and $t'_p \triangleq T$. Since the components of $\boldsymbol{\nu}'$ satisfy (3.7), the components of $\boldsymbol{\theta}'$ are non-negative. Furthermore,

$$\sum_{k=1}^{p} \theta'_k = t'_p - t'_0 = T.$$

Hence, $\boldsymbol{\theta}' \in \Theta$. Now, using equation (3.12), we obtain

$$\mu(\alpha_k|\boldsymbol{\theta}') = \sum_{l=1}^{k} \theta'_l = t'_k, \qquad k = 1, \ldots, p-1.$$

It then follows immediately that $\boldsymbol{\nu}' = \tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}')$. Since $\boldsymbol{\nu}' \in \Gamma$ was chosen arbitrarily, this implies that

$$\Gamma \subset \left\{ \tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}) : \ \boldsymbol{\theta} \in \Theta \right\}. \tag{3.14}$$

By combining inclusions (3.13) and (3.14), we obtain the following equation that links

the sets Γ and Θ:

$$\Gamma = \big\{\, \tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}) :\ \boldsymbol{\theta} \in \Theta \,\big\}. \tag{3.15}$$

Now, recall that for each $\boldsymbol{\theta} \in \Theta$, the function $\mu(\cdot|\boldsymbol{\theta})$ is non-decreasing. Therefore,

$$0 = \mu(0|\boldsymbol{\theta}) \leq \mu(s|\boldsymbol{\theta}) \leq \mu(1|\boldsymbol{\theta}) = T, \qquad s \in [0,1]. \tag{3.16}$$

Equation (3.15) and inequality (3.16) ensure that for each $(\boldsymbol{\theta}, \boldsymbol{\sigma}) \in \Theta \times \Xi$, we may define a corresponding function $\tilde{\boldsymbol{x}}(\cdot|\boldsymbol{\theta}, \boldsymbol{\sigma}) : [0,1] \to \mathbb{R}^n$ as follows:

$$\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta}, \boldsymbol{\sigma}) \triangleq \boldsymbol{x}^p(\mu(s|\boldsymbol{\theta})|\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}), \boldsymbol{\sigma}), \qquad s \in [0,1]. \tag{3.17}$$

The function $\tilde{\boldsymbol{x}}^p(\cdot|\boldsymbol{\theta}, \boldsymbol{\sigma})$ is a new state trajectory defined on the interval $[0,1]$.

Now, note that

$$\dot{\mu}(s|\boldsymbol{\theta}) = p\theta_k, \qquad s \in \mathcal{J}_k, \qquad k = 1, \ldots, p, \tag{3.18}$$

where

$$\mathcal{J}_k \triangleq \begin{cases} [\alpha_{k-1}, \alpha_k), & \text{if } k = 1, \\ (\alpha_{k-1}, \alpha_k), & \text{if } k \in \{2, \ldots, p-1\}, \\ (\alpha_{k-1}, \alpha_k], & \text{if } k = p. \end{cases}$$

By differentiating (3.17) with respect to s, and then using equations (3.9) and (3.18), we obtain

$$\dot{\tilde{\boldsymbol{x}}}^p(s|\boldsymbol{\theta}, \boldsymbol{\sigma}) = p\theta_k \boldsymbol{f}\big(\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta}, \boldsymbol{\sigma}), \boldsymbol{\sigma}^k\big), \qquad s \in \mathcal{J}_k, \qquad k = 1, \ldots, p. \tag{3.19}$$

We also have

$$\tilde{\boldsymbol{x}}^p(\alpha_k|\boldsymbol{\theta}, \boldsymbol{\sigma}) = \lim_{s \to \alpha_k} \tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta}, \boldsymbol{\sigma}), \qquad k = 1, \ldots, p-1, \tag{3.20}$$

and

$$\tilde{\boldsymbol{x}}^p(0|\boldsymbol{\theta}, \boldsymbol{\sigma}) = \boldsymbol{x}^0. \tag{3.21}$$

The new dynamic system (3.19)-(3.21) is obtained by transforming the time variable from $t \in [0, T]$ to $s \in [0,1]$. It has switching times at the uniformly distributed locations $s = \alpha_k$, $k = 1, \ldots, p-1$, and is therefore much easier to work with than (3.9)-(3.10).

Let Λ denote the set consisting of all pairs $(\boldsymbol{\theta}, \boldsymbol{\sigma}) \in \Theta \times \Xi$ that satisfy the following continuous inequality constraints:

$$\theta_k h_i\big(\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta}, \boldsymbol{\sigma}), \boldsymbol{\sigma}^k\big) \geq 0, \qquad s \in \bar{\mathcal{J}}_k, \qquad k = 1, \ldots, p, \qquad i = 1, \ldots, q, \tag{3.22}$$

where the overhead bar denotes set closure. These new constraints are equivalent, in a sense, to the original constraints (3.11). This equivalence is stated precisely in the following theorem.

Theorem 3.1. *Let* $(\boldsymbol{\theta}, \boldsymbol{\sigma}) \in \Theta \times \Xi$. *Then* $(\boldsymbol{\theta}, \boldsymbol{\sigma}) \in \Lambda$ *if and only if* $(\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}), \boldsymbol{\sigma}) \in \Omega$.

Proof. Let $(\boldsymbol{\theta}, \boldsymbol{\sigma}) \in \Theta \times \Xi$ be arbitrary but fixed. To simplify the notation, we write $\boldsymbol{x}^p$ instead of $\boldsymbol{x}^p(\cdot|\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}), \boldsymbol{\sigma})$ and $\tilde{\boldsymbol{x}}^p$ instead of $\tilde{\boldsymbol{x}}^p(\cdot|\boldsymbol{\theta}, \boldsymbol{\sigma})$.

Define the index sets

$$\mathcal{G}_1 \triangleq \big\{\, k \in \{1, \ldots, p\} : \ \theta_k = 0 \,\big\}$$

and

$$\mathcal{G}_2 \triangleq \{1, \ldots, p\} \setminus \mathcal{G}_1.$$

Recall from equation (3.12) that

$$\tilde{\nu}_k(\boldsymbol{\theta}) = \sum_{l=1}^{k} \theta_l, \qquad k = 0, \ldots, p.$$

Hence,

$$\tilde{\nu}_k(\boldsymbol{\theta}) - \tilde{\nu}_{k-1}(\boldsymbol{\theta}) = \theta_k, \qquad k = 1, \ldots, p.$$

From this equation, we obtain the following important implication:

$$\tilde{\nu}_k(\boldsymbol{\theta}) = \tilde{\nu}_{k-1}(\boldsymbol{\theta}) \qquad \Longleftrightarrow \qquad k \in \mathcal{G}_1. \tag{3.23}$$

Now, suppose that $(\boldsymbol{\theta}, \boldsymbol{\sigma}) \in \Lambda$. Then

$$\theta_k h_i\big(\tilde{\boldsymbol{x}}^p(s), \boldsymbol{\sigma}^k\big) \geq 0, \qquad s \in \bar{\mathcal{J}}_k, \qquad k = 1, \ldots, p, \qquad i = 1, \ldots, q. \tag{3.24}$$

It follows from (3.23) that for each $k \in \mathcal{G}_1$,

$$[\tilde{\nu}_{k-1}(\boldsymbol{\theta}), \tilde{\nu}_k(\boldsymbol{\theta})) = \emptyset.$$

Consequently, the constraints

$$h_i\big(\boldsymbol{x}^p(t), \boldsymbol{\sigma}^k\big) \geq 0, \qquad t \in [\tilde{\nu}_{k-1}(\boldsymbol{\theta}), \tilde{\nu}_k(\boldsymbol{\theta})), \qquad k \in \mathcal{G}_1, \qquad i = 1, \ldots, q, \tag{3.25}$$

are satisfied vacuously.

On the other hand, if $k \in \mathcal{G}_2$, then we can divide both sides of (3.24) by θ_k to obtain

$$h_i\big(\tilde{\boldsymbol{x}}^p(s), \boldsymbol{\sigma}^k\big) \geq 0, \qquad s \in \bar{\mathcal{J}}_k, \qquad k \in \mathcal{G}_2, \qquad i = 1, \ldots, q. \tag{3.26}$$

Let $\mu_k(\cdot|\boldsymbol{\theta}) : \bar{\mathcal{J}}_k \to \mathbb{R}$ denote the restriction of $\mu(\cdot|\boldsymbol{\theta})$ to the subinterval $\bar{\mathcal{J}}_k$. It follows

from the definition of $\mu(\cdot|\boldsymbol{\theta})$ that

$$\mu_k(s|\boldsymbol{\theta}) = \sum_{l=1}^{k-1} \theta_l + \theta_k(ps - k + 1), \qquad s \in \bar{\mathcal{J}}_k.$$

Since $\theta_k > 0$, $\mu_k(\cdot|\boldsymbol{\theta})$ is a bijection from $\bar{\mathcal{J}}_k$ onto $[\tilde{\nu}_{k-1}(\boldsymbol{\theta}), \tilde{\nu}_k(\boldsymbol{\theta})]$. Hence, there exists an inverse function $\mu_k^{-1} : [\tilde{\nu}_{k-1}(\boldsymbol{\theta}), \tilde{\nu}_k(\boldsymbol{\theta})] \to \bar{\mathcal{J}}_k$ such that

$$t = \mu_k\big(\mu_k^{-1}(t|\boldsymbol{\theta})\big), \qquad t \in [\tilde{\nu}_{k-1}(\boldsymbol{\theta}), \tilde{\nu}_k(\boldsymbol{\theta})]. \tag{3.27}$$

Using (3.26) and (3.27), we obtain

$$h_i\big(\boldsymbol{x}^p(t), \boldsymbol{\sigma}^k\big) = h_i\big(\tilde{\boldsymbol{x}}^p(\mu_k^{-1}(t|\boldsymbol{\theta})), \boldsymbol{\sigma}^k\big) \geq 0, \qquad t \in [\tilde{\nu}_{k-1}(\boldsymbol{\theta}), \tilde{\nu}_k(\boldsymbol{\theta})), \\ k \in \mathcal{G}_2, \qquad i = 1, \ldots, q. \tag{3.28}$$

Inequalities (3.25) and (3.28) show that $(\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}), \boldsymbol{\sigma}) \in \Omega$.

Conversely, suppose that $(\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}), \boldsymbol{\sigma}) \in \Omega$. Then

$$h_i\big(\boldsymbol{x}^p(t), \boldsymbol{\sigma}^k\big) \geq 0, \qquad t \in [\tilde{\nu}_{k-1}(\boldsymbol{\theta}), \tilde{\nu}_k(\boldsymbol{\theta})), \qquad k = 1, \ldots, p, \qquad i = 1, \ldots, q. \tag{3.29}$$

If $k \in \mathcal{G}_1$, then $\theta_k = 0$. Therefore,

$$\theta_k h_i\big(\tilde{\boldsymbol{x}}^p(s), \boldsymbol{\sigma}^k\big) = 0, \qquad s \in \bar{\mathcal{J}}_k, \qquad k \in \mathcal{G}_1, \qquad i = 1, \ldots, q. \tag{3.30}$$

On the other hand, if $k \in \mathcal{G}_2$, then $\mu(\cdot|\boldsymbol{\theta})$ is strictly increasing on $\bar{\mathcal{J}}_k$. Hence,

$$\mu(s|\boldsymbol{\theta}) \in [\tilde{\nu}_{k-1}(\boldsymbol{\theta}), \tilde{\nu}_k(\boldsymbol{\theta})), \qquad s \in \bar{\mathcal{J}}_k \setminus \{\alpha_k\}. \tag{3.31}$$

Using (3.29) and (3.31), we obtain

$$\theta_k h_i\big(\tilde{\boldsymbol{x}}^p(s), \boldsymbol{\sigma}^k\big) = \theta_k h_i\big(\boldsymbol{x}^p(\mu(s|\boldsymbol{\theta})), \boldsymbol{\sigma}^k\big) \geq 0, \\ s \in \bar{\mathcal{J}}_k \setminus \{\alpha_k\}, \qquad k \in \mathcal{G}_2, \qquad i = 1, \ldots, q. \tag{3.32}$$

Since $\tilde{\boldsymbol{x}}^p$ and h_i, $i = 1, \ldots, q$, are continuous functions, (3.32) also holds at $s = \alpha_k$. Thus,

$$\theta_k h_i\big(\tilde{\boldsymbol{x}}^p(s), \boldsymbol{\sigma}^k\big) \geq 0, \qquad s \in \bar{\mathcal{J}}_k, \qquad k \in \mathcal{G}_2, \qquad i = 1, \ldots, q. \tag{3.33}$$

Equation (3.30) and inequality (3.33) show that $(\boldsymbol{\theta}, \boldsymbol{\sigma}) \in \Lambda$. □

We now define a new optimization problem.

Problem $\tilde{\mathrm{P}}_p$. *Find a pair* $(\boldsymbol{\theta}, \boldsymbol{\sigma}) \in \Lambda$ *that minimizes the cost function*

$$\tilde{G}_0^p(\boldsymbol{\theta}, \boldsymbol{\sigma}) \triangleq G_0^p(\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}), \boldsymbol{\sigma}) = \Phi\big(\tilde{\boldsymbol{x}}^p(1|\boldsymbol{\theta}, \boldsymbol{\sigma})\big)$$

over Λ.

Problems P_p and $\tilde{\mathrm{P}}_p$ are equivalent. Indeed, $(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*) \in \Lambda$ is optimal for Problem $\tilde{\mathrm{P}}_p$ if and only if $(\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}^*), \boldsymbol{\sigma}^*) \in \Omega$ is optimal for Problem P_p. To see why, suppose that $(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*) \in \Lambda$ is an optimal solution of Problem $\tilde{\mathrm{P}}_p$ and let $(\boldsymbol{\nu}', \boldsymbol{\sigma}') \in \Omega$ be arbitrary. Then by equation (3.15), there exists a vector $\boldsymbol{\theta}' \in \Theta$ such that $\boldsymbol{\nu}' = \tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}')$. It follows from Theorem 3.1 that $(\boldsymbol{\theta}', \boldsymbol{\sigma}') \in \Lambda$ and $(\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}^*), \boldsymbol{\sigma}^*) \in \Omega$. Hence,

$$G_0^p(\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}^*), \boldsymbol{\sigma}^*) = \tilde{G}_0^p(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*) \leq \tilde{G}_0^p(\boldsymbol{\theta}', \boldsymbol{\sigma}') = G_0^p(\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}'), \boldsymbol{\sigma}') = G_0^p(\boldsymbol{\nu}', \boldsymbol{\sigma}').$$

Since $(\boldsymbol{\nu}', \boldsymbol{\sigma}') \in \Omega$ was chosen arbitrarily, this inequality shows that $(\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}^*), \boldsymbol{\sigma}^*)$ is an optimal solution of Problem P_p.

Conversely, let $(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*) \in \Theta \times \Xi$ be such that $(\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}^*), \boldsymbol{\sigma}^*)$ is an optimal solution of Problem P_p. Furthermore, let $(\boldsymbol{\theta}', \boldsymbol{\sigma}') \in \Lambda$ be arbitrary but fixed. Then it follows from Theorem 3.1 that $(\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}'), \boldsymbol{\sigma}') \in \Omega$ and $(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*) \in \Lambda$. Thus,

$$\tilde{G}_0^p(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*) = G_0^p(\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}^*), \boldsymbol{\sigma}^*) \leq G_0^p(\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}'), \boldsymbol{\sigma}') = \tilde{G}_0^p(\boldsymbol{\theta}', \boldsymbol{\sigma}').$$

Since $(\boldsymbol{\theta}', \boldsymbol{\sigma}') \in \Lambda$ was chosen arbitrarily, this inequality shows that $(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*)$ is an optimal solution of Problem $\tilde{\mathrm{P}}_p$.

Remark 3.3. If $(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*) \in \Lambda$ is an optimal solution of Problem $\tilde{\mathrm{P}}_p$, then $(\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}^*), \boldsymbol{\sigma}^*) \in \Omega$ is optimal for Problem P_p and $\boldsymbol{u}^p(\cdot|\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}^*), \boldsymbol{\sigma}^*)$ is a suboptimal control for Problem P (see Remark 3.2).

3.5 Solving Problem $\tilde{\mathrm{P}}_p$

Notice that (3.22) defines an infinite number of constraints—one for each point in $[0, 1]$. Hence, Problem $\tilde{\mathrm{P}}_p$ is an optimization problem with a finite number of decision variables, but an infinite number of constraints. Optimization problems of this type are called *semi-infinite programming problems*. Semi-infinite programming problems are more complicated than nonlinear programming problems, which have at most a finite number of constraints. Nevertheless, semi-infinite programming problems can be solved using the penalty function algorithm developed in [106]. We now describe how this algorithm can be used to solve Problem $\tilde{\mathrm{P}}_p$.

First, define functions $\tilde{g}^p_{i,k,\epsilon} : \Theta \times \Xi \to \mathbb{R}$, $k = 1, \ldots, p$, $i = 1, \ldots, q$, as follows:

$$\tilde{g}^p_{i,k,\epsilon}(\boldsymbol{\theta}, \boldsymbol{\sigma}) \triangleq \int_{\mathcal{J}_k} \varphi_\epsilon\big(\theta_k h_i(\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta}, \boldsymbol{\sigma}), \boldsymbol{\sigma}^k)\big) ds, \qquad k = 1, \ldots, p, \qquad i = 1, \ldots, q, \quad (3.34)$$

where $\epsilon > 0$ and $\varphi_\epsilon : \mathbb{R} \to \mathbb{R}$ is defined by

$$\varphi_\epsilon(\eta) \triangleq \begin{cases} \eta, & \text{if } \eta < -\epsilon, \\ -(\eta - \epsilon)^2/4\epsilon, & \text{if } -\epsilon \le \eta \le \epsilon, \\ 0, & \text{otherwise.} \end{cases}$$

Next, consider the following auxiliary optimization problem.

Problem $\tilde{\mathrm{P}}_{p,\epsilon,\vartheta}$. *Find a pair $(\boldsymbol{\theta}, \boldsymbol{\sigma}) \in \Theta \times \Xi$ that minimizes the cost function*

$$\tilde{J}^p_{\epsilon,\vartheta}(\boldsymbol{\theta}, \boldsymbol{\sigma}) \triangleq \tilde{G}^p_0(\boldsymbol{\theta}, \boldsymbol{\sigma}) - \vartheta \sum_{i=1}^{q} \sum_{k=1}^{p} \tilde{g}^p_{i,k,\epsilon}(\boldsymbol{\theta}, \boldsymbol{\sigma}),$$

where $\epsilon > 0$ and $\vartheta > 0$, over $\Theta \times \Xi$.

Problem $\tilde{\mathrm{P}}_{p,\epsilon,\vartheta}$ has a finite number of decision variables and a finite number of constraints (the constraints consist of simple bounds on the decision variables and a linear equality constraint). Furthermore, since φ_ϵ is continuously differentiable, the partial derivatives of $\tilde{J}^p_{\epsilon,\vartheta}$ can be computed using the standard formulae given in [100]. We state these formulae below.

First, for each $k = 1, \ldots, p$, define a corresponding function $H_k : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$ as follows:

$$H_k(\theta_k, \boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\sigma}^k) \triangleq -\vartheta \sum_{i=1}^{q} \varphi_\epsilon\big(\theta_k h_i(\boldsymbol{x}, \boldsymbol{\sigma}^k)\big) + \theta_k \boldsymbol{\lambda}^T \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{\sigma}^k),$$
$$(\theta_k, \boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\sigma}^k) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^r.$$

This function is called the *Hamiltonian*.

Next, define the following auxiliary dynamic system:

$$\dot{\boldsymbol{\lambda}}(s) = -\left[\frac{\partial H_k\big(\theta_k, \tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta}, \boldsymbol{\sigma}), \boldsymbol{\lambda}(s), \boldsymbol{\sigma}^k\big)}{\partial \boldsymbol{x}}\right]^T, \qquad s \in \mathcal{J}_k, \qquad k = 1, \ldots, p,$$

and

$$\boldsymbol{\lambda}(1) = \left[\frac{\partial \Phi\big(\tilde{\boldsymbol{x}}^p(1|\boldsymbol{\theta}, \boldsymbol{\sigma})\big)}{\partial \boldsymbol{x}}\right]^T,$$

where $(\boldsymbol{\theta}, \boldsymbol{\sigma}) \in \Theta \times \Xi$. This auxiliary dynamic system is called the *costate system*. Let $\boldsymbol{\lambda}(\cdot|\boldsymbol{\theta}, \boldsymbol{\sigma})$ denote the solution of the costate system corresponding to $(\boldsymbol{\theta}, \boldsymbol{\sigma}) \in \Theta \times \Xi$.

Then it follows from Theorem 5.2.1 of [100] that for each pair $(\boldsymbol{\theta}, \boldsymbol{\sigma}) \in \Theta \times \Xi$,

$$\frac{\partial \tilde{J}^p_{\epsilon,\vartheta}(\boldsymbol{\theta}, \boldsymbol{\sigma})}{\partial \theta_k} = \int_{\mathcal{J}_k} \frac{\partial H_k\big(\theta_k, \tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta}, \boldsymbol{\sigma}), \boldsymbol{\lambda}(s|\boldsymbol{\theta}, \boldsymbol{\sigma}), \boldsymbol{\sigma}^k\big)}{\partial \theta_k}, \qquad k = 1, \ldots, p, \tag{3.35}$$

and

$$\frac{\partial \tilde{J}^p_{\epsilon,\vartheta}(\boldsymbol{\theta}, \boldsymbol{\sigma})}{\partial \sigma^k_\varsigma} = \int_{\mathcal{J}_k} \frac{\partial H_k\big(\theta_k, \tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta}, \boldsymbol{\sigma}), \boldsymbol{\lambda}(s|\boldsymbol{\theta}, \boldsymbol{\sigma}), \boldsymbol{\sigma}^k\big)}{\partial \sigma^k_\varsigma},$$
$$k = 1, \ldots, p, \qquad \varsigma = 1, \ldots, r. \tag{3.36}$$

Equations (3.35) and (3.36) can be used in conjunction with a gradient-based nonlinear programming algorithm to solve Problem $\tilde{\mathrm{P}}_{p,\epsilon,\vartheta}$. Problem $\tilde{\mathrm{P}}_{p,\epsilon,\vartheta}$ can also be solved directly using the optimal control software MISER3 [47].

What is the relationship between Problems $\tilde{\mathrm{P}}_p$ and $\tilde{\mathrm{P}}_{p,\epsilon,\vartheta}$? First, notice that since h_i, $i = 1, \ldots, q$, are continuous (see Assumption 3.3), the equality constraints

$$\int_{\mathcal{J}_k} \min\big\{\theta_k h_i(\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta}, \boldsymbol{\sigma}), \boldsymbol{\sigma}^k), 0\big\} ds = 0, \qquad k = 1, \ldots, p, \qquad i = 1, \ldots, q, \tag{3.37}$$

are actually equivalent to (3.22). However, the left-hand side of (3.37) is non-smooth, and thus gradient-based nonlinear programming algorithms cannot handle such constraints directly.

Now, we see from Figure 3.1 that φ_ϵ is a smooth approximation of $\min\{\cdot, 0\}$. Thus, when ϵ is small,

$$\varphi_\epsilon(\eta) \approx \min\{\eta, 0\}, \qquad \eta \in \mathbb{R},$$

and

$$\tilde{g}^p_{i,k,\epsilon}(\boldsymbol{\theta}, \boldsymbol{\sigma}) \approx \int_{\mathcal{J}_k} \min\big\{\theta_k h_i(\tilde{\boldsymbol{x}}^p(s|\boldsymbol{\theta}, \boldsymbol{\sigma}), \boldsymbol{\sigma}^k), 0\big\} ds.$$

It follows that we can approximate (3.37)—and therefore (3.22)—by the constraints

$$\tilde{g}^p_{i,k,\epsilon}(\boldsymbol{\theta}, \boldsymbol{\sigma}) = 0, \qquad k = 1, \ldots, p, \qquad i = 1, \ldots, q. \tag{3.38}$$

These approximate constraints are smooth because φ_ϵ is continuously differentiable. Furthermore, notice that violations of the approximate constraints (3.38) are penalized in the cost function of Problem $\tilde{\mathrm{P}}_{p,\epsilon,\vartheta}$. Hence, we expect that Problem $\tilde{\mathrm{P}}_{p,\epsilon,\vartheta}$ is a good approximation of Problem $\tilde{\mathrm{P}}_p$ when ϵ is small and ϑ is large. More precisely, we have the following two results, which are proved in [106].

Theorem 3.2. *For each $\epsilon > 0$, there exists a corresponding $\vartheta(\epsilon) > 0$ such that if $\vartheta > \vartheta(\epsilon)$, then the optimal solution of Problem $\tilde{P}_{p,\epsilon,\vartheta}$ is feasible for Problem $\tilde{P}_p$.*

Theorem 3.3. *Suppose that $(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*) \in \Theta \times \Xi$ is an optimal solution of Problem $\tilde{P}_p$. For*

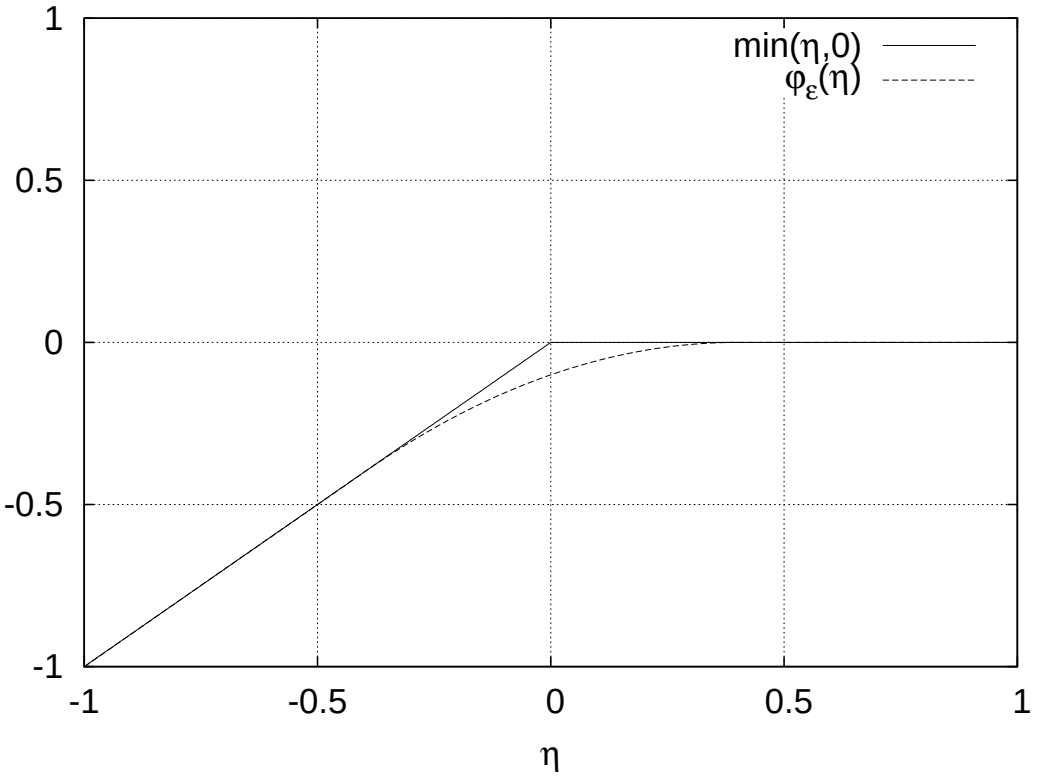

Figure 3.1: The smoothing function φ_ϵ for $\epsilon = 0.4$.

each $\epsilon > 0$, *let* $(\boldsymbol{\theta}^*_{\epsilon,\vartheta}, \boldsymbol{\sigma}^*_{\epsilon,\vartheta})$ *denote the solution of Problem* $\tilde{P}_{p,\epsilon,\vartheta}$, *where* $\vartheta > 0$ *is chosen to ensure that* $(\boldsymbol{\theta}^*_{\epsilon,\vartheta}, \boldsymbol{\sigma}^*_{\epsilon,\vartheta}) \in \Lambda$ (*Theorem 3.2 guarantees that this can always be done*). *Then*

$$\lim_{\epsilon \to 0} \tilde{G}_0^p(\boldsymbol{\theta}^*_{\epsilon,\vartheta}, \boldsymbol{\sigma}^*_{\epsilon,\vartheta}) = \tilde{G}_0^p(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*).$$

Theorems 3.2 and 3.3 suggest the following method for solving Problem $\tilde{\mathrm{P}}_p$. First, choose an initial positive value for ϵ. Second, repeatedly solve Problem $\tilde{\mathrm{P}}_{p,\epsilon,\vartheta}$ for increasing values of ϑ until the solution obtained is feasible for Problem $\tilde{\mathrm{P}}_p$. According to Theorem 3.2, Problem $\tilde{\mathrm{P}}_p$ only needs to be solved a finite number of times here. Next, decrease ϵ and repeat this procedure, using the solution obtained in the previous step as the new starting point. We terminate this loop when ϵ is sufficiently small. It follows from Theorem 3.3 that the solution of Problem $\tilde{\mathrm{P}}_{p,\epsilon,\vartheta}$ at this stage is a good approximation for the solution of Problem $\tilde{\mathrm{P}}_p$.

The method described above is summarized in the following algorithm.

Algorithm 3.1. Input $\epsilon_{\min} > 0$, $\epsilon_0 > \epsilon_{\min}$, $\vartheta_{\max} > 0$, $\vartheta_0 < \vartheta_{\max}$, and $(\boldsymbol{\theta}^0, \boldsymbol{\sigma}^0) \in \Theta \times \Xi$.

(i) Initialize $\epsilon_0 \to \epsilon$ and $\vartheta_0 \to \vartheta$.

(ii) Using $(\boldsymbol{\theta}^0, \boldsymbol{\sigma}^0)$ as the initial guess, solve Problem $\tilde{\mathrm{P}}_{p,\epsilon,\vartheta}$. Let $(\boldsymbol{\theta}^{\epsilon,\vartheta,*}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*})$ denote the solution obtained.

(iii) If $(\boldsymbol{\theta}^{\epsilon,\vartheta,*}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*}) \in \Lambda$, then go to Step (iv). Otherwise, go to Step (v).

(iv) If $\epsilon > \epsilon_{\min}$, then set $\epsilon/10 \to \epsilon$ and $(\boldsymbol{\theta}^{\epsilon,\vartheta,*}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*}) \to (\boldsymbol{\theta}^0, \boldsymbol{\sigma}^0)$ and go to Step (ii). Otherwise, stop; take $(\boldsymbol{\theta}^{\epsilon,\vartheta,*}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*})$ as the optimal solution of Problem $\tilde{\mathrm{P}}_p$.

(v) If $\vartheta < \vartheta_{\max}$, then set $10\vartheta \to \vartheta$ and go to Step (ii). Otherwise, stop; ϑ is too large (in this case, the feasible region of Problem $\tilde{\mathrm{P}}_p$ is probably empty).

3.6 Convergence results

Problem $\tilde{\mathrm{P}}_p$ can be solved using Algorithm 3.1, after which a suboptimal control for Problem P can be constructed according to Remark 3.3. By repeating these steps for increasing values of p, we obtain a sequence of suboptimal controls. A natural question to ask is: does this sequence converge to an optimal control? This important question is the focus of this section.

Since p is no longer fixed, we now denote Γ by Γ^p, Ξ by Ξ^p, Ω by Ω^p, and $\mathcal{I}_k(\boldsymbol{\nu})$ by $\mathcal{I}_k^p(\boldsymbol{\nu})$. Our first result is given below.

Theorem 3.4. *Let $\boldsymbol{u} \in \mathcal{U}$ be an admissible control for Problem P. Then there exists a sequence $\{(\boldsymbol{\nu}^p, \boldsymbol{\sigma}^p)\}_{p=2}^{\infty}$, where $(\boldsymbol{\nu}^p, \boldsymbol{\sigma}^p) \in \Gamma^p \times \Xi^p$, such that $\boldsymbol{u}^p(\cdot|\boldsymbol{\nu}^p, \boldsymbol{\sigma}^p) \to \boldsymbol{u}$ uniformly on $[0, T)$ as $p \to \infty$.*

Proof. Let $\boldsymbol{u} \in \mathcal{U}$ be arbitrary but fixed. Furthermore, let $\{\tau_l\}_{l=0}^d \subset [0, T]$ be a finite set containing all of $\boldsymbol{u}$'s discontinuities (such a set exists because $\boldsymbol{u}$ is piecewise continuous). Without loss of generality, we assume that $\tau_0 = 0$, $\tau_d = T$, and

$$\tau_{l-1} < \tau_l, \qquad l = 1, \ldots, d.$$

Now, let $\{\boldsymbol{\nu}^p\}_{p=2}^{\infty}$ be a sequence of vectors with the following properties:

(i) For each $p \geq 2$, $\boldsymbol{\nu}^p = [t_1^p, \ldots, t_{p-1}^p]^T \in \Gamma^p$ and $t_{k-1}^p < t_k^p$, $k = 1, \ldots, p$;*

(ii) For each $p \geq d$, there exists corresponding integers $\kappa(p, l)$, $l = 1, \ldots, d-1$, such that

$$\tau_l = t_{\kappa(p,l)}^p, \qquad l = 1, \ldots, d-1.$$

(iii) $\max\limits_{1 \leq k \leq p} (t_k^p - t_{k-1}^p) \to 0$ as $p \to \infty$.

It is easy to see that such a sequence exists. Note that when $p \geq d$, each τ_l, $l = 1, \ldots, d-1$, coincides with one of the switching times in the vector $\boldsymbol{\nu}^p$ (property (ii)).

*Note that $t_0^p = 0$ and $t_p^p = T$.

Now, for each integer $p \geq 2$, define

$$\boldsymbol{\sigma}^{p,k} \triangleq \frac{1}{t_k^p - t_{k-1}^p} \int_{t_{k-1}^p}^{t_k^p} \boldsymbol{u}(\eta) d\eta, \qquad k = 1, \ldots, p,$$

and

$$\boldsymbol{\sigma}^p \triangleq (\boldsymbol{\sigma}^{p,1}, \ldots, \boldsymbol{\sigma}^{p,p}).$$

Since $\boldsymbol{u}$ is an admissible control, $\boldsymbol{\sigma}^p \in \Xi^p$ for each $p \geq 2$. Hence, $(\boldsymbol{\nu}^p, \boldsymbol{\sigma}^p) \in \Gamma^p \times \Xi^p$ for each $p \geq 2$. By Lemma 6.4.1 of [100], $\boldsymbol{u}^p(\cdot|\boldsymbol{\nu}^p, \boldsymbol{\sigma}^p) \to \boldsymbol{u}$ almost everywhere on $[0, T]$ as $p \to \infty$. We will show that this convergence is actually uniform on $[0, T)$.

Let $\delta > 0$ be arbitrary but fixed. Since $\boldsymbol{u}$ is continuous from the right, the restriction of $\boldsymbol{u}$ to $[\tau_{l-1}, \tau_l)$, $l = 1, \ldots, d$, is uniformly continuous. Thus, there exists a real number $\omega > 0$ such that the following implication holds uniformly with respect to $l \in \{1, \ldots, d\}$:

$$\eta_1, \eta_2 \in [\tau_{l-1}, \tau_l) \text{ and } |\eta_1 - \eta_2| < \omega \quad \Longrightarrow \quad \big|\boldsymbol{u}(\eta_1) - \boldsymbol{u}(\eta_2)\big|_r < \delta. \tag{3.39}$$

Now, suppose that $t \in [0, T)$. Clearly, $t \in [\tau_{l-1}, \tau_l)$ for some $l \in \{1, \ldots, d\}$. Moreover, for each integer $p \geq 2$, there exists a corresponding $k(p) \in \{1, \ldots, p\}$ such that $t \in \mathcal{I}_{k(p)}^p(\boldsymbol{\nu}^p)$. By property (ii) above,

$$\mathring{\mathcal{I}}_{k(p)}^p(\boldsymbol{\nu}^p) \subset [\tau_{l-1}, \tau_l), \qquad p \geq d, \tag{3.40}$$

where $\mathring{\mathcal{I}}_{k(p)}^p(\boldsymbol{\nu}^p)$ is the interior of $\mathcal{I}_{k(p)}^p(\boldsymbol{\nu}^p)$.

Furthermore, for each integer $p \geq 2$,

$$\begin{aligned} \big|\boldsymbol{u}^p(t|\boldsymbol{\nu}^p, \boldsymbol{\sigma}^p) - \boldsymbol{u}(t)\big| &= \big|\boldsymbol{\sigma}^{p,k(p)} - \boldsymbol{u}(t)\big| \\ &\leq \frac{1}{t_{k(p)}^p - t_{k(p)-1}^p} \int_{t_{k(p)-1}^p}^{t_{k(p)}^p} \big|\boldsymbol{u}(\eta) - \boldsymbol{u}(t)\big| d\eta \\ &= \frac{1}{t_{k(p)}^p - t_{k(p)-1}^p} \int_{\mathring{\mathcal{I}}_{k(p)}^p(\boldsymbol{\nu}^p)} \big|\boldsymbol{u}(\eta) - \boldsymbol{u}(t)\big| d\eta. \end{aligned} \tag{3.41}$$

By property (iii), there exists an integer $p' \geq 2$ such that

$$\max_{1 \leq k \leq p} (t_k^p - t_{k-1}^p) < \omega, \qquad p \geq p', \tag{3.42}$$

where ω is as defined in (3.39). Define

$$p'' \triangleq \max\{p', d, 2\}.$$

Then it follows from (3.42) that for each integer $p \geq p''$,

$$|\eta - t| \leq t^p_{k(p)} - t^p_{k(p)-1} < \omega, \qquad \eta \in \mathring{\mathcal{I}}^p_{k(p)}(\boldsymbol{\nu}^p). \tag{3.43}$$

In view of (3.40) and (3.43), we can invoke implication (3.39) to obtain

$$\big|\boldsymbol{u}(\eta) - \boldsymbol{u}(t)\big| < \delta, \qquad \eta \in \mathring{\mathcal{I}}^p_{k(p)}, \qquad p \geq p''.$$

Substituting this inequality into (3.41) gives

$$\big|\boldsymbol{u}^p(t|\boldsymbol{\nu}^p, \boldsymbol{\sigma}^p) - \boldsymbol{u}(t)\big| < \frac{1}{t^p_{k(p)} - t^p_{k(p)-1}} \int_{\mathring{\mathcal{I}}^p_{k(p)}(\boldsymbol{\nu}^p)} \delta d\eta = \delta, \qquad p \geq p''. \tag{3.44}$$

Since p'' is independent of $t \in [0, T)$, and δ was chosen arbitrarily, inequality (3.44) shows that $\boldsymbol{u}^p(\cdot|\boldsymbol{\nu}^p, \boldsymbol{\sigma}^p) \to \boldsymbol{u}$ uniformly on $[0, T)$ as $p \to \infty$. □

Before continuing, we recall the following two results from Chapter 6 of [100].

Lemma 3.1. *There exists a constant $L_2 > 0$ such that*

$$\big|\boldsymbol{x}(t|\boldsymbol{u})\big| \leq L_2, \qquad t \in [0, T], \qquad \boldsymbol{u} \in \mathcal{U}.$$

Lemma 3.2. *Let $\{\boldsymbol{u}^p\}_{p=2}^{\infty}$ be a sequence of admissible controls converging to $\boldsymbol{u} \in \mathcal{U}$ almost everywhere on $[0, T]$. Then the following two results hold:*

(i) $\boldsymbol{x}(\cdot|\boldsymbol{u}^p) \to \boldsymbol{x}(\cdot|\boldsymbol{u})$ *uniformly on* $[0, T]$ *as* $p \to \infty$*; and*

(ii) $G_0(\boldsymbol{u}^p) \to G_0(\boldsymbol{u})$ *as* $p \to \infty$.

Define the set

$$\Psi \triangleq \big\{\, \boldsymbol{v} \in \mathbb{R}^n : \ |\boldsymbol{v}| \leq L_2 \,\big\},$$

where L_2 is the constant from Lemma 3.1. Furthermore, let $\mathring{\mathcal{F}}$ denote the set consisting of all admissible controls $\boldsymbol{u} \in \mathcal{U}$ such that

$$\inf_{t \in [0,T)} h_i\big(\boldsymbol{x}(t|\boldsymbol{u}), \boldsymbol{u}(t)\big) > 0, \qquad i = 1, \ldots, q.$$

Clearly, $\mathring{\mathcal{F}} \subset \mathcal{F}$.

We assume that the following regularity condition is satisfied.

Assumption 3.4. If $\boldsymbol{u}^* \in \mathcal{F}$ is an optimal control for Problem P, then there exists a corresponding $\bar{\boldsymbol{u}} \in \mathring{\mathcal{F}}$ such that

$$\zeta \bar{\boldsymbol{u}} + (1 - \zeta)\boldsymbol{u}^* \in \mathring{\mathcal{F}}, \qquad \zeta \in (0, 1].$$

Similar assumptions are made in [48,100,101,106]. We are now ready to prove the following important convergence result.

Theorem 3.5. *Suppose that $\boldsymbol{u}^*$ is an optimal control for Problem P. Furthermore, for each integer $p \geq 2$, let $\boldsymbol{u}^{p,*}$ denote the suboptimal control constructed from the solution of Problem P_p according to Remark 3.2. Then*

$$\lim_{p\to\infty} G_0(\boldsymbol{u}^{p,*}) = G_0(\boldsymbol{u}^*).$$

Proof. By Assumption 3.4, there exists a $\bar{\boldsymbol{u}} \in \mathring{\mathcal{F}}$ such that

$$\bar{\boldsymbol{u}}^j \triangleq \boldsymbol{u}^* + \frac{1}{j}(\bar{\boldsymbol{u}} - \boldsymbol{u}^*) \in \mathring{\mathcal{F}}, \qquad j \geq 1. \tag{3.45}$$

Hence, for each integer $j \geq 1$, there is a corresponding real number $\upsilon_j > 0$ such that

$$h_i\big(\boldsymbol{x}(t|\bar{\boldsymbol{u}}^j), \bar{\boldsymbol{u}}^j(t)\big) \geq \upsilon_j, \qquad t \in [0,T), \qquad i = 1,\ldots,q. \tag{3.46}$$

We now temporarily fix j. Let $\{(\bar{\boldsymbol{\nu}}^{j,p}, \bar{\boldsymbol{\sigma}}^{j,p})\}_{p=2}^{\infty}$ denote the sequence from Theorem 3.4 corresponding to the admissible control $\bar{\boldsymbol{u}}^j$. For convenience, we will write $\bar{\boldsymbol{u}}^{j,p}$ instead of $\boldsymbol{u}^p(\cdot|\bar{\boldsymbol{\nu}}^{j,p}, \bar{\boldsymbol{\sigma}}^{j,p})$. Observe the following:

(i) $\bar{\boldsymbol{u}}^{j,p}(t) \in \mathcal{W}$ and $\bar{\boldsymbol{u}}^j(t) \in \mathcal{W}$ for each $t \in [0,T]$ (definition of $\mathcal{U}$);

(ii) $\boldsymbol{x}(t|\bar{\boldsymbol{u}}^{j,p}) \in \Psi$ and $\boldsymbol{x}(t|\bar{\boldsymbol{u}}^j) \in \Psi$ for each $t \in [0,T]$ (Lemma 3.1);

(iii) $\bar{\boldsymbol{u}}^{j,p} \to \bar{\boldsymbol{u}}^j$ uniformly on $[0,T)$ as $p \to \infty$ (Theorem 3.4);

(iv) $\boldsymbol{x}(\cdot|\bar{\boldsymbol{u}}^{j,p}) \to \boldsymbol{x}(\cdot|\bar{\boldsymbol{u}}^j)$ uniformly on $[0,T]$ as $p \to \infty$ (part (i) of Lemma 3.2); and

(v) The functions h_i, $i = 1,\ldots,q$, are uniformly continuous on the compact set $\Psi \times \mathcal{W}$ (Assumption 3.3).

These facts imply the existence an integer $p'_j \geq 2$ such that for each $p \geq p'_j$,

$$\big|h_i(\boldsymbol{x}(t|\bar{\boldsymbol{u}}^{j,p}), \bar{\boldsymbol{u}}^{j,p}(t)) - h_i(\boldsymbol{x}(t|\bar{\boldsymbol{u}}^j), \bar{\boldsymbol{u}}^j(t))\big| < \frac{\upsilon_j}{2}, \quad t \in [0,T), \quad i = 1,\ldots,q. \tag{3.47}$$

It follows from inequalities (3.46) and (3.47) that for each integer $p \geq p'_j$,

$$h_i\big(\boldsymbol{x}(t|\bar{\boldsymbol{u}}^{j,p}), \bar{\boldsymbol{u}}^{j,p}(t)\big) > \frac{\upsilon_j}{2}, \qquad t \in [0,T), \qquad i = 1,\ldots,q. \tag{3.48}$$

Furthermore, by part (ii) of Lemma 3.2, there exists another integer $p''_j \geq 2$ such that

$$\big|G_0(\bar{\boldsymbol{u}}^{j,p}) - G_0(\bar{\boldsymbol{u}}^j)\big| < \frac{1}{j}, \qquad p \geq p''_j, \tag{3.49}$$

Define

$$p_j \triangleq \max\{p_j', p_j''\}.$$

Inequality (3.48) shows that $\bar{\boldsymbol{u}}^{j,p_j} \in \mathcal{F}$. Hence, $(\bar{\boldsymbol{\nu}}^{j,p_j}, \bar{\boldsymbol{\sigma}}^{j,p_j}) \in \Omega^{p_j}$.

Now, let $\delta > 0$. We see from (3.45) that $\bar{\boldsymbol{u}}^j \to \boldsymbol{u}^*$ pointwise on $[0,T]$ as $j \to \infty$. Hence, by part (ii) of Lemma 3.2, there exists an integer $j' \geq 1$ such that

$$\left|G_0(\bar{\boldsymbol{u}}^j) - G_0(\boldsymbol{u}^*)\right| < \frac{\delta}{2}, \qquad j \geq j'. \tag{3.50}$$

Choose a fixed integer $j \geq \max\{j', 2/\delta\}$. Then it follows from (3.49) and (3.50) that

$$\begin{aligned}\left|G_0^{p_j}(\bar{\boldsymbol{\nu}}^{j,p_j}, \bar{\boldsymbol{\sigma}}^{j,p_j}) - G_0(\boldsymbol{u}^*)\right| &= \left|G_0(\bar{\boldsymbol{u}}^{j,p_j}) - G_0(\boldsymbol{u}^*)\right| \\ &\leq \left|G_0(\bar{\boldsymbol{u}}^{j,p_j}) - G_0(\bar{\boldsymbol{u}}^j)\right| + \left|G_0(\bar{\boldsymbol{u}}^j) - G_0(\boldsymbol{u}^*)\right| \\ &< \delta. \end{aligned} \tag{3.51}$$

Now, suppose that $p \geq p_j$. Clearly,

$$G_0(\boldsymbol{u}^*) \leq G_0(\boldsymbol{u}^{p,*}) \leq G_0(\boldsymbol{u}^{p_j,*}).$$

Hence,

$$G_0(\boldsymbol{u}^*) \leq G_0^p(\boldsymbol{\nu}^{p,*}, \boldsymbol{\sigma}^{p,*}) \leq G_0^{p_j}(\boldsymbol{\nu}^{p_j,*}, \boldsymbol{\sigma}^{p_j,*}), \tag{3.52}$$

where $(\boldsymbol{\nu}^{p,*}, \boldsymbol{\sigma}^{p,*})$ and $(\boldsymbol{\nu}^{p_j,*}, \boldsymbol{\sigma}^{p_j,*})$ are optimal solutions of Problems P_p and P_{p_j}, respectively. Since $(\bar{\boldsymbol{\nu}}^{j,p_j}, \bar{\boldsymbol{\sigma}}^{j,p_j}) \in \Omega^{p_j}$, inequality (3.52) gives

$$G_0(\boldsymbol{u}^*) \leq G_0(\boldsymbol{u}^{p,*}) \leq G_0^{p_j}(\bar{\boldsymbol{\nu}}^{j,p_j}, \bar{\boldsymbol{\sigma}}^{j,p_j}). \tag{3.53}$$

Finally, combining (3.51) and (3.53) yields

$$G_0(\boldsymbol{u}^*) \leq G_0(\boldsymbol{u}^{p,*}) \leq G_0^{p_j}(\bar{\boldsymbol{\nu}}^{j,p_j}, \bar{\boldsymbol{\sigma}}^{j,p_j}) < G_0(\boldsymbol{u}^*) + \delta.$$

Since $\delta > 0$ was chosen arbitrarily, this inequality shows that $G_0(\boldsymbol{u}^{p,*}) \to G_0(\boldsymbol{u}^*)$ as $p \to \infty$. □

Theorem 3.5 states that the costs of the suboptimal controls converge to the minimum cost as $p \to \infty$. Although there is no guarantee that the controls themselves converge, we do have the following result.

Theorem 3.6. *Let $\boldsymbol{u}^*$ and $\boldsymbol{u}^{p,*}$ be as defined in Theorem 3.5, and suppose that $\{\boldsymbol{u}^{p,*}\}_{p=2}^{\infty}$ converges almost everywhere on $[0,T]$ to an admissible control $\hat{\boldsymbol{u}}$. Then $\hat{\boldsymbol{u}}$ is an optimal control for Problem P.*

Proof. From part (ii) of Lemma 3.2, we have

$$\lim_{p\to\infty} G_0(\boldsymbol{u}^{p,*}) = G_0(\hat{\boldsymbol{u}}).$$

Therefore, by Theorem 3.5,

$$G_0(\hat{\boldsymbol{u}}) = G_0(\boldsymbol{u}^*).$$

It remains to show that $\hat{\boldsymbol{u}}$ is a feasible control. Suppose, to the contrary, that $\hat{\boldsymbol{u}}$ is infeasible. Then there exists an integer $\iota \in \{1, \ldots, q\}$ and a time point $\eta \in [0, T)$ such that

$$h_\iota\big(\boldsymbol{x}(\eta|\hat{\boldsymbol{u}}), \hat{\boldsymbol{u}}(\eta)\big) < 0.$$

Since $\hat{\boldsymbol{u}}$ is continuous from the right, there exists an $\omega > 0$ such that

$$h_\iota\big(\boldsymbol{x}(t|\hat{\boldsymbol{u}}), \hat{\boldsymbol{u}}(t)\big) < 0, \qquad t \in [\eta, \eta + \omega). \tag{3.54}$$

For each integer $j \geq 1$, define

$$\mathcal{A}_j \triangleq \big\{\, t \in [0, T) :\ h_\iota\big(\boldsymbol{x}(t|\hat{\boldsymbol{u}}), \hat{\boldsymbol{u}}(t)\big) \leq -1/j \,\big\}.$$

Clearly, $\{\mathcal{A}_j\}_{j=1}^{\infty}$ is an increasing sequence of measurable sets. Thus, by Theorem D in Section 9 of [37],

$$M_L\bigg\{\bigcup_{j=1}^{\infty} \mathcal{A}_j\bigg\} = \lim_{j\to\infty} M_L(\mathcal{A}_j), \tag{3.55}$$

where $M_L(\cdot)$ denotes the Lebesgue measure.

Now, it follows from (3.54) that

$$[\eta, \eta + \omega) \subset \bigcup_{j=1}^{\infty} \mathcal{A}_j. \tag{3.56}$$

By combining (3.55) and (3.56), we obtain

$$\omega \leq \lim_{j\to\infty} M_L(\mathcal{A}_j).$$

Hence, there exists an integer $j' \geq 1$ such that

$$0 < \frac{\omega}{2} \leq M_L(\mathcal{A}_{j'}). \tag{3.57}$$

It follows immediately from the definition of $\mathcal{A}_{j'}$ that

$$h_\iota\big(\boldsymbol{x}(t|\hat{\boldsymbol{u}}), \hat{\boldsymbol{u}}(t)\big) \leq -\frac{1}{j'}, \qquad t \in \mathcal{A}_{j'}. \tag{3.58}$$

Observe the following:

(i) $\boldsymbol{u}^{p,*}(t) \in \mathcal{W}$ and $\hat{\boldsymbol{u}}(t) \in \mathcal{W}$ for each $t \in [0,T]$ (definition of $\mathcal{U}$);

(ii) $\boldsymbol{x}(t|\boldsymbol{u}^{p,*}) \in \Psi$ and $\boldsymbol{x}(t|\hat{\boldsymbol{u}}) \in \Psi$ for each $t \in [0,T]$ (Lemma 3.1);

(iii) There exists a set $\mathcal{C} \subset [0,T]$ of measure $M_L(\mathcal{C}) < M_L(\mathcal{A}_{j'})/2$ such that $\boldsymbol{u}^{p,*} \to \hat{\boldsymbol{u}}$ uniformly on $[0,T] \setminus \mathcal{C}$ as $p \to \infty$ (Egoroff's Theorem [5, 37]);

(iv) $\boldsymbol{x}(\cdot|\boldsymbol{u}^{p,*}) \to \boldsymbol{x}(\cdot|\hat{\boldsymbol{u}})$ uniformly on $[0,T]$ as $p \to \infty$ (part (i) of Lemma 3.2); and

(v) h_ι is uniformly continuous on $\Psi \times \mathcal{W}$.

These facts imply that there exists an integer $p' \geq 2$ such that

$$\left|h_\iota(\boldsymbol{x}(t|\boldsymbol{u}^{p',*}), \boldsymbol{u}^{p',*}(t)) - h_\iota(\boldsymbol{x}(t|\hat{\boldsymbol{u}}), \hat{\boldsymbol{u}}(t))\right| < \frac{1}{2j'}, \qquad t \in [0,T] \setminus \mathcal{C}. \tag{3.59}$$

Combining (3.58) and (3.59) gives

$$h_\iota\big(\boldsymbol{x}(t|\boldsymbol{u}^{p',*}), \boldsymbol{u}^{p',*}(t)\big) < -\frac{1}{2j'}, \qquad t \in \mathcal{A}_{j'} \setminus \mathcal{C}. \tag{3.60}$$

We now show that $\mathcal{A}_{j'} \setminus \mathcal{C}$ is non-empty. Indeed,

$$M_L(\mathcal{A}_{j'} \setminus \mathcal{C}) = M_L(\mathcal{A}_{j'}) - M_L(\mathcal{A}_{j'} \cap \mathcal{C}) \geq M_L(\mathcal{A}_{j'}) - M_L(\mathcal{C}) > M_L(\mathcal{A}_{j'})/2 > 0,$$

where the last inequality follows from (3.57). Hence, $\mathcal{A}_{j'} \setminus \mathcal{C}$ is a set of positive measure, and therefore cannot be empty. This means that inequality (3.60) contradicts the feasibility of $\boldsymbol{u}^{p',*}$. Thus, $\hat{\boldsymbol{u}} \in \mathcal{F}$ as required. □

Remark 3.4. Theorems 3.5 and 3.6 suggest the following method for solving Problem P. First, choose an integer $p \geq 2$ and solve Problem $\tilde{\mathrm{P}}_p$ using Algorithm 3.1. Then, double p and re-solve Problem $\tilde{\mathrm{P}}_p$, using the optimal solution from the previous step as the initial guess. Repeat this step until the change in the optimal value of the cost function is within a desired tolerance. A suboptimal control for Problem P can then be constructed according to Remark 3.3.

3.7 Numerical examples

For illustration, we consider two numerical examples.

3.7.1 Rayleigh's optimal control problem

The following optimal control problem appears in [31]: find a control $u : [0, 4.5] \to \mathbb{R}$ that minimizes the cost function

$$\int_0^{4.5} \left\{u^2(t) + x_1^2(t)\right\} dt$$

subject to the dynamics

$$\begin{aligned} \dot{x}_1(t) &= x_2(t), && t \in [0, 4.5], \\ \dot{x}_2(t) &= -x_1(t) + x_2(t)\big(1.4 - 0.14x_2^2(t)\big) + 4u(t), && t \in [0, 4.5], \end{aligned}$$

and

$$\begin{aligned} x_1(0) &= -5, \\ x_2(0) &= -5, \end{aligned}$$

and the continuous inequality constraint

$$-u(t) - \frac{1}{6}x_1(t) \geq 0, \qquad t \in [0, 4.5].$$

To solve this problem, we applied the discretization procedure described in Sections 3.3 and 3.4 (with $p = 10$). We then solved the resulting Problem $\tilde{\mathrm{P}}_p$ via Algorithm 3.1, with MISER3 used to solve Problem $\tilde{\mathrm{P}}_{p,\epsilon,\vartheta}$ in Step (ii). Note that MISER3 invokes NLPQLP (see [93]) to solve Problem $\tilde{\mathrm{P}}_{p,\epsilon,\vartheta}$ as a nonlinear programming problem. The gradients required by NLPQLP are generated automatically by MISER3 using the gradient formulae in Section 3.5 (see equations (3.35) and (3.36)).

The smoothing and penalty parameters in Algorithm 3.1 were initially selected as $\epsilon = 0.1$ and $\vartheta = 10.0$, respectively. They were subsequently adjusted according to Steps (iii)-(v) of Algorithm 3.1. Recall that for each value of ϵ, the penalty parameter is increased until the solution of Problem $\tilde{\mathrm{P}}_{p,\epsilon,\vartheta}$ is feasible for Problem $\tilde{\mathrm{P}}_p$. We terminated Algorithm 3.1 when $\epsilon = 1.0 \times 10^{-6}$ and $\vartheta = 1.0 \times 10^5$. The initial ϵ ($\epsilon = 0.1$) required a large value of ϑ to ensure feasibility, but after that ϑ hardly changed as ϵ was decreased.

The suboptimal control constructed from the final solution of Problem $\tilde{\mathrm{P}}_{p,\epsilon,\vartheta}$ is shown, along with the state variables and the continuous inequality constraint, in Figure 3.2. Note that the continuous inequality constraint is satisfied everywhere. Also note that only a small improvement (less than 1%) was obtained by re-solving the problem with $p = 20$.

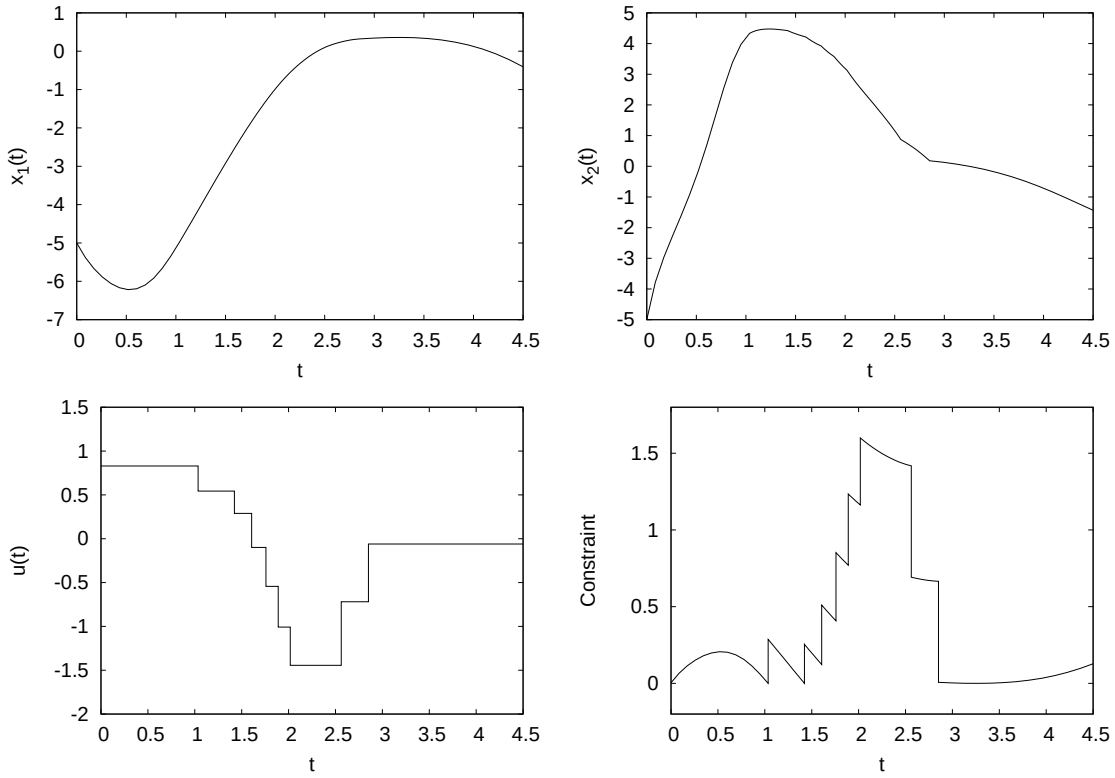

Figure 3.2: The optimal control in Example 3.7.1 and the corresponding state variables and constraint profile.

3.7.2 Optimal control of a container crane

The following optimal control problem is a modified version of the one in [91]: find control functions $u_1 : [0, 10] \to \mathbb{R}$ and $u_2 : [0, 10] \to \mathbb{R}$ that minimize

$$\frac{1}{2}\int_0^{10} \big\{x_3^2(t) + x_6^2(t)\big\}dt + (x_1(10) - 10)^2 + (x_2(10) - 14)^2 + x_3^2(10) \\ + (x_4(10) - 2)^2 + x_5^2(10) + x_6^2(10)$$

subject to the dynamics

$$\begin{aligned}
\dot{x}_1(t) &= x_4(t), && t \in [0, 10],\\
\dot{x}_2(t) &= x_5(t), && t \in [0, 10],\\
\dot{x}_3(t) &= x_6(t), && t \in [0, 10],\\
\dot{x}_4(t) &= u_1(t) + 17.27x_3(t), && t \in [0, 10],\\
\dot{x}_5(t) &= u_2(t), && t \in [0, 10],\\
\dot{x}_6(t) &= -\frac{1}{x_2(t)}\big(u_1(t) + 27.08x_3(t) + 2x_5(t)x_6(t)\big), && t \in [0, 10],
\end{aligned}$$

and

$$\begin{aligned}
x_1(0) &= 0,\\
x_2(0) &= 22,\\
x_3(0) &= 0,\\
x_4(0) &= 0,\\
x_5(0) &= -0.85,\\
x_6(0) &= 0,
\end{aligned}$$

and the continuous inequality constraints

$$\begin{aligned}
-u_1(t) - 17.27x_3(t) + 10\big(2 - x_4(t)\big) &\geq 0, \qquad t \in [0, 10],\\
u_1(t) + 17.27x_3(t) + 10\big(2 + x_4(t)\big) &\geq 0, \qquad t \in [0, 10],\\
-u_2(t) + 10\big(0.85 + x_5(t)\big) &\geq 0, \qquad t \in [0, 10],\\
u_2(t) + 10\big(0.85 - x_5(t)\big) &\geq 0, \qquad t \in [0, 10],
\end{aligned}$$

and the control constraints

$$\begin{aligned}
-2.83 \leq u_1(t) &\leq 2.83, \qquad t \in [0, 10],\\
-0.20 \leq u_2(t) &\leq 0.71, \qquad t \in [0, 10].
\end{aligned}$$

The dynamics in this problem model the motion of a sea container as it is transported via crane from a cargo ship to a truck (or vice versa). The cost function penalizes the container swing angle (large container swings are dangerous).

As in Example 3.7.1, we discretized this optimal control problem to obtain Problem $\tilde{\mathrm{P}}_p$ (with $p = 10$). We then solved Problem $\tilde{\mathrm{P}}_p$ via Algorithm 3.1, with MISER3 used to solve Problem $\tilde{\mathrm{P}}_{p,\epsilon,\vartheta}$ in Step (ii). Initially, $\epsilon = 0.1$ and $\vartheta = 1.0$; Algorithm 3.1 was terminated when $\epsilon = 1.0 \times 10^{-5}$ and $\vartheta = 1.0 \times 10^4$. A large value of ϑ was required initially, but once feasibility was attained, ϑ did not change as ϵ was decreased. The optimal control and optimal state variables are shown in Figures 3.3 and 3.4, respectively.

3.8 Conclusion

In this chapter, we developed a new computational method for solving nonlinear optimal control problems with continuous inequality constraints. This method has several major advantages over the ϵ-τ algorithm discussed in [100, 101]. In particular, it is capable of handling continuous inequality constraints that include the control function explicitly.

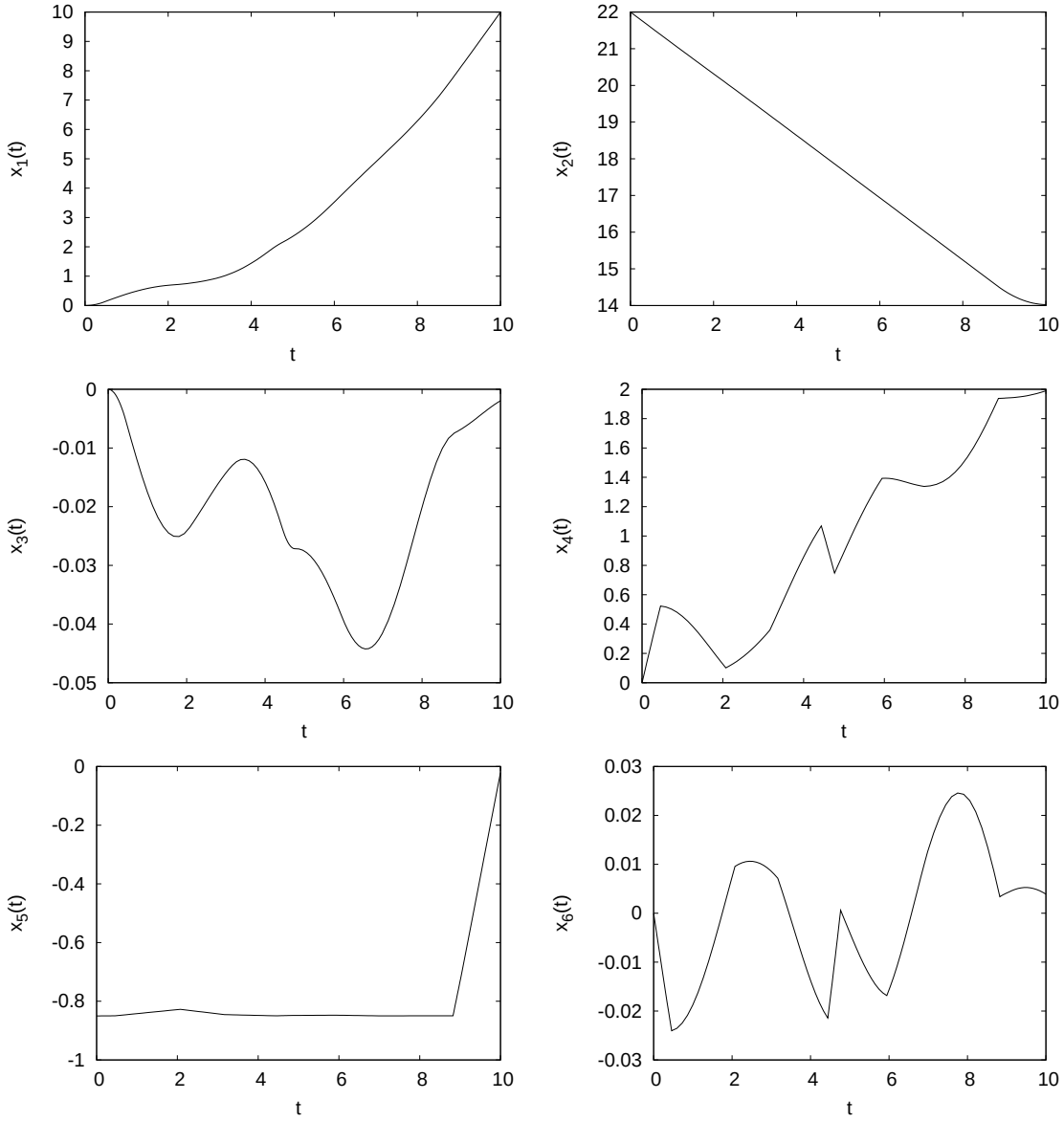

Figure 3.3: The optimal state variables for Example 3.7.2.

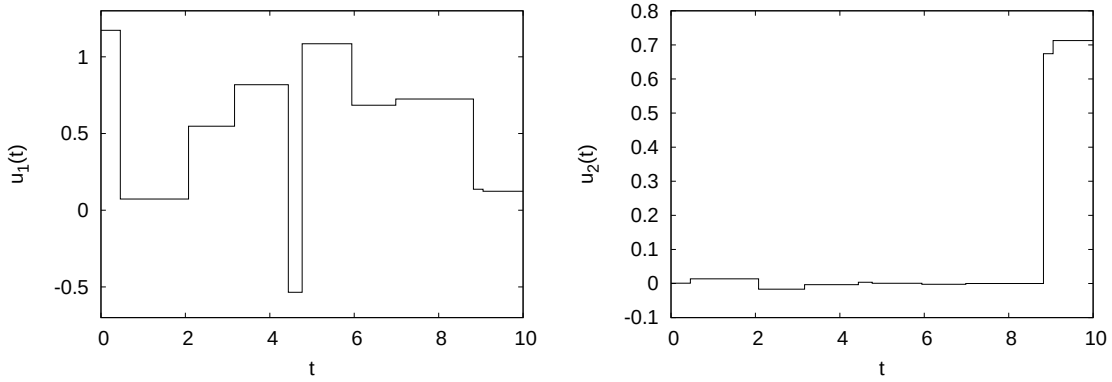

Figure 3.4: The optimal controls for Example 3.7.2.

Since the constraint functions (3.3) are potentially discontinuous in time, deriving the convergence results in Section 3.6 was difficult (continuity is exploited in [100, 101] to prove that the ϵ-τ algorithm converges). Nevertheless, Theorem 3.5 guarantees that the costs of the suboptimal controls converge to the optimal cost. Furthermore, Theorem 3.6 ensures that if the sequence of suboptimal controls converges to an admissible control almost everywhere, then this admissible control is optimal.

We point out that the admissible controls in this chapter are restricted to piecewise continuous functions. In [100, 101] (and also in Chapter 2), the controls are selected from a larger class of bounded measurable functions. The arguments used to establish Theorem 6.1 are not valid in this more general setting. Nevertheless, realistic control inputs are invariably piecewise continuous.

CHAPTER 4

Optimal control of a switched-capacitor DC-DC power converter*

4.1 Introduction

A *DC-DC power converter* is an electrical circuit that transforms a DC input voltage, which is supplied by a battery, into a different DC output voltage. In this chapter, we will consider *switched-capacitor DC-DC power converters*, which are constructed primarily from capacitors and switches [17,18,46]. Because of their small size and high power density, switched-capacitor DC-DC power converters are ideal voltage transformers for mobile electronic appliances such as laptop computers, cellular phones, and portable gaming consoles. As such, their popularity has soared over the past decade.

The capacitors in a switched-capacitor DC-DC power converter are capable of both storing and supplying energy. Whether a particular capacitor absorbs energy from the source or delivers energy to the load depends on the switch configuration. The switch configuration—and therefore the circuit topology—is actually changed periodically while the power converter operates. This ensures that each capacitor alternates between energy absorb mode and energy supply mode. To explain further, at any given time, some of the capacitors are supplying energy to the load, while the others are absorbing energy from the source. When the circuit topology is changed (via the switch configuration), these roles are reversed: the capacitors that were previously supplying energy begin to charge up, and the capacitors that were previously absorbing energy begin to discharge.

If a particular circuit topology is active for too long, then the capacitors connected to the load will run out of energy, and the power converter's output voltage will drop considerably. This must be avoided, because large variations in the output voltage can damage the devices attached to the power converter. Hence, the circuit topology should be changed frequently. However, because each topology switch causes an energy leak, excessive switching is very inefficient [3]. It is therefore imperative that the topology switching times be chosen judiciously, so that switching energy losses and output voltage

*This chapter is based on [74].

variation are curbed.

Many methods have been proposed for controlling the switching mechanism in a switched-capacitor DC-DC power converter—see, for example, [16, 27, 52, 63] and the references cited therein. Most of these methods are based on a linear time-invariant dynamic model, which is derived via averaging and/or linearization. However, although its individual topologies are linear, a switched-capacitor DC-DC power converter is actually a strongly nonlinear system. This is because it does not remain in one particular topology, but rather switches between several of them. Moreover, the voltage across each of its capacitors drops suddenly when the topology is changed. Conventional control methods ignore such behavior, and consequently their performance is only guaranteed under a so-called *small signal assumption*.

The problem of determining the topology switching times a priori was recently formulated in [42] as an optimal control problem. This optimal control problem can be solved using the software package MISER3 [47]. The major advantage of this approach is that it avoids averaging and linearization. In fact, the power converter is modeled as a *switched system* that switches between several subsystems of differential equations—one for each topology—during the time horizon. The optimization and control of switched systems is currently an active research area with a multitude of interesting applications [9, 14, 40, 94, 115, 132]. In Chapter 5 we will consider another important optimal control problem in this area.

The optimal control problem formulated in [42] has a cost function that penalizes two quantities: (i) the output voltage variation; and (ii) the output voltage sensitivity with respect to uncertainties and disturbances. The output voltage sensitivity is calculated via several complex formulae (one of which stretches over ten lines). To evaluate these formulae, the eigenvalues of certain matrices in the governing switched system need to be derived *analytically* as functions of the load resistance. This is usually a very difficult task; in fact, it is impossible if the matrices have dimension greater than four. This is a serious restriction, and hence it is imperative that a superior method be developed for solving the optimal control problem formulated in [42]. The purpose of this chapter is to develop such a method.

4.2 Problem formulation

We consider a switched-capacitor DC-DC power converter with n capacitors and $m \geq 2$ distinct circuit topologies. Each of these topologies is active once during the operating period $[0, T]$, where $T > 0$ is a given terminal time.

Let t_k, $k = 1, \ldots, m-1$, denote the times at which the power converter changes its topology. These times are called *switching times*. We assume that the difference between each pair of consecutive switching times is at least $\tau > 0$ (the power converter cannot

switch arbitrarily fast). Consequently, the switching times must satisfy the following constraints:

$$t_{k-1} + \tau \leq t_k, \qquad k = 1, \ldots, m, \tag{4.1}$$

where $t_0 \triangleq 0$ and $t_m \triangleq T$. Let Γ denote the set of all vectors $\boldsymbol{\nu} = [t_1, \ldots, t_{m-1}]^T \in \mathbb{R}^{m-1}$ that satisfy (4.1). Each element of Γ is called a *switching-time vector*.

For each $i = 1, \ldots, n$, let $x_i(t)$ denote the voltage across the ith capacitor at time t. The vector $\boldsymbol{x}(t) \in \mathbb{R}^n$, whose ith component is $x_i(t)$, is called the *state voltage vector*. Since each capacitor's initial voltage is fixed,

$$\boldsymbol{x}(0) = \boldsymbol{x}(0^+) = \boldsymbol{x}^0, \tag{4.2}$$

where $\boldsymbol{x}^0 \in \mathbb{R}^n$ is a given vector.

Topology switches are accompanied by sudden energy losses. Consequently, the state voltage vector changes instantaneously at each switching time:

$$\boldsymbol{x}(t_k) = \boldsymbol{x}(t_k^+) = \boldsymbol{x}(t_k^-) + \boldsymbol{z}^k\big(\boldsymbol{x}(t_k^-)\big), \qquad k = 1, \ldots, m-1, \tag{4.3}$$

where $\boldsymbol{z}^k : \mathbb{R}^n \to \mathbb{R}^n$, $k = 1, \ldots, m-1$, are given functions.

The kth circuit topology is active from $t = t_{k-1}$ to $t = t_k$. During this period, the state voltage is governed by the following dynamic system:

$$\dot{\boldsymbol{x}}(t) = A_k(R_L)\boldsymbol{x}(t) + B_k(R_L)\boldsymbol{\sigma}, \qquad t \in (t_{k-1}, t_k), \tag{4.4}$$

where $\boldsymbol{\sigma} \in \mathbb{R}^r$ is the input voltage vector (whose components are the DC input voltages); $R_L \in \mathbb{R}$ is the load resistance; and $A_k : \mathbb{R} \to \mathbb{R}^{n\times n}$ and $B_k : \mathbb{R} \to \mathbb{R}^{n\times r}$ are given functions of the load resistance. These functions are derived using Kirchhoff's voltage laws.

The power converter's output voltage at time t is given by

$$y(t) = C_k(R_L)\boldsymbol{x}(t) + D_k(R_L)\boldsymbol{\sigma}, \qquad t \in \mathcal{I}_k, \qquad k = 1, \ldots, m, \tag{4.5}$$

where $C_k : \mathbb{R} \to \mathbb{R}^{1\times n}$ and $D_k : \mathbb{R} \to \mathbb{R}^{1\times r}$ are given functions of the load resistance and the subintervals $\mathcal{I}_k$, $k = 1, \ldots, m$, are defined as

$$\mathcal{I}_k \triangleq \begin{cases} [t_{k-1}, t_k), & \text{if } k \in \{1, \ldots, m-1\}, \\ [t_{k-1}, t_k], & \text{if } k = m. \end{cases}$$

Equations (4.2)-(4.5) can be combined to form the following *switched system* with m

subsystems:

$$\dot{\boldsymbol{x}}(t) = A_k(R_L)\boldsymbol{x}(t) + B_k(R_L)\boldsymbol{\sigma}, \qquad t \in (t_{k-1}, t_k), \qquad k = 1, \ldots, m, \tag{4.6}$$

$$y(t) = C_k(R_L)\boldsymbol{x}(t) + D_k(R_L)\boldsymbol{\sigma}, \qquad t \in \mathcal{I}_k, \qquad k = 1, \ldots, m, \tag{4.7}$$

and

$$\boldsymbol{x}(t_k) = \boldsymbol{x}(t_k^+) = \begin{cases} \boldsymbol{x}^0, & \text{if } k = 0, \qquad (4.8a) \\ \boldsymbol{x}(t_k^-) + \boldsymbol{z}^k\big(\boldsymbol{x}(t_k^-)\big), & \text{if } k \in \{1, \ldots, m-1\}. \qquad (4.8b) \end{cases}$$

Let $\boldsymbol{x}(\cdot|\boldsymbol{\nu})$ and $y(\cdot|\boldsymbol{\nu})$ denote the solution of (4.6)-(4.8) corresponding to the switching-time vector $\boldsymbol{\nu} = [t_1, \ldots, t_{m-1}]^T \in \Gamma$.

The state voltage $\boldsymbol{x}(\cdot|\boldsymbol{\nu})$ evolves as follows. It starts at $\boldsymbol{x}^0$ at time $t = 0$ and is governed by equation (4.6) with $k = 1$ until time $t = t_1$. The power converter then switches to the second topology, which causes the state voltage to change instantaneously from $\boldsymbol{x}(t_1^-)$ to $\boldsymbol{x}(t_1^+)$—see equation (4.8b). Restarting from $\boldsymbol{x}(t_1^+)$, the state voltage evolves smoothly according to equation (4.6) with $k = 2$ until $t = t_2$, at which time the power converter switches to the third topology and the state voltage jumps once more. The state voltage continues to evolve in this way for the remainder of the time horizon.

We define the *output voltage ripple* as the difference between the maximum and minimum output voltage. That is, for each switching-time vector $\boldsymbol{\nu} \in \Gamma$, the output voltage ripple is

$$\sup_{t\in[0,T]} y(t|\boldsymbol{\nu}) - \inf_{t\in[0,T]} y(t|\boldsymbol{\nu}). \tag{4.9}$$

Obviously, a small ripple indicates that the power converter's output voltage is steady. Hence, the switching times should be chosen to minimize (4.9).

The output voltage should also be robust with respect to variations in the input voltage and/or load resistance; otherwise, changing the devices attached to the power converter could cause the output voltage to fluctuate. Accordingly, the switching times should be chosen to minimize

$$\sup_{t\in[0,T]} \left| \frac{\partial y(t|\boldsymbol{\nu})}{\partial R_L} \right| \tag{4.10}$$

and

$$\sup_{t\in[0,T]} \left| \frac{\partial y(t|\boldsymbol{\nu})}{\partial \boldsymbol{\sigma}} \right|_\infty, \tag{4.11}$$

where $|\cdot|_\infty$ denotes the infinity norm in $\mathbb{R}^r$.

We now define the following optimal control problem, whose cost function is the weighted sum of (4.9)-(4.11).

Problem P. *Find a switching-time vector* $\boldsymbol{\nu} \in \Gamma$ *that minimizes the cost function*

$$G_0(\boldsymbol{\nu}) \triangleq \alpha\left\{ \sup_{t\in[0,T]} y(t|\boldsymbol{\nu}) - \inf_{t\in[0,T]} y(t|\boldsymbol{\nu}) \right\} + \beta \sup_{t\in[0,T]} \left| \frac{\partial y(t|\boldsymbol{\nu})}{\partial R_L} \right| + \gamma \sup_{t\in[0,T]} \left| \frac{\partial y(t|\boldsymbol{\nu})}{\partial \boldsymbol{\sigma}} \right|_\infty,$$

where $\alpha \geq 0$, $\beta \geq 0$, *and* $\gamma \geq 0$ *are given real numbers, over* Γ.

The constants α, β, and γ in Problem P are weights used to adjust the relative importance of each term in G_0. Before finishing this section, we make the following assumption.

Assumption 4.1. The given functions A_k, B_k, C_k, and D_k, $k = 1, \ldots, m$, and $\boldsymbol{z}^k$, $k = 1, \ldots, m-1$, are continuously differentiable.

4.3 Problem transformation

The cost function in Problem P is very unusual; it involves infimums, supremums, and the derivative of the output with respect to the load resistance and input voltage. Computing this function is extremely difficult. Furthermore, because it involves norms, it is a non-smooth function. Hence, Problem P cannot be solved using a gradient-based nonlinear programming algorithm. The aim of this section is to transform Problem P into an equivalent problem that is easier to solve.

Consider the following auxiliary switched system:

$$\dot{\boldsymbol{\psi}}(t) = \frac{\partial A_k(R_L)}{\partial R_L}\boldsymbol{x}(t|\boldsymbol{\nu}) + A_k(R_L)\boldsymbol{\psi}(t) + \frac{\partial B_k(R_L)}{\partial R_L}\boldsymbol{\sigma}, \qquad t \in (t_{k-1}, t_k), \qquad k = 1, \ldots, m, \tag{4.12}$$

and

$$\boldsymbol{\psi}(t_k) = \boldsymbol{\psi}(t_k^+) = \begin{cases} \mathbf{0}, & \text{if } k = 0, \quad (4.13a) \\ \boldsymbol{\psi}(t_k^-) + \dfrac{\partial \boldsymbol{z}^k\big(\boldsymbol{x}(t_k^-|\boldsymbol{\nu})\big)}{\partial \boldsymbol{x}}\boldsymbol{\psi}(t_k^-), & \text{if } k \in \{1, \ldots, m-1\}, \quad (4.13b) \end{cases}$$

where $\boldsymbol{\nu} = [t_1, \ldots, t_{m-1}]^T \in \Gamma$. Let $\boldsymbol{\psi}(\cdot|\boldsymbol{\nu})$ denote the solution of (4.12)-(4.13).

We now prove the following important result.

Theorem 4.1. *For each switching-time vector* $\boldsymbol{\nu} \in \Gamma$,

$$\frac{\partial \boldsymbol{x}(t|\boldsymbol{\nu})}{\partial R_L} = \boldsymbol{\psi}(t|\boldsymbol{\nu}), \qquad t \in [0, T].$$

Proof. Let $\boldsymbol{\nu} = [t_1, \ldots, t_{m-1}]^T \in \Gamma$ be arbitrary but fixed. It follows from (4.6) that for each $k = 1, \ldots, m$,

$$\boldsymbol{x}(t|\boldsymbol{\nu}) = \boldsymbol{x}(t_{k-1}^+|\boldsymbol{\nu}) + \int_{t_{k-1}}^{t} \big(A_k(R_L)\boldsymbol{x}(\eta|\boldsymbol{\nu}) + B_k(R_L)\boldsymbol{\sigma}\big)d\eta, \qquad t \in (t_{k-1}, t_k). \tag{4.14}$$

Differentiating this equation with respect to R_L gives

$$\frac{\partial \boldsymbol{x}(t|\boldsymbol{\nu})}{\partial R_L} = \frac{\partial \boldsymbol{x}(t_{k-1}^+|\boldsymbol{\nu})}{\partial R_L} + \int_{t_{k-1}}^{t} \frac{\partial A_k(R_L)}{\partial R_L}\boldsymbol{x}(\eta|\boldsymbol{\nu})d\eta + \int_{t_{k-1}}^{t} \left\{ A_k(R_L)\frac{\partial \boldsymbol{x}(\eta|\boldsymbol{\nu})}{\partial R_L} + \frac{\partial B_k(R_L)}{\partial R_L}\boldsymbol{\sigma} \right\} d\eta, \qquad t \in (t_{k-1}, t_k).$$

By differentiating this equation with respect to time, we obtain

$$\frac{d}{dt}\left\{ \frac{\partial \boldsymbol{x}(t|\boldsymbol{\nu})}{\partial R_L} \right\} = \frac{\partial A_k(R_L)}{\partial R_L}\boldsymbol{x}(t|\boldsymbol{\nu}) + A_k(R_L)\frac{\partial \boldsymbol{x}(t|\boldsymbol{\nu})}{\partial R_L} + \frac{\partial B_k(R_L)}{\partial R_L}\boldsymbol{\sigma}, \quad t \in (t_{k-1}, t_k). \tag{4.15}$$

Now, differentiating (4.8b) with respect to R_L gives

$$\frac{\partial \boldsymbol{x}(t_k|\boldsymbol{\nu})}{\partial R_L} = \frac{\partial \boldsymbol{x}(t_k^+|\boldsymbol{\nu})}{\partial R_L} = \frac{\partial \boldsymbol{x}(t_k^-|\boldsymbol{\nu})}{\partial R_L} + \frac{\partial \boldsymbol{z}^k\big(\boldsymbol{x}(t_k^-|\boldsymbol{\nu})\big)}{\partial \boldsymbol{x}}\frac{\partial \boldsymbol{x}(t_k^-|\boldsymbol{\nu})}{\partial R_L}, \quad k = 1, \ldots, m-1. \tag{4.16}$$

Furthermore, from (4.8a) we have

$$\frac{\partial \boldsymbol{x}(0|\boldsymbol{\nu})}{\partial R_L} = \frac{\partial \boldsymbol{x}(0^+|\boldsymbol{\nu})}{\partial R_L} = \frac{\partial}{\partial R_L}\{\boldsymbol{x}^0\} = \boldsymbol{0}. \tag{4.17}$$

Equations (4.15)-(4.17) show that $\partial \boldsymbol{x}(\cdot|\boldsymbol{\nu})/\partial R_L$ is a solution of (4.12)-(4.13). Since such a solution is unique (see [1, 2]), we must have

$$\frac{\partial \boldsymbol{x}(t|\boldsymbol{\nu})}{\partial R_L} = \boldsymbol{\psi}(t|\boldsymbol{\nu}), \qquad t \in [0, T],$$

as required. □

For each $j = 1, \ldots, r$, consider another auxiliary switched system as follows:

$$\dot{\boldsymbol{\phi}}^j(t) = A_k(R_L)\boldsymbol{\phi}^j(t) + B_{k,j}(R_L), \qquad t \in (t_{k-1}, t_k), \qquad k = 1, \ldots, m, \tag{4.18}$$

and

$$\boldsymbol{\phi}^j(t_k) = \boldsymbol{\phi}^j(t_k^+) = \begin{cases} \boldsymbol{0}, & \text{if } k = 0, \quad (4.19a) \\ \boldsymbol{\phi}^j(t_k^-) + \dfrac{\partial \boldsymbol{z}^k\big(\boldsymbol{x}(t_k^-|\boldsymbol{\nu})\big)}{\partial \boldsymbol{x}}\boldsymbol{\phi}^j(t_k^-), & \text{if } k \in \{1, \ldots, m-1\}, \quad (4.19b) \end{cases}$$

where $\boldsymbol{\nu} = [t_1, \ldots, t_{m-1}]^T \in \Gamma$ and $B_{k,j}(R_L)$ denotes the jth column of the matrix $B_k(R_L)$. Let $\boldsymbol{\phi}^j(\cdot|\boldsymbol{\nu})$ denote the solution of (4.18)-(4.19).

The following result is the analogue of Theorem 4.1 for the derivative of the state voltage with respect to $\boldsymbol{\sigma}$.

Theorem 4.2. *For each switching-time vector* $\boldsymbol{\nu} \in \Gamma$,

$$\frac{\partial \boldsymbol{x}(t|\boldsymbol{\nu})}{\partial \sigma_j} = \boldsymbol{\phi}^j(t|\boldsymbol{\nu}), \qquad t \in [0,T], \qquad j = 1,\ldots,r.$$

Proof. Let $\boldsymbol{\nu} = [t_1,\ldots,t_{m-1}]^T \in \Gamma$ and $j \in \{1,\ldots,r\}$ be arbitrary but fixed. Recall equation (4.14) from the proof of Theorem 4.1:

$$\boldsymbol{x}(t|\boldsymbol{\nu}) = \boldsymbol{x}(t_{k-1}^+|\boldsymbol{\nu}) + \int_{t_{k-1}}^{t} \big(A_k(R_L)\boldsymbol{x}(\eta|\boldsymbol{\nu}) + B_k(R_L)\boldsymbol{\sigma}\big)d\eta, \qquad t \in (t_{k-1},t_k),$$

where $k = 1,\ldots,m$. Differentiating this equation with respect to σ_j gives

$$\frac{\partial \boldsymbol{x}(t|\boldsymbol{\nu})}{\partial \sigma_j} = \frac{\partial \boldsymbol{x}(t_{k-1}^+|\boldsymbol{\nu})}{\partial \sigma_j} + \int_{t_{k-1}}^{t} \left\{ A_k(R_L)\frac{\partial \boldsymbol{x}(\eta|\boldsymbol{\nu})}{\partial \sigma_j} + B_{k,j}(R_L)\right\}d\eta, \qquad t \in (t_{k-1},t_k).$$

By differentiating this equation with respect to time, we obtain

$$\frac{d}{dt}\left\{\frac{\partial \boldsymbol{x}(t|\boldsymbol{\nu})}{\partial \sigma_j}\right\} = A_k(R_L)\frac{\partial \boldsymbol{x}(t|\boldsymbol{\nu})}{\partial \sigma_j} + B_{k,j}(R_L), \qquad t \in (t_{k-1},t_k). \tag{4.20}$$

Now, differentiating (4.8b) with respect to σ_j gives

$$\frac{\partial \boldsymbol{x}(t_k|\boldsymbol{\nu})}{\partial \sigma_j} = \frac{\partial \boldsymbol{x}(t_k^+|\boldsymbol{\nu})}{\partial \sigma_j} = \frac{\partial \boldsymbol{x}(t_k^-|\boldsymbol{\nu})}{\partial \sigma_j} + \frac{\partial \boldsymbol{z}^k\big(\boldsymbol{x}(t_k^-|\boldsymbol{\nu})\big)}{\partial \boldsymbol{x}}\frac{\partial \boldsymbol{x}(t_k^-|\boldsymbol{\nu})}{\partial \sigma_j}, \quad k = 1,\ldots,m-1. \tag{4.21}$$

Furthermore, from (4.8a) we have

$$\frac{\partial \boldsymbol{x}(0|\boldsymbol{\nu})}{\partial \sigma_j} = \frac{\partial \boldsymbol{x}(0^+|\boldsymbol{\nu})}{\partial \sigma_j} = \frac{\partial}{\partial \sigma_j}\big\{\boldsymbol{x}^0\big\} = \boldsymbol{0}. \tag{4.22}$$

Equations (4.20)-(4.22) show that $\partial \boldsymbol{x}(\cdot|\boldsymbol{\nu})/\partial \sigma_j$ is the unique solution of (4.18)-(4.19). Hence,

$$\frac{\partial \boldsymbol{x}(t|\boldsymbol{\nu})}{\partial \sigma_j} = \boldsymbol{\phi}^j(t|\boldsymbol{\nu}), \qquad t \in [0,T].$$

This completes the proof. □

For each switching-time vector $\boldsymbol{\nu} \in \Gamma$, define corresponding functions $u(\cdot|\boldsymbol{\nu}) : [0,T] \to \mathbb{R}$ and $w_j(\cdot|\boldsymbol{\nu}) : [0,T] \to \mathbb{R}$, $j = 1,\ldots,r$, as follows:

$$u(t|\boldsymbol{\nu}) \triangleq \frac{\partial C_k(R_L)}{\partial R_L}\boldsymbol{x}(t|\boldsymbol{\nu}) + C_k(R_L)\boldsymbol{\psi}(t|\boldsymbol{\nu}) + \frac{\partial D_k(R_L)}{\partial R_L}\boldsymbol{\sigma}, \quad t \in \mathcal{I}_k, \quad k = 1,\ldots,m, \tag{4.23}$$

and

$$w_j(t|\boldsymbol{\nu}) \triangleq C_k(R_L)\boldsymbol{\phi}^j(t|\boldsymbol{\nu}) + D_{k,j}(R_L), \quad t \in \mathcal{I}_k, \quad k = 1,\ldots,m, \tag{4.24}$$

where $D_{k,j}(R_L)$ denotes the jth column of the matrix $D_k(R_L)$.

Now, differentiating (4.7) with respect to R_L gives

$$\frac{\partial y(t|\boldsymbol{\nu})}{\partial R_L} = \frac{\partial C_k(R_L)}{\partial R_L}\boldsymbol{x}(t|\boldsymbol{\nu}) + C_k(R_L)\frac{\partial \boldsymbol{x}(t|\boldsymbol{\nu})}{\partial R_L} + \frac{\partial D_k(R_L)}{\partial R_L}\boldsymbol{\sigma}, \qquad t \in \mathcal{I}_k, \qquad k = 1, \ldots, m.$$

By substituting the result of Theorem 4.1 into this equation and then comparing it with equation (4.23), we see that

$$\frac{\partial y(t|\boldsymbol{\nu})}{\partial R_L} = u(t|\boldsymbol{\nu}), \qquad t \in [0, T]. \tag{4.25}$$

Similarly, by Theorem 4.2,

$$\frac{\partial y(t|\boldsymbol{\nu})}{\partial \sigma_j} = w_j(t|\boldsymbol{\nu}), \qquad t \in [0, T], \qquad j = 1, \ldots, r. \tag{4.26}$$

Hence, we can readily calculate the derivative of the output with respect to the load resistance and input voltage by solving the auxiliary systems (4.12)-(4.13) and (4.18)-(4.19). These auxiliary systems must be solved *simultaneously* with (4.6)-(4.8), because the state voltage vector appears in their right-hand sides. Thus, to calculate the output voltage sensitivity, we first integrate the expanded switched system consisting of (4.6)-(4.8), (4.12)-(4.13), (4.18)-(4.19), and then substitute the solution into (4.23)-(4.24). This method is very convenient and straightforward. In contrast, the method proposed in [42] is extremely tedious: it uses several complicated formulas, one of which has five nested summations stretching over ten lines.

We now define a new optimization problem as follows.

Problem Q. *Find a pair* $(\boldsymbol{\nu}, \boldsymbol{\zeta}) \in \Gamma \times \mathbb{R}^4$ *that minimizes the cost function*

$$J_0(\boldsymbol{\zeta}) \triangleq \alpha\zeta_1 + \alpha\zeta_2 + \beta\zeta_3 + \gamma\zeta_4,$$

where $\alpha \geq 0$, $\beta \geq 0$, *and* $\gamma \geq 0$, *are given real numbers, subject to the constraints*

$$y(t|\boldsymbol{\nu}) \leq \zeta_1, \qquad t \in [0, T], \tag{4.27a}$$

$$-y(t|\boldsymbol{\nu}) \leq \zeta_2, \qquad t \in [0, T], \tag{4.27b}$$

$$-\zeta_3 \leq u(t|\boldsymbol{\nu}) \leq \zeta_3, \qquad t \in [0, T], \tag{4.27c}$$

$$-\zeta_4 \leq w_j(t|\boldsymbol{\nu}) \leq \zeta_4, \qquad t \in [0, T], \qquad j = 1, \ldots, r. \tag{4.27d}$$

It turns out that Problem Q is equivalent to Problem P. We state this formally as the following theorem.

Theorem 4.3. *Let $\boldsymbol{\nu}^* \in \Gamma$ be a switching-time vector and define*

$$\zeta_1^* \triangleq \sup_{t\in[0,T]} y(t|\boldsymbol{\nu}^*), \tag{4.28a}$$

$$\zeta_2^* \triangleq -\inf_{t\in[0,T]} y(t|\boldsymbol{\nu}^*), \tag{4.28b}$$

$$\zeta_3^* \triangleq \sup_{t\in[0,T]} \left|\frac{\partial y(t|\boldsymbol{\nu}^*)}{\partial R_L}\right|, \tag{4.28c}$$

$$\zeta_4^* \triangleq \sup_{t\in[0,T]} \left|\frac{\partial y(t|\boldsymbol{\nu}^*)}{\partial \boldsymbol{\sigma}}\right|_\infty. \tag{4.28d}$$

Then $(\boldsymbol{\nu}^*, \boldsymbol{\zeta}^*)$ is optimal for Problem Q if and only if $\boldsymbol{\nu}^*$ is optimal for Problem P.

Proof. First, note that

$$J_0(\boldsymbol{\zeta}^*) = G_0(\boldsymbol{\nu}^*). \tag{4.29}$$

Now, suppose that $\boldsymbol{\nu}^* \in \Gamma$ is an optimal solution of Problem P. By equations (4.25), (4.26), and (4.28),

$$\begin{aligned}
y(t|\boldsymbol{\nu}^*) &\le \zeta_1^*, && t \in [0,T],\\
-y(t|\boldsymbol{\nu}^*) &\le \zeta_2^*, && t \in [0,T],\\
-\zeta_3^* \le u(t|\boldsymbol{\nu}^*) = \frac{\partial y(t|\boldsymbol{\nu}^*)}{\partial R_L} &\le \zeta_3^*, && t \in [0,T],\\
-\zeta_4^* \le w_j(t|\boldsymbol{\nu}^*) = \frac{\partial y(t|\boldsymbol{\nu}^*)}{\partial \sigma_j} &\le \zeta_4^*, && t \in [0,T], \qquad j = 1,\ldots,r.
\end{aligned}$$

These inequalities show that $(\boldsymbol{\nu}^*, \boldsymbol{\zeta}^*)$ satisfies the constraints (4.27), and is therefore feasible for Problem Q. We will show that it is optimal.

Let $(\boldsymbol{\nu}, \boldsymbol{\zeta}) \in \Gamma \times \mathbb{R}^4$ be an arbitrary feasible pair for Problem Q. Then

$$\begin{aligned}
y(t|\boldsymbol{\nu}) &\le \zeta_1, && t \in [0,T],\\
-y(t|\boldsymbol{\nu}) &\le \zeta_2, && t \in [0,T],\\
-\zeta_3 \le u(t|\boldsymbol{\nu}) = \frac{\partial y(t|\boldsymbol{\nu})}{\partial R_L} &\le \zeta_3, && t \in [0,T],\\
-\zeta_4 \le w_j(t|\boldsymbol{\nu}) = \frac{\partial y(t|\boldsymbol{\nu})}{\partial \sigma_j} &\le \zeta_4, && t \in [0,T], \qquad j = 1,\ldots,r.
\end{aligned}$$

These inequalities show that $y(t|\boldsymbol{\nu})$ is bounded above by ζ_1; $y(t|\boldsymbol{\nu})$ is bounded below by $-\zeta_2$; $|\partial y(t|\boldsymbol{\nu})/\partial R_L|$ is bounded above by ζ_3; and $|\partial y(t|\boldsymbol{\nu})/\partial \boldsymbol{\sigma}|_\infty$ is bounded above

by ζ_4. Hence,

$$G_0(\boldsymbol{\nu}) = \alpha\left\{ \sup_{t\in[0,T]} y(t|\boldsymbol{\nu}) - \inf_{t\in[0,T]} y(t|\boldsymbol{\nu}) \right\} + \beta \sup_{t\in[0,T]} \left| \frac{\partial y(t|\boldsymbol{\nu})}{\partial R_L} \right| + \gamma \sup_{t\in[0,T]} \left| \frac{\partial y(t|\boldsymbol{\nu})}{\partial \boldsymbol{\sigma}} \right|_\infty$$
$$\leq \alpha\zeta_1 + \alpha\zeta_2 + \beta\zeta_3 + \gamma\zeta_4.$$

Since $\boldsymbol{\nu}^*$ is optimal for Problem P,

$$G_0(\boldsymbol{\nu}^*) \leq G_0(\boldsymbol{\nu}) \leq \alpha\zeta_1 + \alpha\zeta_2 + \beta\zeta_3 + \gamma\zeta_4 = J_0(\boldsymbol{\zeta}). \tag{4.30}$$

Combining equation (4.29) and inequality (4.30) gives

$$J_0(\boldsymbol{\zeta}^*) \leq J_0(\boldsymbol{\zeta}).$$

Since $(\boldsymbol{\nu}, \boldsymbol{\zeta}) \in \Gamma \times \mathbb{R}^4$ was chosen arbitrarily, this inequality shows that $(\boldsymbol{\nu}^*, \boldsymbol{\zeta}^*)$ is optimal for Problem Q.

Conversely, let $(\boldsymbol{\nu}^*, \boldsymbol{\zeta}^*) \in \Gamma \times \mathbb{R}^4$ be an optimal solution for Problem Q and suppose that $\boldsymbol{\nu}^*$ is *not* optimal for Problem P. Then there exists a $\boldsymbol{\nu}' \in \Gamma$ such that

$$G_0(\boldsymbol{\nu}') < G_0(\boldsymbol{\nu}^*). \tag{4.31}$$

Define a vector $\boldsymbol{\zeta}' \in \mathbb{R}^4$ as follows:

$$\begin{aligned}
\zeta_1' &\triangleq \sup_{t\in[0,T]} y(t|\boldsymbol{\nu}'), \\
\zeta_2' &\triangleq -\inf_{t\in[0,T]} y(t|\boldsymbol{\nu}'), \\
\zeta_3' &\triangleq \sup_{t\in[0,T]} \left| \frac{\partial y(t|\boldsymbol{\nu}')}{\partial R_L} \right|, \\
\zeta_4' &\triangleq \sup_{t\in[0,T]} \left| \frac{\partial y(t|\boldsymbol{\nu}')}{\partial \boldsymbol{\sigma}} \right|_\infty.
\end{aligned}$$

Clearly,

$$J_0(\boldsymbol{\zeta}') = G_0(\boldsymbol{\nu}'). \tag{4.33}$$

Furthermore, similar arguments to those used in the first part of the proof show that $(\boldsymbol{\nu}', \boldsymbol{\zeta}')$ is feasible for Problem Q. Combining equations (4.29) and (4.33) with inequality (4.31) gives

$$J_0(\boldsymbol{\zeta}') < J_0(\boldsymbol{\zeta}^*).$$

Since $(\boldsymbol{\nu}', \boldsymbol{\zeta}')$ is feasible for Problem Q, this contradicts the optimality of $(\boldsymbol{\nu}^*, \boldsymbol{\zeta}^*)$. Hence, $\boldsymbol{\nu}^*$ must be an optimal switching-time vector for Problem P. □

In contrast with Problem P, Problem Q has a very simple cost function. However, Prob-

lem Q is still difficult to solve for two reasons: (i) its switching times are decision variables; and (ii) its constraints (4.27) must be satisfied at *every* point in the time horizon. Recall from Chapter 3 that such constraints are called *continuous inequality constraints*.

We now apply the time-scaling transformation to Problem Q. As in Chapters 2 and 3, the time-scaling transformation maps the variable switching times to fixed points in a new time horizon, thereby eliminating the first difficulty mentioned in the previous paragraph.

First, let

$$\Theta \triangleq \left\{ \boldsymbol{\theta} \in \mathbb{R}^m : \ \theta_k \geq \tau, \ k = 1, \ldots, m; \ \theta_1 + \cdots + \theta_m = T \right\}.$$

For each $\boldsymbol{\theta} \in \Theta$, define a corresponding function $\mu(\cdot|\boldsymbol{\theta}) : [0, m] \rightarrow \mathbb{R}$ by

$$\mu(s|\boldsymbol{\theta}) \triangleq \begin{cases} \displaystyle\sum_{l=1}^{\lfloor s \rfloor} \theta_l + \theta_{\lfloor s \rfloor + 1}(s - \lfloor s \rfloor), & \text{if } s \in [0, m), \\ T, & \text{if } s = m, \end{cases}$$

where $\lfloor \cdot \rfloor$ denotes the floor function.

It is easy to see that $\mu(\cdot|\boldsymbol{\theta})$ is continuous and strictly increasing on $[0, m]$. Moreover,

$$\mu(0|\boldsymbol{\theta}) = 0$$

and

$$\mu(m|\boldsymbol{\theta}) = T.$$

Hence, $\mu(\cdot|\boldsymbol{\theta})$ is a bijection from $[0, m]$ onto $[0, T]$. The time-scaling transformation involves making the following change of variable:

$$t = \mu(s|\boldsymbol{\theta}). \tag{4.34}$$

In the new time horizon $[0, m]$, the switching times occur at the fixed points $s = k$, $k = 1, \ldots, m$. Hence,

$$t_k = \mu(k|\boldsymbol{\theta}) = \sum_{l=1}^{k} \theta_l, \qquad k = 0, \ldots, m. \tag{4.35}$$

Clearly,

$$\left[\mu(1|\boldsymbol{\theta}), \ldots, \mu(m-1|\boldsymbol{\theta})\right]^T \in \Gamma.$$

Moreover, *every* switching-time vector in Γ is generated by an element of Θ in this way.

Let

$$\begin{aligned}\tilde{\boldsymbol{x}}(s) &\triangleq \boldsymbol{x}(\mu(s|\boldsymbol{\theta})),\\ \tilde{\boldsymbol{\psi}}(s) &\triangleq \boldsymbol{\psi}(\mu(s|\boldsymbol{\theta})),\\ \tilde{\boldsymbol{\phi}}^j(s) &\triangleq \boldsymbol{\phi}^j(\mu(s|\boldsymbol{\theta})), \qquad j=1,\dots,r.\end{aligned}$$

By differentiating these equations with respect to s, and using (4.6), (4.12), and (4.18), we obtain

$$\left.\begin{aligned}\dot{\tilde{\boldsymbol{x}}}(s) &= \theta_k A_k(R_L)\tilde{\boldsymbol{x}}(s) + \theta_k B_k(R_L)\boldsymbol{\sigma}\\ \dot{\tilde{\boldsymbol{\psi}}}(s) &= \theta_k \frac{\partial A_k(R_L)}{\partial R_L}\tilde{\boldsymbol{x}}(s) + \theta_k A_k(R_L)\tilde{\boldsymbol{\psi}}(s) + \theta_k \frac{\partial B_k(R_L)}{\partial R_L}\boldsymbol{\sigma}\\ \dot{\tilde{\boldsymbol{\phi}}}^j(s) &= \theta_k A_k(R_L)\tilde{\boldsymbol{\phi}}^j(s) + \theta_k B_{k,j}(R_L), \quad j=1,\dots,r,\end{aligned}\right\} \quad s\in(k-1,k), \tag{4.36}$$

where $k=1,\dots,m$.

It follows from (4.8b), (4.13b), and (4.19b) that

$$\left.\begin{aligned}\tilde{\boldsymbol{x}}(k^+) &= \tilde{\boldsymbol{x}}(k^-) + \boldsymbol{z}^k(\tilde{\boldsymbol{x}}(k^-)),\\ \tilde{\boldsymbol{\psi}}(k^+) &= \tilde{\boldsymbol{\psi}}(k^-) + \frac{\partial \boldsymbol{z}^k(\tilde{\boldsymbol{x}}(k^-))}{\partial \boldsymbol{x}}\tilde{\boldsymbol{\psi}}(k^-),\\ \tilde{\boldsymbol{\phi}}^j(k^+) &= \tilde{\boldsymbol{\phi}}^j(k^-) + \frac{\partial \boldsymbol{z}^k(\tilde{\boldsymbol{x}}(k^-))}{\partial \boldsymbol{x}}\tilde{\boldsymbol{\phi}}^j(k^-), \quad j=1,\dots,r.\end{aligned}\right\} \quad k=1,\dots,m-1. \tag{4.37}$$

Using (4.8a), (4.13a), and (4.19a), we obtain the initial conditions

$$\left.\begin{aligned}\tilde{\boldsymbol{x}}(0) &= \boldsymbol{x}^0,\\ \tilde{\boldsymbol{\psi}}(0) &= \boldsymbol{0},\\ \tilde{\boldsymbol{\phi}}^j(0) &= \boldsymbol{0}, \quad j=1,\dots,r.\end{aligned}\right\} \tag{4.38}$$

Let $\tilde{\boldsymbol{x}}(\cdot|\boldsymbol{\theta})$, $\tilde{\boldsymbol{\psi}}(\cdot|\boldsymbol{\theta})$, and $\tilde{\boldsymbol{\phi}}^j(\cdot|\boldsymbol{\theta})$, $j=1,\dots,r$, denote the solutions of (4.36)-(4.38) corresponding to $\boldsymbol{\theta}\in\Theta$. We also define

$$\begin{aligned}\tilde{y}(s|\boldsymbol{\theta}) &\triangleq y(\mu(s|\boldsymbol{\theta})),\\ \tilde{u}(s|\boldsymbol{\theta}) &\triangleq u(\mu(s|\boldsymbol{\theta})),\\ \tilde{w}_j(s|\boldsymbol{\theta}) &\triangleq w_j(\mu(s|\boldsymbol{\theta})), \qquad j=1,\dots,r.\end{aligned}$$

Hence,

$$\left.\begin{aligned} \tilde{y}(s|\boldsymbol{\theta}) &= C_k(R_L)\tilde{\boldsymbol{x}}(s|\boldsymbol{\theta}) + D_k(R_L)\boldsymbol{\sigma}, \\ \tilde{u}(s|\boldsymbol{\theta}) &= \frac{\partial C_k(R_L)}{\partial R_L}\tilde{\boldsymbol{x}}(s|\boldsymbol{\theta}) + C_k(R_L)\tilde{\boldsymbol{\psi}}(s|\boldsymbol{\theta}) + \frac{\partial D_k(R_L)}{\partial R_L}\boldsymbol{\sigma}, \\ \tilde{w}_j(s|\boldsymbol{\theta}) &= C_k(R_L)\tilde{\boldsymbol{\phi}}^j(s|\boldsymbol{\theta}) + D_{k,j}(R_L), \end{aligned}\right\} \quad s \in [k-1, k), ^* \quad (4.39)$$

where $k = 1, \ldots, m$.

The constraints (4.27) become

$$\left.\begin{aligned} \tilde{y}(s|\boldsymbol{\theta}) &\le \zeta_1, \\ -\tilde{y}(s|\boldsymbol{\theta}) &\le \zeta_2, \\ -\zeta_3 \le \tilde{u}(s|\boldsymbol{\theta}) &\le \zeta_3, \\ -\zeta_4 \le \tilde{w}_j(s|\boldsymbol{\theta}) &\le \zeta_4, \quad j = 1, \ldots, r, \end{aligned}\right\} \quad s \in [0, m]. \quad (4.40)$$

We now define the following optimal control problem, which is equivalent to Problem Q.

Problem $\tilde{\mathrm{Q}}$. *Find a pair* $(\boldsymbol{\theta}, \boldsymbol{\zeta}) \in \Theta \times \mathbb{R}^4$ *that minimizes the cost function*

$$J_0(\boldsymbol{\zeta}) = \alpha\zeta_1 + \alpha\zeta_2 + \beta\zeta_3 + \gamma\zeta_4$$

subject to the continuous inequality constraints (4.40).

Problem $\tilde{\mathrm{Q}}$ is a semi-infinite programming problem with a linear cost function. Its solution $(\boldsymbol{\theta}^*, \boldsymbol{\zeta}^*) \in \Theta \times \mathbb{R}^4$ immediately furnishes the following optimal switching times for Problem Q:

$$t_k^* = \mu(k|\boldsymbol{\theta}^*) = \sum_{l=1}^{k} \theta_l^*, \qquad k = 1, \ldots, m-1.$$

In the next section, we will discuss a method for solving Problem $\tilde{\mathrm{Q}}$.

4.4 Solving Problem $\tilde{\mathrm{Q}}$

Notice that the constraints (4.40) are similar to the continuous inequality constraints considered in Chapter 3. In this section, we will see that the method used to solve Problem $\tilde{\mathrm{P}}_p$ in Chapter 3 is also applicable to Problem $\tilde{\mathrm{Q}}$.

We first rewrite (4.40) as follows:

$$h_i(s|\boldsymbol{\theta}, \boldsymbol{\zeta}) \ge 0, \qquad s \in [0, m], \qquad i = 1, \ldots, 2r+4, \quad (4.41)$$

*When $k = m$, this interval is $[m-1, m]$.

where

$$\begin{aligned}
h_1(s|\boldsymbol{\theta},\boldsymbol{\zeta}) &\triangleq -\tilde{y}(s|\boldsymbol{\theta}) + \zeta_1,\\
h_2(s|\boldsymbol{\theta},\boldsymbol{\zeta}) &\triangleq \tilde{y}(s|\boldsymbol{\theta}) + \zeta_2,\\
h_3(s|\boldsymbol{\theta},\boldsymbol{\zeta}) &\triangleq -\tilde{u}(s|\boldsymbol{\theta}) + \zeta_3,\\
h_4(s|\boldsymbol{\theta},\boldsymbol{\zeta}) &\triangleq \tilde{u}(s|\boldsymbol{\theta}) + \zeta_3,\\
h_{2j+3}(s|\boldsymbol{\theta},\boldsymbol{\zeta}) &\triangleq -\tilde{w}_j(s|\boldsymbol{\theta}) + \zeta_4, \qquad j = 1,\ldots,r,\\
h_{2j+4}(s|\boldsymbol{\theta},\boldsymbol{\zeta}) &\triangleq \tilde{w}_j(s|\boldsymbol{\theta}) + \zeta_4, \qquad j = 1,\ldots,r.
\end{aligned}$$

Next, we define the following optimization problem.

Problem $\tilde{\mathrm{Q}}_{\boldsymbol{\epsilon},\boldsymbol{\vartheta}}$. *Find a pair $(\boldsymbol{\theta},\boldsymbol{\zeta}) \in \Theta \times \mathbb{R}^4$ that minimizes the cost function*

$$\tilde{J}_{\epsilon,\vartheta}(\boldsymbol{\theta},\boldsymbol{\zeta}) \triangleq J_0(\boldsymbol{\zeta}) - \vartheta \sum_{i=1}^{2r+4} \int_0^m \varphi_\epsilon\big(h_i(s|\boldsymbol{\theta},\boldsymbol{\zeta})\big)ds,$$

where $\epsilon > 0$, $\vartheta > 0$, and $\varphi_\epsilon : \mathbb{R} \to \mathbb{R}$ is defined by

$$\varphi_\epsilon(\eta) \triangleq \begin{cases} \eta, & \text{if } \eta < -\epsilon,\\ -(\eta-\epsilon)^2/4\epsilon, & \text{if } -\epsilon \le \eta \le \epsilon,\\ 0, & \text{otherwise.}\end{cases}$$

Problem $\tilde{\mathrm{Q}}_{\epsilon,\vartheta}$ is a nonlinear programming problem with simple bounds on the variables and a linear equality constraint (recall the definition of the set Θ). Computing the gradient of each of these constraints is straightforward. Computing the gradient of $\tilde{J}_{\epsilon,\vartheta}$ is more difficult, but it can be done using the formulae reported in [47,69,125]. For completeness, we state these formulae below.

First, note that the dynamic system (4.36)-(4.38) can be written in the following form:

$$\dot{\boldsymbol{v}}(s) = \theta_k \hat{A}_k(R_L)\boldsymbol{v}(s) + \theta_k \hat{B}_k(R_L)\boldsymbol{\sigma} + \theta_k \hat{O}_k(R_L), \qquad s \in (k-1,k), \qquad k = 1,\ldots,m,$$

and

$$\boldsymbol{v}(k) = \boldsymbol{v}(k^+) = \begin{cases} \boldsymbol{v}^0, & \text{if } k = 0,\\ \boldsymbol{g}^k\big(\boldsymbol{v}(k^-)\big), & \text{if } k \in \{1,\ldots,m-1\},\end{cases}$$

where $\hat{A}_k(R_L) \in \mathbb{R}^{(2+r)n\times(2+r)n}$, $\hat{B}_k(R_L) \in \mathbb{R}^{(2+r)n\times r}$, $\hat{O}_k(R_L) \in \mathbb{R}^{(2+r)n}$, and

$$\boldsymbol{v}(s) \triangleq \Big[\tilde{\boldsymbol{x}}(s), \tilde{\boldsymbol{\psi}}(s), \tilde{\boldsymbol{\phi}}^1(s), \ldots, \tilde{\boldsymbol{\phi}}^r(s)\Big]^T \in \mathbb{R}^{(2+r)n}$$

and

$$\boldsymbol{v}^0 = \Big[\big(\boldsymbol{x}^0\big)^T, \mathbf{0}, \ldots, \mathbf{0}\Big]^T \in \mathbb{R}^{(2+r)n}.$$

Let $\boldsymbol{v}(\cdot|\boldsymbol{\theta})$ denote the solution of this system corresponding to $\boldsymbol{\theta} \in \Theta$. Hence,

$$\boldsymbol{v}(s|\boldsymbol{\theta}) = \Big[\tilde{\boldsymbol{x}}(s|\boldsymbol{\theta}), \tilde{\boldsymbol{\psi}}(s|\boldsymbol{\theta}), \tilde{\boldsymbol{\phi}}^1(s|\boldsymbol{\theta}), \ldots, \tilde{\boldsymbol{\phi}}^r(s|\boldsymbol{\theta})\Big]^T, \qquad s \in [0, m].$$

Define the *Hamiltonian function* $H : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^4 \to \mathbb{R}$ as follows:

$$H(s, \boldsymbol{v}, \boldsymbol{\lambda}, \boldsymbol{\theta}, \boldsymbol{\zeta}) \triangleq -\vartheta \sum_{i=1}^{2r+4} \varphi_\epsilon\big(h_i(s|\boldsymbol{\theta}, \boldsymbol{\zeta})\big) + \boldsymbol{\lambda}^T\big(\theta_k \hat{A}_k(R_L)\boldsymbol{v} + \theta_k \hat{B}_k(R_L)\boldsymbol{\sigma} + \theta_k \hat{O}_k(R_L)\big),$$
$$s \in [k-1, k), \quad k = 1, \ldots, m.$$

Furthermore, define the following auxiliary dynamic system:

$$\dot{\boldsymbol{\lambda}}(s) = -\left[\frac{\partial H\big(s, \boldsymbol{v}(s|\boldsymbol{\theta}), \boldsymbol{\lambda}(s), \boldsymbol{\theta}, \boldsymbol{\zeta}\big)}{\partial \boldsymbol{v}}\right]^T, \qquad s \in [0, m],$$

and

$$\boldsymbol{\lambda}(k^-) = \begin{cases} \mathbf{0}, & \text{if } k = m, \\ \left[\dfrac{\partial \boldsymbol{g}^k(\boldsymbol{v}(k^-|\boldsymbol{\theta}))}{\partial \boldsymbol{v}}\right]^T \boldsymbol{\lambda}(k^+), & \text{if } k \in \{1, \ldots, m-1\}, \end{cases}$$

where $(\boldsymbol{\theta}, \boldsymbol{\zeta}) \in \Theta \times \mathbb{R}^4$. This auxiliary system is called the *costate system*. Let $\boldsymbol{\lambda}(\cdot|\boldsymbol{\theta}, \boldsymbol{\zeta})$ denote the solution of the costate system corresponding to $(\boldsymbol{\theta}, \boldsymbol{\zeta}) \in \Theta \times \mathbb{R}^4$.

Now, it follows from Appendix C of [47] that for each $(\boldsymbol{\theta}, \boldsymbol{\zeta}) \in \Theta \times \mathbb{R}^4$,

$$\frac{\partial \tilde{J}_{\epsilon,\vartheta}(\boldsymbol{\theta}, \boldsymbol{\zeta})}{\partial \theta_k} = \int_0^m \frac{\partial H\big(s, \boldsymbol{v}(s|\boldsymbol{\theta}), \boldsymbol{\lambda}(s|\boldsymbol{\theta}, \boldsymbol{\zeta}), \boldsymbol{\theta}, \boldsymbol{\zeta}\big)}{\partial \theta_k}, \qquad k = 1, \ldots, m,$$

and

$$\frac{\partial \tilde{J}_{\epsilon,\vartheta}(\boldsymbol{\theta}, \boldsymbol{\zeta})}{\partial \zeta_\iota} = \frac{\partial J_0(\boldsymbol{\zeta})}{\partial \zeta_\iota} + \int_0^m \frac{\partial H\big(s, \boldsymbol{v}(s|\boldsymbol{\theta}), \boldsymbol{\lambda}(s|\boldsymbol{\theta}, \boldsymbol{\zeta}), \boldsymbol{\theta}, \boldsymbol{\zeta}\big)}{\partial \zeta_\iota} ds, \qquad \iota = 1, 2, 3, 4.$$

These formulae can be used in conjunction with a gradient-based nonlinear programming algorithm to solve Problem $\tilde{\mathrm{Q}}_{\epsilon,\vartheta}$.

The relationship between Problems $\tilde{\mathrm{Q}}_{\epsilon,\vartheta}$ and $\tilde{\mathrm{Q}}$ is the same as the relationship between Problems $\tilde{\mathrm{P}}_{p,\epsilon,\vartheta}$ and $\tilde{\mathrm{P}}_p$ in Chapter 3. Thus, Problem $\tilde{\mathrm{Q}}_{\epsilon,\vartheta}$ is a good approximation of Problem $\tilde{\mathrm{Q}}$ when ϵ is small and ϑ is large. More precisely, the following analogues of Theorems 3.2 and 3.3 hold:

- For each $\epsilon > 0$, there exists a corresponding $\vartheta(\epsilon) > 0$ such that if $\vartheta > \vartheta(\epsilon)$, then the solution of Problem $\tilde{\mathrm{Q}}_{\epsilon,\vartheta}$ is feasible for Problem $\tilde{\mathrm{Q}}$.

- If $(\boldsymbol{\theta}^*, \boldsymbol{\zeta}^*)$ is a solution of Problem $\tilde{\mathrm{Q}}$ and $(\boldsymbol{\theta}^{\epsilon,\vartheta,*}, \boldsymbol{\zeta}^{\epsilon,\vartheta,*})$ is a solution of Problem $\tilde{\mathrm{Q}}_{\epsilon,\vartheta}$, then
$$J_0(\boldsymbol{\zeta}^{\epsilon,\vartheta,*}) \to J_0(\boldsymbol{\zeta}^*) \quad \text{as} \quad \epsilon \to 0,$$
provided that for each ϵ, $(\boldsymbol{\theta}^{\epsilon,\vartheta,*}, \boldsymbol{\zeta}^{\epsilon,\vartheta,*})$ is feasible for Problem $\tilde{\mathrm{Q}}$ (that is, for each ϵ, the penalty parameter ϑ is sufficiently large).

By virtue of these results, we can solve Problem $\tilde{\mathrm{Q}}$ by repeatedly solving Problem $\tilde{\mathrm{Q}}_{\epsilon,\vartheta}$ for decreasing values of ϵ. See Section 3.5 for more details.

4.5 Existence of an optimal solution

Thus far, we have tacitly assumed that Problem P has an optimal solution. In this section, we will vindicate this assumption.

First, we present two preliminary lemmas, which follow readily from Assumption 4.1 and Lemmas 6.4.2 and 6.4.3 in [100].

Lemma 4.1. *There exists a real number $L_1 > 0$ such that*

$$\left.\begin{aligned} \left|\tilde{y}(s|\boldsymbol{\theta})\right| &\le L_1, \\ \left|\tilde{u}(s|\boldsymbol{\theta})\right| &\le L_1, \\ \left|\tilde{w}_j(s|\boldsymbol{\theta})\right| &\le L_1, \quad j = 1, \dots, r, \end{aligned}\right\} \qquad s \in [0, m], \qquad \boldsymbol{\theta} \in \Theta.$$

Lemma 4.2. *Suppose that $\{\boldsymbol{\theta}^p\}_{p=1}^{\infty} \subset \Theta$ is a sequence converging to $\boldsymbol{\theta} \in \Theta$. Then*

$$\left.\begin{aligned} \lim_{p\to\infty} \tilde{y}(s|\boldsymbol{\theta}^p) &= \tilde{y}(s|\boldsymbol{\theta}), \\ \lim_{p\to\infty} \tilde{u}(s|\boldsymbol{\theta}^p) &= \tilde{u}(s|\boldsymbol{\theta}), \\ \lim_{p\to\infty} \tilde{w}_j(s|\boldsymbol{\theta}^p) &= \tilde{w}_j(s|\boldsymbol{\theta}), \quad j = 1, \dots, r, \end{aligned}\right\} \qquad s \in [0, m].$$

We now state and prove the main result of this section.

Theorem 4.4. *Problem P has an optimal solution.*

Proof. Recall that Problems P and $\tilde{\mathrm{Q}}$ are equivalent. Hence, it is sufficient to prove that Problem $\tilde{\mathrm{Q}}$ has an optimal solution.

Let $\mathcal{F}$ be the set consisting of all pairs $(\boldsymbol{\theta}, \boldsymbol{\zeta}) \in \Theta \times \mathbb{R}^4$ that satisfy the constraints (4.40). In other words, $\mathcal{F}$ is the feasible region for Problem $\tilde{\mathrm{Q}}$.

Now, let $\eta \in [0, m]$ and $\kappa \in \{1, \dots, r\}$ be arbitrary. For each pair $(\boldsymbol{\theta}, \boldsymbol{\zeta}) \in \mathcal{F}$,

$$J_0(\boldsymbol{\zeta}) = \alpha\zeta_1 + \alpha\zeta_2 + \beta\zeta_3 + \gamma\zeta_4 \ge \alpha\tilde{y}(\eta|\boldsymbol{\theta}) - \alpha\tilde{y}(\eta|\boldsymbol{\theta}) + \beta\left|\tilde{u}(\eta|\boldsymbol{\theta})\right| + \gamma\left|\tilde{w}_\kappa(\eta|\boldsymbol{\theta})\right| \ge 0.$$

This shows that the set

$$\left\{\, J_0(\boldsymbol{\zeta}) :\ (\boldsymbol{\theta}, \boldsymbol{\zeta}) \in \mathcal{F} \,\right\} \subset \mathbb{R}$$

is bounded below by zero. Hence, we can find a real number $\omega \geq 0$ that satisfies

$$\omega = \inf \big\{ J_0(\boldsymbol{\zeta}) : \ (\boldsymbol{\theta}, \boldsymbol{\zeta}) \in \mathcal{F} \big\}.$$

Therefore, for each integer $p \geq 1$, there exists a corresponding pair $(\boldsymbol{\theta}^p, \boldsymbol{\zeta}^p) \in \mathcal{F}$ such that

$$J_0(\boldsymbol{\zeta}^p) < \omega + \frac{1}{p}.$$

Clearly,

$$\lim_{p \to \infty} J_0(\boldsymbol{\zeta}^p) = \omega. \tag{4.42}$$

Now, for each integer $p \geq 1$, define another vector $\bar{\boldsymbol{\zeta}}^p \in \mathbb{R}^4$ as follows:

$$\begin{aligned}
\bar{\zeta}_1^p &\triangleq \sup_{s \in [0,m]} \tilde{y}(s|\boldsymbol{\theta}^p), \\
\bar{\zeta}_2^p &\triangleq - \inf_{s \in [0,m]} \tilde{y}(s|\boldsymbol{\theta}^p), \\
\bar{\zeta}_3^p &\triangleq \sup_{s \in [0,m]} \big|\tilde{u}(s|\boldsymbol{\theta}^p)\big|, \\
\bar{\zeta}_4^p &\triangleq \sup_{s \in [0,m]} \max_{1 \leq j \leq r} \big|\tilde{w}_j(s|\boldsymbol{\theta}^p)\big|.
\end{aligned}$$

It is easy to see that

$$(\boldsymbol{\theta}^p, \bar{\boldsymbol{\zeta}}^p) \in \mathcal{F}, \qquad p \geq 1. \tag{4.43}$$

Furthermore, it follows from (4.40) that ζ_1^p is an upper bound for $\tilde{y}(s|\boldsymbol{\theta}^p)$, $-\zeta_2^p$ is a lower bound for $\tilde{y}(s|\boldsymbol{\theta}^p)$, ζ_3^p is an upper bound for $|\tilde{u}(s|\boldsymbol{\theta}^p)|$, and ζ_4^p is an upper bound for $\max_{1 \leq j \leq r} |\tilde{w}_j(s|\boldsymbol{\theta}^p)|$. Hence,

$$J_0(\bar{\boldsymbol{\zeta}}^p) \leq J_0(\boldsymbol{\zeta}^p), \qquad p \geq 1. \tag{4.44}$$

By (4.43) and (4.44), we have

$$\omega \leq J_0(\bar{\boldsymbol{\zeta}}^p) \leq J_0(\boldsymbol{\zeta}^p), \qquad p \geq 1.$$

Thus, by applying the well-known Squeeze Theorem and using equation (4.42), we obtain

$$\lim_{p \to \infty} J_0(\bar{\boldsymbol{\zeta}}^p) = \omega. \tag{4.45}$$

Now, it is clear that

$$0 \leq \theta_k^p \leq T, \qquad k = 1, \ldots, m, \qquad p \geq 1. \tag{4.46}$$

In addition, by Lemma 4.1,

$$-L_1 \leq \bar{\zeta}_\iota^p \leq L_1, \qquad \iota = 1, 2, 3, 4, \qquad p \geq 1. \tag{4.47}$$

Inequalities (4.46) and (4.47) show that the sequence of pairs $\{(\boldsymbol{\theta}^p, \bar{\boldsymbol{\zeta}}^p)\}_{p=1}^{\infty} \subset \mathcal{F}$ is bounded in $\mathbb{R}^m \times \mathbb{R}^4$. Hence, by the Bolzano-Weierstrass Theorem (see Chapter 3 of [6]), there exists a subsequence, which we denote by the original sequence, that converges to a pair $(\boldsymbol{\theta}^*, \boldsymbol{\zeta}^*) \in \Theta \times \mathbb{R}^4$. Inclusion (4.43) and Lemma 4.2 imply that

$$(\boldsymbol{\theta}^*, \boldsymbol{\zeta}^*) \in \mathcal{F}. \tag{4.48}$$

Furthermore, by (4.45),

$$\omega = \lim_{p \to \infty} J_0(\bar{\boldsymbol{\zeta}}^p) = \alpha\zeta_1^* + \alpha\zeta_2^* + \beta\zeta_3^* + \gamma\zeta_4^* = J_0(\boldsymbol{\zeta}^*). \tag{4.49}$$

Inclusion (4.48) and equation (4.49) show that $(\boldsymbol{\theta}^*, \boldsymbol{\zeta}^*)$ is optimal for Problem $\tilde{\mathrm{Q}}$. □

4.6 A numerical example

Consider the switched-capacitor DC-DC power converter discussed in [110]. This switched-capacitor DC-DC power converter has three primary capacitors and four circuit topologies. A circuit schematic for each topology is shown in Figure 4.1. The circuit parameters are as follows:

$$\begin{aligned} C_1 &= C_2 = C_3 = 30.0 \times 10^{-6} \text{ F}, \\ R_1 &= R_2 = R_3 = 0.02 \ \Omega, \\ R_S &= 0.01 \ \Omega, \\ R_L &= 75.0 \ \Omega. \end{aligned}$$

The matrices A_k, B_k, C_k, and D_k, $k = 1, 2, 3, 4$, for this power converter can be derived readily using Kirchhoff's laws. For reference, they are listed in Section 4.A.

We assume that the terminal time here is $T = 2.0 \times 10^{-5}$ seconds and that the minimum duration of each topology is $\tau = 1.0 \times 10^{-6}$ seconds. We also assume that topology switches are accompanied by a 5% voltage leak. Hence,

$$\boldsymbol{z}^k\big(\boldsymbol{x}(t_k^-)\big) = -0.05\boldsymbol{x}(t_k^-), \qquad k = 1, 2, 3.$$

We wrote a Fortran 90 program to solve Problem $\tilde{\mathrm{Q}}$ corresponding to this switched-capacitor DC-DC power converter. This program uses NLPQLP (see [93]) to perform the optimization and LSODA (see [41]) to solve the differential equations.

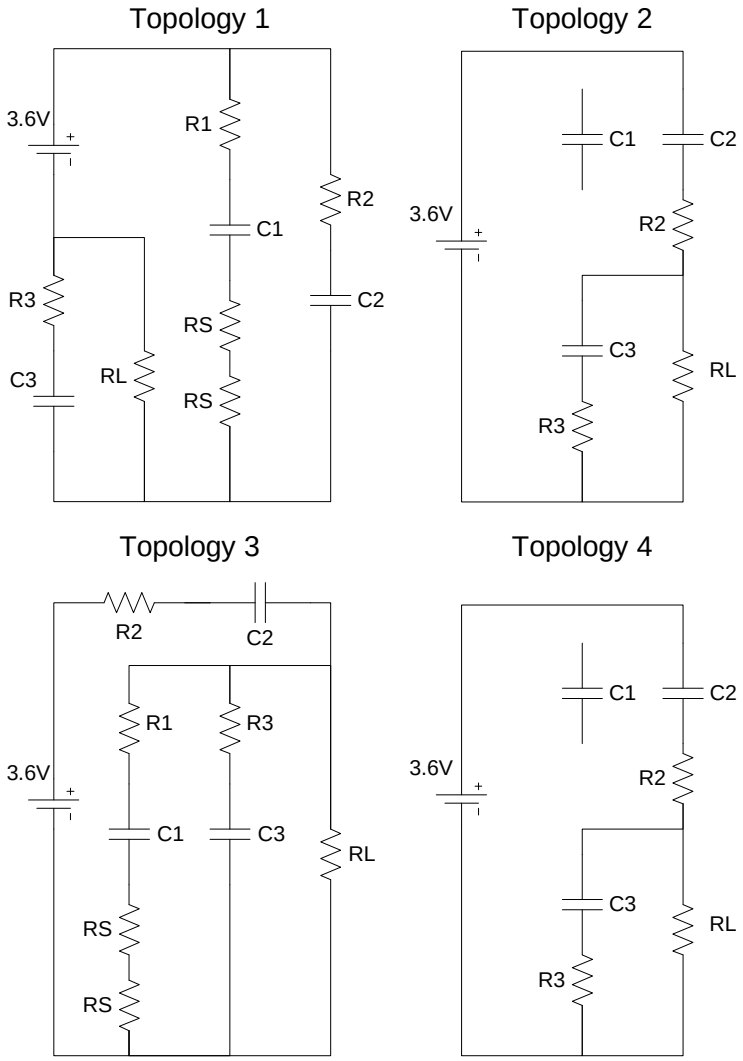

Figure 4.1: The circuit topologies for the switched-capacitor DC-DC power converter in Section 4.6.

We first solved Problem $\tilde{\mathrm{Q}}$ by assuming that the power converter starts from rest ($\boldsymbol{x}^0 = \mathbf{0}$). We then solved Problem $\tilde{\mathrm{Q}}$ for subsequent operating periods, using the final state from the previous problem as the initial state for the next. This was repeated until it was evident that the power converter had reached its steady state. Figure 4.2 shows the evolution of the power converter's output voltage under the optimal switching regime. Figures 4.3-4.5 show the voltage across each of the capacitors.
As expected, the power converter acts as a voltage halver at the steady state. The optimal steady state switching instants are

$$t_1^* = 6.8853 \times 10^{-6},$$
$$t_2^* = 7.8853 \times 10^{-6},$$
$$t_3^* = 8.8853 \times 10^{-6}.$$

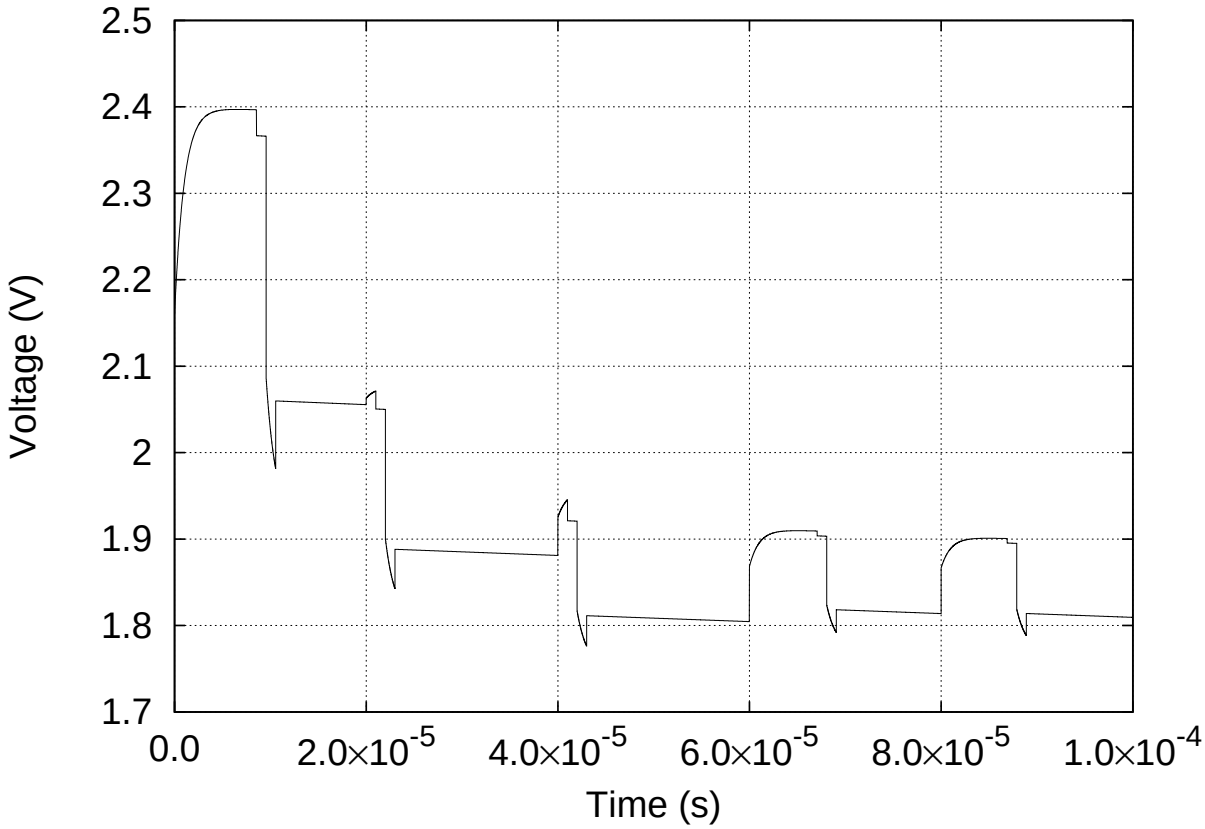

Figure 4.2: Output voltage profile under the optimal switching regime.

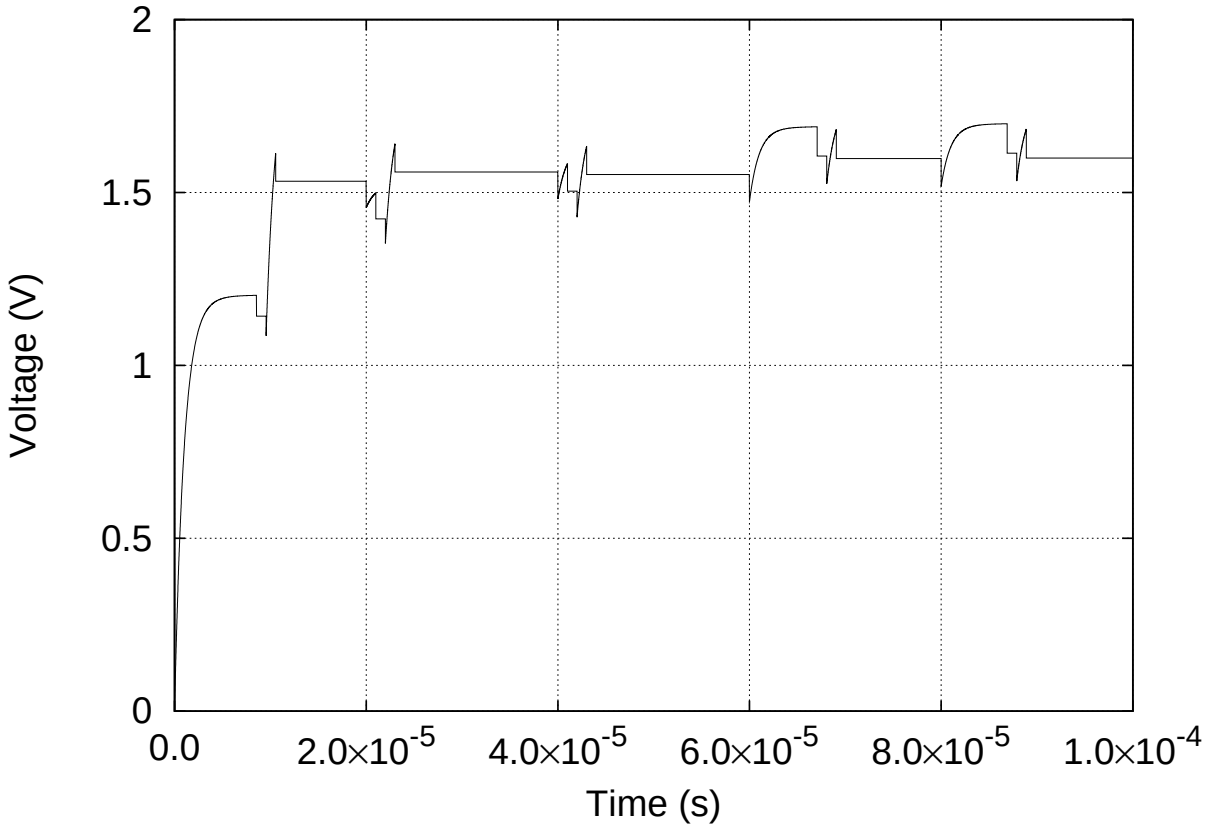

Figure 4.3: Voltage across capacitor 1.

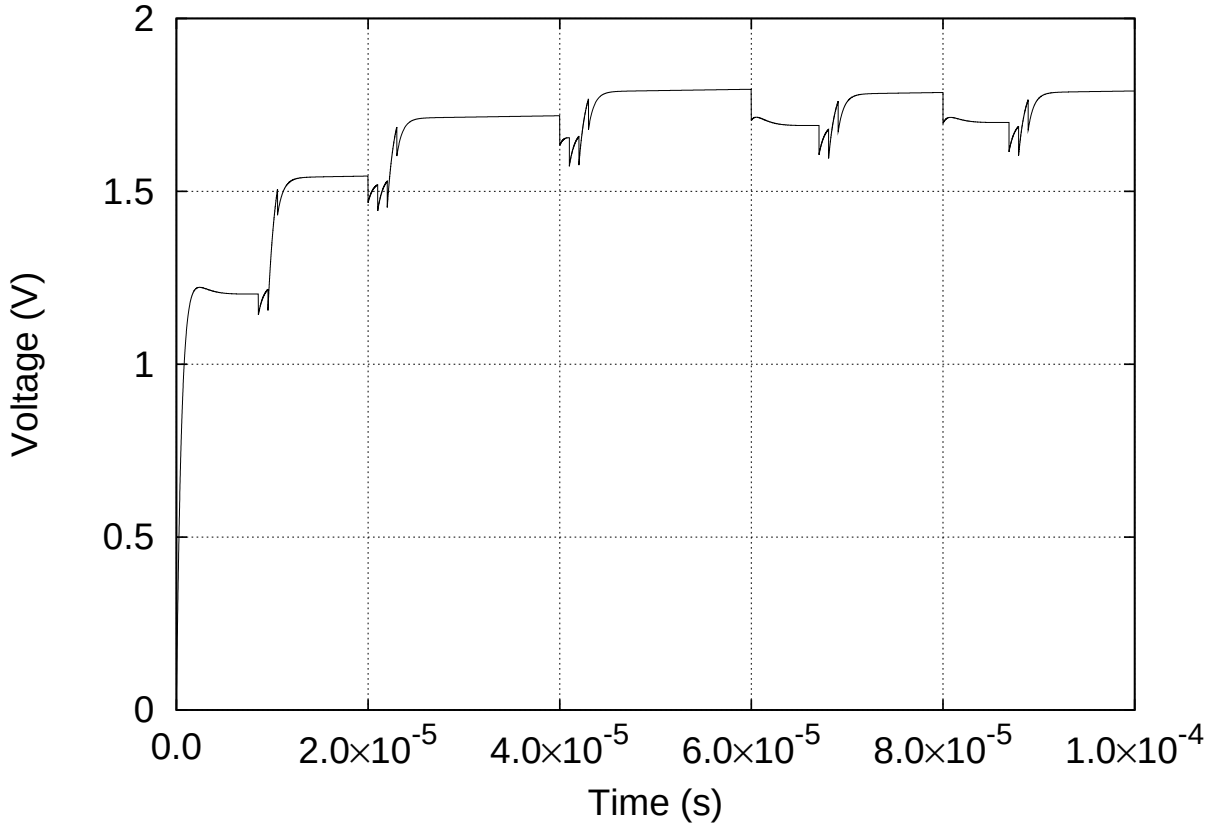

Figure 4.4: Voltage across capacitor 2.

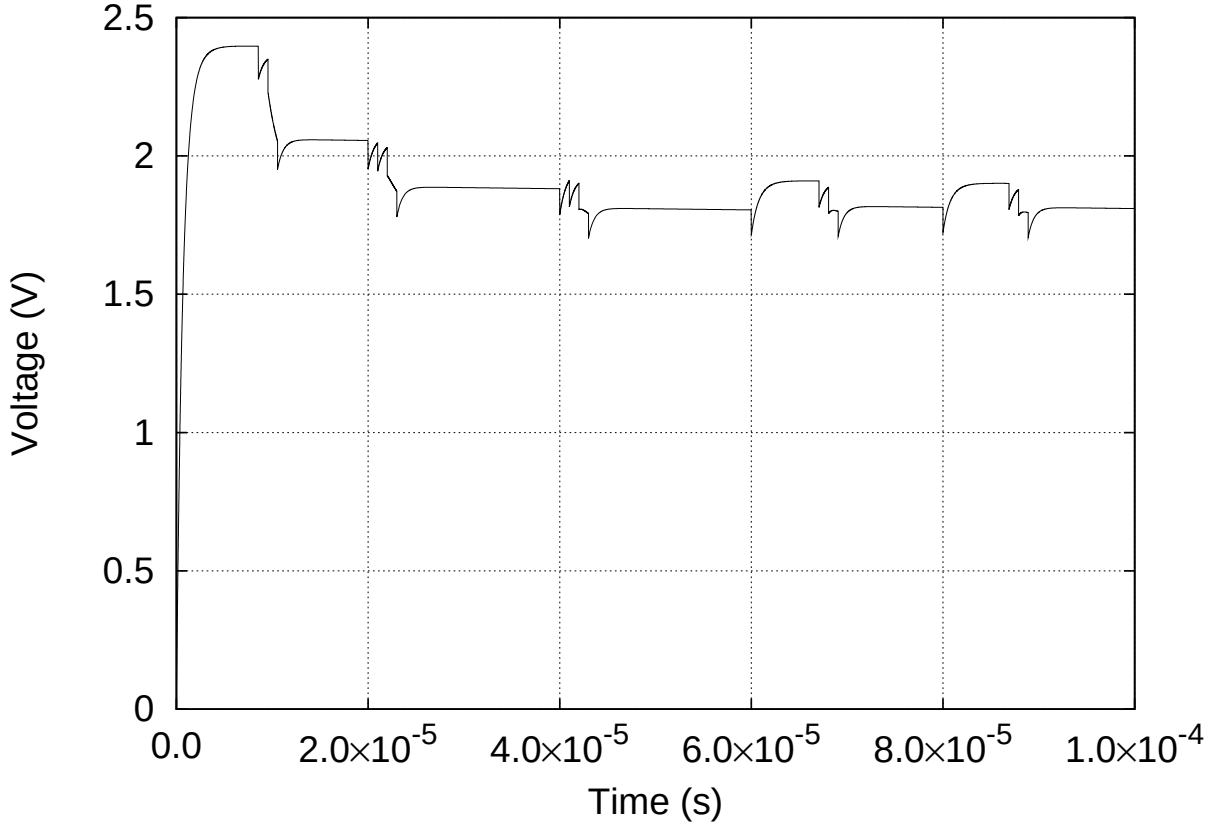

Figure 4.5: Voltage across capacitor 3.

Furthermore, the sensitivities of the output voltage with respect to the load resistance and input are 1.7927×10^{-4} and 6.6595×10^{-1}, respectively. These values, in particular the sensitivity with respect to the load resistance, are very small and thus the output generated by the optimal switching regime is insensitive to changes in the load and input. Also note that the output voltage ripple is 1.1260×10^{-1} V at the steady state.

4.7 Conclusion

In this chapter, we considered the problem of determining optimal switching times for a switched-capacitor DC-DC power converter. Inspired by the work in [42], we first formulated this problem as a switched system optimal control problem, and then developed a new computational method for solving it. We also proved that this optimal control problem has a solution. The main advantage of our new method is that it computes the output voltage sensitivity via an auxiliary switched system, which can be solved simultaneously with the state system. Computing the output voltage sensitivity in this way is much easier than applying the complex procedure given in [42].

4.A System matrices in Section 4.6

Let

$$R_0 \triangleq 2R_S + R_1.$$

Furthermore, define

$$\upsilon_1 = \upsilon_3 \triangleq R_2R_LR_0 + R_2R_3R_0 + R_2R_LR_3 + R_3R_LR_0$$

and

$$\upsilon_2 = \upsilon_4 \triangleq R_2R_L + R_2R_3 + R_LR_3.$$

The matrices in the dynamic model in Section 4.6 are listed below.

$$A_1(R_L) = \begin{bmatrix} \frac{-R_2R_L - R_2R_3 - R_3R_L}{C_1\upsilon_1} & \frac{R_LR_3}{C_1\upsilon_1} & \frac{-R_2R_L}{C_1\upsilon_1} \\ \frac{R_3R_L}{C_2\upsilon_1} & \frac{-R_LR_0 - R_3R_0 - R_3R_L}{C_2\upsilon_1} & \frac{-R_LR_0}{C_2\upsilon_1} \\ \frac{-R_2R_L}{C_3\upsilon_1} & \frac{-R_0R_L}{C_3\upsilon_1} & \frac{-R_2R_L - R_2R_0 - R_0R_L}{C_3\upsilon_1} \end{bmatrix}$$

$$A_2(R_L) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{-R_L - R_3}{C_2\upsilon_2} & \frac{-R_L}{C_2\upsilon_2} \\ 0 & \frac{-R_L}{C_3\upsilon_2} & \frac{-R_2 - R_L}{C_3\upsilon_2} \end{bmatrix}$$

$$A_3(R_L) = \begin{bmatrix} \frac{-R_2R_L - R_2R_3 - R_3R_L}{C_1\upsilon_3} & \frac{-R_3R_L}{C_1\upsilon_3} & \frac{R_2R_L}{C_1\upsilon_3} \\ \frac{-R_3R_L}{C_2\upsilon_3} & \frac{-R_LR_0 - R_3R_0 - R_3R_L}{C_2\upsilon_3} & \frac{-R_LR_0}{C_2\upsilon_3} \\ \frac{R_2R_L}{C_3\upsilon_3} & \frac{-R_LR_0}{C_3\upsilon_3} & \frac{-R_2R_0 - R_LR_0 - R_2R_L}{C_3\upsilon_3} \end{bmatrix}$$

$$A_4(R_L) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{-R_L - R_3}{C_2\upsilon_4} & \frac{-R_L}{C_2\upsilon_4} \\ 0 & \frac{-R_L}{C_3\upsilon_4} & \frac{-R_2 - R_L}{C_3\upsilon_4} \end{bmatrix}$$

$$B_1(R_L) = \begin{bmatrix} \dfrac{R_2R_L + R_2R_3}{C_1\upsilon_1} & \dfrac{R_LR_0 + R_3R_0}{C_2\upsilon_1} & \dfrac{R_2R_L + R_LR_0}{C_3\upsilon_1} \end{bmatrix}$$

$$B_2(R_L) = \begin{bmatrix} 0 & \dfrac{R_L + R_3}{C_2\upsilon_2} & \dfrac{R_L}{C_3\upsilon_2} \end{bmatrix}$$

$$B_3(R_L) = \begin{bmatrix} \dfrac{R_3R_L}{C_1\upsilon_3} & \dfrac{R_LR_0 + R_3R_0 + R_3R_L}{C_2\upsilon_3} & \dfrac{R_LR_0}{C_3\upsilon_3} \end{bmatrix}$$

$$B_4(R_L) = \begin{bmatrix} 0 & \dfrac{R_L + R_3}{C_2\upsilon_4} & \dfrac{R_L}{C_3\upsilon_4} \end{bmatrix}$$

$$C_1(R_L) = \begin{bmatrix} \dfrac{-R_2R_3R_L}{\upsilon_1} & \dfrac{-R_3R_LR_0}{\upsilon_1} & \dfrac{R_2R_LR_0}{\upsilon_1} \end{bmatrix}$$

$$C_2(R_L) = \begin{bmatrix} 0 & \dfrac{-R_3R_L}{\upsilon_2} & \dfrac{R_2R_L}{\upsilon_2} \end{bmatrix}$$

$$C_3(R_L) = \begin{bmatrix} \dfrac{R_2R_3R_L}{\upsilon_3} & \dfrac{-R_3R_LR_0}{\upsilon_3} & \dfrac{R_2R_0R_L}{\upsilon_3} \end{bmatrix}$$

$$C_4(R_L) = \begin{bmatrix} 0 & \dfrac{-R_3R_L}{\upsilon_4} & \dfrac{R_2R_L}{\upsilon_4} \end{bmatrix}$$

$$D_1(R_L) = \frac{R_2R_3R_L + R_3R_LR_0}{\upsilon_1}$$

$$D_2(R_L) = \frac{R_3R_L}{\upsilon_2}$$

$$D_3(R_L) = \frac{R_3R_LR_0}{\upsilon_3}$$

$$D_4(R_L) = \frac{R_3R_L}{\upsilon_4}$$

CHAPTER 5

Optimal control of a switched system*

5.1 Introduction

Many systems operate by switching between different *subsystems* or *modes*. Such systems are called *switched systems*. An example of a switched system is the switched-capacitor DC-DC power converter considered in Chapter 4, which operates by switching between several circuit topologies. Other examples of switched systems include robots [10], locomotives [44,45], bioconversion reactors [25,26], and hybrid power generators [90,122,124].

The switched system in Chapter 4, which models a switched-capacitor DC-DC power converter, has linear subsystems. In this chapter, we consider a more general switched system whose subsystems are *nonlinear*. Each state jump in this switched system (the abrupt change in the state that accompanies a subsystem switch) depends not only on the state immediately before the switch, but also on a set of control variables. These control variables, and the times at which the subsystem switches occur, should be chosen to minimize a given cost function.

We saw in Chapter 4 that the time-scaling transformation proposed in [60] is a useful tool for determining the optimal switching times in a switched system. This transformation is discussed in detail in [64, 89, 132], where it is applied to switched systems *without* state jumps, and in [68,69,125], where it is applied to switched systems *with* state jumps. Its main virtue is that it yields a new optimal control problem whose switching times are fixed points, not decision variables. This new problem is equivalent to, but much easier to solve than, the original optimal control problem.

In Chapter 4 and [68, 69, 125], the difference between consecutive switching times is assumed to be strictly positive. In other words, the switching times are distinct. This requirement is necessary to ensure that the time-scaling transformation does not introduce "fictitious" state jumps—a situation that occurs when two or more consecutive switching times coincide. For example, if t_1 and t_2 coincide, then the state jump at time $t = t_1 = t_2$ corresponds to state jumps at *both* $s = 1$ and $s = 2$ in the new time horizon. In this case, the time-scaling transformation converts a single state jump into two separate state jumps,

*This chapter is based on [73].

and therefore does not preserve the structure of the system. Note that this situation only occurs in switched systems with state jumps; there is no need to assume that the switching times are distinct if the state trajectory is continuous (see [64, 89, 132]).

In practice, excessive switching between subsystems can adversely affect the overall system. We already saw in Chapter 4 that changing the circuit topology in a switched-capacitor DC-DC power converter causes a voltage leak. Thus, it is usually not optimal to control a switched system by applying every available switch; the optimal strategy may instead involve "deleting" switches by merging two or more switching times into a single switch. In this case, some of the optimal switching times coincide, and the methods discussed in [68, 69, 125] (and Chapter 4) are not appropriate. The aim of this chapter is to develop a new method that is capable of handling this case.

5.2 Problem formulation

Consider the following switched system with $m \geq 2$ subsystems:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}^k\big(\boldsymbol{x}(t), \boldsymbol{\sigma}\big), \qquad t \in (t_{k-1}, t_k), \qquad k = 1, \ldots, m, \tag{5.1}$$

and

$$\boldsymbol{x}(t_k^+) = \begin{cases} \boldsymbol{x}^0, & \text{if } k = 0, \quad (5.2a) \\ \boldsymbol{x}(t_k^-) + \boldsymbol{z}^k\big(\boldsymbol{x}(t_k^-), \boldsymbol{\sigma}\big), & \text{if } k \in \{1, \ldots, m-1\} \text{ and } t_{k-1} < t_k < T, \quad (5.2b) \end{cases}$$

where $t_0 \triangleq 0$, $t_m \triangleq T$, and $T > 0$ is a given terminal time; t_k, $k = 1, \ldots, m-1$, are the subsystem switching times; $\boldsymbol{x}(t) \in \mathbb{R}^n$ is the system state at time t; $\boldsymbol{x}^0 \in \mathbb{R}^n$ is a given initial state; $\boldsymbol{\sigma} \in \mathbb{R}^r$ is a vector of control variables; and $\boldsymbol{f}^k : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^n$, $k = 1, \ldots, m$, and $\boldsymbol{z}^k : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^n$, $k = 1, \ldots, m-1$, are given functions.

The subsystem switching times, which are chosen by the system designer, must satisfy the following constraints:

$$t_{k-1} \leq t_k, \qquad k = 1, \ldots, m. \tag{5.3}$$

A vector $\boldsymbol{\nu} = [t_1, \ldots, t_{m-1}]^T \in \mathbb{R}^{m-1}$ that satisfies (5.3) is called a *switching-time vector*. Let Γ denote the set of all such switching-time vectors.

The control variables are also chosen by the system designer. They are subject to the following constraints:

$$a_\varsigma \leq \sigma_\varsigma \leq b_\varsigma, \qquad \varsigma = 1, \ldots, r, \tag{5.4}$$

where a_ς and b_ς, $\varsigma = 1, \ldots, r$, are given real numbers such that $a_\varsigma < b_\varsigma$. A vector $\boldsymbol{\sigma} \in \mathbb{R}^r$ that satisfies (5.4) is called a *control vector*. Let Ξ denote the set of all such control vectors. A pair $(\boldsymbol{\nu}, \boldsymbol{\sigma}) \in \Gamma \times \Xi$ is called a *control pair*.

We assume that the following conditions are satisfied.

Assumption 5.1. The functions $\boldsymbol{f}^k$, $k = 1, \ldots, m$, and $\boldsymbol{z}^k$, $k = 1, \ldots, m-1$, are continuously differentiable.

Assumption 5.2. There exists a real number $L_1 > 0$ such that

$$\left|\boldsymbol{f}^k(\boldsymbol{v}, \boldsymbol{\sigma})\right| \leq L_1(1 + |\boldsymbol{v}|), \qquad k = 1, \ldots, m, \qquad (\boldsymbol{v}, \boldsymbol{\sigma}) \in \mathbb{R}^n \times \Xi.$$

The switched system (5.1)-(5.2) is considerably more general than the one in Chapter 4. In particular, its subsytems are described by nonlinear differential equations, and its state jumps depend explicitly on the control variables. Furthermore, consecutive switching times in (5.1)-(5.2) may coincide if necessary (constraint (5.3) is still satisfied if $t_{k-1} = t_k$). Notice that if the switching times t_{k-1} and t_k coincide, then $(t_{k-1}, t_k) = \emptyset$ and the kth subsystem in (5.1) is effectively deleted. This does not occur in the switched system in Chapter 4, because inequality (4.1) ensures that the difference between consecutive switching times is strictly positive (and thus every subsystem is active for a nontrivial duration of the time horizon).

By merging two or more consecutive switching times into a single switch, the system designer can remove inefficient/redundant subsystems. This is particularly useful when the number of switches in the model is an overestimate of the optimal number of switches, which is often the case in optimal discrete-valued control problems (see [61, 86, 90]).

Given a control pair $(\boldsymbol{\nu}, \boldsymbol{\sigma}) \in \Gamma \times \Xi$, the switched system (5.1)-(5.2) evolves in the following manner. It begins in state $\boldsymbol{x}^0$ at time $t = 0$ with subsystem $\iota(1)$, where

$$\iota(1) \triangleq \min\big\{\, k \in \{1, \ldots, m\} : \; t_k > 0 \,\big\}.$$

Subsystem $\iota(1)$ runs smoothly according to equation (5.1) with $k = \iota(1)$ until time $t = t_{\iota(1)}$. If $t_{\iota(1)} = T$, then the system stops. Otherwise, the state jumps from $\boldsymbol{x}(t^-_{\iota(1)})$ to a new point $\boldsymbol{x}(t^+_{\iota(1)})$, which is given by equation (5.2b) with $k = \iota(1)$. The system then activates subsystem $\iota(2)$, where

$$\iota(2) \triangleq \min\big\{\, k \in \{\iota(1) + 1, \ldots, m\} : \; t_k > t_{\iota(1)} \,\big\}.$$

Subsystem $\iota(2)$ runs smoothly according to (5.1) with $k = \iota(2)$ until $t = t_{\iota(2)}$, at which time the state experiences another jump from $\boldsymbol{x}(t^-_{\iota(2)})$ to $\boldsymbol{x}(t^+_{\iota(2)})$. The system continues to evolve in this way for the remainder of the time horizon.

Let $\boldsymbol{x}(\cdot|\boldsymbol{\nu}, \boldsymbol{\sigma})$ denote the solution of the switched system (5.1)-(5.2) corresponding to the control pair $(\boldsymbol{\nu}, \boldsymbol{\sigma}) \in \Gamma \times \Xi$. Without loss of generality, we assume that this solution is continuous from the right. Thus,

$$\boldsymbol{x}(t_k|\boldsymbol{\nu}, \boldsymbol{\sigma}) = \boldsymbol{x}(t^+_k|\boldsymbol{\nu}, \boldsymbol{\sigma}), \qquad k = 0, \ldots, m-1.$$

Our optimal control problem is defined below.

Problem P. *Find a control pair* $(\boldsymbol{\nu}, \boldsymbol{\sigma}) \in \Gamma \times \Xi$ *that minimizes the cost function*

$$G_0(\boldsymbol{\nu}, \boldsymbol{\sigma}) \triangleq \sum_{k=1}^{m} \int_{t_{k-1}}^{t_k} \mathcal{L}_k\big(\boldsymbol{x}(t|\boldsymbol{\nu}, \boldsymbol{\sigma}), \boldsymbol{\sigma}\big) dt,$$

where $\mathcal{L}_k : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$, $k = 1, \ldots, m$, *are given functions, over* $\Gamma \times \Xi$.

To conclude this section, we make the following assumption.

Assumption 5.3. The functions $\mathcal{L}_k : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$, $k = 1, \ldots, m$, are continuously differentiable.

5.3 Time-scaling transformation

Our first step in tackling Problem P is to apply the time-scaling transformation, so that the subsystem switching times become fixed points in a new time horizon.

Let

$$\Theta \triangleq \big\{ \boldsymbol{\theta} \in \mathbb{R}^m : \ \theta_k \geq 0, \ k = 1, \ldots, m; \ \theta_1 + \cdots + \theta_m = T \big\}.$$

Furthermore, for each $\boldsymbol{\theta} \in \Theta$, define a corresponding function $\mu(\cdot|\boldsymbol{\theta}) : [0, m] \to \mathbb{R}$ as follows:

$$\mu(s|\boldsymbol{\theta}) \triangleq \begin{cases} \displaystyle\sum_{l=1}^{\lfloor s \rfloor} \theta_l + \theta_{\lfloor s \rfloor + 1}(s - \lfloor s \rfloor), & \text{if } s \in [0, m), \\ T, & \text{if } s = m, \end{cases}$$

where $\lfloor \cdot \rfloor$ denotes the floor function. Clearly,

$$\mu(k|\boldsymbol{\theta}) = \sum_{l=1}^{k} \theta_l, \qquad k = 0, \ldots, m. \tag{5.5}$$

It is easy to see that $\mu(\cdot|\boldsymbol{\theta})$ has the following properties:

(a) $\mu(0|\boldsymbol{\theta}) = 0$, $\mu(m|\boldsymbol{\theta}) = T$, and $\mu(s|\boldsymbol{\theta}) \in [0, T]$ for each $s \in [0, m]$;

(b) $\mu(\cdot|\boldsymbol{\theta})$ is a continuous function;

(c) For each $k = 1, \ldots, m$, $\mu(\cdot|\boldsymbol{\theta})$ is constant on $[k-1, k]$ if and only if $\theta_k = 0$; and

(d) For each $k = 1, \ldots, m$, $\mu(\cdot|\boldsymbol{\theta})$ is strictly increasing on $[k-1, k]$ if and only if $\theta_k > 0$.

Properties (c) and (d) imply that $\mu(\cdot|\boldsymbol{\theta})$ is non-decreasing on each subinterval $[k-1, k]$, $k = 1, \ldots, m$. Hence, for each $\boldsymbol{\theta} \in \Theta$,

$$\mu(k-1|\boldsymbol{\theta}) \leq \mu(k|\boldsymbol{\theta}), \qquad k = 1, \ldots, m.$$

By comparing this inequality with (5.3), we see that

$$\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}) \triangleq \big[\mu(1|\boldsymbol{\theta}), \ldots, \mu(m-1|\boldsymbol{\theta})\big]^T \in \mathbb{R}^{m-1} \tag{5.6}$$

is a valid switching-time vector for Problem P. Thus,

$$\big\{\, \tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}) : \ \boldsymbol{\theta} \in \Theta \,\big\} \subset \Gamma. \tag{5.7}$$

The reverse inclusion also holds. To see why, let $\boldsymbol{\nu}' = [t'_1, \ldots, t'_{m-1}]^T \in \Gamma$ and define a corresponding vector $\boldsymbol{\theta}' \in \mathbb{R}^m$ as follows:

$$\theta'_k = t'_k - t'_{k-1}, \qquad k = 1, \ldots, m,$$

where $t'_0 \triangleq 0$ and $t'_m \triangleq T$. Since the components of $\boldsymbol{\nu}'$ satisfy (5.3), the components of $\boldsymbol{\theta}'$ are non-negative. Moreover,

$$\sum_{k=1}^{m} \theta'_k = \sum_{k=1}^{m} (t'_k - t'_{k-1}) = t'_m - t'_0 = T.$$

Hence, $\boldsymbol{\theta}' \in \Theta$. Now, by equation (5.5),

$$\mu(k|\boldsymbol{\theta}') = \sum_{l=1}^{k} \theta'_l = \sum_{l=1}^{k} (t'_l - t'_{l-1}) = t'_k - t'_0 = t'_k, \qquad k = 1, \ldots, m-1.$$

It follows immediately that $\boldsymbol{\nu}' = \tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}')$. Since $\boldsymbol{\nu}' \in \Gamma$ was chosen arbitrarily, this implies that

$$\Gamma \subset \big\{\, \tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}) : \ \boldsymbol{\theta} \in \Theta \,\big\}. \tag{5.8}$$

Combining inclusions (5.7) and (5.8) gives

$$\Gamma = \big\{\, \tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}) : \ \boldsymbol{\theta} \in \Theta \,\big\}. \tag{5.9}$$

This is an important equation; it shows that each switching-time vector is generated by a corresponding vector in Θ.

Now, for each $(\boldsymbol{\theta}, \boldsymbol{\sigma}) \in \Theta \times \Xi$, define a new state variable $\tilde{\boldsymbol{x}}(\cdot|\boldsymbol{\theta}, \boldsymbol{\sigma}) : [0, m] \to \mathbb{R}^n$ as follows:

$$\tilde{\boldsymbol{x}}(s|\boldsymbol{\theta}, \boldsymbol{\sigma}) \triangleq \boldsymbol{x}(\mu(s|\boldsymbol{\theta})|\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}), \boldsymbol{\sigma}), \qquad s \in [0, m]. \tag{5.10}$$

Note that equation (5.9) and property (a) of $\mu(\cdot|\boldsymbol{\theta})$ ensure that this definition is valid.

We will now derive the dynamics governing the behavior of the new state $\tilde{\boldsymbol{x}}(\cdot|\boldsymbol{\theta}, \boldsymbol{\sigma})$. To this end, note that

$$\dot{\mu}(s|\boldsymbol{\theta}) = \theta_k, \qquad s \in (k-1, k), \qquad k = 1, \ldots, m. \tag{5.11}$$

By differentiating (5.10) with respect to s, and then using equations (5.1) and (5.11) and properties (c) and (d) of $\mu(\cdot|\boldsymbol{\theta})$, we obtain

$$\dot{\tilde{\boldsymbol{x}}}(s|\boldsymbol{\theta},\boldsymbol{\sigma}) = \theta_k \boldsymbol{f}^k\big(\tilde{\boldsymbol{x}}(s|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}\big), \qquad s \in (k-1,k), \qquad k = 1,\dots,m. \tag{5.12}$$

Furthermore, it follows from equation (5.2) and property (b) of $\mu(\cdot|\boldsymbol{\theta})$ that for each integer $k = 0,\dots,m-1$,

$$\tilde{\boldsymbol{x}}(k^+|\boldsymbol{\theta},\boldsymbol{\sigma}) = \begin{cases} \boldsymbol{x}^0, & \text{if } k = 0, \quad (5.13a)\\ \tilde{\boldsymbol{x}}(k^-|\boldsymbol{\theta},\boldsymbol{\sigma}) + \boldsymbol{z}^k\big(\tilde{\boldsymbol{x}}(k^-|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}\big), & \text{if } \mu(k-1|\boldsymbol{\theta}) < \mu(k|\boldsymbol{\theta}) < T, \quad (5.13b)\\ \tilde{\boldsymbol{x}}(k^-|\boldsymbol{\theta},\boldsymbol{\sigma}), & \text{otherwise.} \quad (5.13c) \end{cases}$$

We define a new optimal control problem as follows.

Problem $\tilde{\mathrm{P}}$. *Find a pair $(\boldsymbol{\theta},\boldsymbol{\sigma}) \in \Theta \times \Xi$ that minimizes the cost function*

$$\tilde{G}_0(\boldsymbol{\theta},\boldsymbol{\sigma}) \triangleq G_0(\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}),\boldsymbol{\sigma}) = \sum_{k=1}^{m} \int_{k-1}^{k} \theta_k \mathcal{L}_k\big(\tilde{\boldsymbol{x}}(s|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}\big) ds$$

over $\Theta \times \Xi$.

Problems P and $\tilde{\mathrm{P}}$ are equivalent. Indeed, $(\boldsymbol{\theta}^*,\boldsymbol{\sigma}^*) \in \Theta \times \Xi$ is optimal for Problem $\tilde{\mathrm{P}}$ if and only if $(\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}^*),\boldsymbol{\sigma}^*) \in \Gamma \times \Xi$ is optimal for Problem P. To see why, let $(\boldsymbol{\theta}^*,\boldsymbol{\sigma}^*) \in \Theta \times \Xi$ be an optimal pair for Problem $\tilde{\mathrm{P}}$, and let $(\boldsymbol{\nu}',\boldsymbol{\sigma}') \in \Gamma \times \Xi$ be arbitrary but fixed. Then by equation (5.9), there exists a $\boldsymbol{\theta}' \in \Theta$ such that $\boldsymbol{\nu}' = \tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}')$. We have

$$G_0(\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}^*),\boldsymbol{\sigma}^*) = \tilde{G}_0(\boldsymbol{\theta}^*,\boldsymbol{\sigma}^*) \le \tilde{G}_0(\boldsymbol{\theta}',\boldsymbol{\sigma}') = G_0(\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}'),\boldsymbol{\sigma}') = G_0(\boldsymbol{\nu}',\boldsymbol{\sigma}').$$

Since $(\boldsymbol{\nu}',\boldsymbol{\sigma}') \in \Gamma \times \Xi$ was chosen arbitrarily, this inequality shows that $(\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}^*),\boldsymbol{\sigma}^*)$ is optimal for Problem P.

On the other hand, suppose that $(\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}^*),\boldsymbol{\sigma}^*)$, where $(\boldsymbol{\theta}^*,\boldsymbol{\sigma}^*) \in \Theta \times \Xi$, is an optimal control pair for Problem P. Furthermore, let $(\boldsymbol{\theta}',\boldsymbol{\sigma}') \in \Theta \times \Xi$ be arbitrary but fixed. Then by equation (5.9), $\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}') \in \Gamma$. Therefore,

$$\tilde{G}_0(\boldsymbol{\theta}^*,\boldsymbol{\sigma}^*) = G_0(\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}^*),\boldsymbol{\sigma}^*) \le G_0(\tilde{\boldsymbol{\nu}}(\boldsymbol{\theta}'),\boldsymbol{\sigma}') = \tilde{G}_0(\boldsymbol{\theta}',\boldsymbol{\sigma}'),$$

which shows that $(\boldsymbol{\theta}^*,\boldsymbol{\sigma}^*)$ is optimal for Problem $\tilde{\mathrm{P}}$.

5.4 Problem approximation

We expect that Problem $\tilde{\mathrm{P}}$, having fixed switching times, is easier to solve than Problem P. However, we will see in this section that Problem $\tilde{\mathrm{P}}$ still cannot be solved directly using

a conventional optimization method.

We first simplify the state jump conditions (5.13). In fact, it turns out that the condition $\mu(k|\boldsymbol{\theta}) < T$ in (5.13b) is unnecessary and can be removed. To see why, suppose that $(\boldsymbol{\theta}, \boldsymbol{\sigma}) \in \Theta \times \Xi$ satisfies

$$\mu(k|\boldsymbol{\theta}) = T = \mu(m|\boldsymbol{\theta})$$

for some $k \in \{1, \dots, m-1\}$. Then it follows from equation (5.5) that

$$\theta_l = 0, \qquad l = k+1, \dots, m.$$

Therefore,

$$\tilde{G}_0(\boldsymbol{\theta}, \boldsymbol{\sigma}) = \sum_{l=1}^{k} \int_{l-1}^{l} \theta_l \mathcal{L}_l\big(\tilde{\boldsymbol{x}}(s|\boldsymbol{\theta}, \boldsymbol{\sigma}), \boldsymbol{\sigma}\big) ds,$$

which shows that the cost of the pair $(\boldsymbol{\theta}, \boldsymbol{\sigma})$ does not depend on $\tilde{\boldsymbol{x}}(s|\boldsymbol{\theta}, \boldsymbol{\sigma})$, $s \in [k, m]$. In particular, $\tilde{G}_0(\boldsymbol{\theta}, \boldsymbol{\sigma})$ does not depend on whether $\tilde{\boldsymbol{x}}(k^+|\boldsymbol{\theta}, \boldsymbol{\sigma})$ is calculated via equation (5.13b) or via equation (5.13c). Hence, we may remove the condition $\mu(k|\boldsymbol{\theta}) < T$ from (5.13b); the state jump conditions (5.13) then become

$$\tilde{\boldsymbol{x}}(k^+|\boldsymbol{\theta}, \boldsymbol{\sigma}) = \begin{cases} \boldsymbol{x}^0, & \text{if } k = 0, \quad (5.14a) \\ \tilde{\boldsymbol{x}}(k^-|\boldsymbol{\theta}, \boldsymbol{\sigma}) + \boldsymbol{z}^k\big(\tilde{\boldsymbol{x}}(k^-|\boldsymbol{\theta}, \boldsymbol{\sigma}), \boldsymbol{\sigma}\big), & \text{if } \mu(k-1|\boldsymbol{\theta}) < \mu(k|\boldsymbol{\theta}), \quad (5.14b) \\ \tilde{\boldsymbol{x}}(k^-|\boldsymbol{\theta}, \boldsymbol{\sigma}), & \text{otherwise.} \quad (5.14c) \end{cases}$$

Now, it follows from equation (5.5) that for each $k = 1, \dots, m$,

$$\mu(k-1|\boldsymbol{\theta}) = \mu(k|\boldsymbol{\theta}) \qquad \Longleftrightarrow \qquad \theta_k = 0.$$

This implication is clearly equivalent to

$$\mu(k-1|\boldsymbol{\theta}) < \mu(k|\boldsymbol{\theta}) \qquad \Longleftrightarrow \qquad \theta_k > 0.$$

By using these two implications, we can express (5.14) in the following compact form:

$$\tilde{\boldsymbol{x}}(k^+|\boldsymbol{\theta}, \boldsymbol{\sigma}) = \begin{cases} \boldsymbol{x}^0, & \text{if } k = 0, \quad (5.15a) \\ \tilde{\boldsymbol{x}}(k^-|\boldsymbol{\theta}, \boldsymbol{\sigma}) + \varphi(\theta_k)\boldsymbol{z}^k\big(\tilde{\boldsymbol{x}}(k^-|\boldsymbol{\theta}, \boldsymbol{\sigma}), \boldsymbol{\sigma}\big), & \text{if } k \in \{1, \dots, m-1\}, \quad (5.15b) \end{cases}$$

where $\varphi : \mathbb{R} \to \mathbb{R}$ is defined by

$$\varphi(\eta) \triangleq \begin{cases} 1, & \text{if } \eta > 0, \\ 0, & \text{if } \eta = 0. \end{cases}$$

Henceforth, we will use (5.15) instead of (5.13) and (5.14). Thus, for the remainder of this chapter, $\tilde{\boldsymbol{x}}(\cdot|\boldsymbol{\theta}, \boldsymbol{\sigma})$ denotes a solution of the switched system consisting of (5.12) and (5.15).

Equation (5.15) involves φ, a discontinuous function. Consequently, the gradient of $\tilde{G}_0$ does not exist, and Problem $\tilde{\mathrm{P}}$ cannot be solved using a gradient-based nonlinear programming algorithm. Note that the gradient formulae in [47, 69, 125], which we used in Chapter 4 to derive the gradient of Problem $\tilde{\mathrm{Q}}_{\epsilon,\vartheta}$'s cost function, are only applicable when the state jump conditions are governed by smooth functions.

To proceed, we approximate (5.15) by

$$\tilde{\boldsymbol{x}}(k^+|\boldsymbol{\theta},\boldsymbol{\sigma}) = \begin{cases} \boldsymbol{x}^0, & \text{if } k = 0, \qquad (5.16a) \\ \tilde{\boldsymbol{x}}(k^-|\boldsymbol{\theta},\boldsymbol{\sigma}) + \varphi_\epsilon(\theta_k)\boldsymbol{z}^k\big(\tilde{\boldsymbol{x}}(k^-|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}\big), & \text{if } k \in \{1,\dots,m-1\}, \qquad (5.16b) \end{cases}$$

where $\epsilon > 0$ and $\varphi_\epsilon : \mathbb{R} \to \mathbb{R}$ is defined by

$$\varphi_\epsilon(\eta) \triangleq \begin{cases} -\frac{2}{\epsilon^3}\eta^3 + \frac{3}{\epsilon^2}\eta^2, & \text{if } 0 \le \eta \le \epsilon, \\ 1, & \text{if } \eta > \epsilon. \end{cases}$$

Note that φ_ϵ is continuously differentiable and

$$\varphi_\epsilon(\eta) = \varphi(\eta), \qquad \eta \notin (0,\epsilon).$$

This smoothing function is illustrated in Figure 5.1.

Let $\tilde{\boldsymbol{x}}^\epsilon(\cdot|\boldsymbol{\theta},\boldsymbol{\sigma})$ denote the solution of (5.12) and (5.16) corresponding to $(\boldsymbol{\theta},\boldsymbol{\sigma}) \in \Theta \times \Xi$ and $\epsilon > 0$. Furthermore, define

$$\tilde{G}_0^\epsilon(\boldsymbol{\theta},\boldsymbol{\sigma}) \triangleq \sum_{k=1}^{m} \int_{k-1}^{k} \theta_k \mathcal{L}_k\big(\tilde{\boldsymbol{x}}^\epsilon(s|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}\big)ds.$$

Consider the following optimization problem.

Problem $\tilde{\mathrm{P}}_{\epsilon,\vartheta}$. *Find a pair $(\boldsymbol{\theta},\boldsymbol{\sigma}) \in \Theta \times \Xi$ that minimizes the cost function*

$$\tilde{J}_{\epsilon,\vartheta}(\boldsymbol{\theta},\boldsymbol{\sigma}) \triangleq \tilde{G}_0^\epsilon(\boldsymbol{\theta},\boldsymbol{\sigma}) + \vartheta \sum_{k=1}^{m-1} \varphi_\epsilon(\theta_k)\big(1 - \varphi_\epsilon(\theta_k)\big),$$

where $\epsilon > 0$ and $\vartheta > 0$ are given real numbers, over $\Theta \times \Xi$.

Unlike φ, the continuous function φ_ϵ can assume intermediate values in $(0,1)$. Hence, the state jump conditions in Problem $\tilde{\mathrm{P}}_{\epsilon,\vartheta}$ do not always reflect the state jump conditions in Problem $\tilde{\mathrm{P}}$. The last term in $\tilde{J}_{\epsilon,\vartheta}$ penalizes "fractional jumps", so that (5.16) is a good approximation of (5.15) at the optimal solution of Problem $\tilde{\mathrm{P}}_{\epsilon,\vartheta}$. It is also evident that $\varphi_\epsilon \to \varphi$ pointwise on $[0,\infty)$ as $\epsilon \to 0$. We therefore expect that Problem $\tilde{\mathrm{P}}_{\epsilon,\vartheta}$ is a good approximation for Problem $\tilde{\mathrm{P}}$ when ϵ is small and ϑ is large. The relationship between Problems $\tilde{\mathrm{P}}$ and $\tilde{\mathrm{P}}_{\epsilon,\vartheta}$ is examined in more detail in Section 5.5.

Since all functions in (5.16) are smooth, the partial derivatives of $\tilde{J}_{\epsilon,\vartheta}$ exist. We now

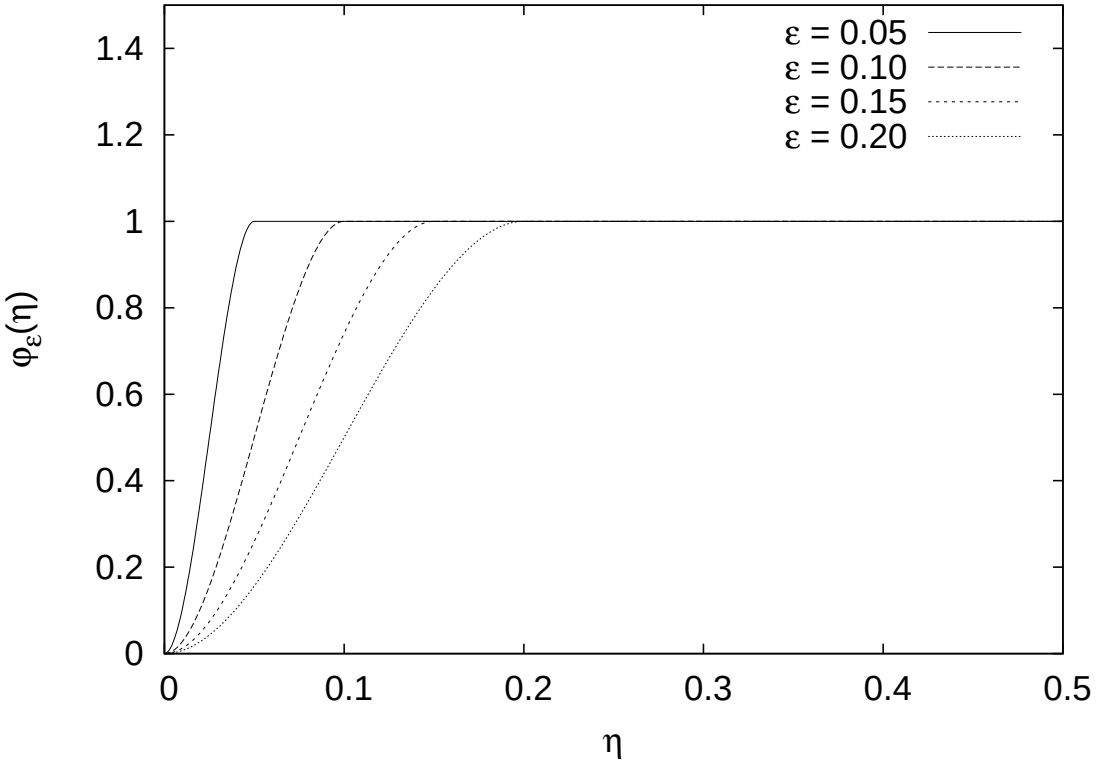

Figure 5.1: The smoothing function φ_ϵ.

derive the formulae for computing these partial derivatives.

First, for each $k = 1, \ldots, m$, define the *Hamiltonian* $H_k : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \times \Xi \to \mathbb{R}$ as follows:

$$H_k(\theta_k, \boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\sigma}) \triangleq \theta_k \mathcal{L}_k(\boldsymbol{x}, \boldsymbol{\sigma}) + \theta_k \boldsymbol{\lambda}^T \boldsymbol{f}^k(\boldsymbol{x}, \boldsymbol{\sigma}), \qquad (\theta_k, \boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\sigma}) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \times \Xi.$$

Now, consider the following auxiliary switched system:

$$\dot{\boldsymbol{\lambda}}(s) = -\left[\frac{\partial H_k\big(\theta_k, \tilde{\boldsymbol{x}}^\epsilon(s|\boldsymbol{\theta}, \boldsymbol{\sigma}), \boldsymbol{\lambda}(s), \boldsymbol{\sigma}\big)}{\partial \boldsymbol{x}}\right]^T, \qquad s \in (k-1, k), \qquad k = 1, \ldots, m, \quad (5.17)$$

and

$$\boldsymbol{\lambda}(k^-) = \begin{cases} \boldsymbol{0}, & \text{if } k = m, \quad (5.18a) \\ \boldsymbol{\lambda}(k^+) + \varphi_\epsilon(\theta_k)\left[\dfrac{\partial \boldsymbol{z}^k\big(\tilde{\boldsymbol{x}}^\epsilon(k^-|\boldsymbol{\theta}, \boldsymbol{\sigma}), \boldsymbol{\sigma}\big)}{\partial \boldsymbol{x}}\right]^T \boldsymbol{\lambda}(k^+), & \\ & \text{if } k \in \{1, \ldots, m-1\}, \quad (5.18b) \end{cases}$$

where $(\boldsymbol{\theta}, \boldsymbol{\sigma}) \in \Theta \times \Xi$ and $\epsilon > 0$. This auxiliary switched system is called the *costate system*. Let $\boldsymbol{\lambda}^\epsilon(\cdot|\boldsymbol{\theta}, \boldsymbol{\sigma})$ denote the solution of the costate system corresponding to the pair $(\boldsymbol{\theta}, \boldsymbol{\sigma}) \in \Theta \times \Xi$.

The partial derivatives of $\tilde{J}_{\epsilon,\vartheta}$ are given in the following two theorems.

Theorem 5.1. *For each* $(\boldsymbol{\theta}, \boldsymbol{\sigma}) \in \Theta \times \Xi$,

$$\frac{\partial \tilde{J}_{\epsilon,\vartheta}(\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial \theta_k} = \hat{\rho}_{k,m-1}\Big\{\dot{\varphi}_\epsilon(\theta_k)\big[\boldsymbol{\lambda}^\epsilon(k^+|\boldsymbol{\theta},\boldsymbol{\sigma})\big]^T \boldsymbol{z}^k\big(\tilde{\boldsymbol{x}}^\epsilon(k^-|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}\big) + \vartheta\dot{\varphi}_\epsilon(\theta_k) - 2\vartheta\varphi_\epsilon(\theta_k)\dot{\varphi}_\epsilon(\theta_k)\Big\} + \int_{k-1}^{k} \frac{\partial H_k\big(\theta_k, \tilde{\boldsymbol{x}}^\epsilon(s|\boldsymbol{\theta},\boldsymbol{\sigma}), \boldsymbol{\lambda}^\epsilon(s|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}\big)}{\partial \theta_k} ds, \qquad k = 1,\dots,m,$$

where $\boldsymbol{\lambda}^\epsilon(m^+|\boldsymbol{\theta},\boldsymbol{\sigma}) \triangleq \mathbf{0}$, $\boldsymbol{z}^m \triangleq \mathbf{0}$, *and*

$$\hat{\rho}_{k,m-1} \triangleq \begin{cases} 1, & \text{if } k \le m-1, \\ 0, & \text{otherwise.} \end{cases}$$

Proof. Let $(\boldsymbol{\theta},\boldsymbol{\sigma}) \in \Theta \times \Xi$ be arbitrary but fixed. Furthermore, let $\boldsymbol{v} : [0,m] \to \mathbb{R}^n$ be an arbitrary piecewise continuous function that is differentiable on $(k-1,k)$, $k = 1,\dots,m$. We will simplify the notation in this proof by writing $\tilde{\boldsymbol{x}}^\epsilon(\cdot)$ instead of $\tilde{\boldsymbol{x}}^\epsilon(\cdot|\boldsymbol{\theta},\boldsymbol{\sigma})$.

Using the Hamiltonian, we can write $\tilde{J}_{\epsilon,\vartheta}(\boldsymbol{\theta},\boldsymbol{\sigma})$ as

$$\tilde{J}_{\epsilon,\vartheta}(\boldsymbol{\theta},\boldsymbol{\sigma}) = \sum_{l=1}^{m} \int_{l-1}^{l} \Big\{H_l\big(\theta_l, \tilde{\boldsymbol{x}}^\epsilon(s), \boldsymbol{v}(s), \boldsymbol{\sigma}\big) - \big[\boldsymbol{v}(s)\big]^T \dot{\tilde{\boldsymbol{x}}}^\epsilon(s)\Big\} ds + \vartheta \sum_{l=1}^{m-1} \varphi_\epsilon(\theta_l)\big(1-\varphi_\epsilon(\theta_l)\big).$$

Applying integration by parts to the last term in the integrand yields

$$\begin{aligned}\tilde{J}_{\epsilon,\vartheta}(\boldsymbol{\theta},\boldsymbol{\sigma}) = \sum_{l=1}^{m} \int_{l-1}^{l} \Big\{H_l\big(\theta_l, \tilde{\boldsymbol{x}}^\epsilon(s), \boldsymbol{v}(s), \boldsymbol{\sigma}\big) + \big[\dot{\boldsymbol{v}}(s)\big]^T \tilde{\boldsymbol{x}}^\epsilon(s)\Big\} ds \\ + \sum_{l=1}^{m-1} \big[\boldsymbol{v}(l^+)\big]^T \tilde{\boldsymbol{x}}^\epsilon(l^+) - \sum_{l=1}^{m-1} \big[\boldsymbol{v}(l^-)\big]^T \tilde{\boldsymbol{x}}^\epsilon(l^-) + \big[\boldsymbol{v}(0^+)\big]^T \boldsymbol{x}^0 \\ - \big[\boldsymbol{v}(m^-)\big]^T \tilde{\boldsymbol{x}}^\epsilon(m^-) + \vartheta \sum_{l=1}^{m-1} \varphi_\epsilon(\theta_l)\big(1-\varphi_\epsilon(\theta_l)\big).\end{aligned}$$

Substituting the state jump conditions (5.16b) into the above equation gives

$$\begin{aligned}\tilde{J}_{\epsilon,\vartheta}(\boldsymbol{\theta},\boldsymbol{\sigma}) = \sum_{l=1}^{m} \int_{l-1}^{l} \Big\{H_l\big(\theta_l, \tilde{\boldsymbol{x}}^\epsilon(s), \boldsymbol{v}(s), \boldsymbol{\sigma}\big) + \big[\dot{\boldsymbol{v}}(s)\big]^T \tilde{\boldsymbol{x}}^\epsilon(s)\Big\} ds \\ + \sum_{l=1}^{m-1} \big[\boldsymbol{v}(l^+)\big]^T \Big\{\tilde{\boldsymbol{x}}^\epsilon(l^-) + \varphi_\epsilon(\theta_l)\boldsymbol{z}^l\big(\tilde{\boldsymbol{x}}^\epsilon(l^-),\boldsymbol{\sigma}\big)\Big\} - \sum_{l=1}^{m-1} \big[\boldsymbol{v}(l^-)\big]^T \tilde{\boldsymbol{x}}^\epsilon(l^-) \\ + \big[\boldsymbol{v}(0^+)\big]^T \boldsymbol{x}^0 - \big[\boldsymbol{v}(m^-)\big]^T \tilde{\boldsymbol{x}}^\epsilon(m^-) + \vartheta \sum_{l=1}^{m-1} \varphi_\epsilon(\theta_l)\big(1-\varphi_\epsilon(\theta_l)\big). \quad (5.19)\end{aligned}$$

Now, using the cumulative Kronecker delta $\hat{\rho}_{l,m-1}$, we can rewrite this equation as follows:

$$\begin{aligned}\tilde{J}_{\epsilon,\vartheta}(\boldsymbol{\theta},\boldsymbol{\sigma}) = \sum_{l=1}^{m}\int_{l-1}^{l}\Big\{H_l\big(\theta_l,\tilde{\boldsymbol{x}}^{\epsilon}(s),\boldsymbol{v}(s),\boldsymbol{\sigma}\big)+\big[\dot{\boldsymbol{v}}(s)\big]^T\tilde{\boldsymbol{x}}^{\epsilon}(s)\Big\}ds \\ +\sum_{l=1}^{m}\hat{\rho}_{l,m-1}\big[\boldsymbol{v}(l^+)\big]^T\Big\{\tilde{\boldsymbol{x}}^{\epsilon}(l^-)+\varphi_{\epsilon}(\theta_l)\boldsymbol{z}^l\big(\tilde{\boldsymbol{x}}^{\epsilon}(l^-),\boldsymbol{\sigma}\big)\Big\} \\ -\sum_{l=1}^{m}\hat{\rho}_{l,m-1}\big[\boldsymbol{v}(l^-)\big]^T\tilde{\boldsymbol{x}}^{\epsilon}(l^-)+\big[\boldsymbol{v}(0^+)\big]^T\boldsymbol{x}^0-\big[\boldsymbol{v}(m^-)\big]^T\tilde{\boldsymbol{x}}^{\epsilon}(m^-), \\ +\vartheta\sum_{l=1}^{m}\hat{\rho}_{l,m-1}\varphi_{\epsilon}(\theta_l)\big(1-\varphi_{\epsilon}(\theta_l)\big),\end{aligned} \tag{5.20}$$

where $\boldsymbol{v}(m^+)\triangleq\boldsymbol{v}(m)$. Differentiating equation (5.20) with respect to θ_k, $k=1,\ldots,m$, yields

$$\begin{aligned}\frac{\partial\tilde{J}_{\epsilon,\vartheta}(\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial\theta_k} = \sum_{l=1}^{m}\int_{l-1}^{l}\left\{\frac{\partial H_l\big(\theta_l,\tilde{\boldsymbol{x}}^{\epsilon}(s),\boldsymbol{v}(s),\boldsymbol{\sigma}\big)}{\partial\boldsymbol{x}}\frac{\partial\tilde{\boldsymbol{x}}^{\epsilon}(s)}{\partial\theta_k}+\big[\dot{\boldsymbol{v}}(s)\big]^T\frac{\partial\tilde{\boldsymbol{x}}^{\epsilon}(s)}{\partial\theta_k}\right\}ds \\ +\int_{k-1}^{k}\frac{\partial H_k\big(\theta_k,\tilde{\boldsymbol{x}}^{\epsilon}(s),\boldsymbol{v}(s),\boldsymbol{\sigma}\big)}{\partial\theta_k}ds-\sum_{l=1}^{m}\hat{\rho}_{l,m-1}\big[\boldsymbol{v}(l^-)\big]^T\frac{\partial\tilde{\boldsymbol{x}}^{\epsilon}(l^-)}{\partial\theta_k} \\ +\sum_{l=1}^{m}\hat{\rho}_{l,m-1}\big[\boldsymbol{v}(l^+)\big]^T\left\{\frac{\partial\tilde{\boldsymbol{x}}^{\epsilon}(l^-)}{\partial\theta_k}+\varphi_{\epsilon}(\theta_l)\frac{\partial\boldsymbol{z}^l\big(\tilde{\boldsymbol{x}}^{\epsilon}(l^-),\boldsymbol{\sigma}\big)}{\partial\boldsymbol{x}}\frac{\partial\tilde{\boldsymbol{x}}^{\epsilon}(l^-)}{\partial\theta_k}\right\} \\ +\hat{\rho}_{k,m-1}\dot{\varphi}_{\epsilon}(\theta_k)\big[\boldsymbol{v}(k^+)\big]^T\boldsymbol{z}^k\big(\tilde{\boldsymbol{x}}^{\epsilon}(k^-),\boldsymbol{\sigma}\big)-\big[\boldsymbol{v}(m^-)\big]^T\frac{\partial\tilde{\boldsymbol{x}}^{\epsilon}(m^-)}{\partial\theta_k} \\ +\hat{\rho}_{k,m-1}\Big\{\vartheta\dot{\varphi}_{\epsilon}(\theta_k)\big(1-\varphi_{\epsilon}(\theta_k)\big)-\vartheta\varphi_{\epsilon}(\theta_k)\dot{\varphi}_{\epsilon}(\theta_k)\Big\}.\end{aligned} \tag{5.21}$$

Since $\boldsymbol{v}$ was chosen arbitrarily, we can set $\boldsymbol{v}=\boldsymbol{\lambda}^{\epsilon}(\cdot|\boldsymbol{\theta},\boldsymbol{\sigma})$. Substituting the costate equations (5.17)-(5.18) into equation (5.21) completes the proof. □

Theorem 5.2. *For each* $(\boldsymbol{\theta},\boldsymbol{\sigma})\in\Theta\times\Xi$,

$$\begin{aligned}\frac{\partial\tilde{J}_{\epsilon,\vartheta}(\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial\sigma_{\varsigma}} = \sum_{k=1}^{m-1}\varphi_{\epsilon}(\theta_k)\big[\boldsymbol{\lambda}^{\epsilon}(k^+|\boldsymbol{\theta},\boldsymbol{\sigma})\big]^T\frac{\partial\boldsymbol{z}^k\big(\tilde{\boldsymbol{x}}^{\epsilon}(k^-|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}\big)}{\partial\sigma_{\varsigma}} \\ +\sum_{k=1}^{m}\int_{k-1}^{k}\frac{\partial H_k\big(\theta_k,\tilde{\boldsymbol{x}}^{\epsilon}(s|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\lambda}^{\epsilon}(s|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}\big)}{\partial\sigma_{\varsigma}}ds,\quad\varsigma=1,\ldots,r.\end{aligned}$$

Proof. Let $(\boldsymbol{\theta},\boldsymbol{\sigma})\in\Theta\times\Xi$ be arbitrary but fixed, and let $\boldsymbol{v}:[0,m]\to\mathbb{R}^n$ be an arbitrary piecewise continuous function that is differentiable on $(k-1,k)$, $k=1,\ldots,m$. As in the proof of Theorem 5.1, we simplify the notation by writing $\tilde{\boldsymbol{x}}^{\epsilon}(\cdot)$ instead of $\tilde{\boldsymbol{x}}^{\epsilon}(\cdot|\boldsymbol{\theta},\boldsymbol{\sigma})$.

First, recall equation (5.19) from the proof of Theorem 5.1:

$$\begin{aligned}\tilde{J}_{\epsilon,\vartheta}(\boldsymbol{\theta},\boldsymbol{\sigma}) = \sum_{k=1}^{m}\int_{k-1}^{k}\Big\{H_k\big(\theta_k,\tilde{\boldsymbol{x}}^{\epsilon}(s),\boldsymbol{v}(s),\boldsymbol{\sigma}\big)+\big[\dot{\boldsymbol{v}}(s)\big]^T\tilde{\boldsymbol{x}}^{\epsilon}(s)\Big\}ds \\ +\sum_{k=1}^{m-1}\big[\boldsymbol{v}(k^+)\big]^T\Big\{\tilde{\boldsymbol{x}}^{\epsilon}(k^-)+\varphi_{\epsilon}(\theta_k)\boldsymbol{z}^k\big(\tilde{\boldsymbol{x}}^{\epsilon}(k^-),\boldsymbol{\sigma}\big)\Big\}-\sum_{k=1}^{m-1}\big[\boldsymbol{v}(k^-)\big]^T\tilde{\boldsymbol{x}}^{\epsilon}(k^-) \\ +\big[\boldsymbol{v}(0^+)\big]^T\boldsymbol{x}^0-\big[\boldsymbol{v}(m^-)\big]^T\tilde{\boldsymbol{x}}^{\epsilon}(m^-)+\vartheta\sum_{k=1}^{m-1}\varphi_{\epsilon}(\theta_k)\big(1-\varphi_{\epsilon}(\theta_k)\big).\end{aligned}$$

Differentiating this equation with respect to σ_ς, $\varsigma = 1,\ldots,r$, gives

$$\begin{aligned}\frac{\partial\tilde{J}_{\epsilon,\vartheta}(\boldsymbol{\theta},\boldsymbol{\sigma})}{\partial\sigma_\varsigma} = \sum_{k=1}^{m}\int_{k-1}^{k}\left\{\frac{\partial H_k\big(\theta_k,\tilde{\boldsymbol{x}}^{\epsilon}(s),\boldsymbol{v}(s),\boldsymbol{\sigma}\big)}{\partial\boldsymbol{x}}\frac{\partial\tilde{\boldsymbol{x}}^{\epsilon}(s)}{\partial\sigma_\varsigma}+\big[\dot{\boldsymbol{v}}(s)\big]^T\frac{\partial\tilde{\boldsymbol{x}}^{\epsilon}(s)}{\partial\sigma_\varsigma}\right\}ds \\ +\sum_{k=1}^{m}\int_{k-1}^{k}\frac{\partial H_k\big(\theta_k,\tilde{\boldsymbol{x}}^{\epsilon}(s),\boldsymbol{v}(s),\boldsymbol{\sigma}\big)}{\partial\sigma_\varsigma}ds-\sum_{k=1}^{m-1}\big[\boldsymbol{v}(k^-)\big]^T\frac{\partial\tilde{\boldsymbol{x}}^{\epsilon}(k^-)}{\partial\sigma_\varsigma} \\ +\sum_{k=1}^{m-1}\big[\boldsymbol{v}(k^+)\big]^T\left\{\frac{\partial\tilde{\boldsymbol{x}}^{\epsilon}(k^-)}{\partial\sigma_\varsigma}+\varphi_{\epsilon}(\theta_k)\frac{\partial\boldsymbol{z}^k\big(\tilde{\boldsymbol{x}}^{\epsilon}(k^-),\boldsymbol{\sigma}\big)}{\partial\boldsymbol{x}}\frac{\partial\tilde{\boldsymbol{x}}^{\epsilon}(k^-)}{\partial\sigma_\varsigma}\right\} \\ -\big[\boldsymbol{v}(m^-)\big]^T\frac{\partial\tilde{\boldsymbol{x}}^{\epsilon}(m^-)}{\partial\sigma_\varsigma}+\sum_{k=1}^{m-1}\varphi_{\epsilon}(\theta_k)\big[\boldsymbol{v}(k^+)\big]^T\frac{\partial\boldsymbol{z}^k\big(\tilde{\boldsymbol{x}}^{\epsilon}(k^-),\boldsymbol{\sigma}\big)}{\partial\sigma_\varsigma}. \qquad (5.22)\end{aligned}$$

Setting $\boldsymbol{v} = \boldsymbol{\lambda}^{\epsilon}(\cdot|\boldsymbol{\theta},\boldsymbol{\sigma})$ and substituting the costate equations (5.17)-(5.18) into equation (5.22) establishes the result. □

Theorems 5.1 and 5.2 show that the partial derivatives of $\tilde{J}_{\epsilon,\vartheta}$ can be computed by solving the state and costate systems. Note that the state system must be solved first, because its solution appears on the right-hand side of (5.17)-(5.18).

The following algorithm can be used to compute the value and gradient of $\tilde{J}_{\epsilon,\vartheta}$ for a given pair $(\boldsymbol{\theta},\boldsymbol{\sigma}) \in \Theta \times \Xi$.

Algorithm 5.1. Input $(\boldsymbol{\theta},\boldsymbol{\sigma}) \in \Theta \times \Xi$.

(i) Obtain $\tilde{\boldsymbol{x}}^{\epsilon}(\cdot|\boldsymbol{\theta},\boldsymbol{\sigma})$ by solving the switched system consisting of (5.12) and (5.16).

(ii) Use $\tilde{\boldsymbol{x}}^{\epsilon}(\cdot|\boldsymbol{\theta},\boldsymbol{\sigma})$ to compute $\tilde{J}_{\epsilon,\vartheta}(\boldsymbol{\theta},\boldsymbol{\sigma})$.

(iii) Obtain $\boldsymbol{\lambda}^{\epsilon}(\cdot|\boldsymbol{\theta},\boldsymbol{\sigma})$ by solving the costate system (5.17)-(5.18).

(iv) Use $\tilde{\boldsymbol{x}}^{\epsilon}(\cdot|\boldsymbol{\theta},\boldsymbol{\sigma})$ and $\boldsymbol{\lambda}^{\epsilon}(\cdot|\boldsymbol{\theta},\boldsymbol{\sigma})$ to compute the partial derivatives $\partial\tilde{J}_{\epsilon,\vartheta}(\boldsymbol{\theta},\boldsymbol{\sigma})/\partial\theta_k$, $k = 1,\ldots,m$, and $\partial\tilde{J}_{\epsilon,\vartheta}(\boldsymbol{\theta},\boldsymbol{\sigma})/\partial\sigma_\varsigma$, $\varsigma = 1,\ldots,r$, according to the formulae in Theorems 5.1 and 5.2.

Algorithm 5.1 can be used in conjunction with a gradient-based global optimization technique, such as the filled function method (see [67, 128–130, 134]), to solve Problem $\tilde{\mathrm{P}}_{\epsilon,\vartheta}$.

Alternatively, Problem $\tilde{\mathrm{P}}_{\epsilon,\vartheta}$ can be solved by repeatedly applying a nonlinear programming algorithm from different starting points, and selecting the best local solution that is obtained.

It is important to note that $\tilde{J}_{\epsilon,\vartheta}$ is non-convex in general. In fact, since $\tilde{J}_{\epsilon,\vartheta}$ contains a penalty term that approximates a discontinuous function, Problem $\tilde{\mathrm{P}}_{\epsilon,\vartheta}$ likely has many local solutions. Accordingly, it is imperative that a global optimization strategy, such as those suggested above, be used to solve Problem $\tilde{\mathrm{P}}_{\epsilon,\vartheta}$.

5.5 Convergence results and algorithm

In the previous section, we introduced Problem $\tilde{\mathrm{P}}_{\epsilon,\vartheta}$ and showed that it can be solved using existing optimization techniques. We will now investigate the relationship between Problems $\tilde{\mathrm{P}}$ and $\tilde{\mathrm{P}}_{\epsilon,\vartheta}$.

We begin by establishing some preliminary results.

Lemma 5.1. *There exists a real number $L_2 > 0$ such that*

$$\big|\tilde{\boldsymbol{x}}^{\epsilon}(s|\boldsymbol{\theta},\boldsymbol{\sigma})\big| \leq L_2, \qquad s \in [0,m], \qquad (\boldsymbol{\theta},\boldsymbol{\sigma}) \in \Theta \times \Xi, \qquad \epsilon > 0, \tag{5.23}$$

and

$$\big|\boldsymbol{\lambda}^{\epsilon}(s|\boldsymbol{\theta},\boldsymbol{\sigma})\big| \leq L_2, \qquad s \in [0,m], \qquad (\boldsymbol{\theta},\boldsymbol{\sigma}) \in \Theta \times \Xi, \qquad \epsilon > 0. \tag{5.24}$$

Proof. Let $\epsilon > 0$ and $(\boldsymbol{\theta},\boldsymbol{\sigma}) \in \Theta \times \Xi$ be arbitrary but fixed. It follows from (5.12) and (5.16a) that

$$\tilde{\boldsymbol{x}}^{\epsilon}(s|\boldsymbol{\theta},\boldsymbol{\sigma}) = \boldsymbol{x}^0 + \int_0^s \theta_1 \boldsymbol{f}^1\big(\tilde{\boldsymbol{x}}^{\epsilon}(\eta|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma}\big)d\eta, \qquad s \in [0,1).$$

Taking the norm of both sides gives

$$\big|\tilde{\boldsymbol{x}}^{\epsilon}(s|\boldsymbol{\theta},\boldsymbol{\sigma})\big| \leq \big|\boldsymbol{x}^0\big| + \int_0^s T\big|\boldsymbol{f}^1(\tilde{\boldsymbol{x}}^{\epsilon}(\eta|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma})\big|d\eta, \qquad s \in [0,1).$$

By using Assumption 5.2, we obtain

$$\big|\tilde{\boldsymbol{x}}^{\epsilon}(s|\boldsymbol{\theta},\boldsymbol{\sigma})\big| \leq \big|\boldsymbol{x}^0\big| + L_1T + \int_0^s L_1T\big|\tilde{\boldsymbol{x}}^{\epsilon}(\eta|\boldsymbol{\theta},\boldsymbol{\sigma})\big|d\eta, \qquad s \in [0,1).$$

Thus, by Gronwall's Lemma,

$$\big|\tilde{\boldsymbol{x}}^{\epsilon}(s|\boldsymbol{\theta},\boldsymbol{\sigma})\big| \leq \alpha_1 \exp(L_1T), \qquad s \in [0,1), \tag{5.25}$$

where

$$\alpha_1 \triangleq \big|\boldsymbol{x}^0\big| + L_1T.$$

In particular,

$$\big|\tilde{\boldsymbol{x}}^\epsilon(1^-|\boldsymbol{\theta},\boldsymbol{\sigma})\big| \le \alpha_1 \exp(L_1 T). \tag{5.26}$$

Now, define the set

$$\mathcal{A} \triangleq \big\{\, \boldsymbol{v} \in \mathbb{R}^n :\ |\boldsymbol{v}| \le \alpha_1 \exp(L_1 T) \,\big\}.$$

Since $\mathcal{A}$ and Ξ are compact sets, it follows from Assumption 5.1 that there exists a real number $\alpha_2 > 0$ such that

$$\sup_{(\boldsymbol{v},\boldsymbol{w})\in\mathcal{A}\times\Xi} \big|\boldsymbol{z}^1(\boldsymbol{v},\boldsymbol{w})\big| \le \alpha_2.$$

By taking the norm of both sides of (5.16b) (with $k = 1$) and then using (5.26), we obtain

$$\begin{aligned}\big|\tilde{\boldsymbol{x}}^\epsilon(1|\boldsymbol{\theta},\boldsymbol{\sigma})\big| = \big|\tilde{\boldsymbol{x}}^\epsilon(1^+|\boldsymbol{\theta},\boldsymbol{\sigma})\big| &\le \big|\tilde{\boldsymbol{x}}^\epsilon(1^-|\boldsymbol{\theta},\boldsymbol{\sigma})\big| + \big|\varphi_\epsilon(\theta_1)\big| \cdot \big|\boldsymbol{z}^1(\tilde{\boldsymbol{x}}^\epsilon(1^-|\boldsymbol{\theta},\boldsymbol{\sigma}),\boldsymbol{\sigma})\big| \\ &\le \alpha_1 \exp(L_1 T) + \alpha_2. \end{aligned} \tag{5.27}$$

Since $\epsilon > 0$ and $(\boldsymbol{\theta},\boldsymbol{\sigma}) \in \Theta \times \Xi$ were chosen arbitrarily, inequalities (5.25) and (5.27) show that

$$\big|\tilde{\boldsymbol{x}}^\epsilon(s|\boldsymbol{\theta},\boldsymbol{\sigma})\big| \le \alpha_1 \exp(L_1 T) + \alpha_2, \qquad s \in [0,1], \qquad (\boldsymbol{\theta},\boldsymbol{\sigma}) \in \Theta \times \Xi, \qquad \epsilon > 0.$$

The arguments used to establish this inequality can be repeated for $s \in [1,2]$, $s \in [2,3]$, and so on for the remainder of the time horizon, which ultimately proves inequality (5.23). Inequality (5.24) is proved in a similar manner. □

Note that Lemma 5.1 can be restated as follows: the family of functions

$$\big\{\, \tilde{\boldsymbol{x}}^\epsilon(\cdot|\boldsymbol{\theta},\boldsymbol{\sigma}),\ \boldsymbol{\lambda}^\epsilon(\cdot|\boldsymbol{\theta},\boldsymbol{\sigma}) :\ (\boldsymbol{\theta},\boldsymbol{\sigma}) \in \Theta \times \Xi,\ \epsilon > 0 \,\big\}$$

is equibounded on $[0, m]$.

Our next result is proved below.

Lemma 5.2. *There exists a function $\gamma : (0,\infty) \to (0,\infty)$ with the following properties:*

(i) *γ is of order $\mathcal{O}(1/\epsilon)$; and*

(ii) *For all $\boldsymbol{\theta}', \boldsymbol{\theta}'' \in \Theta$ and $\boldsymbol{\sigma} \in \Xi$,*

$$\big|\tilde{G}_0^\epsilon(\boldsymbol{\theta}',\boldsymbol{\sigma}) - \tilde{G}_0^\epsilon(\boldsymbol{\theta}'',\boldsymbol{\sigma})\big| \le \gamma(\epsilon)\big|\boldsymbol{\theta}' - \boldsymbol{\theta}''\big|.$$

Proof. Let

$$\Psi \triangleq \big\{\, \boldsymbol{v} \in \mathbb{R}^n : |\boldsymbol{v}| \le L_2 \,\big\},$$

where L_2 is as defined in Lemma 5.1. It follows from Assumptions 5.1 and 5.3 that we

can find two real numbers $\beta_1 > 0$ and $\beta_2 > 0$ such that

$$\sup\Big\{\,\big|\mathcal{L}_k(\boldsymbol{v},\boldsymbol{\sigma}) + \boldsymbol{\lambda}^T\boldsymbol{f}^k(\boldsymbol{v},\boldsymbol{\sigma})\big| : (\boldsymbol{v},\boldsymbol{\lambda},\boldsymbol{\sigma}) \in \Psi\times\Psi\times\Xi,\ k\in\{1,\ldots,m\}\,\Big\} \le \beta_1 \tag{5.28}$$

and

$$\sup\Big\{\,\big|\boldsymbol{z}^k(\boldsymbol{v},\boldsymbol{\sigma})\big| :\ (\boldsymbol{v},\boldsymbol{\sigma}) \in \Psi\times\Xi,\ k\in\{1,\ldots,m-1\}\,\Big\} \le \beta_2. \tag{5.29}$$

Define the function $\gamma : (0,\infty) \to (0,\infty)$ as follows:

$$\gamma(\epsilon) \triangleq \sqrt{m}\left[\beta_1 + \frac{3\beta_2 L_2}{2\epsilon}\right], \qquad \epsilon > 0.$$

Clearly, γ is of order $\mathcal{O}(1/\epsilon)$. It remains to show that γ also satisfies property (ii).

Let $\boldsymbol{\theta}', \boldsymbol{\theta}'' \in \Theta$, $\boldsymbol{\sigma} \in \Xi$, and $\epsilon > 0$ be arbitrary but fixed. By Taylor's Theorem, there exists a $\bar{\boldsymbol{\theta}} \in \Theta$ such that

$$\tilde{G}_0^\epsilon(\boldsymbol{\theta}',\boldsymbol{\sigma}) - \tilde{G}_0^\epsilon(\boldsymbol{\theta}'',\boldsymbol{\sigma}) = \frac{\partial \tilde{G}_0^\epsilon(\bar{\boldsymbol{\theta}},\boldsymbol{\sigma})}{\partial \boldsymbol{\theta}}(\boldsymbol{\theta}' - \boldsymbol{\theta}''). \tag{5.30}$$

By Theorem 5.1, we have

$$\begin{aligned}\frac{\partial \tilde{G}_0^\epsilon(\bar{\boldsymbol{\theta}},\boldsymbol{\sigma})}{\partial \theta_k} &= \hat{\rho}_{k,m-1}\dot{\varphi}_\epsilon(\bar{\theta}_k)\big[\boldsymbol{\lambda}^\epsilon(k^+|\bar{\boldsymbol{\theta}},\boldsymbol{\sigma})\big]^T\boldsymbol{z}^k\big(\tilde{\boldsymbol{x}}^\epsilon(k^-|\bar{\boldsymbol{\theta}},\boldsymbol{\sigma}),\boldsymbol{\sigma}\big)\\ &\quad + \int_{k-1}^{k}\frac{\partial H_k\big(\bar{\theta}_k,\tilde{\boldsymbol{x}}^\epsilon(s|\bar{\boldsymbol{\theta}},\boldsymbol{\sigma}),\boldsymbol{\lambda}^\epsilon(s|\bar{\boldsymbol{\theta}},\boldsymbol{\sigma}),\boldsymbol{\sigma}\big)}{\partial \theta_k}ds, \quad k = 1,\ldots,m.\end{aligned} \tag{5.31}$$

Taking the norm of both sides in (5.31) yields

$$\begin{aligned}\left|\frac{\partial \tilde{G}_0^\epsilon(\bar{\boldsymbol{\theta}},\boldsymbol{\sigma})}{\partial \theta_k}\right| &\le \big|\dot{\varphi}_\epsilon(\bar{\theta}_k)\big|\cdot\big|\boldsymbol{\lambda}^\epsilon(k^+|\bar{\boldsymbol{\theta}},\boldsymbol{\sigma})\big|\cdot\big|\boldsymbol{z}^k(\tilde{\boldsymbol{x}}^\epsilon(k^-|\bar{\boldsymbol{\theta}},\boldsymbol{\sigma}),\boldsymbol{\sigma})\big|\\ &\quad + \int_{k-1}^{k}\left|\frac{\partial H_k\big(\bar{\theta}_k,\tilde{\boldsymbol{x}}^\epsilon(s|\bar{\boldsymbol{\theta}},\boldsymbol{\sigma}),\boldsymbol{\lambda}^\epsilon(s|\bar{\boldsymbol{\theta}},\boldsymbol{\sigma}),\boldsymbol{\sigma}\big)}{\partial \theta_k}\right|ds, \qquad k = 1,\ldots,m.\end{aligned}$$

Therefore, by Lemma 5.1 and (5.28)-(5.29),

$$\left|\frac{\partial \tilde{G}_0^\epsilon(\bar{\boldsymbol{\theta}},\boldsymbol{\sigma})}{\partial \theta_k}\right| \le \beta_1 + \beta_2 L_2\big|\dot{\varphi}_\epsilon(\bar{\theta}_k)\big|, \qquad k = 1,\ldots,m. \tag{5.32}$$

It is easy to see that

$$\big|\dot{\varphi}_\epsilon(\eta)\big| = \dot{\varphi}_\epsilon(\eta) \le \frac{3}{2\epsilon}, \qquad \eta \ge 0.$$

Hence, (5.32) becomes

$$\left|\frac{\partial \tilde{G}_0^\epsilon(\bar{\boldsymbol{\theta}},\boldsymbol{\sigma})}{\partial \theta_k}\right| \le \beta_1 + \frac{3\beta_2 L_2}{2\epsilon}, \qquad k = 1,\ldots,m.$$

Therefore,

$$\left|\frac{\partial \tilde{G}_0^\epsilon(\bar{\boldsymbol{\theta}}, \boldsymbol{\sigma})}{\partial \boldsymbol{\theta}}\right| \leq \sqrt{m}\left[\beta_1 + \frac{3\beta_2 L_2}{2\epsilon}\right] = \gamma(\epsilon). \tag{5.33}$$

Finally, by taking the norm of (5.30) and then using (5.33), we obtain

$$\left|\tilde{G}_0^\epsilon(\boldsymbol{\theta}', \boldsymbol{\sigma}) - \tilde{G}_0^\epsilon(\boldsymbol{\theta}'', \boldsymbol{\sigma})\right| \leq \gamma(\epsilon)\left|\boldsymbol{\theta}' - \boldsymbol{\theta}''\right|,$$

as required. □

Now, define

$$\hat{\epsilon} \triangleq \min\left\{1, T/m^2\right\}.$$

For the remainder of this section, $(\boldsymbol{\theta}^{\epsilon,\vartheta,*}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*}) \in \Theta \times \Xi$ denotes an optimal solution of Problem $\tilde{\mathrm{P}}_{\epsilon,\vartheta}$.

Theorem 5.3. *For each $\epsilon \in (0, \hat{\epsilon})$, there exists a corresponding real number $\vartheta(\epsilon) > 0$ such that if $\vartheta > \vartheta(\epsilon)$, then*

$$\theta_k^{\epsilon,\vartheta,*} < \frac{\epsilon^{3/2}}{2} \qquad \text{or} \qquad \theta_k^{\epsilon,\vartheta,*} > \epsilon - \frac{\epsilon^{3/2}}{2}, \qquad k = 1, \ldots, m-1. \tag{5.34}$$

Proof. Let $\epsilon \in (0, \hat{\epsilon})$. Furthermore, let $(\boldsymbol{\theta}^{\epsilon,*}, \boldsymbol{\sigma}^{\epsilon,*}) \in \Theta \times \Xi$ denote a minimizer of $\tilde{G}_0^\epsilon$ on $\Theta \times \Xi$. Then

$$\tilde{G}_0^\epsilon(\boldsymbol{\theta}^{\epsilon,*}, \boldsymbol{\sigma}^{\epsilon,*}) \leq \tilde{G}_0^\epsilon(\boldsymbol{\theta}^{\epsilon,\vartheta,*}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*}), \qquad \vartheta > 0.$$

By appending the penalty term in $\tilde{J}_{\epsilon,\vartheta}$ to both sides of this inequality, we obtain

$$\tilde{G}_0^\epsilon(\boldsymbol{\theta}^{\epsilon,*}, \boldsymbol{\sigma}^{\epsilon,*}) + \vartheta \sum_{k=1}^{m-1} \varphi_\epsilon(\theta_k^{\epsilon,\vartheta,*})\big(1 - \varphi_\epsilon(\theta_k^{\epsilon,\vartheta,*})\big) \leq \tilde{J}_{\epsilon,\vartheta}(\boldsymbol{\theta}^{\epsilon,\vartheta,*}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*}), \qquad \vartheta > 0.$$

Thus,

$$\tilde{G}_0^\epsilon(\boldsymbol{\theta}^{\epsilon,*}, \boldsymbol{\sigma}^{\epsilon,*}) + \vartheta \sum_{k=1}^{m-1} \varphi_\epsilon(\theta_k^{\epsilon,\vartheta,*})\big(1 - \varphi_\epsilon(\theta_k^{\epsilon,\vartheta,*})\big) \leq \tilde{J}_{\epsilon,\vartheta}(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\sigma}}), \qquad \vartheta > 0, \tag{5.35}$$

where $\bar{\boldsymbol{\sigma}} \in \Xi$ and

$$\bar{\theta}_k \triangleq \frac{T}{m}, \qquad k = 1, \ldots, m.$$

Since $\epsilon < \hat{\epsilon} < T/m$,

$$\varphi_\epsilon(\bar{\theta}_k) = 1, \qquad k = 1, \ldots, m-1.$$

Thus, the penalty term in $\tilde{J}_{\epsilon,\vartheta}(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\sigma}})$ vanishes and so inequality (5.35) becomes

$$\tilde{G}_0^\epsilon(\boldsymbol{\theta}^{\epsilon,*}, \boldsymbol{\sigma}^{\epsilon,*}) + \vartheta \sum_{k=1}^{m-1} \varphi_\epsilon(\theta_k^{\epsilon,\vartheta,*})\big(1 - \varphi_\epsilon(\theta_k^{\epsilon,\vartheta,*})\big) \leq \tilde{G}_0^\epsilon(\bar{\boldsymbol{\theta}}, \bar{\boldsymbol{\sigma}}), \qquad \vartheta > 0.$$

Rearranging this inequality gives

$$\varphi_\epsilon(\theta_k^{\epsilon,\vartheta,*})\big(1-\varphi_\epsilon(\theta_k^{\epsilon,\vartheta,*})\big) \le \frac{\tilde{G}_0^\epsilon(\bar{\boldsymbol{\theta}},\bar{\boldsymbol{\sigma}}) - \tilde{G}_0^\epsilon(\boldsymbol{\theta}^{\epsilon,*},\boldsymbol{\sigma}^{\epsilon,*})}{\vartheta}, \quad k=1,\dots,m-1, \quad \vartheta>0. \tag{5.36}$$

Since $(\boldsymbol{\theta}^{\epsilon,*},\boldsymbol{\sigma}^{\epsilon,*})$ is a minimizer of $\tilde{G}_0^\epsilon$, the right-hand side of (5.36) is non-negative. Hence, it is clear that increasing ϑ forces the penalty term on the left-hand side to approach zero. Thus, it follows that (5.34) is satisfied when ϑ is sufficiently large. □

Theorem 5.4. *Let $\epsilon \in (0,\hat{\epsilon})$ and $\vartheta > \vartheta(\epsilon)$, where $\vartheta(\epsilon)$ is as defined in Theorem 5.3. Then there exists a corresponding vector $\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta} \in \Theta$ such that*

$$\big|\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta} - \boldsymbol{\theta}^{\epsilon,\vartheta,*}\big| \le (m-1)\epsilon^{3/2} \tag{5.37}$$

and

$$\varphi_\epsilon(\hat{\theta}_k^{\epsilon,\vartheta}) = \varphi(\hat{\theta}_k^{\epsilon,\vartheta}), \qquad k=1,\dots,m-1. \tag{5.38}$$

Proof. Define

$$\mathcal{G}_1^{\epsilon,\vartheta} \triangleq \big\{\, k\in\{1,\dots,m-1\} :\ 0 \le \theta_k^{\epsilon,\vartheta,*} < \epsilon^{3/2}/2 \,\big\} \tag{5.39}$$

and

$$\mathcal{G}_2^{\epsilon,\vartheta} \triangleq \big\{\, k\in\{1,\dots,m-1\} :\ \epsilon - \epsilon^{3/2}/2 < \theta_k^{\epsilon,\vartheta,*} \le \epsilon \,\big\}. \tag{5.40}$$

Since $\epsilon < \hat{\epsilon} \le 1$,

$$\frac{\epsilon^{3/2}}{2} < \epsilon - \frac{\epsilon^{3/2}}{2}.$$

Thus, the index sets $\mathcal{G}_1^{\epsilon,\vartheta}$ and $\mathcal{G}_2^{\epsilon,\vartheta}$ are disjoint.

Now, let $\kappa(\epsilon,\vartheta) \in \{1,\dots,m\}$ be such that

$$\theta_{\kappa(\epsilon,\vartheta)}^{\epsilon,\vartheta,*} = \max_{1\le k\le m} \theta_k^{\epsilon,\vartheta,*}.$$

Since $\boldsymbol{\theta}^{\epsilon,\vartheta,*} \in \Theta$,

$$\theta_{\kappa(\epsilon,\vartheta)}^{\epsilon,\vartheta,*} \ge \frac{T}{m} > \hat{\epsilon} > \epsilon. \tag{5.41}$$

Therefore,

$$\kappa(\epsilon,\vartheta) \notin \mathcal{G}_1^{\epsilon,\vartheta} \cup \mathcal{G}_2^{\epsilon,\vartheta}.$$

Now, for each $k=1,\dots,m$, define

$$\hat{\theta}_k^{\epsilon,\vartheta} \triangleq \begin{cases} 0, & \text{if } k \in \mathcal{G}_1^{\epsilon,\vartheta}, \\ \epsilon, & \text{if } k \in \mathcal{G}_2^{\epsilon,\vartheta}, \\ \theta_k^{\epsilon,\vartheta,*} - \alpha_{\epsilon,\vartheta}, & \text{if } k = \kappa(\epsilon,\vartheta), \\ \theta_k^{\epsilon,\vartheta,*}, & \text{otherwise}, \end{cases}$$

where

$$\alpha_{\epsilon,\vartheta} \triangleq \sum_{k\in\mathcal{G}_2^{\epsilon,\vartheta}} \left(\epsilon - \theta_k^{\epsilon,\vartheta,*}\right) - \sum_{k\in\mathcal{G}_1^{\epsilon,\vartheta}} \theta_k^{\epsilon,\vartheta,*}.$$

We will show that

$$\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta} = \left[\hat{\theta}_1^{\epsilon,\vartheta}, \ldots, \hat{\theta}_m^{\epsilon,\vartheta}\right]^T \in \mathbb{R}^m$$

is an element of Θ that satisfies (5.37) and (5.38).

First, it follows from (5.39) and (5.40) that

$$|\alpha_{\epsilon,\vartheta}| < \frac{(m-1)\epsilon^{3/2}}{2}. \tag{5.42}$$

By using (5.41) and (5.42), we obtain

$$\hat{\theta}_{\kappa(\epsilon,\vartheta)}^{\epsilon,\vartheta} = \theta_{\kappa(\epsilon,\vartheta)}^{\epsilon,\vartheta,*} - \alpha_{\epsilon,\vartheta} > \frac{T}{m} - \frac{(m-1)\epsilon^{3/2}}{2} > \frac{T}{m} - (m-1)\epsilon^{3/2}. \tag{5.43}$$

Since $\epsilon < \hat{\epsilon}$,

$$m\epsilon < \frac{T}{m}. \tag{5.44}$$

Substituting (5.44) into (5.43) gives

$$\hat{\theta}_{\kappa(\epsilon,\vartheta)}^{\epsilon,\vartheta} > m\epsilon - (m-1)\epsilon^{3/2} = \epsilon + (m-1)(\epsilon - \epsilon^{3/2}) > \epsilon, \tag{5.45}$$

which shows that $\hat{\theta}_{\kappa(\epsilon,\vartheta)}^{\epsilon,\vartheta}$ is non-negative. The other components of $\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta}$ are clearly non-negative. Moreover,

$$\sum_{k=1}^{m} \hat{\theta}_k^{\epsilon,\vartheta} = \sum_{k\in\mathcal{G}_2^{\epsilon,\vartheta}} \epsilon + \theta_{\kappa(\epsilon,\vartheta)}^{\epsilon,\vartheta,*} - \alpha_{\epsilon,\vartheta} + \sum_{\substack{k\notin\mathcal{G}_1^{\epsilon,\vartheta}\cup\mathcal{G}_2^{\epsilon,\vartheta} \\ k\neq\kappa(\epsilon,\vartheta)}} \theta_k^{\epsilon,\vartheta,*} = \sum_{k=1}^{m} \theta_k^{\epsilon,\vartheta,*} = T. \tag{5.46}$$

Thus, $\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta} \in \Theta$.

Now,

$$\left|\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta} - \boldsymbol{\theta}^{\epsilon,\vartheta,*}\right|^2 = \alpha_{\epsilon,\vartheta}^2 + \sum_{k\in\mathcal{G}_1^{\epsilon,\vartheta}} \left(\theta_k^{\epsilon,\vartheta,*}\right)^2 + \sum_{k\in\mathcal{G}_2^{\epsilon,\vartheta}} \left(\epsilon - \theta_k^{\epsilon,\vartheta,*}\right)^2.$$

Hence, using (5.39), (5.40), and (5.42), we obtain

$$\left|\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta} - \boldsymbol{\theta}^{\epsilon,\vartheta,*}\right|^2 < \frac{(m-1)^2\epsilon^3}{4} + \sum_{k\in\mathcal{G}_1^{\epsilon,\vartheta}\cup\mathcal{G}_2^{\epsilon,\vartheta}} \frac{\epsilon^3}{4} \leq \frac{(m-1)^2\epsilon^3}{4} + \frac{(m-1)\epsilon^3}{4} < (m-1)^2\epsilon^3,$$

which proves (5.37).

Clearly,

$$\varphi_\epsilon(\hat{\theta}_k^{\epsilon,\vartheta}) = \varphi(\hat{\theta}_k^{\epsilon,\vartheta}), \qquad k \in \mathcal{G}_1^{\epsilon,\vartheta} \cup \mathcal{G}_2^{\epsilon,\vartheta}. \tag{5.47}$$

Furthermore, by inequality (5.45),

$$\varphi_\epsilon(\hat{\theta}_k^{\epsilon,\vartheta}) = \varphi(\hat{\theta}_k^{\epsilon,\vartheta}), \qquad k = \kappa(\epsilon,\vartheta). \tag{5.48}$$

Finally, since ϑ was chosen to ensure that (5.34) holds,

$$\theta_k^{\epsilon,\vartheta,*} > \epsilon, \qquad k \in \{1,\dots,m-1\} \setminus \big(\mathcal{G}_1^{\epsilon,\vartheta} \cup \mathcal{G}_2^{\epsilon,\vartheta}\big).$$

Therefore,

$$\varphi_\epsilon(\hat{\theta}_k^{\epsilon,\vartheta}) = \varphi_\epsilon(\theta_k^{\epsilon,\vartheta,*}) = 1 = \varphi(\hat{\theta}_k^{\epsilon,\vartheta}), \qquad k \in \{1,\dots,m-1\} \setminus \hat{\mathcal{G}}^{\epsilon,\vartheta}, \tag{5.49}$$

where

$$\hat{\mathcal{G}}^{\epsilon,\vartheta} \triangleq \{\kappa(\epsilon,\vartheta)\} \cup \mathcal{G}_1^{\epsilon,\vartheta} \cup \mathcal{G}_2^{\epsilon,\vartheta}.$$

Equations (5.47)-(5.49) prove equation (5.38). □

Note that the proof of Theorem 5.4 is constructive: the vector $\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta} = \big[\hat{\theta}_1^{\epsilon,\vartheta},\dots,\hat{\theta}_m^{\epsilon,\vartheta}\big]^T$ can be computed from a solution of Problem $\tilde{\mathrm{P}}_{\epsilon,\vartheta}$ via the following formula:

$$\hat{\theta}_k^{\epsilon,\vartheta} \triangleq \begin{cases} 0, & \text{if } k \in \mathcal{G}_1^{\epsilon,\vartheta}, \\ \epsilon, & \text{if } k \in \mathcal{G}_2^{\epsilon,\vartheta}, \\ \theta_k^{\epsilon,\vartheta,*} - \alpha_{\epsilon,\vartheta}, & \text{if } k = \kappa(\epsilon,\vartheta), \\ \theta_k^{\epsilon,\vartheta,*}, & \text{otherwise}, \end{cases} \tag{5.50}$$

where

$$\mathcal{G}_1^{\epsilon,\vartheta} = \big\{\, k \in \{1,\dots,m-1\} :\ 0 \le \theta_k^{\epsilon,\vartheta,*} < \epsilon^{3/2}/2 \,\big\},$$

$$\mathcal{G}_2^{\epsilon,\vartheta} = \big\{\, k \in \{1,\dots,m-1\} :\ \epsilon - \epsilon^{3/2}/2 < \theta_k^{\epsilon,\vartheta,*} \le \epsilon \,\big\},$$

$$\kappa(\epsilon,\vartheta) = \operatorname*{arg\,max}_{1\le k\le m} \theta_k^{\epsilon,\vartheta,*},$$

and

$$\alpha_{\epsilon,\vartheta} = \sum_{k\in\mathcal{G}_2^{\epsilon,\vartheta}} \big(\epsilon - \theta_k^{\epsilon,\vartheta,*}\big) - \sum_{k\in\mathcal{G}_1^{\epsilon,\vartheta}} \theta_k^{\epsilon,\vartheta,*}.$$

(Recall that $\mathcal{G}_1^{\epsilon,\vartheta}$ and $\mathcal{G}_2^{\epsilon,\vartheta}$ are disjoint and $\kappa(\epsilon,\vartheta) \notin \mathcal{G}_1^{\epsilon,\vartheta} \cup \mathcal{G}_2^{\epsilon,\vartheta}$.)

The next theorem is the main result of this section.

Theorem 5.5. *Suppose that $(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*) \in \Theta \times \Xi$ is an optimal solution of Problem $\tilde{P}$. For each $\epsilon \in (0, \hat{\epsilon})$, let ϑ be sufficiently large so that* (5.34) *is satisfied (Theorem 5.3 guarantees that this can always be done), and let $\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta}$ be as defined in Theorem 5.4. Then*

$$\lim_{\epsilon\to 0} \tilde{G}_0(\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*}) = \tilde{G}_0(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*).$$

Proof. Define

$$\bar{\epsilon} \triangleq \min_{1\le k\le m-1} \{\, \theta_k^* :\ \theta_k^* > 0 \,\}.$$

Let $\epsilon < \min(\hat{\epsilon}, \bar{\epsilon})$ be arbitrary. Then for each $k = 1, \ldots, m-1$, either

$$\theta_k^* > \epsilon$$

or

$$\theta_k^* = 0.$$

Hence,

$$\varphi_\epsilon(\theta_k^*) = \varphi(\theta_k^*), \qquad k = 1, \ldots, m-1. \tag{5.51}$$

Consequently, when $(\boldsymbol{\theta}, \boldsymbol{\sigma}) = (\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*)$, equations (5.15) and (5.16) coincide. This implies that

$$\tilde{\boldsymbol{x}}(s|\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*) = \tilde{\boldsymbol{x}}^\epsilon(s|\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*), \qquad s \in [0, m],$$

and

$$\tilde{G}_0(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*) = \tilde{G}_0^\epsilon(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*). \tag{5.52}$$

Equation (5.51) also shows that

$$\varphi_\epsilon(\theta_k^*) \in \{0, 1\}, \qquad k = 1, \ldots, m-1. \tag{5.53}$$

From equations (5.52) and (5.53), we obtain

$$\tilde{J}_{\epsilon,\vartheta}(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*) = \tilde{G}_0^\epsilon(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*) = \tilde{G}_0(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*). \tag{5.54}$$

Now, since the penalty term in $\tilde{J}_{\epsilon,\vartheta}$ is non-negative and the pair $(\boldsymbol{\theta}^{\epsilon,\vartheta,*}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*})$ is optimal for Problem $\tilde{\mathrm{P}}_{\epsilon,\vartheta}$, we have

$$\tilde{G}_0^\epsilon(\boldsymbol{\theta}^{\epsilon,\vartheta,*}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*}) \le \tilde{J}_{\epsilon,\vartheta}(\boldsymbol{\theta}^{\epsilon,\vartheta,*}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*}) \le \tilde{J}_{\epsilon,\vartheta}(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*).$$

Thus, by (5.54),

$$\tilde{G}_0^\epsilon(\boldsymbol{\theta}^{\epsilon,\vartheta,*}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*}) \le \tilde{J}_{\epsilon,\vartheta}(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*) = \tilde{G}_0(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*). \tag{5.55}$$

Furthermore, it follows from equation (5.38) that

$$\tilde{\boldsymbol{x}}(s|\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*}) = \tilde{\boldsymbol{x}}^\epsilon(s|\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*}), \qquad s \in [0, m],$$

and

$$\tilde{G}_0(\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*}) = \tilde{G}_0^\epsilon(\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*}). \tag{5.56}$$

We now combine inequality (5.55) with equation (5.56) to obtain

$$0 \leq \tilde{G}_0(\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*}) - \tilde{G}_0(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*) \leq \tilde{G}_0^{\epsilon}(\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*}) - \tilde{G}_0^{\epsilon}(\boldsymbol{\theta}^{\epsilon,\vartheta,*}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*}).$$

(The lower bound is zero here because $(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*)$ is a minimizer of $\tilde{G}_0$.) Thus, by Lemma 5.2,

$$0 \leq \tilde{G}_0(\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*}) - \tilde{G}_0(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*) \leq \gamma(\epsilon)\big|\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta} - \boldsymbol{\theta}^{\epsilon,\vartheta,*}\big|.$$

Applying inequality (5.37) yields

$$0 \leq \tilde{G}_0(\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*}) - \tilde{G}_0(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*) \leq (m-1)\epsilon^{3/2}\gamma(\epsilon), \tag{5.57}$$

Since γ is of order $\mathcal{O}(1/\epsilon)$, inequality (5.57) implies that

$$\tilde{G}_0(\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*}) - \tilde{G}_0(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*) = \mathcal{O}(\sqrt{\epsilon}).$$

Thus,

$$\lim_{\epsilon \to 0} \tilde{G}_0(\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*}) = \tilde{G}_0(\boldsymbol{\theta}^*, \boldsymbol{\sigma}^*),$$

as required. □

Theorem 5.5 suggests the following method for solving Problem $\tilde{\mathrm{P}}$. First, choose an initial $\epsilon \in (0, \hat{\epsilon})$ and solve Problem $\tilde{\mathrm{P}}_{\epsilon,\vartheta}$ for increasing values of $\vartheta > 0$ until (5.34) is satisfied (Theorem 5.3 ensures that Problem $\tilde{\mathrm{P}}_{\epsilon,\vartheta}$ only needs to be solved a finite number of times here). Second, construct $\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta}$ from $\boldsymbol{\theta}^{\epsilon,\vartheta,*}$ using equation (5.50). Third, decrease ϵ and repeat the first two steps. We eventually terminate this loop when ϵ is sufficiently small, taking $(\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*})$ as an approximate solution of Problem $\tilde{\mathrm{P}}$.

The method described above is summarized in the following algorithm.

Algorithm 5.2. Input $\epsilon_{\min} \in (0, \hat{\epsilon})$, $\epsilon_0 \in (\epsilon_{\min}, \hat{\epsilon})$, $\vartheta_{\max} > 0$, $\vartheta_0 \in (0, \vartheta_{\max})$, and a pair $(\boldsymbol{\theta}^0, \boldsymbol{\sigma}^0) \in \Theta \times \Xi$.

(i) Initialize $\epsilon_0 \to \epsilon$ and $\vartheta_0 \to \vartheta$.

(ii) Using $(\boldsymbol{\theta}^0, \boldsymbol{\sigma}^0)$ as the initial guess, solve Problem $\tilde{\mathrm{P}}_{\epsilon,\vartheta}$. Let $(\boldsymbol{\theta}^{\epsilon,\vartheta,*}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*})$ denote the solution obtained.

(iii) If $\boldsymbol{\theta}^{\epsilon,\vartheta,*}$ satisfies condition (5.34), then construct $\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta}$ using equation (5.50) and go to Step (iv). Otherwise, go to Step (v).

(iv) If $\epsilon > \epsilon_{\min}$, then set $\epsilon/10 \to \epsilon$ and $(\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*}) \to (\boldsymbol{\theta}^0, \boldsymbol{\sigma}^0)$ and go to Step (ii). Otherwise, stop; take $(\hat{\boldsymbol{\theta}}^{\epsilon,\vartheta}, \boldsymbol{\sigma}^{\epsilon,\vartheta,*})$ as the optimal solution of Problem $\tilde{\mathrm{P}}$.

(v) If $\vartheta < \vartheta_{\max}$, then set $10\vartheta \to \vartheta$ and go to Step (ii). Otherwise, stop; ϑ is too large (in this case, Problem $\tilde{\mathrm{P}}$ is probably ill-posed).

5.6 A numerical example

Consider the switched-capacitor DC-DC power converter discussed in Section 4.6. Recall that this switched-capacitor DC-DC power converter has three capacitors and four circuit topologies. For each $i = 1, 2, 3$, let $x_i(t)$ denote the voltage across the ith capacitor at time t. Furthermore, let $y(t)$ denote the output voltage at time t. We assume that the DC input voltage is 3.6 V and the load resistance is 75 Ω. The other circuit parameters are as follows:

$$\begin{aligned} R_1 &= R_2 = R_3 = 0.02\ \Omega, \\ R_S &= 0.01\ \Omega, \\ C_1 &= 30 \times 10^{-5}\ \text{F}, \\ C_2 &= 45 \times 10^{-5}\ \text{F}, \\ C_3 &= 60 \times 10^{-5}\ \text{F}. \end{aligned}$$

This power converter only has three distinct circuit topologies (topologies 2 and 4 are the same). For each $k = 1, 2, 3$, the kth circuit topology is modeled by the following dynamics:

$$\begin{aligned} \dot{\boldsymbol{x}}(t) &= A_k \boldsymbol{x}(t) + 3.6 B_k, \\ y(t) &= C_k \boldsymbol{x}(t) + 3.6 D_k, \end{aligned}$$

where $A_k \in \mathbb{R}^{3\times 3}$, $B_k \in \mathbb{R}^{3\times 1}$, $C_k \in \mathbb{R}^{1\times 3}$, and $D_k \in \mathbb{R}$ are given in Section 4.A. The output voltage is regulated to the desired 1.8 V (half of the 3.6 V input) by switching between these topologies in an appropriate manner.

An *operating schedule* for the power converter specifies the order in which the topologies are operated (the switching sequence) and the times at which the topologies are switched (the switching times). Since the ideal output is 1.8 V, the operating schedule should be chosen to minimize

$$\int_0^T (y(t) - 1.8)^2 dt, \tag{5.58}$$

where $T \triangleq 1.0 \times 10^{-4}$ is the terminal time.

We use the method suggested in [89] and model the power converter by the following dynamics:

$$\dot{\boldsymbol{x}}(t) = A_{\iota_k} \boldsymbol{x}(t) + 3.6 B_{\iota_k}, \qquad t \in (t_{k-1}, t_k), \qquad k = 1, \ldots, 9, \tag{5.59}$$

$$y(t) = C_{\iota_k} \boldsymbol{x}(t) + 3.6 D_{\iota_k}, \qquad t \in [t_{k-1}, t_k), \qquad k = 1, \ldots, 9, \tag{5.60}$$

where $t_0 \triangleq 0$, $t_9 \triangleq 1.0 \times 10^{-4}$, and t_k, $k = 1, \ldots, 8$, are switching times such that $t_{k-1} \leq t_k$;

and

$$\iota_k \triangleq \text{mod}(k-1,3)+1, \qquad k=1,\ldots,9.$$

Equations (5.59)-(5.60) can replicate any operating schedule. For example, if

$$0=t_0=t_1<t_2=t_3<t_4=t_5<t_6=t_7=t_8=t_9=1.0\times10^{-4},$$

then the switching sequence is $\{2,1,3\}$ (that is, operate Topology 2, then Topology 1, then Topology 3) and the switching times are $t_2=t_3$ and $t_4=t_5$. Thus, it is clear that unnecessary subsystems in (5.59)-(5.60) can be removed by combining some of the switches.

We assume that each capacitor loses 10% of its voltage when the circuit topology is changed. Hence, we have the following state jump conditions:

$$\boldsymbol{x}(t_k^+)=\begin{cases}[0,0,0]^T, & \text{if } k=0, \qquad (5.61a)\\ 0.9\boldsymbol{x}(t_k^-), & \text{if } k\in\{1,\ldots,8\} \text{ and } t_{k-1}<t_k<1.0\times10^{-4}. \qquad (5.61b)\end{cases}$$

The optimal control problem is as follows: choose the switching times t_k, $k=1,\ldots,8$, to minimize (5.58) subject to the switched system (5.59)-(5.61). We solved this problem by implementing Algorithm 5.2 in Fortran 90. In this implementation, Problem $\tilde{\text{P}}_{\epsilon,\vartheta}$ is solved by starting NLPQLP (see [93]) from ten random points, where Algorithm 5.1 is used to construct the required gradients. LSODA (see [41]) is used to solve the governing switched system numerically. Initially, $\epsilon=0.9\hat{\epsilon}$ (where $m=9$ and $T=1.0\times10^{-4}$) and $\vartheta=2.0$. We terminate Algorithm 5.2 when $\epsilon<9\hat{\epsilon}/10^4$.

The optimal switching times in (5.59)-(5.61) are

$$\begin{aligned}&t_0^*=t_1^*=0.0,\\ &t_2^*=t_3^*=t_4^*=1.0122\times10^{-6},\\ &t_5^*=t_6^*=9.4824\times10^{-6},\\ &t_7^*=1.9633\times10^{-5},\\ &t_8^*=t_9^*=1.0\times10^{-4}.\end{aligned}$$

This solution eliminates subsystems $\{1,3,4,6,9\}$. Hence, the optimal switching sequence is $\{2,1,2\}$. Figure 5.2 shows the voltage across the load and the capacitors when the optimal operating schedule is applied.

5.7 Conclusion

In this chapter, we considered an optimal control problem involving a nonlinear switched system with state jumps. The most interesting aspect of this problem is that apply-

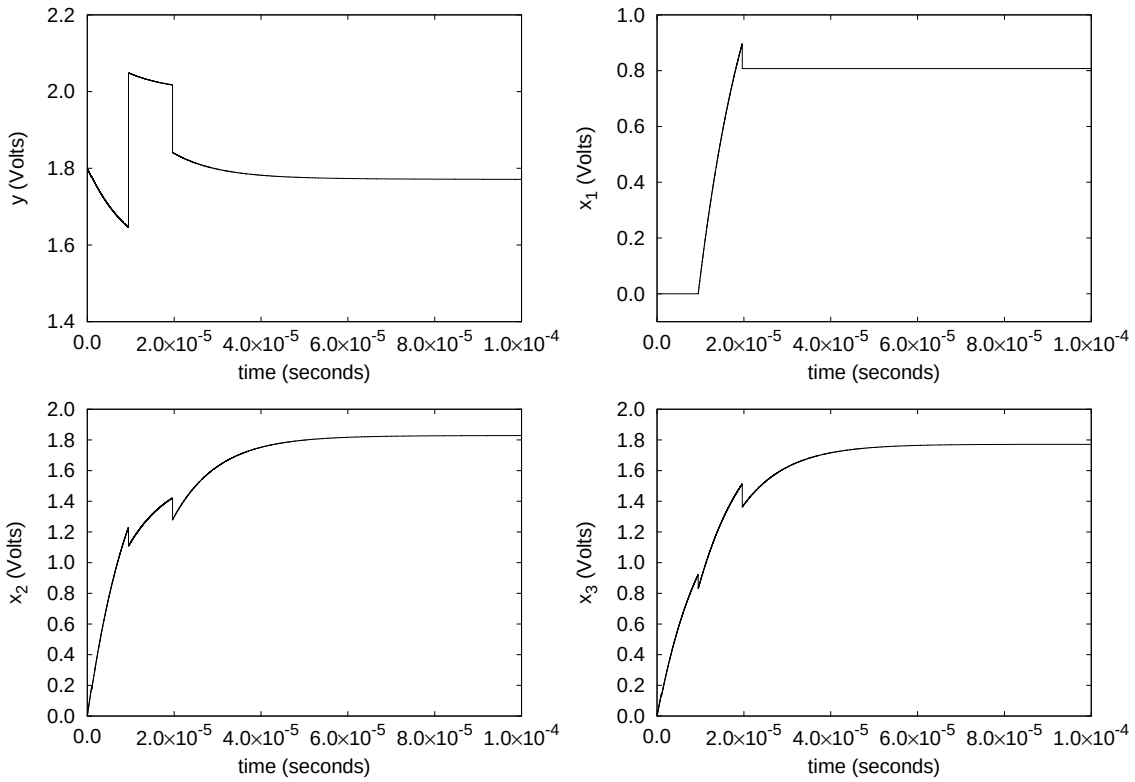

Figure 5.2: The output and state voltages corresponding to the optimal operating schedule.

ing the time-scaling transformation yields an optimal control problem with *discontinuous* state jump conditions. These discontinuities prevent the direct use of a gradient-based optimization method. Instead, we proposed an approximation scheme whereby the discontinuous state jumps are approximated by smooth ones. This yields a class of approximate optimization problems, each of which can be solved using a gradient-based optimization method. We also proved rigorously that this approximation scheme converges.

CHAPTER 6

State-delay identification via optimal control techniques*

6.1 Introduction

A mathematical model for a system is typically constructed as follows. First, the system is embedded within a family of systems having a common structure, and a general model is designed to encapsulate this structure. Second, the general model is tailored to the specific system of interest by choosing appropriate values for the model parameters. This second step is called *parameter identification*. Parameter identification is usually done by comparing the system output observed in practice with the system output predicted by the model, and then adjusting the parameters accordingly.

In this chapter, we consider a parameter identification problem for a nonlinear delay-differential system. This system is quite different from the dynamic systems considered in Chapters 2-5: at each time t, its instantaneous rate of change depends not only on its current state, but also on its state at times $t - \tau_i$, $i = 1, \ldots, r$, where each τ_i is a so-called *state-delay*. These state-delays are model parameters that need to be identified.

Many systems—for example, predator-prey systems [131], continuously-stirred tank reactors [13], zinc production systems [116,117], and aerospace systems [112]—have delays in their dynamics. Such delays are usually not known exactly and therefore need to be estimated. The estimates of the delays should be chosen so that the discrepancy between predicted and observed system output is as small as possible. Accordingly, in this chapter, we formulate the problem of identifying state-delays as an optimal control problem in which the state-delays are control variables and the cost function penalizes the squared difference between predicted and observed system output.

This optimal control problem is very unusual: the state-delays in its governing delay-differential system are not known in advance, and are instead control variables to be determined optimally. Furthermore, the state-delays influence the problem's cost function *implicitly* through the delay-differential system. Thus, determining the gradient of the

*This chapter is based on [70].

cost function is a very difficult task. In this chapter, we will show that the cost function's gradient can be computed by solving a set of auxiliary delay-differential systems. This is a fundamental result; it enables one to solve the optimal control problem using a standard nonlinear programming algorithm, and thereby obtain accurate estimates for the state-delays.

We emphasize that this approach to state-delay identification is applicable to both linear and nonlinear delay-differential systems. In contrast, most previous work on delay identification has focussed on linear systems—see, for example, [81,82,109] and the references cited therein. Indeed, we are not aware of any existing identification methods that are capable of dealing with the broad class of nonlinear delay systems considered here. Thus, this chapter demonstrates that optimal control techniques can be powerful tools for tackling parameter identification problems—especially those with nonlinear dynamics.

6.2 Problem formulation

Consider the following dynamic model:

$$\dot{\boldsymbol{x}}(t) = \sum_{i=1}^{r} \boldsymbol{f}^i\big(\boldsymbol{x}(t), \boldsymbol{x}(t-\tau_i)\big), \qquad t \in (0,T], \tag{6.1}$$

$$\boldsymbol{x}(t) = \boldsymbol{z}(t), \qquad t \in [-\bar{\tau}, 0], \tag{6.2}$$

and

$$\boldsymbol{y}(t) = \boldsymbol{g}(\boldsymbol{x}(t)), \qquad t \in [-\bar{\tau}, T], \tag{6.3}$$

where $T > 0$ and $\bar{\tau} > 0$ are given real numbers; $\boldsymbol{x}(t) \in \mathbb{R}^n$ is the system state at time t; $\boldsymbol{y}(t) \in \mathbb{R}^m$ is the system output at time t; τ_i, $i = 1, \ldots, r$, are unknown state-delays that need to be identified; and $\boldsymbol{f}^i : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$, $i = 1, \ldots, r$, $\boldsymbol{z} : \mathbb{R} \to \mathbb{R}^n$, and $\boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^m$ are given functions.

We assume that the following conditions are satisfied.

Assumption 6.1. The functions $\boldsymbol{f}^i$, $i = 1, \ldots, r$, and $\boldsymbol{g}$ are continuously differentiable.

Assumption 6.2. The function $\boldsymbol{z}$ is twice continuously differentiable.

Assumption 6.3. There exists a real number $L_1 > 0$ such that

$$\big|\boldsymbol{f}^i(\boldsymbol{v}, \boldsymbol{w})\big| \leq L_1(1 + |\boldsymbol{v}| + |\boldsymbol{w}|), \qquad (\boldsymbol{v}, \boldsymbol{w}) \in \mathbb{R}^n \times \mathbb{R}^n, \qquad i = 1, \ldots, r.$$

Suppose that the system modeled by (6.1)-(6.3) has been observed (for example, in an experiment) at times t_j, $j = 1, \ldots, p$. For each $j = 1, \ldots, p$, let $\hat{\boldsymbol{y}}^j \in \mathbb{R}^m$ denote the system output measured at time $t = t_j$. Our goal is to use the experimental data $\{(t_j, \hat{\boldsymbol{y}}^j)\}_{j=1}^p$ to estimate the state-delays in (6.1)-(6.3).

We assume that each state-delay is non-negative. Thus,

$$\tau_i \geq 0, \qquad i = 1, \ldots, r. \tag{6.4}$$

We also assume that each state-delay is bounded above by $\bar{\tau}$:

$$\tau_i \leq \bar{\tau}, \qquad i = 1, \ldots, r. \tag{6.5}$$

A vector $\boldsymbol{\tau} \in \mathbb{R}^r$ that satisfies inequalities (6.4) and (6.5) is called a *candidate state-delay vector*. Let $\mathcal{T}$ denote the set consisting of all candidate state-delay vectors.

By Theorem 3.3.3 of [2], the dynamic system (6.1)-(6.2) has a unique solution corresponding to each candidate state-delay vector $\boldsymbol{\tau} \in \mathcal{T}$. We denote this solution by $\boldsymbol{x}(\cdot|\boldsymbol{\tau})$. Substituting $\boldsymbol{x}(\cdot|\boldsymbol{\tau})$ into equation (6.3) gives $\boldsymbol{y}(\cdot|\boldsymbol{\tau})$, the *predicted system output* corresponding to $\boldsymbol{\tau} \in \mathcal{T}$. That is, for each $\boldsymbol{\tau} \in \mathcal{T}$,

$$\boldsymbol{y}(t|\boldsymbol{\tau}) \triangleq \boldsymbol{g}(\boldsymbol{x}(t|\boldsymbol{\tau})), \qquad t \in [-\bar{\tau}, T].$$

We now state the following optimal control problem, which involves choosing estimates for the state-delays so that the predicted output best fits the experimental data.

Problem P. *Find a candidate state-delay vector $\boldsymbol{\tau} \in \mathcal{T}$ that minimizes the cost function*

$$G_0(\boldsymbol{\tau}) \triangleq \sum_{j=1}^{p} \left|\boldsymbol{y}(t_j|\boldsymbol{\tau}) - \hat{\boldsymbol{y}}^j\right|^2$$

over $\mathcal{T}$.

Notice that the times t_j, $j = 1, \ldots, p$, are characteristic times in the cost function G_0 (see Chapter 2). Also notice that the state-delays in the delay-differential system (6.1)-(6.3) are actually the control variables in Problem P. Thus, Problem P is quite different to the standard time-delay optimal control problems considered in Chapter 5 of [39] and Chapter 12 of [100]. These standard problems are governed by delay systems of the following type:

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}\big(\boldsymbol{x}(t), \boldsymbol{x}(t-h), \boldsymbol{u}(t)\big), \qquad t \in (0, T],$$

and

$$\boldsymbol{x}(t) = \boldsymbol{z}(t), \qquad t \in [-h, 0],$$

where $h > 0$ is given and $\boldsymbol{u}$ is a control function to be determined optimally.

If the output function $\boldsymbol{g}$ does not provide enough information about the state variables, then a solution of Problem P may not yield good estimates of the state-delays. For example, if $\boldsymbol{g}$ is a constant function, then *every* vector in $\mathcal{T}$ is a solution of Problem P. In this case, Problem P is not a sensible mathematical formulation of the state-delay iden-

tification problem. Thus, throughout this chapter, we assume that the output function $\boldsymbol{g}$ provides enough system information so that Problem P is a sensible formulation of the state-delay identification problem. This is usually the case in practice. In fact, many systems have $\boldsymbol{g} \triangleq \boldsymbol{x}$ (that is, the system's output is just its state).

6.3 Preliminary results

For each $(k, \boldsymbol{\tau}) \in \{1, \ldots, r\} \times \mathcal{T}$, define

$$\mathcal{S}(k, \boldsymbol{\tau}) \triangleq \big\{\, \epsilon \in \mathbb{R} : \ \boldsymbol{\tau} + \epsilon \boldsymbol{e}^{r,k} \in \mathcal{T} \,\big\},$$

where $\boldsymbol{e}^{r,k}$ is the kth standard unit basis vector in $\mathbb{R}^r$. Clearly,

$$\mathcal{S}(k, \boldsymbol{\tau}) = [-\tau_k, \bar{\tau} - \tau_k],$$

which shows that $\mathcal{S}(k, \boldsymbol{\tau})$ is a closed interval of positive measure. Moreover, $0 \in \mathcal{S}(k, \boldsymbol{\tau})$. Let $\mathcal{V}$ denote the set of all triples $(k, \boldsymbol{\tau}, \epsilon) \in \{1, \ldots, r\} \times \mathcal{T} \times \mathbb{R}$ such that $\epsilon \in \mathcal{S}(k, \boldsymbol{\tau})$.

For each $(k, \boldsymbol{\tau}, \epsilon) \in \mathcal{V}$, define the following $\mathbb{R}^n$-valued functions on $[-\bar{\tau}, T]$:

$$\boldsymbol{\xi}(t|k, \boldsymbol{\tau}, \epsilon) \triangleq \boldsymbol{x}(t|\boldsymbol{\tau} + \epsilon \boldsymbol{e}^{r,k}) - \boldsymbol{x}(t|\boldsymbol{\tau}),$$

and

$$\boldsymbol{\phi}(t|k, \boldsymbol{\tau}, \epsilon) \triangleq \begin{cases} \dot{\boldsymbol{z}}(t), & \text{if } t \in [-\bar{\tau}, 0], \\ \displaystyle\sum_{i=1}^{r} \boldsymbol{f}^i\big(\boldsymbol{x}(t|\boldsymbol{\tau} + \epsilon \boldsymbol{e}^{r,k}), \boldsymbol{x}(t - \tau_i - \epsilon \rho_{k,i}|\boldsymbol{\tau} + \epsilon \boldsymbol{e}^{r,k})\big), & \text{if } t \in (0, T], \end{cases}$$

where $\rho_{k,i}$ denotes the Kronecker delta.

Furthermore, for each $(k, \boldsymbol{\tau}, \epsilon) \in \mathcal{V}$, define corresponding functions $\boldsymbol{\gamma}^i : [0, T] \to \mathbb{R}^n$, $i = 1, \ldots, r$, as follows:

$$\boldsymbol{\gamma}^i(t|k, \boldsymbol{\tau}, \epsilon) \triangleq \boldsymbol{x}(t - \tau_i - \epsilon \rho_{k,i}|\boldsymbol{\tau} + \epsilon \boldsymbol{e}^{r,k}) - \boldsymbol{x}(t - \tau_i|\boldsymbol{\tau}), \qquad i = 1, \ldots, r.$$

We immediately see that

$$\boldsymbol{\gamma}^i(t|k, \boldsymbol{\tau}, \epsilon) = \boldsymbol{\xi}(t - \tau_i|k, \boldsymbol{\tau}, \epsilon), \qquad t \in [0, T], \qquad i \neq k, \tag{6.6}$$

and

$$\boldsymbol{\xi}(t|k, \boldsymbol{\tau}, \epsilon) = \boldsymbol{0}, \qquad t \in [-\bar{\tau}, 0]. \tag{6.7}$$

Furthermore, if $(k', \boldsymbol{\tau}', \epsilon'), (k'', \boldsymbol{\tau}'', \epsilon'') \in \mathcal{V}$, then

$$\boldsymbol{\phi}(t|k', \boldsymbol{\tau}', \epsilon') - \boldsymbol{\phi}(t|k'', \boldsymbol{\tau}'', \epsilon'') = \mathbf{0}, \qquad t \in [-\bar{\tau}, 0]. \tag{6.8}$$

It is also clear that for almost all $t \in [-\bar{\tau}, T]$,

$$\dot{\boldsymbol{x}}(t|\boldsymbol{\tau} + \epsilon \boldsymbol{e}^{r,k}) = \boldsymbol{\phi}(t|k, \boldsymbol{\tau}, \epsilon).$$

Thus, if t_1 and t_2 are two points in $[-\bar{\tau}, T]$, then

$$\boldsymbol{x}(t_2|\boldsymbol{\tau} + \epsilon \boldsymbol{e}^{r,k}) - \boldsymbol{x}(t_1|\boldsymbol{\tau} + \epsilon \boldsymbol{e}^{r,k}) = \int_{t_1}^{t_2} \boldsymbol{\phi}(t|k, \boldsymbol{\tau}, \epsilon) dt. \tag{6.9}$$

Notice that k does not influence $\boldsymbol{\phi}(t|k, \boldsymbol{\tau}, 0)$. Hence, where appropriate, we will simplify the notation by writing $\boldsymbol{\phi}(\cdot|\boldsymbol{\tau})$ instead of $\boldsymbol{\phi}(\cdot|k, \boldsymbol{\tau}, 0)$. It follows immediately from (6.8) that for each $(k, \boldsymbol{\tau}, \epsilon) \in \mathcal{V}$,

$$\boldsymbol{\phi}(t|k, \boldsymbol{\tau}, \epsilon) - \boldsymbol{\phi}(t|\boldsymbol{\tau}) = \mathbf{0}, \qquad t \in [-\bar{\tau}, 0]. \tag{6.10}$$

We now prove the following result.

Lemma 6.1. *There exists a real number $L_2 > 0$ such that*

$$\big|\boldsymbol{x}(t|\boldsymbol{\tau})\big| \leq L_2, \qquad t \in [-\bar{\tau}, T], \qquad \boldsymbol{\tau} \in \mathcal{T}.$$

Proof. Let $\boldsymbol{\tau} \in \mathcal{T}$ be arbitrary but fixed. By Assumption 6.2, there exists a real number $\alpha_1 > 0$ such that

$$\sup\big\{\, |\boldsymbol{z}(t)| : \ t \in [-\bar{\tau}, 0] \,\big\} \leq \alpha_1.$$

Hence,

$$\big|\boldsymbol{x}(t|\boldsymbol{\tau})\big| = \big|\boldsymbol{z}(t)\big| \leq \alpha_1, \qquad t \in [-\bar{\tau}, 0]. \tag{6.11}$$

On the other hand, if $t \in (0, T]$, then

$$\boldsymbol{x}(t|\boldsymbol{\tau}) = \boldsymbol{z}(0) + \sum_{i=1}^{r} \int_0^t \boldsymbol{f}^i\big(\boldsymbol{x}(s|\boldsymbol{\tau}), \boldsymbol{x}(s - \tau_i|\boldsymbol{\tau})\big) ds.$$

Applying Assumption 6.3 gives

$$\begin{aligned}\big|\boldsymbol{x}(t|\boldsymbol{\tau})\big| &\le \alpha_1 + rL_1T + \int_0^t rL_1\big|\boldsymbol{x}(s|\boldsymbol{\tau})\big|ds + \sum_{i=1}^{r}\int_0^t L_1\big|\boldsymbol{x}(s-\tau_i|\boldsymbol{\tau})\big|ds\\ &= \alpha_1 + rL_1T + \int_0^t rL_1\big|\boldsymbol{x}(s|\boldsymbol{\tau})\big|ds + \sum_{i=1}^{r}\int_{-\tau_i}^{t-\tau_i} L_1\big|\boldsymbol{x}(s|\boldsymbol{\tau})\big|ds\\ &\le \alpha_1 + rL_1T + \int_0^t rL_1\big|\boldsymbol{x}(s|\boldsymbol{\tau})\big|ds + \int_{-\bar{\tau}}^t rL_1\big|\boldsymbol{x}(s|\boldsymbol{\tau})\big|ds\\ &\le \alpha_1 + rL_1T + rL_1\alpha_1\bar{\tau} + \int_0^t 2rL_1\big|\boldsymbol{x}(s|\boldsymbol{\tau})\big|ds.\end{aligned}$$

Thus, by Gronwall's Lemma,

$$\big|\boldsymbol{x}(t|\boldsymbol{\tau})\big| \le \alpha_2\exp(2rL_1T), \qquad t\in(0,T], \tag{6.12}$$

where

$$\alpha_2 \triangleq \alpha_1 + rL_1T + rL_1\alpha_1\bar{\tau}.$$

Since $\boldsymbol{\tau}\in\mathcal{T}$ was chosen arbitrarily, the result follows from (6.11) and (6.12). □

Define

$$\Psi \triangleq \big\{\,\boldsymbol{v}\in\mathbb{R}^n:\ |\boldsymbol{v}|\le L_2\,\big\},$$

where L_2 is as defined in Lemma 6.1. Clearly,

$$\boldsymbol{x}(t|\boldsymbol{\tau})\in\Psi, \qquad t\in[-\bar{\tau},T], \qquad \boldsymbol{\tau}\in\mathcal{T}.$$

Hence, by Assumptions 6.1 and 6.2, we immediately obtain the following result.

Lemma 6.2. *There exists a real number $L_3>0$ such that*

$$\big|\boldsymbol{\phi}(t|k,\boldsymbol{\tau},\epsilon)\big|\le L_3, \qquad t\in[-\bar{\tau},T], \qquad (k,\boldsymbol{\tau},\epsilon)\in\mathcal{V}.$$

We also have the following important result.

Lemma 6.3. *There exists a real number $L_4>0$ such that for each $(k,\boldsymbol{\tau},\epsilon)\in\mathcal{V}$,*

$$\max\Big\{\big|\boldsymbol{\xi}(t|k,\boldsymbol{\tau},\epsilon)\big|,\ \big|\boldsymbol{\gamma}^k(t|k,\boldsymbol{\tau},\epsilon)\big|,\ \big|\boldsymbol{\phi}(t|k,\boldsymbol{\tau},\epsilon)-\boldsymbol{\phi}(t|\boldsymbol{\tau})\big|\Big\}\le L_4|\epsilon|, \quad t\in[0,T].$$

Proof. Let $(k,\boldsymbol{\tau},\epsilon)\in\mathcal{V}$ be arbitrary but fixed. To simplify the notation, we write $\boldsymbol{x}^\epsilon$ instead of $\boldsymbol{x}(\cdot|\boldsymbol{\tau}+\epsilon\boldsymbol{e}^{r,k})$ and $\boldsymbol{x}$ instead of $\boldsymbol{x}(\cdot|\boldsymbol{\tau})$.

It is easy to see that

$$\big|\boldsymbol{\gamma}^k(s|k,\boldsymbol{\tau},\epsilon)\big|\le\big|\boldsymbol{x}^\epsilon(s-\tau_k-\epsilon)-\boldsymbol{x}^\epsilon(s-\tau_k)\big|+\big|\boldsymbol{\xi}(s-\tau_k|k,\boldsymbol{\tau},\epsilon)\big|, \qquad s\in[0,T].$$

Hence, by (6.9),

$$\big|\boldsymbol{\gamma}^k(s|k,\boldsymbol{\tau},\epsilon)\big| \le \int_{a(s)}^{b(s)} \big|\boldsymbol{\phi}(\eta|k,\boldsymbol{\tau},\epsilon)\big|d\eta + \big|\boldsymbol{\xi}(s-\tau_k|k,\boldsymbol{\tau},\epsilon)\big|, \qquad s\in[0,T], \tag{6.13}$$

where

$$a(s) \triangleq \min\{s-\tau_k, s-\tau_k-\epsilon\}$$

and

$$b(s) \triangleq \max\{s-\tau_k, s-\tau_k-\epsilon\}.$$

Clearly,

$$b(s)-a(s) = |\epsilon|, \qquad s\in[0,T].$$

Hence, inequality (6.13) becomes

$$\big|\boldsymbol{\gamma}^k(s|k,\boldsymbol{\tau},\epsilon)\big| \le L_3|\epsilon| + \big|\boldsymbol{\xi}(s-\tau_k|k,\boldsymbol{\tau},\epsilon)\big|, \qquad s\in[0,T], \tag{6.14}$$

where L_3 is as defined in Lemma 6.2.

Now,

$$\big|\boldsymbol{\phi}(s|k,\boldsymbol{\tau},\epsilon)-\boldsymbol{\phi}(s|\boldsymbol{\tau})\big| \le \sum_{i=1}^{r} \big|\boldsymbol{f}^i(\boldsymbol{x}^\epsilon(s),\boldsymbol{x}^\epsilon(s-\tau_i-\epsilon\rho_{k,i})) - \boldsymbol{f}^i(\boldsymbol{x}(s),\boldsymbol{x}(s-\tau_i))\big|, \ s\in(0,T].$$

(Equation (6.10) shows that this inequality also holds at $s=0$.) Assumption 6.1 implies that the functions $\boldsymbol{f}^i$, $i=1,\dots,r$, are Lipschitz continuous on $\Psi\times\Psi$. Consequently, there exists a real number $\alpha_1>0$ such that

$$\big|\boldsymbol{\phi}(s|k,\boldsymbol{\tau},\epsilon)-\boldsymbol{\phi}(s|\boldsymbol{\tau})\big| \le r\alpha_1\big|\boldsymbol{\xi}(s|k,\boldsymbol{\tau},\epsilon)\big| + \sum_{i=1}^{r}\alpha_1\big|\boldsymbol{\gamma}^i(s|k,\boldsymbol{\tau},\epsilon)\big|, \qquad s\in[0,T].$$

Thus, by identity (6.6),

$$\begin{aligned}\big|\boldsymbol{\phi}(s|k,\boldsymbol{\tau},\epsilon)-\boldsymbol{\phi}(s|\boldsymbol{\tau})\big| \le{}& r\alpha_1\big|\boldsymbol{\xi}(s|k,\boldsymbol{\tau},\epsilon)\big| + \alpha_1\big|\boldsymbol{\gamma}^k(s|k,\boldsymbol{\tau},\epsilon)\big| \\ &+ \sum_{\substack{i=1\\ i\neq k}}^{r}\alpha_1\big|\boldsymbol{\xi}(s-\tau_i|k,\boldsymbol{\tau},\epsilon)\big|, \quad s\in[0,T].\end{aligned} \tag{6.15}$$

Substituting (6.14) into (6.15) gives

$$\begin{aligned}\big|\boldsymbol{\phi}(s|k,\boldsymbol{\tau},\epsilon)-\boldsymbol{\phi}(s|\boldsymbol{\tau})\big| \le{}& \alpha_1L_3|\epsilon| + r\alpha_1\big|\boldsymbol{\xi}(s|k,\boldsymbol{\tau},\epsilon)\big| \\ &+ \sum_{i=1}^{r}\alpha_1\big|\boldsymbol{\xi}(s-\tau_i|k,\boldsymbol{\tau},\epsilon)\big|, \quad s\in[0,T].\end{aligned} \tag{6.16}$$

Now, if $t \in [0, T]$, then it follows from (6.9) that

$$\boldsymbol{x}^{\epsilon}(t) = \boldsymbol{z}(0) + \int_0^t \boldsymbol{\phi}(s|k, \boldsymbol{\tau}, \epsilon) ds$$

and

$$\boldsymbol{x}(t) = \boldsymbol{z}(0) + \int_0^t \boldsymbol{\phi}(s|\boldsymbol{\tau}) ds.$$

Hence,

$$\big|\boldsymbol{\xi}(t|k, \boldsymbol{\tau}, \epsilon)\big| = \big|\boldsymbol{x}^{\epsilon}(t) - \boldsymbol{x}(t)\big| \leq \int_0^t \big|\boldsymbol{\phi}(s|k, \boldsymbol{\tau}, \epsilon) - \boldsymbol{\phi}(s|\boldsymbol{\tau})\big| ds.$$

Applying inequality (6.16) yields

$$\begin{aligned}
\big|\boldsymbol{\xi}(t|k, \boldsymbol{\tau}, \epsilon)\big| &\leq \alpha_1 L_3 T|\epsilon| + \int_0^t r\alpha_1 \big|\boldsymbol{\xi}(s|k, \boldsymbol{\tau}, \epsilon)\big| ds + \sum_{i=1}^{r} \int_0^t \alpha_1 \big|\boldsymbol{\xi}(s - \tau_i|k, \boldsymbol{\tau}, \epsilon)\big| ds \\
&= \alpha_1 L_3 T|\epsilon| + \int_0^t r\alpha_1 \big|\boldsymbol{\xi}(s|k, \boldsymbol{\tau}, \epsilon)\big| ds + \sum_{i=1}^{r} \int_{-\tau_i}^{t-\tau_i} \alpha_1 \big|\boldsymbol{\xi}(s|k, \boldsymbol{\tau}, \epsilon)\big| ds \\
&\leq \alpha_1 L_3 T|\epsilon| + \int_0^t r\alpha_1 \big|\boldsymbol{\xi}(s|k, \boldsymbol{\tau}, \epsilon)\big| ds + \int_{-\bar{\tau}}^{t} r\alpha_1 \big|\boldsymbol{\xi}(s|k, \boldsymbol{\tau}, \epsilon)\big| ds.
\end{aligned}$$

Thus, by (6.7),

$$\big|\boldsymbol{\xi}(t|k, \boldsymbol{\tau}, \epsilon)\big| \leq \alpha_1 L_3 T|\epsilon| + \int_0^t 2r\alpha_1 \big|\boldsymbol{\xi}(s|k, \boldsymbol{\tau}, \epsilon)\big| ds.$$

Applying Gronwall's Lemma gives

$$\big|\boldsymbol{\xi}(t|k, \boldsymbol{\tau}, \epsilon)\big| \leq \alpha_2|\epsilon|, \qquad t \in [0, T], \tag{6.17}$$

where

$$\alpha_2 \triangleq \alpha_1 L_3 T \exp(2r\alpha_1 T).$$

We now write (6.7) and (6.17) collectively as

$$\big|\boldsymbol{\xi}(t|k, \boldsymbol{\tau}, \epsilon)\big| \leq \alpha_2|\epsilon|, \qquad t \in [-\bar{\tau}, T], \tag{6.18}$$

The result is finally established by substituting (6.18) into (6.14) and (6.16). □

Lemma 6.3 implies that for each fixed $(k, \boldsymbol{\tau}) \in \{1, \ldots, r\} \times \mathcal{T}$,

$$\boldsymbol{x}(\cdot|\boldsymbol{\tau} + \epsilon \boldsymbol{e}^{r,k}) \to \boldsymbol{x}(\cdot|\boldsymbol{\tau}),$$

$$\boldsymbol{\gamma}^k(\cdot|k, \boldsymbol{\tau}, \epsilon) \to \mathbf{0},$$

and

$$\boldsymbol{\phi}(\cdot|k, \boldsymbol{\tau}, \epsilon) \to \boldsymbol{\phi}(\cdot|\boldsymbol{\tau})$$

uniformly on $[0,T]$ as $\epsilon \to 0$.*

We see from equations (6.7) and (6.10) that $\boldsymbol{\xi}(\cdot|k,\boldsymbol{\tau},\epsilon)$ and $\boldsymbol{\phi}(\cdot|k,\boldsymbol{\tau},\epsilon) - \boldsymbol{\phi}(\cdot|\boldsymbol{\tau})$ also satisfy the inequality in Lemma 6.3 for $t \in [-\bar{\tau},0]$. Therefore,

$$\max\Big\{\big|\boldsymbol{\xi}(t|k,\boldsymbol{\tau},\epsilon)\big|,\ \big|\boldsymbol{\phi}(t|k,\boldsymbol{\tau},\epsilon) - \boldsymbol{\phi}(t|\boldsymbol{\tau})\big|\Big\} \le L_4|\epsilon|, \qquad t \in [-\bar{\tau},T]. \tag{6.19}$$

By combining Lemma 6.3, equation (6.6), and inequality (6.19), we obtain

$$\big|\boldsymbol{\gamma}^i(t|k,\boldsymbol{\tau},\epsilon)\big| \le L_4|\epsilon|, \qquad t \in [0,T], \qquad i = 1,\dots,r. \tag{6.20}$$

We now prove our final preliminary result.

Lemma 6.4. *If $(k,\boldsymbol{\tau}) \in \{1,\dots,r\} \times \mathcal{T}$, then for almost all $t \in [0,T]$,*

$$\lim_{\epsilon\to 0} \frac{\boldsymbol{\gamma}^k(t|k,\boldsymbol{\tau},\epsilon) - \boldsymbol{\xi}(t-\tau_k|k,\boldsymbol{\tau},\epsilon)}{\epsilon} = -\boldsymbol{\phi}(t-\tau_k|\boldsymbol{\tau}).$$

Proof. Let $(k,\boldsymbol{\tau}) \in \{1,\dots,r\} \times \mathcal{T}$ be arbitrary but fixed. To prove Lemma 6.4, it is sufficient to show that

$$\lim_{\epsilon\to 0} \frac{\boldsymbol{\gamma}^k(t|k,\boldsymbol{\tau},\epsilon) - \boldsymbol{\xi}(t-\tau_k|k,\boldsymbol{\tau},\epsilon)}{\epsilon} = -\boldsymbol{\phi}(t-\tau_k|\boldsymbol{\tau}), \qquad t \in [0,T] \setminus \{\tau_k\}. \tag{6.21}$$

We thus focus our attention on proving equation (6.21).

Let $t \in [0,T] \setminus \{\tau_k\}$ be arbitrary but fixed. Then for each $\epsilon \in \mathcal{S}(k,\boldsymbol{\tau}) \setminus \{0\}$,

$$\boldsymbol{\gamma}^k(t|k,\boldsymbol{\tau},\epsilon) - \boldsymbol{\xi}(t-\tau_k|k,\boldsymbol{\tau},\epsilon) = \boldsymbol{x}(t-\tau_k-\epsilon|\boldsymbol{\tau}+\epsilon\boldsymbol{e}^{r,k}) - \boldsymbol{x}(t-\tau_k|\boldsymbol{\tau}+\epsilon\boldsymbol{e}^{r,k}).$$

Hence, by (6.9),

$$\frac{\boldsymbol{\gamma}^k(t|k,\boldsymbol{\tau},\epsilon) - \boldsymbol{\xi}(t-\tau_k|k,\boldsymbol{\tau},\epsilon)}{\epsilon} = \epsilon^{-1}\int_{t-\tau_k}^{t-\tau_k-\epsilon} \boldsymbol{\phi}(s|k,\boldsymbol{\tau},\epsilon)ds.$$

We rewrite this equation as follows:

$$\frac{\boldsymbol{\gamma}^k(t|k,\boldsymbol{\tau},\epsilon) - \boldsymbol{\xi}(t-\tau_k|k,\boldsymbol{\tau},\epsilon)}{\epsilon} = -\boldsymbol{\phi}(t-\tau_k|\boldsymbol{\tau}) + \boldsymbol{\omega}(\epsilon), \tag{6.22}$$

where

$$\boldsymbol{\omega}(\epsilon) \triangleq \epsilon^{-1}\int_{t-\tau_k}^{t-\tau_k-\epsilon} \big\{\boldsymbol{\phi}(s|k,\boldsymbol{\tau},\epsilon) - \boldsymbol{\phi}(t-\tau_k|\boldsymbol{\tau})\big\}ds.$$

We will prove equation (6.21) by showing that $\boldsymbol{\omega}(\epsilon) = \mathcal{O}(\epsilon)$ (then $\boldsymbol{\omega}(\epsilon) \to 0$ as $\epsilon \to 0$ and equation (6.21) follows immediately from equation (6.22)).

*It makes sense to consider these limits because zero is a limit point of $\mathcal{S}(k,\boldsymbol{\tau})$.

By the triangle inequality,

$$\big|\boldsymbol{\omega}(\epsilon)\big| \le |\epsilon|^{-1} \int_{a_\epsilon}^{b_\epsilon} \big|\boldsymbol{\phi}(s|k,\boldsymbol{\tau},\epsilon) - \boldsymbol{\phi}(s|\boldsymbol{\tau})\big| ds + |\epsilon|^{-1} \int_{a_\epsilon}^{b_\epsilon} \big|\boldsymbol{\phi}(s|\boldsymbol{\tau}) - \boldsymbol{\phi}(t-\tau_k|\boldsymbol{\tau})\big| ds, \tag{6.23}$$

where

$$a_\epsilon \triangleq \min\{t-\tau_k-\epsilon, t-\tau_k\}$$

and

$$b_\epsilon \triangleq \max\{t-\tau_k-\epsilon, t-\tau_k\}.$$

It is clear that

$$b_\epsilon - a_\epsilon = |\epsilon|. \tag{6.24}$$

Furthermore,

$$-\bar{\tau} \le a_\epsilon \le b_\epsilon \le T. \tag{6.25}$$

In view of (6.25), we may use (6.19) to simplify the first integral in (6.23), giving

$$\big|\boldsymbol{\omega}(\epsilon)\big| \le L_4(b_\epsilon - a_\epsilon) + |\epsilon|^{-1} \int_{a_\epsilon}^{b_\epsilon} \big|\boldsymbol{\phi}(s|\boldsymbol{\tau}) - \boldsymbol{\phi}(t-\tau_k|\boldsymbol{\tau})\big| ds.$$

Thus, by (6.24),

$$\big|\boldsymbol{\omega}(\epsilon)\big| \le L_4|\epsilon| + |\epsilon|^{-1} \int_{a_\epsilon}^{b_\epsilon} \big|\boldsymbol{\phi}(s|\boldsymbol{\tau}) - \boldsymbol{\phi}(t-\tau_k|\boldsymbol{\tau})\big| ds. \tag{6.26}$$

Note that inequality (6.26) holds for every $\epsilon \in \mathcal{S}(k,\boldsymbol{\tau}) \setminus \{0\}$.

Now, since $t \neq \tau_k$, we consider two cases: $t > \tau_k$ (Case 1) and $t < \tau_k$ (Case 2). For each case, we will show that $\boldsymbol{\omega}(\epsilon) = \mathcal{O}(\epsilon)$, from which equation (6.21) follows.

A Case 1: $t > \tau_k$

In this case, it is easy to verify the following implication:

$$\epsilon \in \mathcal{S}(k,\boldsymbol{\tau}),\ |\epsilon| < t-\tau_k \quad \Longrightarrow \quad [a_\epsilon, b_\epsilon] \subset (0,T]. \tag{6.27}$$

By using similar arguments to those given in the proof of Lemma 6.3, we can show that $\boldsymbol{\phi}(\cdot|\boldsymbol{\tau})$ is Lipschitz continuous on $(0,T]$. Hence, there exists a real number $\alpha_1 > 0$ such that

$$\big|\boldsymbol{\phi}(s|\boldsymbol{\tau}) - \boldsymbol{\phi}(t-\tau_k|\boldsymbol{\tau})\big| \le \alpha_1 |s-t+\tau_k|, \qquad s \in (0,T]. \tag{6.28}$$

It follows from (6.27) and (6.28) that when $\epsilon \in \mathcal{S}(k, \boldsymbol{\tau})$ is of sufficiently small magnitude,

$$\big|\boldsymbol{\phi}(s|\boldsymbol{\tau}) - \boldsymbol{\phi}(t - \tau_k|\boldsymbol{\tau})\big| \leq \alpha_1 |s - t + \tau_k| \leq \alpha_1(b_\epsilon - a_\epsilon) = \alpha_1|\epsilon|, \quad s \in [a_\epsilon, b_\epsilon].$$

By substituting this inequality into the second term of (6.26), we obtain

$$\big|\boldsymbol{\omega}(\epsilon)\big| \leq (L_4 + \alpha_1)|\epsilon|,$$

which holds whenever $\epsilon \in \mathcal{S}(k, \boldsymbol{\tau}) \setminus \{0\}$ is of sufficiently small magnitude. Therefore, $\boldsymbol{\omega}(\epsilon) = \mathcal{O}(\epsilon)$ as required.

B Case 2: $t < \tau_k$

We readily verify the following implication:

$$\epsilon \in \mathcal{S}(k, \boldsymbol{\tau}),\ |\epsilon| < \tau_k - t \quad \Longrightarrow \quad [a_\epsilon, b_\epsilon] \subset [-\bar{\tau}, 0]. \tag{6.29}$$

It follows from Assumption 6.2 that $\boldsymbol{\phi}(\cdot|\boldsymbol{\tau})$ is Lipschitz continuous on $[-\bar{\tau}, 0]$. Consequently, there exists a real number $\alpha_2 > 0$ such that

$$\big|\boldsymbol{\phi}(s|\boldsymbol{\tau}) - \boldsymbol{\phi}(t - \tau_k|\boldsymbol{\tau})\big| \leq \alpha_2 |s - t + \tau_k|, \qquad s \in [-\bar{\tau}, 0]. \tag{6.30}$$

It follows from (6.29) and (6.30) that when $\epsilon \in \mathcal{S}(k, \boldsymbol{\tau})$ is of sufficiently small magnitude,

$$\big|\boldsymbol{\phi}(s|\boldsymbol{\tau}) - \boldsymbol{\phi}(t - \tau_k|\boldsymbol{\tau})\big| \leq \alpha_2 |s - t + \tau_k| \leq \alpha_2(b_\epsilon - a_\epsilon) = \alpha_2|\epsilon|, \quad s \in [a_\epsilon, b_\epsilon].$$

Thus, we see from (6.26) that the inequality

$$\big|\boldsymbol{\omega}(\epsilon)\big| \leq (L_4 + \alpha_2)|\epsilon|$$

is satisfied for every $\epsilon \in \mathcal{S}(k, \boldsymbol{\tau}) \setminus \{0\}$ of sufficiently small magnitude. This shows that $\boldsymbol{\omega}(\epsilon) = \mathcal{O}(\epsilon)$, thereby completing the proof. □

6.4 The main result

For each $\boldsymbol{\tau} \in \mathcal{T}$, the corresponding state trajectory, $\boldsymbol{x}(\cdot|\boldsymbol{\tau})$, is a function of time. In other words, if the candidate state-delay vector is fixed, then the solution of (6.1)-(6.2) is a function defined on the time horizon $[-\bar{\tau}, T]$. Alternatively, we can fix $t \in [-\bar{\tau}, T]$ and consider the function $\boldsymbol{x}(t|\cdot) : \mathcal{T} \to \mathbb{R}^n$ whose value at $\boldsymbol{\tau} \in \mathcal{T}$ is $\boldsymbol{x}(t|\boldsymbol{\tau})$. We will show in this section that $\boldsymbol{x}(t|\cdot)$ is differentiable on $\mathcal{T}$. This is a significant result; it is used later to answer several important questions pertaining to Problem P.

To begin, consider the following auxiliary delay-differential system corresponding to

each $k = 1, \ldots, r$:

$$\dot{\boldsymbol{\psi}}^k(t) = \sum_{i=1}^{r} \left\{ \frac{\partial \boldsymbol{f}^i\big(\boldsymbol{x}(t|\boldsymbol{\tau}), \boldsymbol{x}(t-\tau_i|\boldsymbol{\tau})\big)}{\partial \boldsymbol{x}} \boldsymbol{\psi}^k(t) + \frac{\partial \boldsymbol{f}^i\big(\boldsymbol{x}(t|\boldsymbol{\tau}), \boldsymbol{x}(t-\tau_i|\boldsymbol{\tau})\big)}{\partial \boldsymbol{x}^-} \boldsymbol{\psi}^k(t-\tau_i) \right\}$$
$$- \frac{\partial \boldsymbol{f}^k\big(\boldsymbol{x}(t|\boldsymbol{\tau}), \boldsymbol{x}(t-\tau_k|\boldsymbol{\tau})\big)}{\partial \boldsymbol{x}^-} \boldsymbol{\phi}(t-\tau_k|\boldsymbol{\tau}), \qquad t \in (0, T], \tag{6.31}$$

and

$$\boldsymbol{\psi}^k(t) = \mathbf{0}, \qquad t \in [-\bar{\tau}, 0], \tag{6.32}$$

where $\boldsymbol{\tau} \in \mathcal{T}$ and $\partial/\partial \boldsymbol{x}^-$ denotes partial differentiation with respect to the delayed argument. Let $\boldsymbol{\psi}^k(\cdot|\boldsymbol{\tau})$ denote the solution of (6.31)-(6.32).

The next theorem is the main result in this chapter.

Theorem 6.1. *For each $t \in (0, T]$, the function $\boldsymbol{x}(t|\cdot)$ is differentiable on $\mathcal{T}$. Furthermore,*

$$\frac{\partial \boldsymbol{x}(t|\boldsymbol{\tau})}{\partial \tau_k} = \boldsymbol{\psi}^k(t|\boldsymbol{\tau}), \qquad k = 1, \ldots, r, \qquad \boldsymbol{\tau} \in \mathcal{T},$$

Proof. Let $k \in \{1, \ldots, r\}$ and $\boldsymbol{\tau} \in \mathcal{T}$ be arbitrary but fixed. To prove the result, it is sufficient to show that

$$\lim_{\epsilon \to 0} \frac{\boldsymbol{x}(t|\boldsymbol{\tau} + \epsilon \boldsymbol{e}^{r,k}) - \boldsymbol{x}(t|\boldsymbol{\tau})}{\epsilon} = \lim_{\epsilon \to 0} \epsilon^{-1} \boldsymbol{\xi}(t|k, \boldsymbol{\tau}, \epsilon) = \boldsymbol{\psi}^k(t|\boldsymbol{\tau}), \qquad t \in (0, T]. \tag{6.33}$$

We prove (6.33) in three steps.

A Notation

For simplicity, we will write $\boldsymbol{x}^\epsilon$ instead of $\boldsymbol{x}(\cdot|\boldsymbol{\tau} + \epsilon \boldsymbol{e}^{r,k})$ and $\boldsymbol{x}$ instead of $\boldsymbol{x}(\cdot|\boldsymbol{\tau})$. We will also write $\boldsymbol{\xi}^\epsilon$ instead of $\boldsymbol{\xi}(\cdot|k, \boldsymbol{\tau}, \epsilon)$, $\boldsymbol{\gamma}^{\epsilon,i}$ instead of $\boldsymbol{\gamma}^i(\cdot|k, \boldsymbol{\tau}, \epsilon)$, and $\boldsymbol{\phi}$ instead of $\boldsymbol{\phi}(\cdot|\boldsymbol{\tau})$. Since k and $\boldsymbol{\tau}$ are fixed, these simplifications do not cause confusion.

For each $\epsilon \in \mathcal{S}(k, \boldsymbol{\tau})$, define three corresponding $\mathbb{R}^n$-valued functions on $[0, T]$ as follows:

$$\boldsymbol{\upsilon}^{1,\epsilon}(t) \triangleq \sum_{i=1}^{r} \int_0^1 \left\{ \frac{\partial \boldsymbol{f}^i\big(\boldsymbol{x}(t) + \eta \boldsymbol{\xi}^\epsilon(t), \boldsymbol{x}(t-\tau_i) + \eta \boldsymbol{\gamma}^{\epsilon,i}(t)\big)}{\partial \boldsymbol{x}} - \frac{\partial \boldsymbol{f}^i\big(\boldsymbol{x}(t), \boldsymbol{x}(t-\tau_i)\big)}{\partial \boldsymbol{x}} \right\} \boldsymbol{\xi}^\epsilon(t) d\eta,$$

$$\boldsymbol{\upsilon}^{2,\epsilon}(t) \triangleq \sum_{i=1}^{r} \int_0^1 \left\{ \frac{\partial \boldsymbol{f}^i\big(\boldsymbol{x}(t) + \eta \boldsymbol{\xi}^\epsilon(t), \boldsymbol{x}(t-\tau_i) + \eta \boldsymbol{\gamma}^{\epsilon,i}(t)\big)}{\partial \boldsymbol{x}^-} - \frac{\partial \boldsymbol{f}^i\big(\boldsymbol{x}(t), \boldsymbol{x}(t-\tau_i)\big)}{\partial \boldsymbol{x}^-} \right\} \boldsymbol{\gamma}^{\epsilon,i}(t) d\eta,$$

and

$$\boldsymbol{\upsilon}^{3,\epsilon}(t) \triangleq \frac{\partial \boldsymbol{f}^k\big(\boldsymbol{x}(t), \boldsymbol{x}(t-\tau_k)\big)}{\partial \boldsymbol{x}^-} \big(\boldsymbol{\gamma}^{\epsilon,k}(t) - \boldsymbol{\xi}^\epsilon(t-\tau_k) + \epsilon \boldsymbol{\phi}(t-\tau_k)\big).$$

Furthermore, define a function $\beta : \mathcal{S}(k, \boldsymbol{\tau}) \setminus \{0\} \to \mathbb{R}$ by

$$\beta(\epsilon) \triangleq |\epsilon|^{-1} \int_0^T \Big\{ \big|\boldsymbol{\upsilon}^{1,\epsilon}(t)\big| + \big|\boldsymbol{\upsilon}^{2,\epsilon}(t)\big| + \big|\boldsymbol{\upsilon}^{3,\epsilon}(t)\big| \Big\} dt.$$

Finally, in view of Assumption 6.1, we can find two real numbers $\alpha_1 > 0$ and $\alpha_2 > 0$ such that

$$\sup\left\{ |\partial \boldsymbol{f}^i(\boldsymbol{v},\boldsymbol{w})/\partial \boldsymbol{x}| : \ (\boldsymbol{v},\boldsymbol{w}) \in \Psi \times \Psi, \ i \in \{1,\dots,r\} \right\} \leq \alpha_1$$

and

$$\sup\left\{ |\partial \boldsymbol{f}^i(\boldsymbol{v},\boldsymbol{w})/\partial \boldsymbol{x}^-| : \ (\boldsymbol{v},\boldsymbol{w}) \in \Psi \times \Psi, \ i \in \{1,\dots,r\} \right\} \leq \alpha_2,$$

where Ψ is as defined in Section 6.3.

B Behaviour of β as ϵ approaches zero

We now show that $\beta \to 0$ as $\epsilon \to 0$. By Lebesgue's Dominated Convergence Theorem [5], it is sufficient to prove the following two results:

(i) The family of functions

$$\left\{ |\epsilon^{-1}\boldsymbol{v}^{1,\epsilon}| + |\epsilon^{-1}\boldsymbol{v}^{2,\epsilon}| + |\epsilon^{-1}\boldsymbol{v}^{3,\epsilon}| : \ \epsilon \in \mathcal{S}(k,\boldsymbol{\tau}) \setminus \{0\} \right\}$$

is equibounded on $[0,T]$; and

(ii) For almost all $t \in [0,T]$,

$$\left|\epsilon^{-1}\boldsymbol{v}^{1,\epsilon}(t)\right| + \left|\epsilon^{-1}\boldsymbol{v}^{2,\epsilon}(t)\right| + \left|\epsilon^{-1}\boldsymbol{v}^{3,\epsilon}(t)\right| \to 0 \qquad \text{as} \qquad \epsilon \to 0.$$

We prove (i)-(ii) below.

First, note that Ψ is convex. Hence, for each $\epsilon \in \mathcal{S}(k,\boldsymbol{\tau})$,

$$\boldsymbol{x}(t) + \eta\boldsymbol{\xi}^\epsilon(t) \in \Psi, \qquad \eta \in [0,1], \qquad t \in [0,T], \tag{6.34}$$

and

$$\boldsymbol{x}(t-\tau_i) + \eta\boldsymbol{\gamma}^{\epsilon,i}(t) \in \Psi, \qquad \eta \in [0,1], \qquad t \in [0,T], \qquad i = 1,\dots,r. \tag{6.35}$$

Therefore, by inequalities (6.19) and (6.20),

$$\left|\epsilon^{-1}\boldsymbol{v}^{1,\epsilon}(t)\right| \leq 2r\alpha_1 L_4, \qquad t \in [0,T], \qquad \epsilon \in \mathcal{S}(k,\boldsymbol{\tau}) \setminus \{0\}, \tag{6.36}$$

and

$$\left|\epsilon^{-1}\boldsymbol{v}^{2,\epsilon}(t)\right| \leq 2r\alpha_2 L_4, \qquad t \in [0,T], \qquad \epsilon \in \mathcal{S}(k,\boldsymbol{\tau}) \setminus \{0\}. \tag{6.37}$$

Similarly, by inequalities (6.19) and (6.20) and Lemma 6.2,

$$\left|\epsilon^{-1}\boldsymbol{v}^{3,\epsilon}(t)\right| \leq \alpha_2(2L_4 + L_3), \qquad t \in [0,T], \qquad \epsilon \in \mathcal{S}(k,\boldsymbol{\tau}) \setminus \{0\}. \tag{6.38}$$

Result (i) follows immediately from (6.36)-(6.38).

Now, it is clear from inequalities (6.19) and (6.20) that the following limits exist uniformly with respect to $\eta \in [0,1]$ and $t \in [0,T]$:

$$\boldsymbol{x}(t) + \eta \boldsymbol{\xi}^{\epsilon}(t) \to \boldsymbol{x}(t) \quad \text{as} \quad \epsilon \to 0$$

and

$$\boldsymbol{x}(t-\tau_i) + \eta \boldsymbol{\gamma}^{\epsilon,i}(t) \to \boldsymbol{x}(t-\tau_i) \qquad \text{as} \qquad \epsilon \to 0, \qquad i = 1, \ldots, r.$$

Moreover, this convergence takes place inside the closed ball Ψ—see inclusions (6.34) and (6.35). Also note from Assumption 6.1 that the functions $\partial \boldsymbol{f}^i / \partial \boldsymbol{x}$ and $\partial \boldsymbol{f}^i / \partial \boldsymbol{x}^-$, $i = 1, \ldots, r$, are uniformly continuous on $\Psi \times \Psi$. Hence, the following limits exist uniformly with respect to $\eta \in [0,1]$ and $t \in [0,T]$:

$$\lim_{\epsilon \to 0} \frac{\partial \boldsymbol{f}^i\big(\boldsymbol{x}(t) + \eta \boldsymbol{\xi}^{\epsilon}(t), \boldsymbol{x}(t-\tau_i) + \eta \boldsymbol{\gamma}^{\epsilon,i}(t)\big)}{\partial \boldsymbol{x}} = \frac{\partial \boldsymbol{f}^i\big(\boldsymbol{x}(t), \boldsymbol{x}(t-\tau_i)\big)}{\partial \boldsymbol{x}}, \qquad i = 1, \ldots, r,$$

and

$$\lim_{\epsilon \to 0} \frac{\partial \boldsymbol{f}^i\big(\boldsymbol{x}(t) + \eta \boldsymbol{\xi}^{\epsilon}(t), \boldsymbol{x}(t-\tau_i) + \eta \boldsymbol{\gamma}^{\epsilon,i}(t)\big)}{\partial \boldsymbol{x}^-} = \frac{\partial \boldsymbol{f}^i\big(\boldsymbol{x}(t), \boldsymbol{x}(t-\tau_i)\big)}{\partial \boldsymbol{x}^-}, \qquad i = 1, \ldots, r.$$

These limits, together with inequalities (6.19) and (6.20), imply that $\epsilon^{-1}\boldsymbol{\upsilon}^{1,\epsilon}$ and $\epsilon^{-1}\boldsymbol{\upsilon}^{2,\epsilon}$ converge to zero uniformly on $[0,T]$ as $\epsilon \to 0$. Furthermore, Lemma 6.4 implies that $\epsilon^{-1}\boldsymbol{\upsilon}^{3,\epsilon}$ converges to zero almost everywhere on $[0,T]$ as $\epsilon \to 0$. Result (ii) follows immediately.

Thus, since both (i) and (ii) hold, we conclude that

$$\lim_{\epsilon \to 0} \beta(\epsilon) = 0. \tag{6.39}$$

C Comparing $\epsilon^{-1}\boldsymbol{\xi}^{\epsilon}$ with $\boldsymbol{\psi}^k(\cdot|\boldsymbol{\tau})$

Let $\epsilon \in \mathcal{S}(k,\boldsymbol{\tau}) \setminus \{0\}$ be arbitrary but fixed. By (6.9), we have

$$\begin{aligned}\boldsymbol{\xi}^{\epsilon}(t) &= \int_0^t \big\{\boldsymbol{\phi}(s|k,\boldsymbol{\tau},\epsilon) - \boldsymbol{\phi}(s|\boldsymbol{\tau})\big\} ds \\ &= \sum_{i=1}^{r} \int_0^t \Big\{\boldsymbol{f}^i\big(\boldsymbol{x}^{\epsilon}(s), \boldsymbol{x}^{\epsilon}(s-\tau_i-\epsilon\rho_{k,i})\big) - \boldsymbol{f}^i\big(\boldsymbol{x}(s), \boldsymbol{x}(s-\tau_i)\big)\Big\} ds, \quad t \in (0,T].\end{aligned}$$

Thus, by the mean value theorem,

$$\begin{aligned}\boldsymbol{\xi}^{\epsilon}(t) =& \sum_{i=1}^{r}\int_0^t\int_0^1 \Big\{\frac{\partial \boldsymbol{f}^i\big(\boldsymbol{x}(s)+\eta\boldsymbol{\xi}^{\epsilon}(s),\boldsymbol{x}(s-\tau_i)+\eta\boldsymbol{\gamma}^{\epsilon,i}(s)\big)}{\partial \boldsymbol{x}}\boldsymbol{\xi}^{\epsilon}(s)\\ &+\frac{\partial \boldsymbol{f}^i\big(\boldsymbol{x}(s)+\eta\boldsymbol{\xi}^{\epsilon}(s),\boldsymbol{x}(s-\tau_i)+\eta\boldsymbol{\gamma}^{\epsilon,i}(s)\big)}{\partial \boldsymbol{x}^-}\boldsymbol{\gamma}^{\epsilon,i}(s)\Big\}d\eta ds\\ =&\sum_{i=1}^{r}\int_0^t\Big\{\frac{\partial \boldsymbol{f}^i\big(\boldsymbol{x}(s),\boldsymbol{x}(s-\tau_i)\big)}{\partial \boldsymbol{x}}\boldsymbol{\xi}^{\epsilon}(s)+\frac{\partial \boldsymbol{f}^i\big(\boldsymbol{x}(s),\boldsymbol{x}(s-\tau_i)\big)}{\partial \boldsymbol{x}^-}\boldsymbol{\gamma}^{\epsilon,i}(s)\Big\}ds\\ &+\int_0^t\Big\{\boldsymbol{\upsilon}^{1,\epsilon}(s)+\boldsymbol{\upsilon}^{2,\epsilon}(s)\Big\}ds,\quad t\in(0,T].\end{aligned}\tag{6.40}$$

Substituting equation (6.6) into equation (6.40) gives

$$\begin{aligned}\boldsymbol{\xi}^{\epsilon}(t) =& \sum_{i=1}^{r}\int_0^t\Big\{\frac{\partial \boldsymbol{f}^i\big(\boldsymbol{x}(s),\boldsymbol{x}(s-\tau_i)\big)}{\partial \boldsymbol{x}}\boldsymbol{\xi}^{\epsilon}(s)+\frac{\partial \boldsymbol{f}^i\big(\boldsymbol{x}(s),\boldsymbol{x}(s-\tau_i)\big)}{\partial \boldsymbol{x}^-}\boldsymbol{\xi}^{\epsilon}(s-\tau_i)\Big\}ds\\ &+\int_0^t\frac{\partial \boldsymbol{f}^k\big(\boldsymbol{x}(s),\boldsymbol{x}(s-\tau_k)\big)}{\partial \boldsymbol{x}^-}\big(\boldsymbol{\gamma}^{\epsilon,k}(s)-\boldsymbol{\xi}^{\epsilon}(s-\tau_k)\big)ds\\ &+\int_0^t\Big\{\boldsymbol{\upsilon}^{1,\epsilon}(s)+\boldsymbol{\upsilon}^{2,\epsilon}(s)\Big\}ds,\qquad t\in(0,T].\end{aligned}\tag{6.41}$$

Now, integrating the auxiliary system (6.31)-(6.32) yields

$$\begin{aligned}\boldsymbol{\psi}^k(t|\boldsymbol{\tau}) =& \sum_{i=1}^{r}\int_0^t\Big\{\frac{\partial \boldsymbol{f}^i\big(\boldsymbol{x}(s),\boldsymbol{x}(s-\tau_i)\big)}{\partial \boldsymbol{x}}\boldsymbol{\psi}^k(s|\boldsymbol{\tau})+\frac{\partial \boldsymbol{f}^i\big(\boldsymbol{x}(s),\boldsymbol{x}(s-\tau_i)\big)}{\partial \boldsymbol{x}^-}\boldsymbol{\psi}^k(s-\tau_i|\boldsymbol{\tau})\Big\}ds\\ &-\int_0^t\frac{\partial \boldsymbol{f}^k\big(\boldsymbol{x}(s),\boldsymbol{x}(s-\tau_k)\big)}{\partial \boldsymbol{x}^-}\boldsymbol{\phi}(s-\tau_k)ds,\qquad t\in(0,T].\end{aligned}\tag{6.42}$$

Combining equations (6.41) and (6.42), we obtain

$$\begin{aligned}\big|\epsilon^{-1}\boldsymbol{\xi}^{\epsilon}(t)-\boldsymbol{\psi}^k(t|\boldsymbol{\tau})\big| \le \beta(\epsilon)&+\int_0^t r\alpha_1\big|\epsilon^{-1}\boldsymbol{\xi}^{\epsilon}(s)-\boldsymbol{\psi}^k(s|\boldsymbol{\tau})\big|ds\\ &+\sum_{i=1}^{r}\int_0^t\alpha_2\big|\epsilon^{-1}\boldsymbol{\xi}^{\epsilon}(s-\tau_i)-\boldsymbol{\psi}^k(s-\tau_i|\boldsymbol{\tau})\big|ds,\quad t\in(0,T].\end{aligned}$$

In view of (6.7) and (6.32), this inequality simplifies to

$$\big|\epsilon^{-1}\boldsymbol{\xi}^{\epsilon}(t)-\boldsymbol{\psi}^k(t|\boldsymbol{\tau})\big| \le \beta(\epsilon)+\int_0^t r(\alpha_1+\alpha_2)\big|\epsilon^{-1}\boldsymbol{\xi}^{\epsilon}(s)-\boldsymbol{\psi}^k(s|\boldsymbol{\tau})\big|ds,\quad t\in(0,T].$$

Applying Gronwall's Lemma gives

$$\big|\epsilon^{-1}\boldsymbol{\xi}^{\epsilon}(t)-\boldsymbol{\psi}^k(t|\boldsymbol{\tau})\big| \le \beta(\epsilon)\exp\big[r(\alpha_1+\alpha_2)T\big],\qquad t\in(0,T].$$

Now, since ϵ was chosen arbitrarily, we can take the limit as $\epsilon \to 0$ to obtain

$$\left|\epsilon^{-1}\boldsymbol{\xi}^{\epsilon}(t) - \boldsymbol{\psi}^{k}(t|\boldsymbol{\tau})\right| \leq \exp\left[r(\alpha_1+\alpha_2)T\right] \lim_{\epsilon\to 0}\beta(\epsilon), \qquad t \in (0,T].$$

Since $\beta \to 0$ as $\epsilon \to 0$ (recall equation (6.39)), this proves equation (6.33). □

Remark 6.1. If $t \in [-\bar{\tau}, 0]$ and $(k, \boldsymbol{\tau}) \in \{1,\ldots,r\} \times \mathcal{T}$, then

$$\frac{\partial \boldsymbol{x}(t|\boldsymbol{\tau})}{\partial \tau_k} = \lim_{\epsilon\to 0}\frac{\boldsymbol{x}(t|\boldsymbol{\tau}+\epsilon\boldsymbol{e}^{r,k}) - \boldsymbol{x}(t|\boldsymbol{\tau})}{\epsilon} = \lim_{\epsilon\to 0}\frac{\boldsymbol{z}(t)-\boldsymbol{z}(t)}{\epsilon} = \boldsymbol{0} = \boldsymbol{\psi}^{k}(t|\boldsymbol{\tau}).$$

Hence, Theorem 6.1 actually holds at every point in the time horizon $[-\bar{\tau}, T]$.

6.5 Gradient computation

Recall our state-delay identification problem from Section 6.2: *For the model* (6.1)-(6.3)*, choose values for the unknown state-delays* τ_i*,* $i = 1,\ldots,r$*, such that the discrepancy between predicted and measured output is minimized.* This problem was formulated mathematically as Problem P, an optimal control problem involving a nonlinear delay-differential system. The control variables τ_i, $i = 1,\ldots,r$, in Problem P influence its cost function G_0 *implicitly* through the governing delay-differential system. Hence, three important questions arise:

- Is G_0 continuous?
- If G_0 is continuous, is it also differentiable?
- If G_0 is differentiable, is there a viable method for computing its gradient?

The next result, which follows readily from Theorem 6.1, shows that G_0 is indeed differentiable (and therefore continuous) on $\mathcal{T}$.

Theorem 6.2. *For each* $\boldsymbol{\tau} \in \mathcal{T}$*,*

$$\frac{\partial G_0(\boldsymbol{\tau})}{\partial \tau_k} = 2\sum_{j=1}^{p}\left[\boldsymbol{y}(t_j|\boldsymbol{\tau}) - \hat{\boldsymbol{y}}^{j}\right]^{T}\frac{\partial \boldsymbol{g}(\boldsymbol{x}(t_j|\boldsymbol{\tau}))}{\partial \boldsymbol{x}}\boldsymbol{\psi}^{k}(t_j|\boldsymbol{\tau}), \qquad k = 1,\ldots,r. \tag{6.43}$$

Proof. According to Theorem 6.1, the functions $\boldsymbol{x}(t_j|\cdot)$, $j = 1,\ldots,p$, are differentiable at $\boldsymbol{\tau} \in \mathcal{T}$. Furthermore,

$$\frac{\partial \boldsymbol{x}(t_j|\boldsymbol{\tau})}{\partial \tau_k} = \boldsymbol{\psi}^{k}(t_j|\boldsymbol{\tau}), \qquad k = 1,\ldots,r, \qquad j = 1,\ldots,p. \tag{6.44}$$

Now, we have

$$G_0(\boldsymbol{\tau}) = \sum_{j=1}^{p}\sum_{l=1}^{m}\left[g_l(\boldsymbol{x}(t_j|\boldsymbol{\tau})) - \hat{y}_l^{j}\right]^2.$$

Differentiating this equation with respect to τ_k, $k = 1, \ldots, r$, gives

$$\frac{\partial G_0(\boldsymbol{\tau})}{\partial \tau_k} = 2 \sum_{j=1}^{p} \sum_{l=1}^{m} \left[g_l(\boldsymbol{x}(t_j|\boldsymbol{\tau})) - \hat{y}_l^j \right] \frac{\partial g_l(\boldsymbol{x}(t_j|\boldsymbol{\tau}))}{\partial \boldsymbol{x}} \frac{\partial \boldsymbol{x}(t_j|\boldsymbol{\tau})}{\partial \tau_k}, \qquad k = 1, \ldots, r.$$

Therefore, by equation (6.44),

$$\frac{\partial G_0(\boldsymbol{\tau})}{\partial \tau_k} = 2 \sum_{j=1}^{p} \sum_{l=1}^{m} \left[y_l(t_j|\boldsymbol{\tau}) - \hat{y}_l^j \right] \frac{\partial g_l(\boldsymbol{x}(t_j|\boldsymbol{\tau}))}{\partial \boldsymbol{x}} \boldsymbol{\psi}^k(t_j|\boldsymbol{\tau}), \qquad k = 1, \ldots, r.$$

Equation (6.43) then follows readily. □

Theorem 6.2 shows that the partial derivatives of G_0 can be expressed in terms of the solution of the state system (6.1)-(6.2) and the solutions of the auxiliary systems (6.31)-(6.32). Note that these systems can be combined to form an expanded system of delay-differential equations. On this basis, we propose the following algorithm for computing the value of G_0 and its gradient.

Algorithm 6.1. Input $\boldsymbol{\tau} \in \mathcal{T}$.

(i) Obtain $\boldsymbol{x}(\cdot|\boldsymbol{\tau})$ and $\boldsymbol{\psi}^k(\cdot|\boldsymbol{\tau})$, $k = 1, \ldots, r$, by solving the delay-differential system consisting of (6.1)-(6.2) and (6.31)-(6.32).

(ii) Use $\boldsymbol{x}(\cdot|\boldsymbol{\tau})$ to compute $\boldsymbol{y}(\cdot|\boldsymbol{\tau})$.

(iii) Use $\boldsymbol{y}(t_j|\boldsymbol{\tau})$, $j = 1, \ldots, p$, to compute $G_0(\boldsymbol{\tau})$.

(iv) Use $\boldsymbol{x}(t_j|\boldsymbol{\tau})$, $\boldsymbol{y}(t_j|\boldsymbol{\tau})$, and $\boldsymbol{\psi}^k(t_j|\boldsymbol{\tau})$, $j = 1, \ldots, p$, to compute the partial derivatives $\partial G_0(\boldsymbol{\tau})/\partial \tau_k$, $k = 1, \ldots, r$, according to the formula in Theorem 6.2.

6.6 Numerical examples

Problem P can be solved using Algorithm 6.1 in conjunction with a gradient-based nonlinear programming software such as NLPQLP (see [93]). We illustrate this approach by considering two examples: a predator-prey model from [131], and a continuously-stirred tank reactor model from [13].

6.6.1 A predator-prey model

The following predator-prey model appears in [131]:

$$\dot{x}_1(t) = x_1(t)\left[1 - 2x_1(t) - \frac{x_3(t)}{x_3(t) + x_1(t)}\right] + 0.5(x_2(t) - x_1(t)), \qquad t \in [0, 5], \qquad (6.45a)$$

$$\dot{x}_2(t) = x_2(t)(1 - 2x_2(t)) + 0.5(x_1(t) - x_2(t)), \qquad t \in [0, 5], \qquad (6.45b)$$

$$\dot{x}_3(t) = x_3(t)\left[-3 + \frac{10x_1(t-\tau)}{x_3(t-\tau) + x_1(t-\tau)}\right], \qquad t \in [0, 5], \qquad (6.45c)$$

and

$$x_1(t) = 1, \quad x_2(t) = 1, \quad x_3(t) = 1, \qquad t \leq 0, \qquad (6.46)$$

where $x_1(t)$ and $x_2(t)$ are the prey population sizes at time t, $x_3(t)$ is the predator population size at time t, and τ is an unknown state-delay that needs to be identified.

We assume that each of the state variables can be measured directly. Therefore, the system output is identical to the state:

$$\boldsymbol{y}(t) = \big[x_1(t), x_2(t), x_3(t)\big]^T.$$

To generate the observed output in Problem P, we simulated the system (6.45)-(6.46) with $\tau = 0.5$ and recorded the state at fifty equidistant time points in $[0, 5]$. This data is used as the observed data. Hence,

$$\hat{\boldsymbol{x}}^j \triangleq \boldsymbol{x}(t_j|0.5), \qquad j = 1, \ldots, 50,$$

where

$$t_j = \frac{j}{10}, \qquad j = 1, \ldots, 50.$$

Our state-delay identification problem is as follows: choose τ to minimize the cost function

$$\sum_{j=1}^{50} \left|\boldsymbol{x}(t_j|\tau) - \hat{\boldsymbol{x}}^j\right|^2$$

subject to the dynamics (6.45)-(6.46).

We wrote a Fortran program, which combines Algorithm 6.1 with NLPQLP and the differential equation solver LSODA (see [41]), to solve this state-delay identification problem. The program was run several times with the following initial values for the state-delay: $\tau = 0.1$, $\tau = 0.3$, $\tau = 0.7$, and $\tau = 0.9$. In each case, the program recovered the optimal state-delay $\tau = 0.5$ in less than eleven iterations. States 1 and 3 corresponding to the different initial state-delays are plotted with the observed data in Figures 6.1-6.2. State 2 is insensitive to changes in the state-delay and is therefore not plotted. Notice that

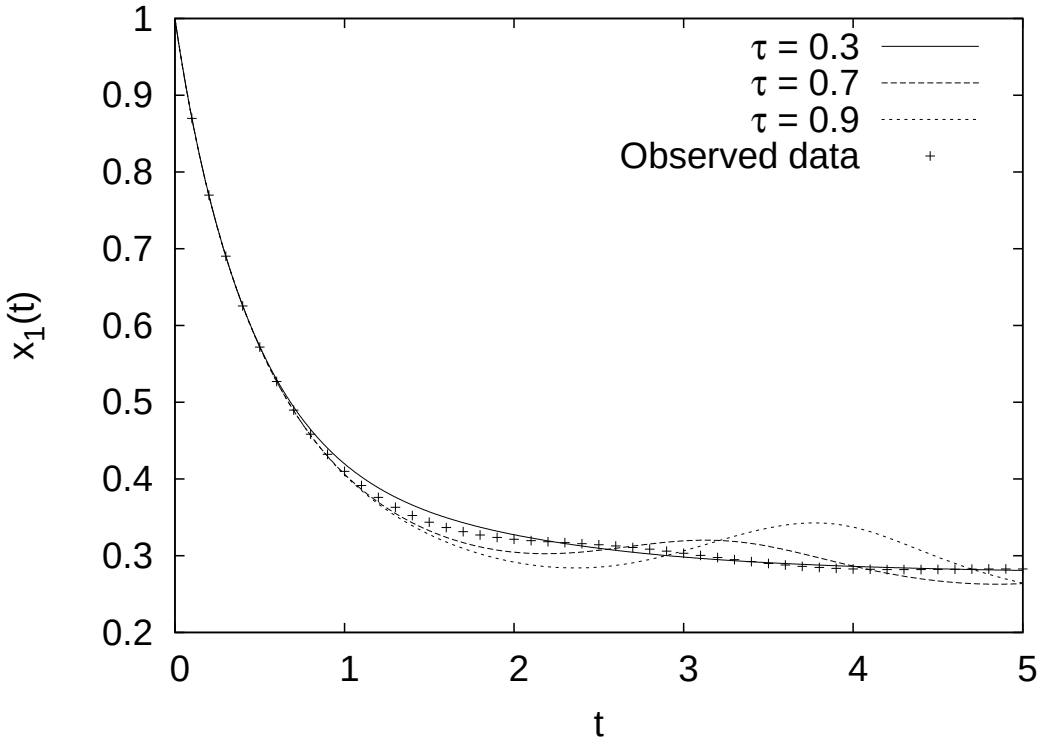

Figure 6.1: State 1 in Example 6.6.1, simulated using different values for the delay. The observed data corresponds to $\tau = 0.5$.

some of the initial state trajectories differ significantly from the observed data. Despite this, our program always converged quickly to the optimal state-delay.

6.6.2 A continuously stirred tank reactor model

The following model of a continuously-stirred tank reactor appears in [13]:

$$\dot{x}_1(t) = -2x_1(t) + 0.1(1 - x_1(t)) \exp\left[\frac{x_2(t)}{1 + 0.05x_2(t)}\right] + x_1(t - \tau), \qquad t \in [0, 10], \quad (6.47a)$$

$$\dot{x}_2(t) = -2.5x_2(t) + 0.8(1 - x_1(t)) \exp\left[\frac{x_2(t)}{1 + 0.05x_2(t)}\right] + x_2(t - \tau), \quad t \in [0, 10], \quad (6.47b)$$

and

$$x_1(t) = 1, \quad x_2(t) = 1, \qquad t \leq 0, \quad (6.48)$$

where $x_1(t)$ is the dimensionless concentration at time t, $x_2(t)$ is the dimensionless temperature at time t, and τ is an unknown state-delay that needs to be identified.

We assume that only the temperature can be measured. Hence,

$$y(t) = x_2(t). \quad (6.49)$$

We generated the observed data by simulating (6.47)-(6.49) with $\tau = 2$ and recording the

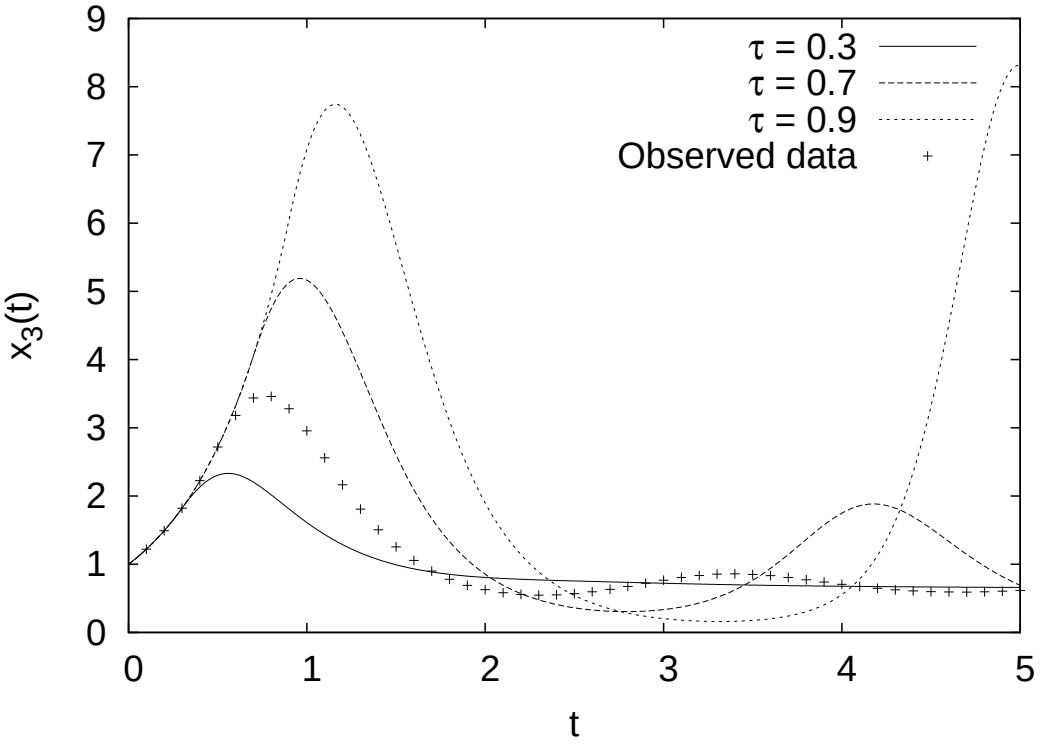

Figure 6.2: State 3 in Example 6.6.1, simulated using different values for the delay. The observed data corresponds to $\tau = 0.5$.

output (that is, the temperature) at twenty equidistant points in $[0, 10]$. Therefore,

$$\hat{y}^j \triangleq y(t_j|2), \qquad j = 1, \ldots, 20,$$

where

$$t_j = \frac{j}{2}, \qquad j = 1, \ldots, 20.$$

Our state-delay identification problem is as follows: choose τ to minimize the cost function

$$\sum_{j=1}^{20} \left| y(t_j|\tau) - \hat{y}^j \right|^2$$

subject to the dynamics (6.47)-(6.49). As in Example 6.6.1, a Fortran program was written to solve this problem. The program was run with the following initial guesses for the unknown state-delay: $\tau = 1.6$, $\tau = 1.8$, $\tau = 2.2$, and $\tau = 2.4$. In each case, the program recovered the optimal solution of $\tau = 2$ successfully in less than seven iterations.

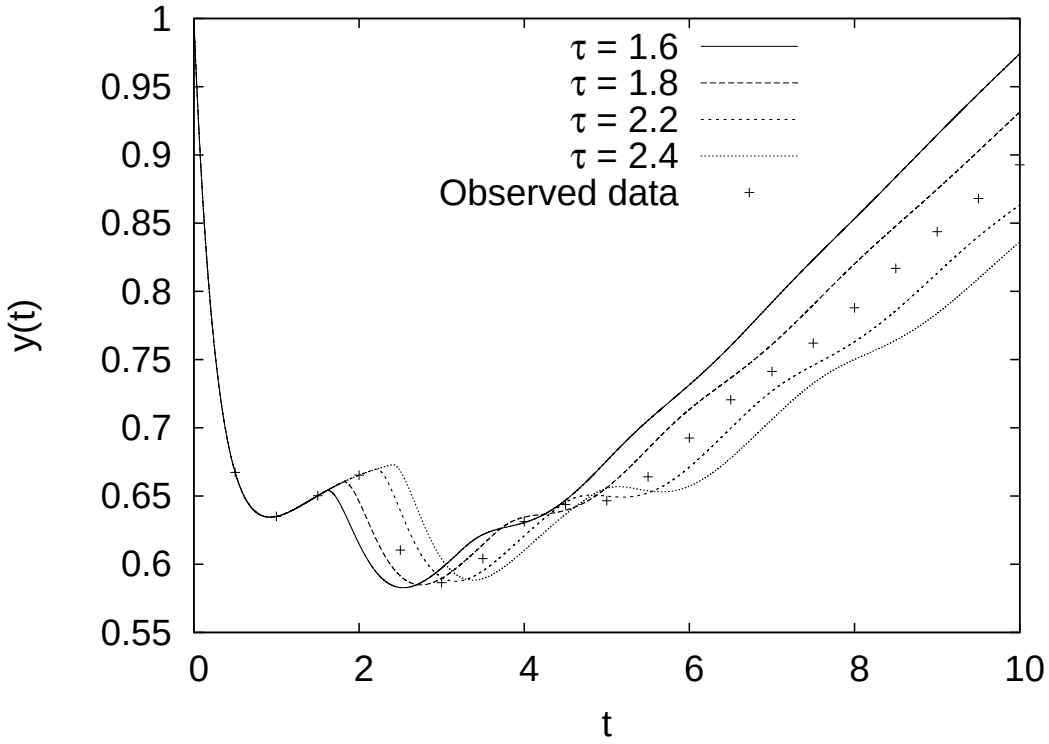

Figure 6.3: Output of the system in Example 6.6.2, simulated using different values for the delay. The observed data corresponds to $\tau = 2$.

6.7 Conclusion

In this chapter, we considered a nonlinear delay-differential system whose state-delays are unknown parameters that need to be identified. We formulated the problem of identifying these state-delays as an optimal control problem in which the state-delays are control variables and the cost function measures the discrepancy between predicted and observed system output. This optimal control problem differs considerably from standard time-delay optimal control problems, because its control variables are the delays themselves. Our main result shows that the gradient of its cost function can be computed by solving a set of auxiliary delay-differential systems. On this basis, we can use a standard nonlinear programming software, such as NLPQLP, to solve the optimal control problem. Our numerical results in Section 6.6 indicate that this approach is very fast; it is therefore ideal for online applications in which efficiency is paramount.

CHAPTER 7

Summary and future research directions

7.1 Main contributions of this thesis

In this thesis, we considered five nonstandard optimal control problems. We developed new methods, which are based on nonlinear programming, for solving these problems numerically. This involved a variety of novel techniques, including the time-scaling transformation and several model transcriptions, as well as complex gradient derivation and detailed convergence analysis. We summarize our main contributions below.

In Chapter 2, we developed a new control parameterization method for solving optimal control problems with characteristic-time inequality constraints. The main idea of this method is to approximate the control by a piecewise constant function whose values and switching times are decision variables to be determined optimally. The approximate control is allowed to change its value at each characteristic time, and also at $p-1$ locations between consecutive characteristic times (p is a fixed integer). This ensures that the order in which the switching times and the characteristic times occur is known in advance (every pth switching time is a characteristic time). Consequently, the time-scaling transformation is able to *simultaneously* map the switching times and the characteristic times to fixed points in a new time horizon. As such, our new method is much easier to implement than the one in [105], which applies the time-scaling transformation twice in succession—once to transform the characteristic times, and once more to transform the switching times (in [105], the characteristic times and the switching times do not coincide). The methods in [77, 78], meanwhile, do not allow the switching times or characteristic times to vary at all; they instead pre-assign them and only choose the control values optimally. In Chapter 2, we also developed a new algorithm for computing the gradient of the characteristic-time inequality constraints. This algorithm involves integrating a set of auxiliary dynamic systems forward in time, simultaneously with the state system. In contrast, the auxiliary systems in [77, 78, 105] are integrated *backwards* in time, and thus the state needs to be interpolated as they are being solved (the auxiliary systems depend on the state). Consequently, our new gradient computation algorithm in Chapter 2 is easier to implement than the ones in [77, 78, 105]. Furthermore, because it

avoids interpolating the state, the algorithm in Chapter 2 is also more accurate.

In Chapter 3, we developed another control parameterization method, this time for optimal control problems with continuous inequality constraints. This new method has two major advantages over its predecessors in [100, 101]: it incorporates the time-scaling transformation so that the control switching times, in addition to the control values, are chosen optimally; and it is guaranteed to converge even when the continuous inequality constraints are explicit functions of the control. In particular, we proved under very mild assumptions that the cost of the suboptimal controls generated by this method converges to the minimum cost (Theorem 3.5). Furthermore, if the suboptimal controls converge almost everywhere to a piecewise continuous function, then this function is an optimal control (Theorem 3.6). The corresponding convergence results in [100, 101] are only valid for optimal control problems with continuous inequality constraints that do not depend on the control function explicitly. To handle the continuous constraints in Chapter 3, we applied the constraint transcription introduced in [48, 106]. This transcription is also used in [100, 101], but in a different way—one that is invalid if the continuous inequality constraints depend explicitly on the control. Our new approach in Chapter 3 is therefore applicable to a much broader class of optimal control problems.

In Chapter 4, we considered the problem of determining an optimal operating schedule for a switched-capacitor DC-DC power converter. The optimal control problem that we formulated is similar to the one in [42]—it involves choosing the topology switching times so that the output voltage ripple and the output voltage sensitivity are minimized. The method proposed in [42] for solving this problem requires that each eigenvalue of the system coefficient matrix be expressed *analytically* in terms of the load resistance. Such expressions are usually very difficult to obtain—in fact, they can only be derived when the dimension of the system coefficient matrix is less than five. We developed an alternative method that is much easier to use. We also proved that the optimal control problem has a solution. To the best of our knowledge, Chapter 4 and [42] are the first attempts at modeling a switched-capacitor DC-DC power converter as a switched system. This model is much more accurate than the linear time-invariant models typically used in the literature, which do not reflect the highly nonlinear and time-varying nature of a switched-capacitor DC-DC power converter.

In Chapter 5, we considered the optimal control of a switched system with nonlinear subsystems and nonlinear state jump conditions. We showed that applying the time-scaling transformation to this problem yields a new optimal control problem that is governed by a switched system with discontinuous state jump conditions. The discontinuities arise because the time-scaling transformation maps each switching time to a distinct integer, even if some of the switching times coincide. For example, the time-scaling transformation always maps $t = t_1$ to $s = 1$ and $t = t_2$ to $s = 2$, regardless of whether $t_1 < t_2$ or $t_1 = t_2$. But if $t_1 = t_2$, then $s = 1$ and $s = 2$ correspond to the same switching time

in the original time horizon, and thus a state jump should be applied at either $s = 1$ or $s = 2$, but not both. This subtlety was neglected in previous work (see [68,69,125]). In Chapter 5, we proposed a new computational approach to address this important issue.

In Chapter 6, we considered the problem of identifying unknown state-delays in a nonlinear delay-differential system. This problem was formulated as an unusual optimal control problem in which the control variables are the state-delays themselves. We showed that the gradient of the cost function in this optimal control problem can be computed by solving an auxiliary delay-differential system. On this basis, the optimal control problem can be solved as a nonlinear programming problem using any gradient-based optimization algorithm. The major advantage of this approach is that it is applicable to a very broad class of nonlinear delay systems; most other delay identification methods are only applicable to linear systems. Furthermore, since this method is based on efficient nonlinear programming techniques, it has excellent potential for real-time applications.

7.2 Future research directions

The work in this thesis has opened several interesting new avenues for future research. We discuss some of them below.

Recall that the admissible controls in Chapter 2 are bounded measurable functions, while the admissible controls in Chapter 3 are restricted to piecewise continuous functions. This restriction is deliberate—the proofs of the convergence results in Section 3.6 are only valid when the controls are piecewise continuous. A question of considerable theoretical interest is whether these convergence results still hold when the class of admissible controls is enlarged to include functions that are not necessarily piecewise continuous, such as functions of bounded variation or even bounded measurable functions. Similar results have been proved in [102] and Chapter 10 of [100] for optimal control problems in which the controls consist of functions of bounded variation and the continuous inequality constraints only restrict the state. The objective function in these problems has a term that penalizes the total variation of the control. The reason for including this term is that in practice there is always some cost associated with changing the input to a system, and thus a control that fluctuates wildly is probably not suitable for implementation. We are currently investigating extending the techniques developed in Chapter 3 to this interesting class of optimal control problems.

The switched systems considered in Chapters 4 and 5 are called *externally-forced switched systems* because their switching mechanisms are under the direct control of the system operator [132]. *Internally-forced switched systems* are very different: their switching times are not chosen beforehand, and are instead determined implicitly by the state trajectory. More specifically, the subsystem switches occur when a given switching criterion—usually an equation depending on the system state—is satisfied. Many systems,

including robots [10] and hybrid power systems [90], are of this type. Internally-forced switched systems are more complicated than externally-forced switched systems because their switching times are not known in advance. They can be viewed as an extension of the so-called *free terminal time* dynamic system (see [65, 99, 104]), whose terminal time is determined by a stopping condition that depends on the state. Extending the optimal control methods developed in Chapters 4 and 5 to internally-forced switched systems is an interesting topic for future research.

The governing dynamic system in the time-delay optimal control problem considered in Chapter 6 is an example of a so-called *input-dependent delay system*—a delay system whose delays are influenced by the control variables. The optimal control of such systems is a difficult topic that has been neglected by the research community [88]. Many important industrial processes, however, can be modeled by input-dependent delay systems. An example is the crushing process described in [88]. In this process, raw material is taken from a repository and delivered to a crusher via a conveyer belt. After crushing, some of the processed material is returned, via another conveyor belt, to the repository, where it is stored before being delivered to the crusher once again. Recycling the output in this way ensures that most of the material undergoes several rounds of crushing (the number of rounds required depends on the desired consistency of the material). Obviously, the recycling mechanism is not instantaneous; there is a delay while the crushed material is transported from the crusher back to the repository. The system controller can influence this delay by varying the speed of the conveyor belts.

Another example of an input-dependent delay system is the continuously-stirred tank reactor described in [21]. This system consists of a water tank with an impeller, an inlet for adding salt, and a conductivity probe for measuring salt concentration. Since salt does not dissolve instantaneously, there is a delay between the time at which the salt is added and the time at which it is detected by the probe. The system controller can influence this delay by varying the speed of the impeller (the faster the impeller rotates, the quicker the salt dissolves and is subsequently detected by the probe).

The computational method developed in Chapter 6 is only applicable to the state-delay identification problem, an optimal control problem governed by an input-dependent delay system with *time-invariant* delays. More research is needed to extend this method to optimal control problems governed by more complicated input-dependent delay systems, such as the crushing system and the continuously-stirred tank reactor described above, whose delays are time-varying. Some systems even have delays that depend on the state (see [111]). We are currently investigating optimal control methods for such systems.

Bibliography

[1] N. U. Ahmed, *Elements of finite-dimensional systems and control theory.* Essex: Longman Scientific and Technical, 1988.

[2] ——, *Dynamic systems and control with applications.* Singapore: World Scientific, 2006.

[3] B. Arntzen and D. Maksimović, "Switched-capacitor DC/DC converters with resonant gate drive," *IEEE Transactions on Power Electronics*, vol. 13, no. 5, pp. 892–902, 1998.

[4] M. Bardi and I. Capuzzo-Dolcetta, *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations.* Boston: Birkhäuser, 2008.

[5] R. G. Bartle, *The elements of integration and Lebesgue measure*, Wiley Classics Library ed. New York: John Wiley, 1995.

[6] R. G. Bartle and D. R. Sherbert, *Introduction to real analysis*, 3rd ed. New York: John Wiley, 2000.

[7] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty, *Nonlinear programming: Theory and algorithms.* New Jersey: John Wiley, 2006.

[8] R. Bellman, *Dynamic programming.* New York: Dover, 2003.

[9] S. C. Bengea and R. A. DeCarlo, "Optimal control of switching systems," *Automatica*, vol. 41, no. 1, pp. 11–27, 2005.

[10] M. Boccadoro, Y. Wardi, M. Egerstedt, and E. Verriest, "Optimal control of switching surfaces in hybrid dynamical systems," *Discrete Event Dynamic Systems: Theory and Applications*, vol. 15, no. 4, pp. 433–448, 2005.

[11] Büskens, C. and Maurer, H., "SQP-methods for solving optimal control problems with control and state constraints: adjoint variables, sensitivity analysis and real-time control," *Journal of Computational and Applied Mathematics*, vol. 120, no. 1-2, pp. 85–108, 2000.

[12] L. Caccetta, I. Loosen, and V. Rehbock, "Computational aspects of the optimal transit path problem," *Journal of Industrial and Management Optimization*, vol. 4, no. 1, pp. 95–105, 2008.

[13] Y. Y. Cao and P. M. Frank, "Analysis and synthesis of nonlinear time-delay systems via fuzzy control approach," *IEEE Transactions on Fuzzy Systems*, vol. 8, no. 2, pp. 200–211, 2000.

[14] D. Chatterjee and D. Liberzon, "Stability analysis of deterministic and stochastic switched systems via a comparison principle and multiple Lyapunov functions," *SIAM Journal on Control and Optimization*, vol. 45, no. 1, pp. 174–206, 2006.

[15] T. W. C. Chen and V. S. Vassiliadis, "Inequality path constraints in optimal control: a finite iteration ε-convergent scheme based on pointwise discretization," *Journal of Process Control*, vol. 15, no. 3, pp. 353–362, 2005.

[16] B. Choi, W. Lim, and S. Choi, "Control design and closed-loop analysis of a switched-capacitor DC-to-DC converter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 3, pp. 1099–1107, 2001.

[17] H. Chung and Y. K. Mok, "Development of a switched-capacitor DC/DC boost converter with continuous input current waveform," *IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications*, vol. 46, no. 6, pp. 756–759, 1999.

[18] H. S. H. Chung, W. C. Chow, S. Y. R. Hui, and S. T. S. Lee, "Development of a switched-capacitor DC-DC converter with bidirectional power flow," *IEEE Transactions on Circuits and Systems—I: Fundamental Theory and Applications*, vol. 47, no. 9, pp. 1383–1389, 2000.

[19] V. Costanza, "Finding initial costates in finite-horizon nonlinear-quadratic optimal control problems," *Optimal Control Applications and Methods*, vol. 29, no. 3, pp. 225–242, 2008.

[20] V. Costanza and P. S. Rivadeneira, "Finite-horizon dynamic optimization of nonlinear systems in real-time," *Automatica*, vol. 44, no. 9, pp. 2427–2434, 2008.

[21] J. Y. Dieulot and J. P. Richard, "Tracking control of a nonlinear system with input-dependent delay," in *Proceedings of the 40th IEEE Conference on Decision and Control*, December 2001, pp. 4027–4031.

[22] B. Farhadinia, K. L. Teo, and R. C. Loxton, "A computational method for a class of non-standard time optimal control problems involving multiple time horizons," *Mathematical and Computer Modelling*, vol. 49, no. 7-8, pp. 1682–1691, 2009.

[23] A. V. Fiacco and G. P. McCormick, *Nonlinear programming: Sequential unconstrained minimization techniques*. Philadelphia: Society for Industrial and Applied Mathematics, 1990.

[24] T. Furukawa, "Time-optimal cooperative control of multiple robot vehicles," in *Proceedings of the 2003 IEEE International Conference on Robotics and Automation*, September 2003, pp. 944–950.

[25] C. Gao, E. Feng, Z. Wang, and Z. Xiu, "Nonlinear dynamical systems of biodissimilation of glycerol to 1,3-Propanediol and their optimal controls," *Journal of Industrial and Management Optimization*, vol. 1, no. 3, pp. 377–388, 2005.

[26] C. Gao, K. Li, E. Feng, and Z. Xiu, "Nonlinear impulsive system of fed-batch culture in fermentative production and its properties," *Chaos, Solitons and Fractals*, vol. 28, no. 1, pp. 271–277, 2006.

[27] F. Garofalo, P. Marino, S. Scala, and F. Vasca, "Control of DC-DC converters with linear optimal feedback and nonlinear feedforward," *IEEE Transactions on Power Electronics*, vol. 9, no. 6, pp. 607–615, 1994.

[28] W. L. Garrard and J. M. Jordan, "Design of nonlinear automatic control systems," *Automatica*, vol. 13, no. 5, pp. 497–505, 1977.

[29] D. J. Gates and M. Westcott, "Solar cars and variational problems equivalent to shortest paths," *SIAM Journal on Control and Optimization*, vol. 34, no. 2, pp. 428–436, 1996.

[30] M. Gerdts, "A nonsmooth Newton's method for control-state constrained optimal control problems," *Mathematics and Computers in Simulation*, vol. 79, no. 4, pp. 925–936, 2008.

[31] ——, "Global convergence of a nonsmooth Newton method for control-state constrained optimal control problems," *SIAM Journal on Optimization*, vol. 19, no. 1, pp. 326–350, 2008.

[32] M. Gerdts and M. Kunkel, "A nonsmooth Newton's method for discretized optimal control problems with state and control constraints," *Journal of Industrial and Management Optimization*, vol. 4, no. 2, pp. 247–270, 2008.

[33] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright, *User's guide for NPSOL 5.0: A Fortran package for nonlinear programming*, Stanford University, Stanford, July 1998.

[34] C. J. Goh and K. L. Teo, "Control parametrization: a unified approach to optimal control problems with general constraints," *Automatica*, vol. 24, no. 1, pp. 3–18, 1988.

[35] B. Z. Guo and T. T. Wu, "Approximation of optimal feedback control: a dynamic programming approach," *Journal of Global Optimization*, to appear.

[36] W. W. Hager, "Runge-Kutta methods in optimal control and the transformed adjoint system," *Numerische Mathematik*, vol. 87, no. 2, pp. 247–282, 2000.

[37] P. R. Halmos, *Measure theory*. New York: Springer, 1974.

[38] R. F. Hartl, S. P. Sethi, and R. G. Vickson, "A survey of the maximum principles for optimal control problems with state constraints," *SIAM Review*, vol. 37, no. 2, pp. 181–218, 1995.

[39] L. Hasdorff, *Gradient optimization and nonlinear control*. New York: John Wiley, 1976.

[40] J. P. Hespanha and S. Morse, "Switching between stabilizing controllers," *Automatica*, vol. 38, no. 11, pp. 1905–1917, 2002.

[41] A. Hindmarsh, "Large ordinary differential equation systems and software," *IEEE Control Systems Magazine*, vol. 2, no. 4, pp. 24–30, 1982.

[42] C. Y. F. Ho, B. W. K. Ling, Y. Q. Liu, P. K. S. Tam, and K. L. Teo, "Optimal PWM control of switched-capacitor DC-DC power converters via model transformation and enhancing control techniques," *IEEE Transactions on Circuits and Systems—I: Regular papers*, vol. 55, no. 5, pp. 1382–1391, 2008.

[43] M. J. Hounslow, R. L. Ryall, and V. R. Marshall, "A discretized population balance for nucleation, growth, and aggregation," *AIChE Journal*, vol. 34, no. 11, pp. 1821–1832, 1988.

[44] P. Howlett, "Optimal strategies for the control of a train," *Automatica*, vol. 32, no. 4, pp. 519–532, 1996.

[45] ——, "The optimal control of a train," *Annals of Operations Research*, vol. 98, no. 1-4, pp. 65–87, 2000.

[46] A. Ioinovici, "Switched-capacitor power electronics circuits," *IEEE Circuits and Systems Magazine*, vol. 1, no. 3, pp. 37–42, 2001.

[47] L. S. Jennings, M. E. Fisher, K. L. Teo, and C. J. Goh, *MISER3 Optimal control software: Theory and user manual*, The University of Western Australia, Perth, July 2004.

[48] L. S. Jennings and K. L. Teo, "A computational algorithm for functional inequality constrained optimization problems," *Automatica*, vol. 26, no. 2, pp. 371–375, 1990.

[49] K. Kaji and K. H. Wong, "Nonlinearly constrained time-delayed optimal control problems," *Journal of Optimization Theory and Applications*, vol. 82, no. 2, pp. 295–313, 1994.

[50] C. Y. Kaya and J. M. Martínez, "Euler discretization and inexact restoration for optimal control," *Journal of Optimization Theory and Applications*, vol. 134, no. 2, pp. 191–206, 2007.

[51] C. Y. Kaya and J. L. Noakes, "Computational method for time-optimal switching control," *Journal of Optimization Theory and Applications*, vol. 117, no. 1, pp. 69–92, 2003.

[52] A. Khayatian and D. G. Taylor, "Multirate modeling and control design for switched-mode power converters," *IEEE Transactions on Automatic Control*, vol. 39, no. 9, pp. 1848–1852, 1994.

[53] D. E. Kirk, *Optimal control theory: An introduction*. New York: Dover, 2004.

[54] A. Y. Lee, "Hereditary optimal control problems: Numerical method based upon a Padé Approximation," *Journal of Optimization Theory and Applications*, vol. 56, no. 1, pp. 157–166, 1988.

[55] ——, "Numerical solution of time-delayed optimal control problems with terminal inequality constraints," *Optimal Control Applications and Methods*, vol. 14, no. 3, pp. 203–210, 1993.

[56] H. W. J. Lee and K. L. Teo, "Control parametrization enhancing technique for solving a special ODE class with state dependent switch," *Journal of Optimization Theory and Applications*, vol. 118, no. 1, pp. 55–66, 2003.

[57] H. W. J. Lee, K. L. Teo, and X. Q. Cai, "An optimal control approach to nonlinear mixed integer programming problems," *Computers and Mathematics with Applications*, vol. 36, no. 3, pp. 87–105, 1998.

[58] H. W. J. Lee, K. L. Teo, and L. S. Jennings, "On optimal control of multi-link vertical planar robot arms systems moving under the effect of gravity," *Journal of the Australian Mathematical Society—Series B*, vol. 39, no. 2, pp. 195–213, 1997.

[59] H. W. J. Lee, K. L. Teo, and A. E. B. Lim, "Sensor scheduling in continuous time," *Automatica*, vol. 37, no. 12, pp. 2017–2023, 2001.

[60] H. W. J. Lee, K. L. Teo, V. Rehbock, and L. S. Jennings, "Control parametrization enhancing technique for time optimal control problems," *Dynamic Systems and Applications*, vol. 6, pp. 243–262, 1997.

[61] ——, "Control parametrization enhancing technique for optimal discrete-valued control problems," *Automatica*, vol. 35, no. 8, pp. 1401–1407, 1999.

[62] H. W. J. Lee and K. H. Wong, "Semi-infinite programming approach to nonlinear time-delayed optimal control problems with linear continuous constraints," *Optimization Methods and Software*, vol. 21, no. 5, pp. 679–691, 2006.

[63] F. H. F. Leung, P. K. S. Tam, and C. K. Li, "An improved LQR-based controller for switching Dc-dc converters," *IEEE Transactions on Industrial Electronics*, vol. 40, no. 5, pp. 521–528, 1993.

[64] R. Li, K. L. Teo, K. H. Wong, and G. R. Duan, "Control parameterization enhancing transform for optimal control of switched systems," *Mathematical and Computer Modelling*, vol. 43, no. 11-12, pp. 1393–1403, 2006.

[65] Q. Lin, R. C. Loxton, K. L. Teo, and Y. H. Wu, "A new computational method for a class of free terminal time optimal control problems," *Pacific Journal of Optimization*, to appear.

[66] C. Liu, Z. Gong, E. Feng, and H. Yin, "Modelling and optimal control for nonlinear multistage dynamical system of microbial fed-batch culture," *Journal of Industrial and Management Optimization*, vol. 5, no. 4, pp. 835–850, 2009.

[67] X. Liu, "A class of continuously differentiable filled functions for global optimization," *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, vol. 38, no. 1, pp. 38–47, 2008.

[68] Y. Liu, A. Eberhard, and K. L. Teo, "A numerical method for a class of mixed switching and impulsive optimal control problems," *Computers and Mathematics with Applications*, vol. 52, no. 5, pp. 625–636, 2006.

[69] Y. Liu, K. L. Teo, L. S. Jennings, and S. Wang, "On a class of optimal control problems with state jumps," *Journal of Optimization Theory and Applications*, vol. 98, no. 1, pp. 65–82, 1998.

[70] R. C. Loxton, K. L. Teo, and V. Rehbock, "An optimization approach to state-delay identification," *IEEE Transactions on Automatic Control*, conditionally accepted.

[71] ——, "On a class of optimal control problems with variable time points in the objective and constraint functionals," in *Proceedings of The 7th International Conference on Optimization Techniques and Applications*, December 2007.

[72] ——, "Optimal control problems with multiple characteristic time points in the objective and constraints," *Automatica*, vol. 44, no. 11, pp. 2923–2929, 2008.

[73] ——, "Computational method for a class of switched system optimal control problems," *IEEE Transactions on Automatic Control*, vol. 54, no. 10, pp. 2455–2460, 2009.

[74] R. C. Loxton, K. L. Teo, V. Rehbock, and W. K. Ling, "Optimal switching instants for a switched-capacitor DC/DC power converter," *Automatica*, vol. 45, no. 4, pp. 973–980, 2009.

[75] R. C. 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*, vol. 45, no. 10, pp. 2250–2257, 2009.

[76] D. G. Luenberger and Y. Ye, *Linear and nonlinear programming*, 3rd ed. New York: Springer, 2008.

[77] R. Martin and K. L. Teo, *Optimal control of drug administration in cancer chemotherapy*. Singapore: World Scientific, 1994.

[78] R. B. Martin, "Optimal control drug scheduling of cancer chemotherapy," *Automatica*, vol. 28, no. 6, pp. 1113–1123, 1992.

[79] J. Nocedal and S. J. Wright, *Numerical optimization*, 2nd ed. New York: Springer, 2006.

[80] H. J. Oberle and B. Sothmann, "Numerical computation of optimal feed rates for a fed-batch fermentation model," *Journal of Optimization Theory and Applications*, vol. 100, no. 1, pp. 1–13, 1999.

[81] Y. Orlov, L. Belkoura, J. P. Richard, and M. Dambrine, "On identifiability of linear time-delay systems," *IEEE Transactions on Automatic Control*, vol. 47, no. 8, pp. 1319–1324, 2002.

[82] ——, "Adaptive identification of linear time-delay systems," *International Journal of Robust and Nonlinear Control*, vol. 13, no. 9, pp. 857–872, 2003.

[83] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, and E. F. Mishchenko, *The mathematical theory of optimal processes*, ser. L. S. Pontryagin selected works. Montreux: Gordon and Breach, 1986, vol. 4.

[84] R. Pytlak and R. B. Vinter, "A feasible directions algorithm for optimal control problems with state and control constraints: Convergence analysis," *SIAM Journal on Control and Optimization*, vol. 36, no. 6, pp. 1999–2019, 1998.

[85] ——, "Feasible direction algorithm for optimal control problems with state and control constraints: Implementation," *Journal of Optimization Theory and Applications*, vol. 101, no. 3, pp. 623–649, 1999.

[86] V. Rehbock and L. Caccetta, "Two defence applications involving discrete valued optimal control," *ANZIAM Journal*, vol. 44, pp. 33–54, 2002.

[87] V. Rehbock and I. Livk, "Optimal control of a batch crystallization process," *Journal of Industrial and Management Optimization*, vol. 3, no. 3, pp. 585–596, 2007.

[88] J. P. Richard, "Time-delay systems: an overview of some recent advances and open problems," *Automatica*, vol. 39, no. 10, pp. 1667–1694, 2003.

[89] 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. New York: Springer, 2005, pp. 447–459.

[90] T. Ruby, V. Rehbock, and W. B. Lawrance, "Optimal control of hybrid power systems," *Dynamics of Continuous, Discrete and Impulsive systems*, vol. 10, pp. 429–439, 2003.

[91] Y. Sakawa and Y. Shindo, "Optimal control of container cranes," *Automatica*, vol. 18, no. 3, pp. 257–266, 1982.

[92] K. Schittkowski, "NLPQL: A Fortran subroutine solving constrained nonlinear programming problems," *Annals of Operations Research*, vol. 5, no. 1-4, pp. 485–500, 1986.

[93] ——, *NLPQLP: A Fortran implementation of a sequential quadratic programming algorithm with distributed and non-monotone line search - User's guide, version 2.24*, University of Bayreuth, Bayreuth, June 2007.

[94] C. Seatzu, D. Corona, A. Giua, and A. Bemporad, "Optimal control of continuous-time switched affine systems," *IEEE Transactions on Automatic Control*, vol. 51, no. 5, pp. 726–741, 2006.

[95] A. Siburian and V. Rehbock, "Numerical procedure for solving a class of singular optimal control problems," *Optimization Methods and Software*, vol. 19, no. 3-4, pp. 413–426, 2004.

[96] E. D. Sontag, *Mathematical control theory: Deterministic finite dimensional systems*, 2nd ed. New York: Springer, 1998.

[97] K. L. Teo, "Control parametrization enhancing transform to optimal control problems," *Nonlinear Analysis*, vol. 63, no. 5-7, pp. 2223–2236, 2005.

[98] K. L. Teo and C. J. Goh, "A simple computational procedure for optimization problems with functional inequality constraints," *IEEE Transactions on Automatic Control*, vol. 32, no. 10, pp. 940–941, 1987.

[99] K. L. Teo, C. J. Goh, and C. C. Lim, "A computational method for a class of dynamical optimization problems in which the terminal time is conditionally free," *IMA Journal of Mathematical Control and Information*, vol. 6, no. 1, pp. 81–95, 1989.

[100] K. L. Teo, C. J. Goh, and K. H. Wong, *A unified computational approach to optimal control problems*. Essex: Longman Scientific and Technical, 1991.

[101] K. L. Teo and L. S. Jennings, "Nonlinear optimal control problems with continuous state inequality constraints," *Journal of Optimization Theory and Applications*, vol. 63, no. 1, pp. 1–22, 1989.

[102] ——, "Optimal control with a cost on changing control," *Journal of Optimization Theory and Applications*, vol. 68, no. 2, pp. 335–357, 1991.

[103] K. L. Teo, L. S. Jennings, H. W. J. Lee, and V. Rehbock, "The control parameterization enhancing transform for constrained optimal control problems," *Journal of the Australian Mathematical Society—Series B*, vol. 40, no. 3, pp. 314–335, 1999.

[104] K. L. Teo, G. Jepps, E. J. Moore, and S. Hayes, "A computational method for free time optimal control problems, with application to maximizing the range of an aircraft-like projectile," *Journal of the Australian Mathematical Society—Series B*, vol. 28, no. 3, pp. 393–413, 1987.

[105] K. L. Teo, W. R. Lee, L. S. Jennings, S. Wang, and Y. Liu, "Numerical solution of an optimal control problem with variable time points in the objective function," *ANZIAM Journal*, vol. 43, pp. 463–478, 2002.

[106] K. L. Teo, V. Rehbock, and L. S. Jennings, "A new computational algorithm for functional inequality constrained optimization problems," *Automatica*, vol. 29, no. 3, pp. 789–792, 1993.

[107] K. L. Teo and K. H. Wong, "A computational method for time-lag control problems with control and terminal inequality constraints," *Optimal Control Applications and Methods*, vol. 8, no. 4, pp. 377–395, 1987.

[108] K. L. Teo, K. H. Wong, and D. J. Clements, "Optimal control computation for linear time-lag systems with linear terminal constraints," *Journal of Optimization Theory and Applications*, vol. 44, no. 3, pp. 509–526, 1984.

[109] J. Tuch, A. Feuer, and Z. J. Palmor, "Time delay estimation in continuous linear time-invariant systems," *IEEE Transactions on Automatic Control*, vol. 39, no. 4, pp. 823–827, 1994.

[110] T. Umeno, K. Takahashi, I. Oota, F. Ueno, and T. Inoue, "New switched-capacitor DC-DC converter with low input current ripple and its hybridization," in *Proceedings of the 33rd Midwest Symposium on Circuits and Systems*, August 1990, pp. 1091–1094.

[111] E. I. Verriest, "Stability of systems with state-dependent and random delays," *IMA Journal of Mathematical Control and Information*, vol. 19, no. 1-2, pp. 103–114, 2002.

[112] L. D. Vidal, C. Jauberthie, and G. J. Blanchard, "Identifiability of a nonlinear delayed-differential aerospace model," *IEEE Transactions on Automatic Control*, vol. 51, no. 1, pp. 154–158, 2006.

[113] T. L. Vincent and W. J. Grantham, *Optimality in parametric systems.* New York: John Wiley, 1981.

[114] O. von Stryk, "Numerical solution of optimal control problems by direct collocation," in *Optimal Control—Calculus of Variations, Optimal Control Theory and Numerical Methods*, R. Bulirsch, A. Miele, J. Stoer, and K. H. Well, Eds. Birkhäuser, 1993, vol. 111, pp. 129–143.

[115] L. Vu and D. Liberzon, "Common Lyapunov functions for families of commuting nonlinear systems," *Systems and Control Letters*, vol. 54, no. 5, pp. 405–416, 2005.

[116] L. Y. Wang, W. H. Gui, K. L. Teo, and R. C. Loxton, "Optimal control problems arising in the zinc sulphate electrolyte purification process," *Journal of Global Optimization*, to appear.

[117] L. Y. Wang, W. H. Gui, K. L. Teo, R. C. Loxton, and C. H. Yang, "Time delayed optimal control problems with multiple characteristic time points: Computation and industrial applications," *Journal of Industrial and Management Optimization*, vol. 5, no. 4, pp. 705–718, 2009.

[118] S. Wang, F. Gao, and K. L. Teo, "An upwind finite-difference method for the approximation of viscosity solutions to Hamilton-Jacobi-Bellman equations," *IMA Journal of Mathematical Control and Information*, vol. 17, no. 2, pp. 167–178, 2000.

[119] S. Wang, L. S. Jennings, and K. L. Teo, "Numerical solution of Hamilton-Jacobi-Bellman equations by an upwind finite volume method," *Journal of Global Optimization*, vol. 27, no. 2-3, pp. 177–192, 2003.

[120] K. H. Wong, D. J. Clements, and K. L. Teo, "Optimal control computation for nonlinear time-lag systems," *Journal of Optimization Theory and Applications*, vol. 47, no. 1, pp. 91–107, 1985.

[121] K. H. Wong, L. S. Jennings, and F. Benyah, "The control parametrization enhancing transform for constrained time-delayed optimal control problems," *ANZIAM Journal*, vol. 43, pp. 154–185, 2002.

[122] S. F. Woon, "Global algorithms for nonlinear discrete optimization and discrete-valued optimal control problems," Ph.D. dissertation, Curtin University of Technology, Perth, December 2009.

[123] S. F. Woon, V. Rehbock, and R. C. Loxton, "Global optimization method for continuous time sensor scheduling," *Nonlinear Dynamics and Systems Theory*, to appear.

[124] ——, "Optimal operation for a hybrid power system," submitted.

[125] C. Z. Wu and K. L. Teo, "Global impulsive optimal control computation," *Journal of Industrial and Management Optimization*, vol. 2, no. 4, pp. 435–450, 2006.

[126] C. Z. Wu, K. L. Teo, R. Li, and Y. Zhao, "Optimal control of switched systems with time delay," *Applied Mathematics Letters*, vol. 19, no. 10, pp. 1062–1067, 2006.

[127] C. Z. Wu, K. L. Teo, Y. Zhao, and W. Y. Yan, "Solving an identification problem as an impulsive optimal parameter selection problem," *Computers and Mathematics with Applications*, vol. 50, no. 1-2, pp. 217–229, 2005.

[128] Z. Y. Wu, F. S. Bai, H. W. J. Lee, and Y. J. Yang, "A filled function method for constrained global optimization," *Journal of Global Optimization*, vol. 39, no. 4, pp. 495–507, 2007.

[129] Z. Y. Wu, H. W. J. Lee, L. S. Zhang, and X. M. Yang, "A novel filled function method and quasi-filled function method for global optimization," *Computational Optimization and Applications*, vol. 34, no. 2, pp. 249–272, 2005.

[130] Z. Y. Wu, L. S. Zhang, K. L. Teo, and F. S. Bai, "New modified function method for global optimization," *Journal of Optimization Theory and Applications*, vol. 125, no. 1, pp. 181–203, 2005.

[131] R. Xu and L. Chen, "Persistence and stability for a two-species ratio-dependent predator-prey system with time delay in a two-patch environment," *Computers and Mathematics with Applications*, vol. 40, no. 4-5, pp. 577–588, 2000.

[132] X. Xu and P. J. Antsaklis, "Optimal control of switched systems based on parameterization of the switching instants," *IEEE Transactions on Automatic Control*, vol. 49, no. 1, pp. 2–16, 2004.

[133] J. Zabczyk, *Mathematical control theory: An introduction.* Boston: Birkhäuser, 2008.

[134] L. S. Zhang, C. K. Ng, D. Li, and W. W. Tian, "A new filled function method for global optimization," *Journal of Global Optimization*, vol. 28, no. 1, pp. 17–43, 2004.

[135] J. L. Zhou, A. L. Tits, and C. T. Lawrence, *User's guide for FFSQP version 3.7: A Fortran code for solving constrained nonlinear (minimax) optimization problems, generating iterates satisfying all inequality and linear constraints*, University of Maryland, Maryland.