2018-10-24

Perspectives in control of conditionally controllable problems

Siamak Ghorbani Faal

Perspectives in control of conditionally controllable problems

by

Siamak Ghorbani Faal

A Thesis

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

In partial fulfillment of the requirements for the

Degree of Doctor of Philosophy

in

Robotics Engineering

October 2018

APPROVED:

Professor Cagdas D. Onal, Advisor

Professor Gregory S. Fischer, Committee member

Professor Umberto Mosco, Committee member

Professor Marcus Sarkis, Committee member

Abstract

Limitations imposed on control functions can significantly affect the performance of a linear controller. When applied to the real physical system, such limitations convert a linear function to a nonlinear input signal that alters the convergence or stability of the solution. The main focus of this study is to identify, classify and propose appropriate techniques to overcome such problems. In this regard, we propose an exact definition for a conditionally controllable problem and investigate control function formulations for such problems under the lenses of planning-based and optimization-based methods.

Acknowledgments

Firstly, I would like to thank all my mentors throughout my academic life, In particular, my PhD advisor, Professor Cagdas D. Onal for his patience, motivation and support of my Ph.D study. In addition, I would like to express my sincere gratitude to my thesis committee, Professors Gregory S. Fischer, Umberto Mosco and Marcus Sarkis, for their insightful comments and encouragement.

I would like to take this chance to express my appreciation for the guidance and help provided by Professors Umberto Mosco and Marcus Sarkis at the Department of Mathematical Sciences in WPI. Owing to their invaluable counsel, I could gain the required mathematical background that allowed me to conduct this research. They have a significant role in my academic path and I will forever be grateful for their generous help, patience and consideration.

It is also well worth mentioning that the initial sparks of this project were formed through exciting conversations that I had with Dr. Stephen Nestinger. He helped me to develop my academic interest and set the foundations of my research career through long discussions that would take until hours after midnight. I specially would like to thank him for his patience, understanding, critics, guidance and most of all, his invaluable friendship.

My sincere thanks also goes to Mr. Siamak Najafi, Dr. Adriana Hera and Dr. Raffaele Potami for their continuous emotional, intellectual and financial support through my PhD. They have been my family away from home and without their support, I could not pursue my studies.

I would like to thank my friends for accepting nothing less than excellence from me. For their unlimited support and understanding. Among all, I would like to specially express my gratitude to Shadi Tasdighi Kalat, Ross Desmond, Paul Heslinga, Sabah Razavi and Pooya Yousefi for their emotional and intellectual support, help

and guidance.

Last but not the least, I would like to thank my beloved family, my parents Maghsoud and Nasrin , my brother Babak and my sister Parisa for supporting me in every possible way throughout my studies and my life in general.

Contents

List of Figures

Chapter 1

Introduction

A sparrow perching on a narrow vibrating twig, a squirrel jumping over branches of an old oak tree in pursuit of a tasty acorn, or a seagull diving into the water for an afternoon meal are few examples of complex maneuvers performed in nature that exploit nonlinear dynamics. Such nonlinearities, that are nightmares for control engineers, seem to be the key to develop elegant solutions for control problems. Unfortunately, due to unruly behavior of nonlinear systems, there are only a few systematic approaches available that help with synthesizing nonlinear control functions. As a result, numerous control problems are solved by linearizing the associated dynamics about an operating point and using the vastly developed techniques for linear systems to formulate a linear controller. Bearing in mind that the real system is still nonlinear, the performance of the developed linear controller may drastically vary from the original criteria. Owing to their simplicity, such linearized models and controllers are desirable, and to some extent effective, for industrial systems. However, they are not advisable for control of modern robotic systems with convoluted dynamics, limited actuations, and unavoidable interactions with the environment.

1.0.1 Outline of contributions

Since one can easily get lost in the labyrinth of nonlinear systems without an Ariadne's thread, we will tie our rope around a subset of control problems that provide a suitable framework for a systematic cascading of linear and nonlinear control approaches. Given this set of problems, namely conditionally controllable, we then investigate appropriate techniques to synthesize control functions. The definition provided for conditionally controllability is rooted in the concept of regions of attraction of dynamic systems [1, 2] and could be assumed as an extension of small time stability in nonlinear dynamics [3].

In the second chapter of this manuscript, we give an exact definition for conditional controllability and show how a problem can morph to being conditionally controllable as we change the limitations imposed on the control function. In particular, we investigate swing-up control of a simple point-mass pendulum and the effect of input saturations on the performance of the control function. Then, in what follows, we investigate possible techniques to formulate a control function for conditionally controllable problems by means of planning-based and optimization theories. In Chapter 3 we propose a method to construct controller for conditionally controllable problems by cascading projected linear and piecewise constant functions. The associated parameters of the piecewise functions are obtained by solving a planning problem in the state space of the system via Ariadne's clew framework. To construct details of this approach, we present a formal setting for exploring trees as a building block for planning based methods in normed vector spaces. We further extend the setting to include time and solving planning problems on spatio-temporal grids. The obtained tools are then used to construct a method for synthesizing control functions to solve regularization problems and the proposed method is used to synthesize control functions for three case studies of: point-mass pendulum, double-pendulum

and cart-pole. Chapter 4 is dedicated to some preliminaries of optimal control theory and presentation of a method to synthesize control functions for conditionally controllable problems by solving an optimization problem. In particular, the control function is constructed through composition of a piecewise constant function with a linear controller. The coefficients of the piecewise function are determined by solving an optimization problem for which the cost is defined as the norm of state vector at a given finite time. If the constructed signal for specific discretization in time does not satisfies a given convergence criteria, the process will start again with a finer time intervals for the piecewise terms. The dissertation is concluded with the materials presented in Chapter 5 at which we also discuss the application of energy-based control in solving conditionally controllable problems.

Chapter 2

Conditional controllability

The current chapter is dedicated to the definition of a conditional controllable problem and in depth study of an example that highlight the motivations of this research. The presented discussions will provide the necessary foundation for the methodologies discussed in the proceeding chapters. In what follows, we present the preliminaries required to give a formal definition for conditional controllability. In order to capture the essence of the proposed definition, we proceed with a case study of a simple point-mass pendulum and investigate the effect of control input saturations on the performance of a controller that is designed based on the linearized system model.

2.1 Formal setting

Throughout this text, $\mathbb{N}$ and $\mathbb{R}$ represent the sets of natural and real numbers, respectively. The set of positive real numbers is denoted with $\mathbb{R}^+ := \{x \in \mathbb{R} : x > 0\}$. All vectors, matrices and vector-valued functions are denoted with boldface letters or symbols. The space of n-dimensional vectors $\mathbf{x} = [x_1, \cdots, x_n]$ where $x_i \in \mathbb{R}$ for every $i \in \{1, 2, \cdots, n\}$ is denoted by $\mathbb{R}^n$. The inner product of two vectors $\mathbf{x}$

and $\mathbf{y}$ in $\mathbb{R}^n$ is defined as

$$\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + \cdots + x_n y_n, \tag{2.1}$$

and the norm of a vector $\mathbf{x} \in \mathbb{R}^n$ is $\|\mathbf{x}\| = \langle \mathbf{x}, \mathbf{x} \rangle^{1/2}$. Let $\mathbf{f} : X \to Y$ to be a function, and $W \subset Y$, then $\mathbf{f} \subset W$ is an equivalent statement to $\mathbf{f}(\mathbf{x}) \in W$, $\forall \mathbf{x} \in X$. Moreover, to simplify notations, we will use $\mathbf{f}$ to indicate the function itself and $\mathbf{f}(\mathbf{x})$ to indicate a point in the image of $\mathbf{f}$ for an arbitrary $\mathbf{x} \in X$. Let A to be a set, the characteristic function for A is denoted with χ_A such that $\chi_A(\mathbf{x}) = 1$ if $\mathbf{x} \in A$ and $= 0$ if $\mathbf{x} \notin A$. A property is said to hold *almost everywhere* (a.e.) if the measure of set of points where it fails to hold is zero.

Despite the fact that analogous definitions for control systems based on concepts of differential geometry and vector fields on manifolds exists (such as the definition presented in [4]), here we will work with a statement of control system that is in line with the definitions presented in [5] as follows.

Definition 2.1.1 (Control system)**.** An ordinary differential equation of the form

$$\dot{\mathbf{y}}(t) = \mathbf{f}(t, \mathbf{y}(t), \mathbf{u}(t)), \tag{2.2}$$

with $\mathbf{f} : \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ is a *control system* if the initial value problem

$$\begin{cases} \dot{\mathbf{y}}(t) = \mathbf{f}(t, \mathbf{y}(t), \mathbf{u}(t)), \\ \mathbf{y}(t_0) = \boldsymbol{\eta}_0, \end{cases} \tag{2.3}$$

has a unique solution in the class of absolutely continuous functions[1] for any initial state $\boldsymbol{\eta}_0 \in \mathbb{R}^n$ and any arbitrarily assigned *control function* $\mathbf{u} : \mathbb{R} \to \mathbb{R}^m$. We will

[1] A real-valued function f on a compact interval $I = [a, b]$ is *absolutely continuous* if f has a

denote a control system with function $\mathbf{f}$. Moreover, let $\mathbf{y}$ be the solution of (2.3) for specific initial condition $\mathbf{y}(t_0) = \boldsymbol{\eta}_0 \in \mathbb{R}^n$ and control $\mathbf{u}$, then $\mathbf{y}$ is the *trajectory* of the control system[2] associated to $\mathbf{u}$ and is denoted by $\mathbf{y}_\mathbf{u}$.

Definition 2.1.2 (Equilibrium point and equilibrium input)**.** Point $(\boldsymbol{\eta}_e,\ \mathbf{w}_e) \in \mathbb{R}^n \times \mathbb{R}^m$ is an equilibrium point of a control system, if $\mathbf{f}(t, \boldsymbol{\eta}_e,\ \mathbf{w}_e) = \mathbf{0}$ for every $t \geq t_0$. Accordingly, $\boldsymbol{\eta}_e$ is the *equilibrium state* and $\mathbf{w}_e$ is the *equilibrium input* associated with $\boldsymbol{\eta}_e$.

From a practical point of view, the differential equation (2.2) represents the time evolution of a real physical system. Accordingly, a trajectory of the system may require to satisfy constraints arising from system's physical characteristics. Such constrains could be categorized into three groups: (i) achievable state vectors, (ii) attainable range of $\mathbf{f}(t, \mathbf{y}(t), \mathbf{u}(t))$, and (iii) feasible control functions applicable to the system. Considering such constraints, we can define control and regularization problems as:

Definition 2.1.3 (Control problem)**.** Given a control system $\mathbf{f}$, a time interval $T = [t_0, t_f] \subset (\mathbb{R} \cup \infty)$ and a set-valued function $t \mapsto Y_d(t) \subset \mathbb{R}^n$ to denote the set of desired state vectors in time. Let $t \mapsto Y(t)$ be a set-valued function representing set of feasible state vectors in time such that for $t \in T$, $Y_d(t) \subset Y(t)$ and $Y(t)$ is a connected subset of $\mathbb{R}^n$ with usual topology[3]. Moreover, let $t \mapsto \widehat{Y}(t) \subset \mathbb{R}^n$ indicate the feasible range of $\mathbf{f}$ in time and U to be the set of feasible control functions. A

Lebesgue integrable derivative f' and

$$f(b) = f(a) + \int_a^b f'(t)\, dt.$$

[2]The terms "system" and "control system" are used interchangeably throughout the text.

[3]Clearly, $\mathbb{R}^n$ with the usual (open ball) topology $\mathscr{T} := \{\|\mathbf{x} - \mathbf{x}_0\| < \epsilon : \forall \mathbf{x}_0 \in \mathbb{R}^n, \forall \epsilon \in \mathbb{R}^+\}$ is a topological space. A connected subset of a topological space is a set that cannot be partitioned into two nonempty subsets that are open in the relative induced topology.

control problem, denoted by $(\mathbf{f}, U, Y, Y_d, \widehat{Y}, T)$, is to find a set of control functions[4] $W \subset U$ such that for any $\boldsymbol{\eta} \in Y(t_0)$, exists $\mathbf{u}_{\boldsymbol{\eta}} \in W$ satisfying

$$\begin{aligned}
&\mathbf{y}_{\mathbf{u}_{\boldsymbol{\eta}}}(t) := \boldsymbol{\eta} + \int_{t_0}^{t} \mathbf{f}(\tau, \mathbf{y}(\tau), \mathbf{u}_{\boldsymbol{\eta}}(\tau)) d\tau \ \in Y(t), \ \forall t \in T, \\
&\mathbf{f}(t, \mathbf{y}_{\mathbf{u}_{\boldsymbol{\eta}}}(t), \mathbf{u}_{\boldsymbol{\eta}}(t)) \in \widehat{Y}(t), \ \forall t \in T, \text{ and} \\
&\mathbf{y}_{\mathbf{u}_{\boldsymbol{\eta}}}(t_f) \in Y_d(t_f).
\end{aligned}$$

Definition 2.1.4 (Regularization problem)**.** Given a control system $\mathbf{f}$, an initial time $t_0 \in \mathbb{R}$; Let $t \mapsto Y(t)$ be a set-valued function representing set of feasible state vectors in time such that for $t \in T$, $Y(t)$ is a connected subset of $\mathbb{R}^n$ with usual topology. Moreover, let $t \mapsto \widehat{Y}(t) \subset \mathbb{R}^n$ indicate the feasible range of $\mathbf{f}$ in time and U to be the set of feasible control functions. Given a desired trajectory $\mathbf{y}_d : \mathbb{R} \to \mathbb{R}^n$, $\mathbf{y}_d \in C^1([t_0, \infty])$, such that $\mathbf{y}_d(t) \in Y(t)$ and $\dot{\mathbf{y}}_d(t) \in \widehat{Y}(t)$ for $t \geq t_0$, a *regularization problem*, denoted by $(\mathbf{f}, U, Y, \mathbf{y}_d, \widehat{Y}, t_0)$, is to find a set of control functions $W \subset U$ such that for any $\boldsymbol{\eta} \in Y(t_0)$, there exists $\mathbf{u}_{\boldsymbol{\eta}} \in W$ satisfying

$$\begin{aligned}
&\mathbf{y}_{\mathbf{u}_{\boldsymbol{\eta}}}(t) := \boldsymbol{\eta} + \int_{t_0}^{t} \mathbf{f}(\tau, \mathbf{y}(\tau), \mathbf{u}_{\boldsymbol{\eta}}(\tau)) d\tau \ \in Y(t), \ \forall t \geq t_0, \\
&\mathbf{f}(t, \mathbf{y}_{\mathbf{u}_{\boldsymbol{\eta}}}(t), \mathbf{u}_{\boldsymbol{\eta}}(t)) \in \widehat{Y}(t), \ \forall t \geq t_0, \text{ and} \\
&\lim_{t \to \infty} \|\mathbf{y}_d(t) - \mathbf{y}_{\mathbf{u}_{\boldsymbol{\eta}}}(t)\| = 0.
\end{aligned}$$

We can explore the connection between control and regularization problems through the following lemmas.

Lemma 2.1.1. *If W is a solution to regularization problem $R = (\mathbf{f}, U, Y, \mathbf{y}_d, \widehat{Y}, t_0)$,*

[4] Alternatively, specific $\boldsymbol{\eta}_0 \in Y$ could be assigned to be the only initial state of the problem; that is, the control problem is to find a function $\mathbf{u} \in U$ such that $\mathbf{y}_u(t_0) = \boldsymbol{\eta}_0$ and $\mathbf{y}_u(t_f) \in Y_d$. Such a definition may be interpreted as a trajectory planning problem. Here, to avoid such specification, we define the control problem of finding $\mathbf{u}_{\boldsymbol{\eta}}$ for every $\boldsymbol{\eta} \in Y$.

then there exists a control problem $C = (\mathbf{f}, U, Y, Y_d, \widehat{Y}, T)$ *such that* W *is also a solution of* C*.*

Proof. Let $\epsilon > 0$ be given. Since W is a solution to R, then $\forall \boldsymbol{\eta} \in Y(t_0)$, $\exists \mathbf{u} \in W$ and $t_{\mathbf{u},\boldsymbol{\eta}}$ such that $\|\mathbf{y}_{\mathbf{u}}(t) - \mathbf{y}_d(t)\| < \epsilon$ for all $t \geq t_{\mathbf{u},\boldsymbol{\eta}}$. Let $W_{\boldsymbol{\eta}} := \{u \in W : \lim_{t\to\infty} \|\mathbf{y}_{\mathbf{u}}(t) - \mathbf{y}_d(t)\| = 0\}$ and define $\tau(\boldsymbol{\eta}) := \inf_{\mathbf{u}\in W_{\boldsymbol{\eta}}} t_{\mathbf{u},\boldsymbol{\eta}}$. Then W is a solution to C with $t_f := \sup_{\boldsymbol{\eta}\in Y} \tau(\boldsymbol{\eta})$, $T = [t_0, t_f]$ and $Y_d(t) = \{\boldsymbol{\eta} \in Y(t_f) : \|\boldsymbol{\eta} - \mathbf{y}_d(t)\| \leq \epsilon\}$. □

Lemma 2.1.2. *If* W *is a solution to a control problem* $C = (\mathbf{f}, U, Y, Y_d, \widehat{Y}, T)$ *such that* $Y_d(t) = \{\boldsymbol{\eta}_d\}$ *for all* $t \geq t_f$ *and* $(\boldsymbol{\eta}_d, \mathbf{u}_{\boldsymbol{\eta}}(t))$ *is an equilibrium point of* $\mathbf{f}$ *for all* $t \geq t_f$*, then* W *is a solution to regularization problem* $S = (\mathbf{f}, U, Y, \mathbf{y}_d, \widehat{Y}, t_0)$ *with* $\mathbf{y}_d(t) = \boldsymbol{\eta}_d$*.*

Proof. Since W is a solution of C, then for all $\boldsymbol{\eta} \in Y(t_0)$, there exists a $\mathbf{u} \in W$ such that

$$\boldsymbol{\eta} + \int_{t_0}^{t_f} \mathbf{f}(\tau, \mathbf{y}(\tau), \mathbf{u}(\tau))\, d\tau \in Y_d = \{\boldsymbol{\eta}_d\}.$$

On the other hand, since $(\boldsymbol{\eta}_d, \mathbf{u}(t))$ is an equilibrium point of $\mathbf{f}$ for all $t \geq t_f$, then $\mathbf{f}(t, \boldsymbol{\eta}_d, \mathbf{u}(t_f)) = 0 = \dot{\mathbf{y}}(t)$. Consequently, the value of $\mathbf{y}_{\mathbf{u}}(t)$ remains at $\boldsymbol{\eta}_d$ and $\mathbf{f}(t, \mathbf{y}(t), \mathbf{u}(t)) = 0$ for all $t \geq t_f$. Thus, for all $t \geq t_f$ we have

$$\|\mathbf{y}_d(t) - \mathbf{y}_{\mathbf{u}}(t)\| = \|\boldsymbol{\eta}_d - \boldsymbol{\eta}_d - \int_{t_f}^{t} \mathbf{f}(\tau, \mathbf{y}(\tau), \mathbf{u}(\tau))\, d\tau\| = 0,$$

that implies $\lim_{t\to\infty} \|\mathbf{y}_d(t) - \mathbf{y}_{\mathbf{u}}(t)\| = 0$. □

Since the objective of a regularization problem is to reduce the error between $\mathbf{y}_{\mathbf{u}}(t)$ and $\mathbf{y}_d(t)$, it is natural to define $\mathbf{e}(t) := \mathbf{y}_d(t) - \mathbf{y}(t)$. In addition, let $\mathbf{u}_d : \mathbb{R} \to \mathbb{R}^m$ to be a given operating control function ($\mathbf{u}_d(t)$ could represent the equilibrium

input associated with $\mathbf{y}_d(t)$), we can define $\mathbf{v}(t) := \mathbf{u}_d(t) - \mathbf{u}(t)$. Writing $\mathbf{f}$ in terms of $\mathbf{e}$ and $\mathbf{v}$ yields

$$\dot{\mathbf{e}}(t) = \dot{\mathbf{y}}_d(t) - \mathbf{f}\big(t, \mathbf{y}_d(t) - \mathbf{e}(t), \mathbf{u}_d(t) - \mathbf{v}(t)\big). \tag{2.4}$$

Equation (2.4) represents the *error dynamics* for the corresponding regularization problem. Moreover, applying Taylor series expansion to (2.4) leads to

$$\begin{aligned} \dot{\mathbf{y}}_d(t) - \dot{\mathbf{e}}(t) &= \mathbf{f}\big(t, \mathbf{y}_d(t) - \mathbf{e}(t), \mathbf{u}_d(t) - \mathbf{v}(t)\big) && (2.5)\\ &\approx \mathbf{f}\big(t, \mathbf{y}_d(t), \mathbf{u}_d(t)\big) - \frac{\partial \mathbf{f}}{\partial \mathbf{y}(t)}\Big|_{\substack{\mathbf{y}_d(t)\\ \mathbf{u}_d(t)}} \mathbf{e}(t) - \frac{\partial \mathbf{f}}{\partial \mathbf{u}(t)}\Big|_{\substack{\mathbf{y}_d(t)\\ \mathbf{u}_d(t)}} \mathbf{v}(t). && (2.6) \end{aligned}$$

Noting that $\dot{\mathbf{y}}_d(t) = \mathbf{f}\big(t, \mathbf{y}_d(t), \mathbf{u}_d(t)\big)$, equation (2.6) simplifies to

$$\dot{\mathbf{e}}(t) \approx \frac{\partial \mathbf{f}}{\partial \mathbf{y}(t)}\Big|_{\substack{\mathbf{y}_d(t)\\ \mathbf{u}_d(t)}} \mathbf{e}(t) + \frac{\partial \mathbf{f}}{\partial \mathbf{u}(t)}\Big|_{\substack{\mathbf{y}_d(t)\\ \mathbf{u}_d(t)}} \mathbf{v}(t) = \mathbf{A}(t)\mathbf{e}(t) + \mathbf{B}(t)\mathbf{v}(t). \tag{2.7}$$

Owing to its linear nature and significant developments of techniques for linear systems, using a linear approximation of the system can significantly simplify solving control and in particular regularization problems. Moreover, if $\mathbf{v}(t)$ is defined as a linear function of $\mathbf{e}(t)$, that is $\mathbf{v}(t) := \mathbf{K}(t)\mathbf{e}(t)$ for a specific $\mathbf{K}(t) \in \mathbb{R}^{m\times n}$, then

$$\dot{\mathbf{e}}(t) = \mathbf{A}(t)\mathbf{e}(t) + \mathbf{B}(t)\mathbf{v}(t) = \big(\mathbf{A}(t) + \mathbf{B}(t)\mathbf{K}(t)\big)\mathbf{e}(t) = \mathbf{A}_{\mathbf{c}}(t)\mathbf{e}(t), \tag{2.8}$$

where $\mathbf{A}_c(t) \in \mathbb{R}^{n\times n}$ defines the closed-loop form of the system. Since (2.8) could be solved as a simple initial value problem (without the explicit dependency on $\mathbf{u}$), then solving the corresponding regularization problem reduces to finding $\mathbf{K}(t)$ such that eigenvalues of $\mathbf{A}_c(t)$ coincide with some desired values (further discussions on linear systems and corresponding linear control synthesis are available in [6]).

Linear controllers of the form $\mathbf{v}(t) := \mathbf{K}(t)\mathbf{e}(t)$ are effective under two assumptions: (i) if $\|\mathbf{e}(t)\|$ is small enough such that the linear model (2.7) is a valid approximation of (2.4); (ii) if the output of $\mathbf{v}(t)$ could be directly applied to the system[5]. In real physical systems, due to the limited power of the actuators and system properties, range of $\mathbf{u}(t)$ may be constrained to only a specific subset of $\mathbb{R}^m$, denoted by D. Consequently, we can define $\mathbf{u}$ as

$$\mathbf{u}(t) = \underset{D}{\text{Proj}}\big(\mathbf{u}_d(t) - \mathbf{v}(t)\big) := \arg\min_{\mathbf{w}\in D} \|\mathbf{u}_d(t) - \mathbf{K}(t)\mathbf{e}(t) - \mathbf{w}\|. \tag{2.9}$$

Let U denote the space of feasible control functions and D to be the acceptable range for $\mathbf{u} \in U$, then we can define two subsets of U that are: the set of realizable functions, U_N, and the set of projected linear functions, U_L, defined as:

$$U_N := \big\{\mathbf{u} \in U : \mathbf{u}(t) \in D \text{ for a.e. } t \in \mathbb{R}\big\}, \tag{2.10}$$

$$U_L := \big\{\mathbf{u} \in U : \mathbf{u}(t) = \underset{D}{\text{Proj}}\big(\mathbf{u}_d(t) - \mathbf{K}(t)\mathbf{e}(t)\big),\ \mathbf{u}_d \in U,\ \mathbf{K}(t) \in \mathbb{R}^{m\times n}\big\}. \tag{2.11}$$

In this research, we will focus our discussions on a specific class of control problems, named as conditionally controllable problems, defined as the following.

Definition 2.1.5 (Conditionally controllable problem)**.** Let $R = (\mathbf{f}, U, Y, \mathbf{y}_d, \widehat{Y}, t_0)$ be a regularization problem and $D \subset \mathbb{R}^m$ be the feasible range for $\mathbf{u} \in U$. Let U_N and U_L be defined as (2.10) and (2.11) for the given D. Then, R is conditionally controllable if the following conditions hold:

(i) The regularization problem $(\mathbf{f}, U_L, Y, \mathbf{y}_d, \widehat{Y}, t_0)$ does not have a solution; However, there exists $t \mapsto Y_N(t) \subset Y(t)$ such that the modified regularization

[5]Clearly if conditions (i) and (ii) are not satisfied, the real representation of error dynamics (2.4) diverges drastically from its linear approximation (2.7). Consequently, the linear techniques used in synthesizing the control functions may be no longer valid.

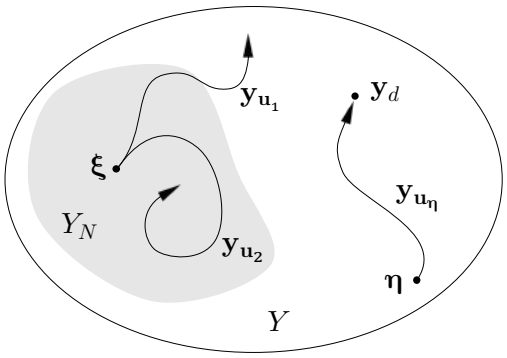

Figure 2.1: A portrait of a conditionally controllable problem. As illustrated, since $\boldsymbol{\eta} \in Y$, $\exists \mathbf{u}_{\boldsymbol{\eta}} \in U_L$ such that $\lim_{t\to\infty} \|\mathbf{y}_d(t) - \mathbf{y}_{\mathbf{u}_{\boldsymbol{\eta}}}(t)\| = 0$. On the other hand, since $\boldsymbol{\xi} \in Y_N$, $\mathbf{y}_{\mathbf{u}_2}$ does not converge to $\mathbf{y}_d$ for any $\mathbf{u}_2 \in U_L$. However, we can find a function $\mathbf{u}_1 \in U_N$ such that $\mathbf{y}_{\mathbf{u}_1}(t_f) \in Y$ for some finite t_f.

problem $(\mathbf{f}, U_L, Y \setminus Y_N, \mathbf{y}_d, \widehat{Y}, t_0)$ has a solution.

(ii) The control problem $(\mathbf{f}, U_N, Y, Y \setminus Y_N, \widehat{Y}, [t_0, t_f])$ has a solution for a finite t_f.

In other words, for conditionally controllable problems, a subset of saturated linear controllers in U_L, which could be synthesized based on relatively simple and general techniques of linear control theory, can solve the control problem in a neighborhood of desired states, but there also exists certain subsets of the state space, for which no function in U_L can lead the system to any point in the set of desired states. As it is discussed in this manuscript, extending the control functions to U_N can lead to more interesting solutions that harness the inherent complexities of nonlinear systems. A conceptual portrait of a conditionally controllable problem is illustrated in Figure 2.1.

Theorem 2.1.3. *Let $R = (\mathbf{f}, U, Y, \mathbf{y}_d, \widehat{Y}, t_0)$ with $D \subset \mathbb{R}^m$, be a conditionally controllable problem. Moreover, let $\mathbf{u}_d : \mathbb{R} \to \mathbb{R}^m$, $\mathbf{u}_d \in U$ to be a given operating control function. If $R_e = (\mathbf{f}_e, V, E, \mathbf{0}, \widehat{E}, t_0)$ with $\mathbf{e}(t) := \mathbf{y}_d(t) - \mathbf{y}(t)$, $\mathbf{v}(t) := \mathbf{u}_d(t) - \mathbf{u}(t)$, $\mathbf{f}_e := \dot{\mathbf{y}}_d(t) - \mathbf{f}(t, \mathbf{y}_d(t) - \mathbf{e}(t), \mathbf{u}_d(t) - \mathbf{v}(t))$, $E(t) := \mathbf{y}_d(t) - Y(t)$, $\widehat{E}(t) := \dot{\mathbf{y}}_d(t) - \widehat{Y}(t)$ and $V := \{\mathbf{u}_d(t) - \mathbf{u}(t) : \mathbf{u} \in U\}$ be the error regularization problem associated with R, then R_e is conditionally controllable if and only if R is conditionally controllable.*

Proof. Based on the definition of E, it is immediate that if $\mathbf{y}(t), \mathbf{y}_d(t) \in Y(t)$ and $\dot{\mathbf{y}}_d(t), \dot{\mathbf{y}} \in \widehat{Y}(t)$ if and only if $\mathbf{e}(t) \in E(t)$ and $\dot{\mathbf{e}}(t) \in \widehat{E}(t)$. First we show that if R is

conditionally controllable, then R_e is also conditionally controllable. Let $W_L \subset U_L$ be a solution for $R_1 = (\mathbf{f}, U_L, Y \setminus Y_N, \mathbf{y}_d, \widehat{Y}, t_0)$, first we show that $Q_L \subset V_L := \{\mathbf{v} \in V : \mathbf{v}(t) = \mathbf{u}_d(t) - \mathbf{u}(t),\ \mathbf{u} \in U_L\}$ is a solution for $R_e = (\mathbf{f}_e, V_L, E \setminus E_N, \mathbf{0}, \widehat{E}, t_0)$, where $E_N := \mathbf{y}_d(0) - Y_N$. Let $\boldsymbol{\eta} \in Y$, since W_L is a solution for R_1, then $\exists \mathbf{u} \in W_L$ such that

$$\lim_{t\to\infty} \|\mathbf{y}_d(t) - \mathbf{y}_\mathbf{u}(t)\| = \|\mathbf{y}_d(t) - \boldsymbol{\eta} - \int_{t_0}^{t} \mathbf{f}(t, \mathbf{y}(t), \mathbf{u}(t))dt\| = 0.$$

On the other hand, we have

$$\mathbf{e}(t) = \boldsymbol{\xi} + \int_{t_0}^{t} \dot{\mathbf{y}}_d(t) - \mathbf{f}(t, \mathbf{y}_d(t) - \mathbf{e}(t), \mathbf{u}_d(t) - \mathbf{v}(t))\, dt,$$

where $\boldsymbol{\xi} \in E$, $\boldsymbol{\xi} = \mathbf{y}_d(t_0) - \boldsymbol{\eta}$. Based on the definition of Q_L, $\exists \mathbf{v} \in Q_L$ such that $\mathbf{v}(t) = \mathbf{u}_d(t) - \mathbf{u}(t)$. Substituting $\mathbf{e}(t) = \mathbf{y}_d(t) - \mathbf{y}(t)$ and $\mathbf{v}(t) = \mathbf{u}_d(t) - \mathbf{u}(t)$ in the definition for $\mathbf{e}(t)$, we get

$$-\int_{t_0}^{t} \mathbf{f}(t, \mathbf{y}(t), \mathbf{u}(t))\, dt = \mathbf{e}(t) - \boldsymbol{\xi} - \int_{t_0}^{t} \dot{\mathbf{y}}_d(t)\, dt.$$

Knowing that $\boldsymbol{\eta} = \mathbf{y}_d(0) - \boldsymbol{\xi}$ and using the obtained expression for $\int_{t_0}^{t} \mathbf{f}(t, \mathbf{y}(t), \mathbf{u}(t))\, dt$ in $\|\mathbf{y}_d(t) - \mathbf{y}_\mathbf{u}(t)\|$ yields

$$\begin{aligned}
\|\mathbf{y}_d(t) - \mathbf{y}_\mathbf{u}(t)\| &= \|\mathbf{y}_d(t) - \boldsymbol{\eta} - \int_{t_0}^{t} \mathbf{f}(t, \mathbf{y}(t), \mathbf{u}(t))dt\| \\
&= \|\mathbf{y}_d(t) - \boldsymbol{\eta} + \mathbf{e}(t) - \boldsymbol{\xi} - \int_{t_0}^{t} \dot{\mathbf{y}}_d(t)\, dt\| \\
&= \|\mathbf{y}_d(t) - \Big(\mathbf{y}_d(0) + \int_{t_0}^{t} \dot{\mathbf{y}}_d(t)\, dt\Big) + \mathbf{e}(t)\| = \|\mathbf{e}(t)\|.
\end{aligned}$$

Consequently, $\lim_{t\to\infty} \|\mathbf{y}_d(t) - \mathbf{y}_\mathbf{u}(t)\| = 0 \implies \lim_{t\to\infty} \|\mathbf{e}(t)\| = 0$. To show that Q_L cannot solve R_e, assume by contradiction that for $\boldsymbol{\xi} \in E_N$, $\exists \mathbf{v} \in Q_L$

such that $\lim_{t\to\infty} \|\mathbf{e_v}(t)\| = 0$. Based on the definition of Q_L, $\exists \mathbf{u} \in W_L$ such $\mathbf{u}(t) = \mathbf{u}_d(t) - \mathbf{v}(t)$; Based on the derivation in the first part, we get

$$\begin{aligned}
\lim_{t\to\infty} \|\mathbf{e_v}(t)\| &= \lim_{t\to\infty} \|\boldsymbol{\xi} + \int_{t_0}^{t} \dot{\mathbf{y}}_d(t) - \mathbf{f}(t, \mathbf{y}_d(t) - \mathbf{e}(t), \mathbf{u}_d(t) - \mathbf{v}(t))\, dt\| \\
&= \lim_{t\to\infty} \|\boldsymbol{\xi} + \int_{t_0}^{t} \dot{\mathbf{y}}_d(t)\, dt - \int_{t_0}^{t} \mathbf{f}(t, \mathbf{y}, \mathbf{u}(t))\, dt\| \\
&= \lim_{t\to\infty} \|\Big(\mathbf{y}_d(0) + \int_{t_0}^{t} \dot{\mathbf{y}}_d(t)\, dt\Big) - \Big(\boldsymbol{\eta} + \int_{t_0}^{t} \mathbf{f}(t, \mathbf{y}, \mathbf{u}(t))\, dt\Big)\| \\
&= \lim_{t\to\infty} \|\mathbf{y}_d(t) - \mathbf{y_u}(t)\| = 0,
\end{aligned}$$

that is contradictions since $\boldsymbol{\xi} \in E_N \implies \boldsymbol{\eta} \in Y_N$ and $\nexists \mathbf{u} \in W_L$ such that $\lim_{t\to\infty} \|\mathbf{y}_d(t) - \mathbf{y_u}(t)\| = 0$.

To show that R_e also satisfies the second requirement for conditional controllability, let $W_N \subset U_N$ to be a solution for $C_1 = (\mathbf{f}, U_N, Y, Y_d, \widehat{Y}, [t_0, t_f])$, $\boldsymbol{\xi} \in E_N$ and $Q_N \subset V_N := \{\mathbf{v} \in V : \mathbf{v}(t) = \mathbf{u}_d(t) - \mathbf{u}(t), \mathbf{u} \in U_N\}$. Moreover, let $E_d(t) := \mathbf{y}_d(t) - Y_d(t)$; we claim that Q_N is a solution for $C_e = (\mathbf{f}_e, V_N, E, E_d, \widehat{E}, [t_0, t_f])$. Since $\boldsymbol{\xi} \in E_N$ then $\boldsymbol{\eta} = \mathbf{y}_d(0) - \boldsymbol{\xi}$ is in Y_N, thus $\exists \mathbf{u} \in W_N$ such that $\mathbf{y_u}(t_f) \in Y_d$. Pick $\mathbf{v} \in Q_N$ such that $\mathbf{v}(t) = \mathbf{u}_d(t) - \mathbf{u}(t)$, then

$$\begin{aligned}
\mathbf{e_v}(t_f) &= \boldsymbol{\xi} + \int_{t_0}^{t_f} \dot{\mathbf{y}}_d(t) - \mathbf{f}(t, \mathbf{y}_d(t) - \mathbf{e}(t), \mathbf{u}_d(t) - \mathbf{v}(t))\, dt \\
&= \boldsymbol{\xi} + \int_{t_0}^{t_f} \dot{\mathbf{y}}_d(t)\, dt - \int_{t_0}^{t_f} \mathbf{f}(t, \mathbf{y}(t), \mathbf{u}\, dt \\
&= \Big(\mathbf{y}_d(0) + \int_{t_0}^{t_f} \dot{\mathbf{y}}_d(t)\, dt\Big) - \Big(\boldsymbol{\eta} + \int_{t_0}^{t_f} \mathbf{f}(t, \mathbf{y}(t), \mathbf{u}\, dt\Big) = \mathbf{y}_d(t_f) - \mathbf{y_u}(t_f).
\end{aligned}$$

Since $\mathbf{y_u}(t_f) \in Y_d$, thus $\mathbf{y}_d(t_f) - \mathbf{y_u}(t_f) \in E_d$, that implies Q_N is a solution for $C_e = (\mathbf{f}_e, V_N, E, E_d, \widehat{E}, [t_0, t_f])$. Using the same steps as presented here, we can show that conditionally controllability of R_e implies conditional controllability of R. $\square$

There are many examples of conditionally controllable problems in physical systems with nonlinear dynamics. In particular, such problems are frequently observed in control of multi rigid body systems. As it is highlighted in the following sections and chapters, in the case of rigid body mechanics, such solutions are often related to controlling the energy or momentum of the system that allows driving state vectors from Y_N to Y_L.

To further examine conditionally controllable problems, we can proceed with the case study of a simple point-mass pendulum model as discussed in the following section.

2.2 Case study: swinging up a pendulum

Consider the point-mass pendulum system as depicted in Figure 2.2. The equation of motion for the pendulum is

$$\ddot{q}(t) = \frac{g}{l}\sin(q(t)) + \frac{u(t)}{ml^2}, \tag{2.12}$$

where $q(t) \in \mathbb{R}$ is the angle measured from an axis parallel to the gravitational (free-fall) acceleration, $g \in \mathbb{R}$, to the pendulum link in counterclockwise direction. $l \in \mathbb{R}^+$ is the length of the pendulum, $m \in \mathbb{R}^+$ is the concentrated mass at the bob. $u(t) \in \mathbb{R}$ is the torque applied at the joint of the pendulum. To simplify the equations and without loss of generality, let $g = l = m = 1$ (with appropriate units). Let $\mathbf{y}(t) := [q(t),\ \dot{q}(t)]^T$ be the state vector, then (2.12) leads to $\dot{\mathbf{y}}(t) = \mathbf{f}(t, \mathbf{y}(t), u(t))$ where

$$\mathbf{f}(t, \mathbf{y}(t), u(t)) = \begin{bmatrix} y_2(t) \\ \sin\big(y_1(t)\big) + u(t) \end{bmatrix}. \tag{2.13}$$

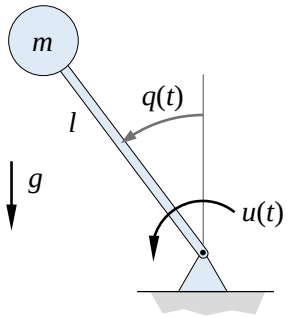

Figure 2.2: The pendulum system and the corresponding parameters.

Let $R = (\mathbf{f}, U, Y, \mathbf{y}_d, \widehat{Y}, t_0)$ to be a regularization problem defined for $\mathbf{f}$ in (2.13) with the objective of moving and holding the pendulum at upward configuration where $y_1(t) = y_2(t) = 0$. Since $\mathbf{f}$ does not explicitly depend on time, we may simply let $t_0 = 0$; moreover, $y_1(t) = y_2(t) = 0$, implies $\mathbf{y}_d(t) \equiv \mathbf{0}$. For simplicity, we assume that there are no physical restrictions on the mechanical structure of the pendulum, therefor $Y = \widehat{Y} = \mathbb{R}^2$. Moreover, we follow the assumption that the motor used to generate the input is just limited by its maximum applicable torque. Accordingly, we define $U := L^\infty$ and $D := [-w_{max}, w_{max}]$, where $w_{max} \geq 0$ is the maximum torque output of the motor. Thus, the exact definition of R is $R = (\mathbf{f}, L^\infty, \mathbb{R}^2, \mathbf{0}, \mathbb{R}^2, 0)$ for $\mathbf{f}$ as in (2.13).

Let $\mathbf{e}(t) := \mathbf{y}_d(t) - \mathbf{y}(t) = -\mathbf{y}(t)$ be the error vector. To set $\mathbf{y}(t) = \mathbf{0}$ be an equilibrium state for the system, we need $\mathbf{f}(t, \mathbf{0}, u_d(t))$ to be zero for all $t \in \mathbb{R}$, which implies $u_d(t) \equiv 0$. Thus, let $v(t) = u_d(t) - u(t) = -u(t)$, then the error dynamics of the system $\dot{\mathbf{e}}(t) = \mathbf{0} - \mathbf{f}\big(t, \mathbf{0} - \mathbf{e}(t), 0 - v(t)\big)$ simplifies to

$$\begin{bmatrix} \dot{e}_1(t) \\ \dot{e}_2(t) \end{bmatrix} = \begin{bmatrix} e_2(t) \\ \sin\big(e_1(t)\big) - u(t) \end{bmatrix} = \mathbf{f}_e(t, \mathbf{e}(t), u(t)), \tag{2.14}$$

with the corresponding linear approximation $\dot{\mathbf{e}}(t) = \mathbf{A}\mathbf{e}(t) + \mathbf{B}u(t)$ with matrices $\mathbf{A}$

and $\mathbf{B}$ equal to

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 0 \\ -1 \end{bmatrix}. \tag{2.15}$$

To explore the behavior of the linear system with linear controllers, let $u(t) = \langle \mathbf{k}, \mathbf{e}(t) \rangle$ for $\mathbf{k} = [k_1,\ k_2]^T \in \mathbb{R}^2$. Then, the closed-loop representation of the system simplifies to $\dot{\mathbf{e}}(t) = \mathbf{A}_c\, \mathbf{e}(t)$, where

$$\mathbf{A}_c = \begin{bmatrix} 0 & 1 \\ 1 - k_1 & -k_2 \end{bmatrix}, \tag{2.16}$$

with eigenvalues

$$\sigma(\mathbf{A}_c) = \{\lambda_1(\mathbf{k}),\, \lambda_2(\mathbf{k})\} = \Big\{ \frac{1}{2} \big(-k_2 \pm \sqrt{k_2^2 + 4(1 - k_1)}\, \big) \Big\}. \tag{2.17}$$

Since $\mathbf{A}_c$ has two distinct eigenvalues in $\mathbb{C}$, it is diagonalizable [7]. Therefor, consider the decomposition $\mathbf{A}_c = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{\mathbf{T}}$, where $\mathbf{Q} \in \mathbb{C}^{2\times 2}$ is the matrix of eigenvectors and $\mathbf{\Lambda} = \operatorname{diag}(\lambda_1(\mathbf{k}), \lambda_1(\mathbf{k}))$. Using such decomposition, we can compute the solution of $\dot{\mathbf{e}}(t) = \mathbf{A}_c\, \mathbf{e}(t)$ for a specific $\mathbf{k}$ as

$$\mathbf{e}_{\mathbf{k}}(t) = \mathbf{Q} \begin{bmatrix} e^{\lambda_1(\mathbf{k})t} & 0 \\ 0 & e^{\lambda_2(\mathbf{k})t} \end{bmatrix} \mathbf{Q}^{-1}\, \mathbf{e}_0, \tag{2.18}$$

where $\mathbf{e}_0 = -\mathbf{y}(0)$ is the initial error value. Since our objective is to find a linear control function that ensures $\lim_{t\to\infty} \|\mathbf{e}_{\mathbf{k}}(t)\| = 0$, we can choose k_1 and k_2 such that $e^{\lambda_1(\mathbf{k})t}$ and $e^{\lambda_2(\mathbf{k})t} \to 0$ as $t \to \infty$, that is $\Re(\lambda_1)$ and $\Re(\lambda_2)$ are negative. Let Ω to be the set of $\mathbf{k}$ vectors such that $\Re(\lambda_1(\mathbf{k}))$ and $\Re(\lambda_2(\mathbf{k}))$ are negative, then

$$\Omega = \{\mathbf{k} \in \mathbb{R}^2 :\ k_1 > 1 \text{ and } k_2 > 0\}. \tag{2.19}$$

Consequently, for every $\mathbf{k} \in \Omega$, $\lim_{t\to\infty} \|\mathbf{e}_{\mathbf{k}}(t)\| = 0$. However, we have not yet explored the response of (2.13) when the input is constrained to set $D = [-w_{max}, w_{max}]$.

Based on Theorem 2.1.3 if $R_e = (\mathbf{f}_e, L^\infty, \mathbb{R}^2, \mathbf{0}, \mathbb{R}^2, 0)$ is conditionally controllable, then R is also conditionally controllable. Accordingly, we can focus our discussions solely on the error dynamics of the system. In what follows we will develop the tools required to investigate how R_e morphs into a conditionally controllable problem as w_{max} decreases. First we need to identify functions in U_N and U_L. Since $D \subset \mathbb{R}$ is an interval, then U_L is

$$U_L = \Big\{ u \in L^\infty : u(t) = \operatorname*{sat}_D \langle \mathbf{k}, \mathbf{e}(t) \rangle, \ \mathbf{k} \text{ and } \mathbf{e}(t) \in \mathbb{R}^2 \Big\}, \tag{2.20}$$

where for a given $A \subset \mathbb{R}$ and $x \in \mathbb{R}$, $\operatorname{sat}_A x$ is

$$\operatorname*{sat}_A x := \min \Big\{ \max \big\{ x, \ \inf(A) \big\}, \sup(A) \Big\}. \tag{2.21}$$

Similarly, we can identify functions in U_N as

$$U_N = \Big\{ u \in L^\infty : u(t) = D \text{ for a.e. } t \in \mathbb{R} \Big\}. \tag{2.22}$$

Let $\mathbf{k} \in \mathbb{R}^2$, we define $\Gamma := \{ \boldsymbol{\xi} \in \mathbb{R}^2 : |\langle \mathbf{k}, \boldsymbol{\xi} \rangle| \leq w_{max} \}$ that corresponds to the set of points between two affine lines $\langle \mathbf{k}, \boldsymbol{\xi} \rangle = -w_{\max}$ and $\langle \mathbf{k}, \boldsymbol{\xi} \rangle = w_{\max}$. Moreover, let $\Sigma_L := \{ \boldsymbol{\xi} \in \mathbb{R}^2 : \langle \mathbf{k}, \boldsymbol{\xi} \rangle \leq -w_{\max} \}$ and $\Sigma_R := \{ \boldsymbol{\xi} \in \mathbb{R}^2 : \langle \mathbf{k}, \boldsymbol{\xi} \rangle \geq w_{\max} \}$ [6]. For every $u \in U_L$, if $\mathbf{e}(t) \in \Gamma$, then $u(t) = \langle \mathbf{k}, \mathbf{e}(t) \rangle$ with the error dynamic defined as

$$\begin{aligned} \dot{e}_1(t) &= e_2(t), \\ \dot{e}_2(t) &= \sin(e_1(t)) - k_1 e_1(t) - k_2 e_2(t). \end{aligned} \tag{2.23}$$

[6]Note that based on the definitions of Γ, Σ_L and Σ_R, we have $\Sigma_L \cap \Gamma = \{ \boldsymbol{\xi} \in \mathbb{R}^2 : \langle \mathbf{k}, \boldsymbol{\xi} \rangle = -w_{max} \}$, $\Sigma_R \cap \Gamma = \{ \boldsymbol{\xi} \in \mathbb{R}^2 : \langle \mathbf{k}, \boldsymbol{\xi} \rangle = w_{max} \}$ and $\Sigma_L \cap \Sigma_R = \emptyset$ for any $w_{max} > 0$.

Let $E_\Gamma(\mathbf{e}(t))$ to denote the total energy of the system for $\mathbf{e}(t) \in \Gamma$, then

$$E_\Gamma(\mathbf{e}(t)) = \frac{1}{2}e_2^2(t) + \cos(e_1(t)) - 1 + \frac{1}{2}k_1 e_1^2(t). \tag{2.24}$$

Taking derivative of $E_\Gamma(\mathbf{e}(t))$ with respect to time leads to

$$\dot{E}_\Gamma(\mathbf{e}(t)) = e_2(t)\dot{e}_2(t) - \sin(e_1(t))\dot{e}_1(t) + k_1 e_1(t)\dot{e}_1(t). \tag{2.25}$$

Substituting $\dot{e}_1(t)$ and $\dot{e}_2(t)$ from (2.23) in (2.26) yields

$$\dot{E}_\Gamma(\mathbf{e}(t)) = -k_2 e_2^2(t), \tag{2.26}$$

which is negative for every $\mathbf{k} \in \Omega$ and zero if $e_2(t) = 0$. On the other hand, for every $u \in U_L$, if $\mathbf{e}(t) \in \Sigma_L \cup \Sigma_R$, then $u(t) = \pm w_{max}$. Thus, the input of the system serves as a conservative force and provides the possibility of characterizing the trajectories in the phase portrait[7] of the error dynamics using work-energy principle. Let $t_a \leq t_b \in \mathbb{R}$, then for every $\mathbf{e} \subset \Sigma_L$ or $\subset \Sigma_R$ based on the work-energy principle we have

$$E_{\Sigma_L}(\mathbf{e}(t_b)) = E_{\Sigma_L}(\mathbf{e}(t_a)) + W_{\mathbf{e}(t_a)\to\mathbf{e}(t_b)}, \tag{2.27}$$

$$E_{\Sigma_R}(\mathbf{e}(t_b)) = E_{\Sigma_R}(\mathbf{e}(t_a)) + W_{\mathbf{e}(t_a)\to\mathbf{e}(t_b)}, \tag{2.28}$$

where $E_{\Sigma_L}(\mathbf{e}(t))$ and $E_{\Sigma_L}(\mathbf{e}(t))$ denote the energy of the system for $\mathbf{e}(t)$ in Σ_L and Σ_R, respectively. $W_{\mathbf{e}(t_a)\to\mathbf{e}(t_b)}$ is the work required to take the system from $\mathbf{e}(t_a)$ to $\mathbf{e}(t_b)$. The total energy associated with the simplified pendulum model in (2.13)

[7] A *phase portrait* is a set of trajectories in the phase plane of the system that illustrates evolution of the dynamics form various initial conditions.

with an arbitrary input is

$$E(\mathbf{y}(t)) = \frac{1}{2}y_2^2(t) + \cos(y_1(t)), \tag{2.29}$$

that implies

$$E_{\Sigma_L}(\mathbf{e}(t)) = E_{\Sigma_R}(\mathbf{e}(t)) = E_{\Sigma}(\mathbf{e}(t)) = \frac{1}{2}e_2^2(t) + \cos(-e_1(t)) = \frac{1}{2}e_2^2(t) + \cos(e_1(t)). \tag{2.30}$$

Moreover, the work done by constant external force $u(t) = \pm w_{max}$ on the pendulum is

$$W_{\mathbf{e}(t_a)\to\mathbf{e}(t_b)} = -\int_{e_1(t_a)}^{e_1(t_b)} u(t)\, de_1 = \begin{cases} w_{max}\big(e_1(t_b) - e_1(t_a)\big), & \text{if } \mathbf{e}(t) \in \Sigma_L, \\ w_{max}\big(e_1(t_a) - e_1(t_b)\big), & \text{if } \mathbf{e}(t) \in \Sigma_R, \end{cases} \tag{2.31}$$

for all $t \in [t_a, t_b]$. Substituting (2.31) and (2.30) in (2.27) and (2.28) yields

for Σ_L :

$$\frac{e_2^2(t_b)}{2} + \cos(e_1(t_b)) - w_{max}e_1(t_b) = \frac{e_2^2(t_a)}{2} + \cos(e_1(t_a)) - w_{max}e_1(t_a), \tag{2.32}$$

for Σ_R :

$$\frac{e_2^2(t_b)}{2} + \cos(e_1(t_b)) + w_{max}e_1(t_b) = \frac{e_2^2(t_a)}{2} + \cos(e_1(t_a)) + w_{max}e_1(t_a). \tag{2.33}$$

Noting that (2.32) and (2.33) are true for any initial condition $\mathbf{e}(t) \in \Sigma_L$ and Σ_R, respectively; we can define

$$H_{\Sigma_L}(\mathbf{e}(t)) := \frac{1}{2}e_2^2(t) + \cos(e_1(t)) - w_{max}e_1(t), \tag{2.34}$$

$$H_{\Sigma_R}(\mathbf{e}(t)) := \frac{1}{2}e_2^2(t) + \cos(e_1(t)) + w_{max}e_1(t). \tag{2.35}$$

The following lemma highlights the connection between $H_{\Sigma_L}(\mathbf{e}(t))$ and $H_{\Sigma_R}(\mathbf{e}(t))$ with trajectories in Σ_L and Σ_R, respectively [8].

Lemma 2.2.1. *If $\mathbf{e}(t) \in \Sigma_L$ for $t \in [t_0, t_1]$, then $\mathbf{e}(t)$ coincides with a level set of $H_{\Sigma_L}(\mathbf{e}(t))$ for all $t \in [t_0, t_1]$. Similarly, if $\mathbf{e}(t) \in \Sigma_R$ for $t \in [t_0, t_1]$, then $\mathbf{e}(t)$ coincides with a level set of $H_{\Sigma_R}(\mathbf{e}(t))$ for all $t \in [t_0, t_1]$.*

Proof. If a level set of $H_{\Sigma_L}(\mathbf{e}(t))$ coincides with a trajectory in Σ_L for $t \in [t_0, t_1]$, then $H_{\Sigma_L}(\mathbf{e}(t))$ must be constant for every point of that trajectory; that is $H_{\Sigma_L(\mathbf{e}(t))=C}$ for all $t \in [t_0, t_1]$, where $C \in \mathbb{R}$ is a constant. Taking the derivative of $H_{\Sigma_L}(\mathbf{e}(t))$ with respect to time leads to

$$\frac{d}{dt}H_{\Sigma_L}(\mathbf{e}(t)) = e_2(t)\dot{e}_2(t) - \sin(e_1(t))\dot{e}_1(t) - w_{max}\dot{e}_1(t).$$

Moreover, for every trajectory in Σ_L we have

$$\begin{aligned}\dot{e}_1(t) &= e_2(t),\\ \dot{e}_2(t) &= \sin(e_1(t)) + w_{max}.\end{aligned}$$

Substituting $\dot{e}_1(t)$ and $\dot{e}_2(t)$ in $\frac{d}{dt}H_{\Sigma_L}(\mathbf{e}(t))$ yields

$$\begin{aligned}\frac{d}{dt}H_{\Sigma_L}(\mathbf{e}(t)) &= e_2(t)\dot{e}_2(t) - \sin(e_1(t))\dot{e}_1(t) - w_{max}\dot{e}_1(t)\\ &= e_2(t)\Big(\sin(e_1(t)) + w_{max} - \sin(e_1(t)) - w_{max}\Big) = 0,\end{aligned}$$

that implies $H_{\Sigma_L}(\mathbf{e}(t))$ is constant for all $t \in [t_0, t_1]$. Similarly, substituting $\dot{e}_1(t) =$

[8] It is interesting to note that $H_{\Sigma_L}(\mathbf{e}(t))$ and $H_{\Sigma_R}(\mathbf{e}(t))$ are the Hamiltonian functions [8] of the system for $u(t) = \pm w_{max}$. As an example

$$\dot{e}_1(t) = \frac{\partial H_{\Sigma_L}(\mathbf{e}(t))}{\partial e_2(t)} = e_2(t), \quad \dot{e}_2(t) = -\frac{\partial H_{\Sigma_L}(\mathbf{e}(t))}{\partial e_1(t)} = \sin(e_1(t)) + w_{max},$$

which is exactly the dynamic equation that represents the trajectories in Σ_L.

$e_2(t)$ and $\dot{e}_2(t) = \sin(e_1(t)) - w_{max}$ in $\frac{d}{dt}H_{\Sigma_R}(\mathbf{e}(t))$ leads to $\frac{d}{dt}H_{\Sigma_R}(\mathbf{e}(t)) = 0$. $\square$

Let $\nabla H_{\Sigma_L}(\mathbf{e}(t))$ and $\nabla H_{\Sigma_R}(\mathbf{e}(t))$ denote gradients of $H_{\Sigma_L}(\mathbf{e}(t))$ and $H_{\Sigma_R}(\mathbf{e}(t))$ with respect to $\mathbf{e}(t)$, respectively; then

$$\nabla H_{\Sigma_L}(\mathbf{e}(t)) = \begin{bmatrix} -\sin(e_1(t)) - w_{max} \\ e_2(t) \end{bmatrix}, \quad \nabla H_{\Sigma_R}(\mathbf{e}(t)) = \begin{bmatrix} -\sin(e_1(t)) + w_{max} \\ e_2(t) \end{bmatrix}. \tag{2.36}$$

Let S_L and S_R to be the set of stationary points of $H_{\Sigma_L}(\mathbf{e}(t))$ and $H_{\Sigma_R}(\mathbf{e}(t))$, respectively. Then

$$S_L = \{\boldsymbol{\xi} = [\xi_1, 0]^T \in \Sigma_L : \sin(\xi_1) + w_{max} = 0\}, \tag{2.37}$$

and

$$S_R = \{\boldsymbol{\xi} = [\xi_1, 0]^T \in \Sigma_R : \sin(\xi_1) - w_{max} = 0\}. \tag{2.38}$$

We can further decompose both S_L and S_R into disjoint subset S_L^1, S_L^2, S_R^1 and S_R^2 defined as

$$S_L^1 := \Big\{ \big[-2j\pi - \sin^{-1}(w_{max}),\, 0 \big]^T : j \in \{0\} \cup \mathbb{N} \Big\}, \tag{2.39}$$

$$S_L^2 := \Big\{ \big[-(2j+1)\pi + \sin^{-1}(w_{max}),\, 0 \big]^T : j \in \{0\} \cup \mathbb{N} \Big\}, \tag{2.40}$$

$$S_R^1 := \Big\{ \big[2j\pi + \sin^{-1}(w_{max}),\, 0 \big]^T : j \in \{0\} \cup \mathbb{N} \Big\}, \tag{2.41}$$

$$S_R^2 := \Big\{ \big[(2j+1)\pi - \sin^{-1}(w_{max}),\, 0 \big]^T : j \in \{0\} \cup \mathbb{N} \Big\}. \tag{2.42}$$

Let $\nabla^2 H_{\Sigma_L}(\mathbf{e}(t))$ and $\nabla^2 H_{\Sigma_R}(\mathbf{e}(t))$ denote Hessians of $H_{\Sigma_L}(\mathbf{e}(t))$ and $H_{\Sigma_R}(\mathbf{e}(t))$ with

respect to $\mathbf{e}(t)$, respectively; then

$$\nabla^2 H_{\Sigma_L}(\mathbf{e}(t)) = \nabla^2 H_{\Sigma_R}(\mathbf{e}(t)) = \nabla^2 H_{\Sigma}(\mathbf{e}(t)) = \begin{bmatrix} -\cos(e_1(t)) & 0 \\ 0 & 1 \end{bmatrix}. \tag{2.43}$$

Note that for every $\boldsymbol{\xi} \in S_L^1 \cup S_R^1$ and $0 \leq w_{max} < 1$,

$$\nabla^2 H_{\Sigma}(\boldsymbol{\xi}) = \begin{bmatrix} -\sqrt{1 - w_{max}^2} & 0 \\ 0 & 1 \end{bmatrix}, \tag{2.44}$$

is indefinite, revealing that every stationary point in $S_L^1 \cup S_R^1$ is a saddle point if $w_{max} \in [0, 1)$. On the contrary, for every $\boldsymbol{\xi} \in S_L^2 \cup S_R^2$ and $0 \leq w_{max} < 1$,

$$\nabla^2 H_{\Sigma}(\boldsymbol{\xi}) = \begin{bmatrix} \sqrt{1 - w_{max}^2} & 0 \\ 0 & 1 \end{bmatrix}, \tag{2.45}$$

is a real positive definite matrix, indicating that every stationary point in $S_L^2 \cup S_R^2$ is a local minimum for $\big(\chi_{\Sigma_L} H_{\Sigma_L} + \chi_{\Sigma_R} H_{\Sigma_R}\big)(\mathbf{e}(t))$ if $w_{max} \in [0, 1)$. Given the characteristics of singular points in S_L and S_R, we define

$$Q_L^j := \Big\{ \boldsymbol{\xi} \in \Sigma_L : \; H_{\Sigma_L}(\boldsymbol{\xi}) < H_{\Sigma_L}([-\alpha(j), 0]^T) \text{ and } \xi_1 < -\alpha(j) \Big\}, \tag{2.46}$$

$$Q_R^j := \Big\{ \boldsymbol{\xi} \in \Sigma_R : \; H_{\Sigma_R}(\boldsymbol{\xi}) < H_{\Sigma_R}([\alpha(j), 0]^T) \text{ and } \xi_1 > \alpha(j) \Big\}, \tag{2.47}$$

where $\alpha(j) := 2j\pi + \sin^{-1}(w_{max})$. The sets Q_L^j and Q_R^j represent the points in Σ_L and Σ_R with smaller H_{Σ_L} and H_{Σ_R} with respect to the points in S_L^j and S_R^j that are located at the left and right side of the points in S_L^1 and S_R^1, respectively. Provide the above statements, we proceed with the following theorem. The corresponding lemmas and corollaries that are used in the proof of Theorem 2.2.2 are presented

Appendix A.

Theorem 2.2.2. *Let* $w_{max} \in (0,1)$ *and* $\mathbf{k} \in \Omega$ *such that* $\Gamma \cap \Big(\bigcup_{j=0}^{\infty} Q_L^j \Big) = \emptyset$. *Then,* $\lim_{t\to\infty} \|\mathbf{e_k}(t)\| = 0$ *if* $\mathbf{e_k}(0) \notin \bigcup_{j=0}^{\infty}(Q_L^j \cup Q_R^j)$.

Proof. Based on Lemmas A.0.6 and A.0.7, if $\mathbf{e_k}(0) \notin \bigcup_{j=0}^{\infty}(Q_L^j \cup Q_R^j)$, then $\exists t' > t_0$ such that $\mathbf{e_k}(t') \in \partial\Gamma$. Moreover, by Lemmas A.0.2 and A.0.3 we have that ever time the trajectory passes through Γ it enters into an orbit with lower H_{Σ_L} or H_{Σ_R} (depending on the region). This decrease of Hamiltonian continues every time the trajectory passes through Γ until it resides in Γ completely. Then, based on Lemma A.0.1, $\lim_{t\to\infty} \|\mathbf{e_k}(t)\| = 0$. □

Now we have all the tools necessary to prove the following theorem which states that $R_e = (\mathbf{f}_e, L^\infty, \mathbb{R}^2, \mathbf{0}, \mathbb{R}^2, 0)$ for $D \subset (-1,1)$, $D \neq \{0\}$ is a conditionally controllable problem.

Theorem 2.2.3. *Let* $R_e = (\mathbf{f}_e, L^\infty, \mathbb{R}^2, \mathbf{0}, \mathbb{R}^2, 0)$ *be a regularization problem where*

$$\mathbf{f}_e(t, \mathbf{e}(t), u) = \begin{bmatrix} e_2(t) \\ \sin\big(e_1(t)\big) - u(t) \end{bmatrix},$$

Moreover, Let U_L *and* U_N *defined as*

$$U_L = \Big\{ u \in L^\infty : u(t) = \operatorname*{sat}_D \langle \mathbf{k}, \mathbf{e}(t) \rangle, \ \mathbf{k} \text{ and } \mathbf{e}(t) \in \mathbb{R}^2 \Big\},$$
$$U_N = \Big\{ u \in L^\infty : u(t) \in D \text{ for a.e. } t \in \mathbb{R} \Big\},$$

where $D := [-w_{max}, w_{max}]$ *for* $w_{max} > 0$. *Then,* R_e *is conditionally controllable if* $w_{max} \in (0,1)$.

Proof. First we show that a solution for $(\mathbf{f}_e, U_L, \mathbb{R}^2 \setminus Y_N, \mathbf{0}, \mathbb{R}^2, 0)$ exists as a subset

of U_L, where $Y_N := \bigcup_{j=0}^{\infty}(Q_L^j \cup Q_R^j)$. Let $\alpha = \sin^{-1}(w_{max})$ and

$$f(\theta) := \sqrt{2\Big(H_{\Sigma_L}(-\alpha) + w_{max}\theta - \cos(\theta)\Big)}$$
$$= \sqrt{2\Big(w_{max}\theta - \cos(\theta) + w_{max}\alpha + \sqrt{1 - w_{max}^2}\Big)}.$$

Moreover, let

$$m(\theta) := \frac{f(\theta)}{\frac{w_{max}}{k_1} + \theta},$$

and θ^* to be a zero of $g(\theta) := \frac{d}{d\theta}f(\theta) - m(\theta)$ in the interval $\theta \in [-\pi + \alpha,\ -\alpha)$ that is $g(\theta^*) = 0$. Note that such zero exists since $g(\theta)$ is continuous for $\theta \in [-\pi + \alpha,\ -\alpha)$, $g(-\pi + \alpha) > 0$ and $\lim_{\theta \to (-\alpha)^-} g(\theta) = -(1 - w_{max}^2)^{\frac{1}{4}} < 0$. Let $\Psi \subset \Omega$ be $\Psi := \{\mathbf{k} \in \Omega :\ -k_1 < m(\theta^*)k_2\}$, then for every $\mathbf{k} \in \Psi$, $\Gamma \cap Y_N = \emptyset$. Thus, based on Theorem 2.2.2, for every $u \in W_L := \{u \in U_L : \mathbf{k} \in \Psi\}$, $\lim_{t\to\infty} \|\mathbf{e}_u(t)\| = 0$ if $\mathbf{e}_u(0) \in \mathbb{R}^2 \setminus Y_N$. In addition, based on Lemma A.0.4 and A.0.5, if $\mathbf{e}_u(0) \in Y_N$, for any $u \in U_L$, $\mathbf{e}_u(t) \in Y_N$ for all $t \geq t_0$. Since $Y_N \cap \mathbf{0} = \emptyset$ for every $w_{max} > 0$, then there is no subset of U_L that can be a solution for $(\mathbf{f}_e, U_L, \mathbb{R}^2, \mathbf{0}, \mathbb{R}^2, 0)$.

As the final step, we need to show that there exists a subset of U_N that solves $(\mathbf{f}_e, U_N, \mathbb{R}^2, Y \setminus Y_N, \mathbb{R}^2, [0, t_f])$ with $t_f < \infty$. Let $j \in \{0\} \cup \mathbb{N}$, $\mathbf{s}_j^1 := [-2j\pi - \alpha,\ 0]^T$, $\mathbf{s}_j^2 := [-(2j+1)\pi + \alpha,\ 0]^T$ and $\mathbf{e}(0) \in Q_L^j$, then

$$H_{\Sigma_L}(\mathbf{s}_j^2) \leq H_{\Sigma_L}(\mathbf{e}(0)) < H_{\Sigma_L}(\mathbf{s}_j^1).$$

Consequently, we need to find a $u \in U_N$ such that $H_{\Sigma_L}(\mathbf{e}_u(t_f)) \geq H_{\Sigma_L}(\mathbf{s}_j^1)$. Noting that $H_{\Sigma_L}(\mathbf{e}(t)) = E_\Sigma(\mathbf{e}(t)) - w_{max}e_1(t)$ we can define

$$\Delta H := H_{\Sigma_L}(\mathbf{s}_j^1) - H_{\Sigma_L}(\mathbf{e}(0)) = E_\Sigma(\mathbf{s}_j^1) - E_\Sigma(\mathbf{e}(0)) + w_{max}(e_1(0) - (s_1)_j^1).$$

Based on the results of Lemma A.0.4, every trajectory in Q_L^j forms an orbit, thus we can simply take $e_1(0) = (s_1)_j^1$ that leads to

$$\Delta H = E_\Sigma(\mathbf{s}_j^1) - E_\Sigma(\mathbf{e}(0)).$$

Since Q_L^j are disjoint, if u can increase the $E_\Sigma(\mathbf{e}_u(t))$ by ΔH in t_f, then $\mathbf{e}_u(t_f) \in Y \setminus Y_N$. To do so, we propose $u(t) := -k_3 w_{max} \operatorname{sign}(e_2(t))$ with $0 < k_3 \leq 1$ that yields

$$\dot{E}_\Sigma(\mathbf{e}(t)) = k_3 w_{max} e_2(t) \operatorname{sign}(e_2(t)) \geq 0.$$

Substituting $u(t)$ in $\mathbf{f}_e$ leads to

$$\dot{e}_2(t) = \sin\big(e_1(t)\big) + k_3 w_{max} \operatorname{sign}(e_2(t)),$$

that implies $\{[n\pi, 0]^T : n \in \{0\} \cup \mathbb{N}\}$ is the set of equilibrium points for the error dynamics endowed with control function u. Since for $w_{max} \in (0, 1)$, $Y_N \cup \{[n\pi, 0]^T : n \in \{0\} \cup \mathbb{N}\} = \emptyset$, using u we can constantly increase $E_\Sigma(\mathbf{e}(t))$. Moreover, for any $j \in \{0\} \cup \mathbb{N}$,

$$H_{\Sigma_L}(\mathbf{s}_j^1) - H_{\Sigma_L}(\mathbf{s}_j^2) = w_{max}(2\alpha - \pi) + 2\sqrt{1 - w_{max}^2}.$$

Therefore, we can let t_f be the time such that

$$\int_0^{t_f} k_3 w_{max} e_2(t) \operatorname{sign}(e_2(t)) = w_{max}(2\alpha - \pi) + 2\sqrt{1 - w_{max}^2}.$$

Finally, since for every $\mathbf{e}(0) \in Q_L^j$, $-\mathbf{e}(0) \in Q_R^j$, we can extend the same conclusion to the trajectories with $\mathbf{e}(0) \in Q_R^j$.

□

Figure 2.3 illustrates Γ, Q_L^j, Q_R^j, S_L, S_R, level sets of H_{Σ_L} and H_{Σ_R} and trajectories in Γ for $w_{max} = 0.50$, 0.75, 1.00 and 1.25 with $\mathbf{k} = [2, 1]^T$. Note that as w_{max} increases to 1, points in S_L^1 and S_R^1 converge to points in S_L^2 and S_R^2, respectively. Consequently, the area of Q_L^j and Q_R^j decreases as $w_{max} \to 1$ from below. For $w_{max} = 1$, $Q_L^j = \{-\pi(2j + \frac{1}{2})\}$ and $Q_R^j = \{\pi(2j + \frac{1}{2})\}$. Moreover Q_L^j and Q_R^j cease to exist for $w_{max} > 1$, indicating that the problem is no longer conditionally controllable.

2.3 Overview

As the presented discussion suggests, finding an exact definition of Y_N may be a tedious task. In practice, such regions could be explored through numerous simulations (or experiments) of the system with different initial conditions. However, the nature of conditionally controllable problems provides the possibility of simplifying nonlinear control synthesis into two subtasks of (i) finding a linear controller for the linearized model using the well known techniques of linear control design and (ii) developing a nonlinear control law that can lead the system trajectories from Y_N to Y_L. Although such decomposition is achievable with various techniques, in this study, we will focus our attention on two methods of Planning-based and Optimal control.

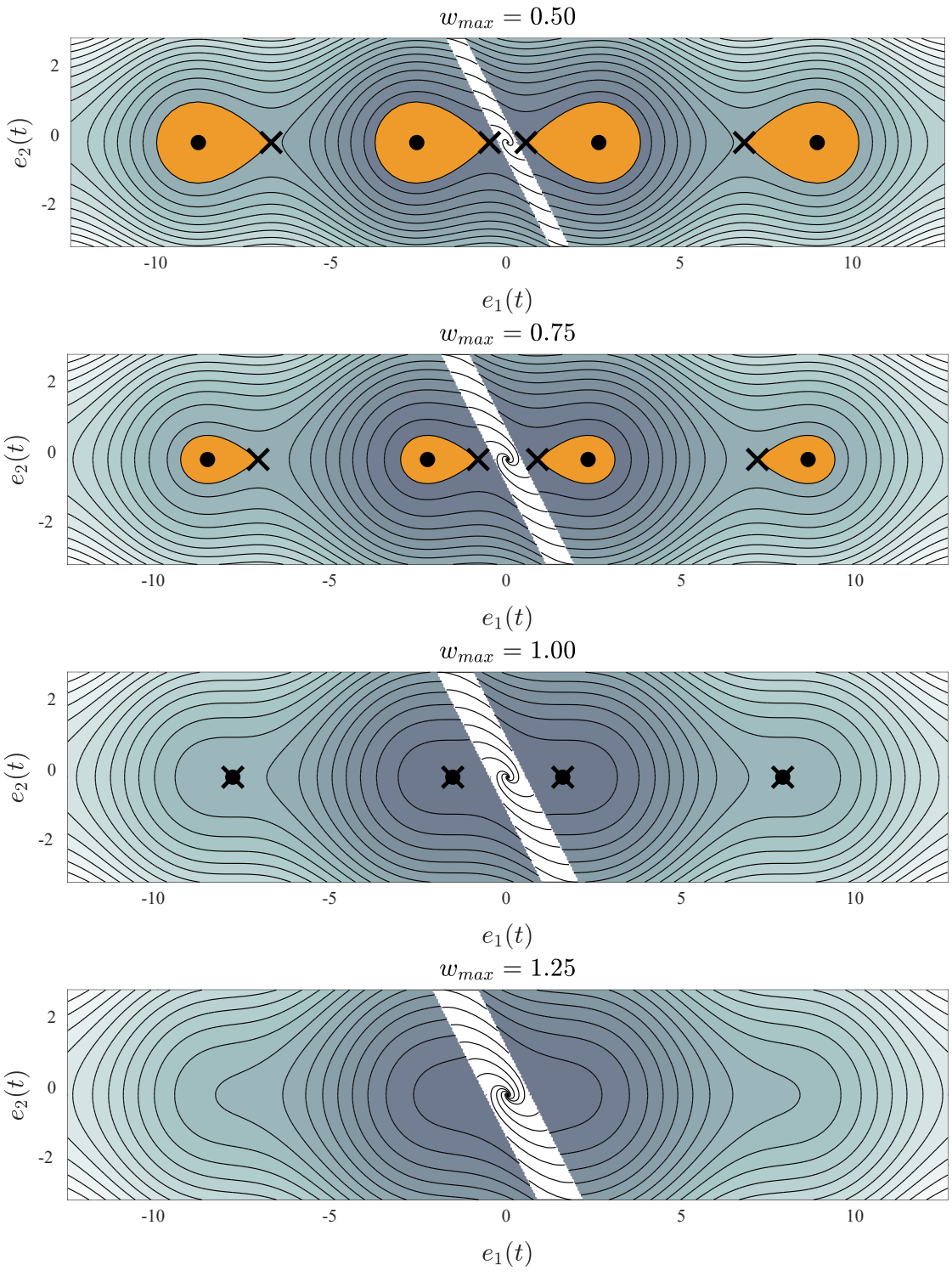

Figure 2.3: Phase portrait of $\mathbf{f}_e$ for $w_{max} \in \{0.50, 0.75, 1.00, 1.25\}$ with $\mathbf{k} = [2,\ 1]^T$. Points in $S_L^1 \cup S_R^1$ and $S_L^2 \cup S_R^2$ are depicted with ($\times$) and (o) markers, respectively. $Q_L^j \cup Q_R^j$ regions are highlighted with orange color and the level sets of $\chi_{\Sigma_L} H_{\Sigma_L} + \chi_{\Sigma_R} H_{\Sigma_R}$ are illustrated with color gradient changing from gray (low) to white (high).

Chapter 3

Planning-based control

As an alternative to classical approaches of designing control functions, we can formulate synthesizing a controller as solution of a motion-planning problem. Two common approaches in the literature that follow this technique are: Linear Quadratic Regulators (LQR)[1] trees [10, 11][2] and randomized kinodynamic planning [12].

LQR trees, and similar approaches, are used in controlling of a fixed-wing glider to perform a bird-like perching maneuvers [13, 14], stabilizing a torque limited double pendulum in the upward configuration [15], and demonstrating aggressive maneuvers with quadrotors [16]. As an alternative, randomized kinodynamic planning approaches [12] construct a Rapidly-exploring Random Tree (RRT) [17] in the state space of the system to find feasible trajectories connecting an initial state to a given

[1] An LQR control function is a linear function $\mathbf{v}(t) := \mathbf{K}\mathbf{e}(t)$ that minimizes the cost functional $J(\mathbf{e}, \mathbf{v}) := \int_t \mathbf{e}^T(t)\mathbf{Q}\mathbf{e}(t) + \mathbf{v}^T(t)\mathbf{R}\mathbf{v}(t)\,dt$ for a system with linear dynamics of the form $\dot{\mathbf{e}}(t) = \mathbf{A}\mathbf{e}(t) + \mathbf{B}\mathbf{v}(t)$. Positive definite matrices $\mathbf{Q}$ and $\mathbf{R}$ are used to define the gains on state and input vectors, respectively. Although some of the preliminaries of optimal control theory is covered in Chapter 4. an interested reader may refer to [9] for detailed derivation of LQR controllers.

[2] It must be noted that, while both methods discussed in [10] and [11] share similar motivations, there are a few major differences associated with them. In [10], the control function synthesis problem is formulated as finding an optimal switching sequence of control gains for a discrete time linear systems with respect to a quadratic cost function; while [11] discusses design of an acyclic connected graph of LQR controllers that are constructed by considering corresponding controllability regions in such a way that the union of controllable regions of the nodes covers a desired subset of the state space.

goal set.

In what follows, we start with discussing a set of tools needed to solve a planning problem in a normed vector space and extend their application to synthesizing control function. In particular, we present a generalization of exploring trees in a normed vector space and their extension to include time. In the presented developments, we have considered paths and trajectories in the set of continuous and absolutely continues functions, respectively. Such consideration allows direct utilization of the proposed setting in solving path planning and control problems. Lastly, based on the presented materials, we propose an algorithm to synthesize control functions for conditionally controllable systems and test its effectiveness in three different case studies.

3.1 Exploring trees

We start with a definition for paths connecting two points in a vector space.

Definition 3.1.1 (Path)**.** Let $(X, \|\cdot\|)$ be a normed vector space, $\mathbf{x}_1, \mathbf{x}_2 \in X$ and $a, b \in \mathbb{R}$ such that $a < b$. A *path* connecting $\mathbf{x}_1$ to $\mathbf{x}_2$ is a function $\boldsymbol{\phi} \in C(\mathbb{R})$, $\boldsymbol{\phi} : \mathbb{R} \to X$ such that $\boldsymbol{\phi}(a) = \mathbf{x}_1$ and $\boldsymbol{\phi}(b) = \mathbf{x}_2$. We will also use $\boldsymbol{\phi}_{\mathbf{x}_1 \to \mathbf{x}_2}$ or more concisely $\boldsymbol{\phi}_{1\to 2}$ as alternative notations when it is required to emphasize the points that are connected via path $\boldsymbol{\phi}$.

In the remaining sections of this chapter, and without loss of generality, we assume that every path starts at its source when the input is 0 and reaches to its destination when the input is 1; that is, we substitute $a = 0$ with and $b = 1$ in Definition 3.1.1.

Lemma 3.1.1 (Path compositions)**.** *Let $(X, \|\cdot\|)$ be a normed vector space and $\mathbf{x}_1, \ldots, \mathbf{x}_p \in X$, $p < \infty$. Let $g_k(s) := (p-1)s - k + 1$ and $\boldsymbol{\phi}_k$ be a path connecting*

$\mathbf{x}_k$ *to* $\mathbf{x}_{k+1}$. *Then* $\boldsymbol{\phi}_{1\to p} : \mathbb{R} \to X$ *defined as*

$$\boldsymbol{\phi}_{1\to p}(s) := \sum_{k=1}^{p-1} \chi_{I_k}(s)\boldsymbol{\phi}_k(g_k(s)),$$

where $I_1 = (-\infty, \frac{1}{p-1})$, $I_{p-1} = [1 - \frac{1}{p-1}, \infty)$ *and* $I_k = \frac{1}{p-1}[k-1, k)$ *for* $1 < k < p-1$, *is a path between* $\mathbf{x}_1$ *and* $\mathbf{x}_p$.

Proof. Since $0 \in I_1$, $1 \in I_{p-1}$ and $\bigcup_{k=1}^{p-1} I_k = \emptyset$, then $\boldsymbol{\phi}_{1\to p}(0) = \boldsymbol{\phi}_1(0) = \mathbf{x}_1$ and $\boldsymbol{\phi}_{1\to p}(1) = \boldsymbol{\phi}_{p-1}(1) = \mathbf{x}_p$. Thus, $\boldsymbol{\phi}_{1\to p}$ satisfies the required boundary conditions. Now we need to show $\boldsymbol{\phi}_{1\to p}$ is in $C(\mathbb{R})$. Let $J := \{\frac{k}{p-1} : 1 \leq k < p-1\}$, for every $s \in L := \bigcup_{k=1}^{p-1} I_k \setminus J$, $\boldsymbol{\phi}_{1\to p}(s) = \boldsymbol{\phi}_k(g_k(s)) + \mathbf{0}$. Since g_k and $\boldsymbol{\phi}_k$ are both continuous, then the composition $\boldsymbol{\phi}_k \circ g_k$ is also continuous that implies continuity of $\boldsymbol{\phi}_{1\to p}(s)$ for every $s \in L$. Now, let $s \in J$, then $s = \frac{k}{p-1}$ for some $1 \leq k < p-1$. Thus

$$\boldsymbol{\phi}_{1\to p}(s) = \boldsymbol{\phi}_k(g_k(s)) = \boldsymbol{\phi}_k(0) = \mathbf{x}_k,$$

On the other hand $\boldsymbol{\phi}_{k-1}(g_{k-1}(s)) = \boldsymbol{\phi}_{k-1}(1) = \mathbf{x}_k$. Let $\epsilon > 0$, since $\boldsymbol{\phi}_{k-1} \circ g_{k-1}$ and $\boldsymbol{\phi}_k \circ g_k$ are both continuous, then there exists $\delta_{k-1} > 0$ and $\delta_k > 0$ such that

$$\|\boldsymbol{\phi}_{k-1}(g_{k-1}(s)) - \boldsymbol{\phi}_{k-1}(g_{k-1}(t))\| = \|\mathbf{x}_k - \boldsymbol{\phi}_{k-1}(g_{k-1}(t))\| < \epsilon \text{ if } |s-t| < \delta_{k-1},$$
$$\|\boldsymbol{\phi}_k(g_k(s)) - \boldsymbol{\phi}_k(g_k(t))\| = \|\mathbf{x}_k - \boldsymbol{\phi}_k(g_k(t))\| < \epsilon \text{ if } |s-t| < \delta_k.$$

Let $\delta := \min\{\delta_{k-1}, \delta_k\}$, then for every t such that $|s-t| < \delta$, $\|\boldsymbol{\phi}_{1\to m}(s) - \boldsymbol{\phi}_{1\to m}(t)\| < \epsilon$ that implies continuity of $\boldsymbol{\phi}_{1\to m}$ for every $s = J$. Since $J \cup L = R$, thus $\boldsymbol{\phi}_{1\to p} \in C(\mathbb{R})$. □

Given the definition of a path and a tool to compose paths, we can proceed with a definition of spatial directed rooted trees, which will form the foundation for the

definition of spatial exploring trees.

Definition 3.1.2 (Spatial directed rooted tree)**.** Let $(X, \|\cdot\|)$ be a normed vector space and V be a countable subset of X. Moreover, let

$$E := \Big\{ \boldsymbol{\phi}_{\mathbf{x}_1 \to \mathbf{x}_2} : \mathbf{x}_1,\, \mathbf{x}_2 \in V,\, \mathbf{x}_1 \neq \mathbf{x}_2 \Big\},$$

be a set of paths between the points in V. Given $\mathbf{r} \in V$, the triple $(V, E, \mathbf{r})$ is *spatial directed rooted tree* with *root* $\mathbf{r}$, if for every $\mathbf{x} \in V$, one can construct a unique path

$$\boldsymbol{\phi}_{\mathbf{r} \to \mathbf{x}}(s) := \sum_k \chi_{I_k}(s) \boldsymbol{\phi}_k(g_k(s)),$$

as established in Lemma 3.1.1 such that $\boldsymbol{\phi}_k \in E$ for every k. If such condition holds, it is easy to verify[3] that $|E| < |V|$.

Definition 3.1.3 (Spatial exploring tree)**.** Let $(X, \|\cdot\|)$ be a normed vector space, $Y \subset X$, $\mathbf{r} \in Y$ and $N \in \mathbb{N}$. Then, a *spatial exploring tree* of size N, donated by Ψ_N, is a spatial directed rooted tree (V, E, r) such that $|V| = N$, which is incrementally constructed based on the following algorithm:

[3]Assume by contradiction that $|E| \geq |V|$. Then there exists at least one pair $\mathbf{x}_1 \neq \mathbf{x}_2$ in V and two paths $\boldsymbol{\phi}_a \neq \boldsymbol{\phi}_b$ in E such that $\boldsymbol{\phi}_a(0) = \boldsymbol{\phi}_b(0) = \mathbf{x}_1$ and $\boldsymbol{\phi}_a(1) = \boldsymbol{\phi}_b(1) = \mathbf{x}_2$. Consequently, one can construct two not equal paths from $\mathbf{r}$ to $\mathbf{x}_2$ by composing $\boldsymbol{\phi}_{\mathbf{r} \to \mathbf{x}_1}$ with $\boldsymbol{\phi}_a$ or $\boldsymbol{\phi}_b$, that leads to a contradiction with uniqueness of $\boldsymbol{\phi}_{\mathbf{r} \to \mathbf{x}}$ for every $\mathbf{x} \in V$.

Data: $\mathbf{r}$, Y

Result: Ψ_N

$V := \{\mathbf{r}\}$, $E := \emptyset$;

while $|V| < N$ **do**

  $(\mathbf{x}, \boldsymbol{\phi}_{\mathbf{v}\to\mathbf{x}}) := \text{Expand}(V, Y)$;

  $V := V \cup \{\mathbf{x}\}$, $E := E \cup \{\boldsymbol{\phi}_{\mathbf{v}\to\mathbf{x}}\}$;

end

$\Psi_N := (V, E, r)$;

where $\text{Expand}(V, Y)$ is a sub-algorithm that returns a new vertex $\mathbf{x} \in Y$ and a path from $\mathbf{v} \in V$ to $\mathbf{x}$. The set Y represents a given workable[4] subset of X.

The following method illustrates how a planning problem could be solved by utilizing spatial exploring trees.

Method 3.1.1. Let $Y \subset (X, \|\cdot\|)$, $\mathbf{r} \in Y$, $\Psi = (V, E, \mathbf{r})$ and $N \in \mathbb{N}$ be a maximum given size for Ψ. A path from the source $\mathbf{r}$ to δ neighborhood of a destination $\mathbf{g} \in Y$ could be constructed by iteratively exploring Y via the following algorithm.

Data: $\mathbf{g}$, Ψ, N

Result: $\boldsymbol{\phi}$

if $\exists \mathbf{v} \in V$ *such that* $\|\mathbf{g} - \mathbf{v}\| \leq \delta$ **then**

  return $\boldsymbol{\phi}_{\mathbf{o}\to\mathbf{v}}$;

else if $|V| < N$ **then**

  $(\mathbf{x}, \boldsymbol{\phi}_{\mathbf{v}\to\mathbf{x}}) := \text{Expand}(V, Y)$;

  $V := V \cup \{\mathbf{x}\}$, $E := E \cup \{\boldsymbol{\phi}_{\mathbf{v}\to\mathbf{x}}\}$;

else

  return $\emptyset$;

end

[4]The set workable subset is also refereed to as the *free* space, denoted as $\mathcal{C}_{free}$, in some planning related literature such as [18] and [19].

where $0 < \delta \in \mathbb{R}$ is a predefined acceptable distance from the goal point. Note that, if $|V|$ exceeds N and there is no vertex in V within a ball of radius δ around $\mathbf{g}$, then the solver will return $\boldsymbol{\phi} = \emptyset$ to indicate incapability of finding a path from $\mathbf{r}$ to $\mathbf{g}$ for the defined Y, N and δ values.

Depending on details of Expand(V, Y) in Definition 3.1.3, we can construct different spatial search trees. As an example, we proceed with a definition for rapidly-exploring random trees, which are commonly abbreviated as RRT.

Definition 3.1.4 (Rapidly-exploring random tree)**.** Let D be a probability density function for $Y \subset (X, \|\cdot\|)$ and let $\mathbf{x} \in Y \sim D$ denote choosing a random point in Y based on D. A *rapidly expanding random tree* is a spatial exploring tree that is constructed based on the Expand(V, Y) algorithm defined as

Data: V, Y

Result: $(\mathbf{x}, \boldsymbol{\phi})$

$\mathbf{x} \in Y \sim D$;

$\mathbf{v} := \arg\min_{\mathbf{w} \in V} \|\mathbf{x} - \mathbf{w}\|$;

if $\exists \boldsymbol{\phi} \in C$ *such that* $\boldsymbol{\phi}(0) = \mathbf{v}$, $\boldsymbol{\phi}(1) = \mathbf{x}$, *and* $\boldsymbol{\phi}(s) \in Y$, $\forall s \in \mathbb{R}$ **then**

  return $(\mathbf{x}, \boldsymbol{\phi})$;

else

  return $(\emptyset, \emptyset)$;

end

In an alternative implementation of rapidly-exploring random trees, instead of discarding paths for which $\boldsymbol{\phi}([0,1]) \cap Y^c \neq \emptyset$, we pick a new vertex $\widehat{\mathbf{x}}$ by finding $t^* = \min\{t \in [0,1] : \boldsymbol{\phi}(t) \notin \mathring{Y}\}$[5] and defining $\widehat{\mathbf{x}} := \boldsymbol{\phi}(t^*)$, and $\widehat{\boldsymbol{\phi}}(s) := \boldsymbol{\phi}(\frac{s}{t^*})$. if $t^* \neq 0$, we return $\widehat{\mathbf{x}}$ and $\widehat{\boldsymbol{\phi}}$ as the new vertex and path.

[5]Let $\mathbf{v} = \boldsymbol{\phi}(0) \in \mathring{Y}$, then $\exists \epsilon > 0$ such that $\{\mathbf{x} \in X : \|\mathbf{x} - \mathbf{v}\| < \epsilon\} \subset \mathring{Y}$. Also, since $\boldsymbol{\phi}$ is continuous, $\exists \delta > 0$ such that $\|\boldsymbol{\phi}(\delta) - \boldsymbol{\phi}(0)\| = \|\boldsymbol{\phi}(\delta) - \mathbf{v}\| < \epsilon \implies \boldsymbol{\phi}(\delta) \in \mathring{Y}$. Thus, for every $t \in Q := \{s \in (0,1) : \boldsymbol{\phi}(s) \in \mathring{Y}\}$, $\exists \delta_t > 0$ such that $\{s \in (0,1) : |t - s| < \delta_t\} \in Q$ that implies Q is open and consequently its complement with respect to the interval $[0,1]$, that is

In order to solve a control problem using rapidly-exploring random trees, we need to employ a strategy to synthesize a path between every two points in the feasible subset of the state space. Note that, developing such strategy is equivalent to finding a solution set $\{W_{\boldsymbol{\eta}_d}\}$ for the set of control problems $\{(\mathbf{f}, U, Y, \boldsymbol{\eta}_d, \widehat{Y}, T) : \forall \boldsymbol{\eta}_d \in Y\}$ that is clearly a complicated and computationally expensive task, specifically for systems with convoluted dynamics; However, some possible workarounds discussed in the literature are:

i) Using an iterative algorithm, such as shooting method, dynamic programming or a method derived from techniques in optimal control theory, to synthesize a control function that generates a trajectory connecting two vertices of the spatial exploring tree. Examples of such techniques are explored in [20] and [21]. In general, due to their iterative nature, application of such algorithms are laborious.

ii) Decomposing vertex-to-vertex transition problem into two subproblems of (I) finding a path between two vertices of the tree based on system kinematics (by computing $\dot{\mathbf{y}}(t)$ as time derivative of the path $\mathbf{y}(t)$, which ignores $\dot{\mathbf{y}}(t) = \mathbf{f}(t, \mathbf{y}(t), \mathbf{u}(t))$ as a dynamic constraint); and (II) using a control function to follow the obtained path using the linear approximation of the dynamics. Assuming that the initial point of the path coincides with the initial state of the system, which is a consequence of solving the kinematics based planning, then $\mathbf{e}(0) = \mathbf{0}$, that implies possibility of using a linear controller. However, if the problem is conditionally controllable and the source vertex belongs to Y_N, then this controller will fail to follow the given path. Accordingly, an implementation of the method needs to handle possibilities of diverging from the desired

$Q^c = \{s \in [0,1] : \boldsymbol{\phi}(s) \notin \mathring{Y}\}$, is a close set. Since Q^c is a bounded, close and not empty (based on the initial assumption that $\boldsymbol{\phi}([0,1]) \cap Y^c \neq \emptyset$), it has a minimum. If $\mathbf{v} \in Y \setminus \mathring{Y}$, then $\mathbf{v} \in \overline{Y}$ (Y is a close set) and $\mathbf{v} = \Phi(0) \notin \mathring{Y}$, thus $t^* = 0$.

path. An example of such implementation of rapidly-exploring random trees is presented in [22].

As the above discussions suggest, a more natural way of employing trees to solve a control problem is to include time and dynamic constraints in the growth of the spatial trees. In this regard, we introduce dynamic-based expanding trees as a special subset of spatio-temporal exploring trees, as presented in the following section.

3.2 Spatio-temporal exploring trees

To provide a definition for spatio-temporal exploring trees, we need to include time in the definition of spatial directed rooted tree as presented in Section 3.1. Thus, we start with the following Lemma

Lemma 3.2.1 (Trajectory compositions)**.** *Let* $(X, \|\cdot\|)$ *be a normed vector space,* $\mathbf{v}_k = (t_k, \mathbf{x}_k) \in \mathbb{R} \times X$ *for* $k = 1, \ldots, p < \infty \in \mathbb{N}$ *such that* $t_k < t_{k+1}$ *for all* k*, and* $\boldsymbol{\varphi}_k : \mathbb{R} \to X$ *be a trajectory connecting* $\mathbf{v}_k$ *to* $\mathbf{v}_{k+1}$*, that is,* $\boldsymbol{\varphi}_k \in AC(\mathbb{R})$*,* $\boldsymbol{\varphi}_k(t_k) = \mathbf{x}_k$ *and* $\boldsymbol{\varphi}_k(t_{k+1}) = \mathbf{x}_{k+1}$ *for every* $k = 1, \ldots, p-1$*. Then*

$$\boldsymbol{\varphi}_{1\to p}(t) := \mathbf{x}_1 + \int_{t_1}^{t} \Big(\sum_{k=1}^{p-1} \chi_{I_k}(\tau) \boldsymbol{\varphi}_k'(\tau) \Big) d\tau,$$

where $I_k = [t_k, t_{k+1}]$ *for* $1 \leq k < p-1$ *and* $I_{p-1} = [t_{k-1}, \infty)$*, is a trajectory*[6] *that connects* $\mathbf{v}_1 = (t_1, \mathbf{x}_1)$ *to* $\mathbf{v}_p = (t_p, \mathbf{x}_p)$*.*

Proof. Based on the assumptions of the lemma, $\boldsymbol{\varphi}_k \in AC(\mathbb{R})$, $\boldsymbol{\varphi}_k(t_k) = \mathbf{x}_k$ and $\boldsymbol{\varphi}_k(t_{k+1}) = \mathbf{x}_{k+1}$ for every $k = 1, \ldots, p-1$. Thus for every $k \in \{1, \ldots, p-1\}$, there

[6]Note that, since we are choosing $\boldsymbol{\varphi}_k$ to be an absolutely continuous function, there is no contradiction with the given definition of a trajectory as in Definition 2.1.1.

exists a Lebesgue integrable derivative of $\boldsymbol{\varphi}_k$, namely $\boldsymbol{\varphi}'_k$, such that

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \int_{t_k}^{t_{k+1}} \boldsymbol{\varphi}'_k(\tau)d\tau.$$

Moreover, if $\boldsymbol{\varphi}'_k$ is integrable, then $\chi_{I_k}(\tau)\boldsymbol{\varphi}'_k(\tau)$ is also integrable. Finally, since a finite sum of Lebesgue integrable functions is also Lebesgue integrable, we get $\sum_{k=1}^{p-1} \chi_{I_k}(\tau)\boldsymbol{\varphi}'_k(\tau)$ is integrable that implies $\boldsymbol{\varphi}_{1\to p} \in AC(\mathbb{R})$.

To check for the boundary values of $\boldsymbol{\varphi}_{1\to p}$ at $t = t_1$ and t_p we can proceed as what follows. For $t = t_1$ we have

$$\boldsymbol{\varphi}_{1\to p}(t_1) = \mathbf{x}_1 + \int_{t_1}^{t_1} \Big(\sum_{k=1}^{p-1} \chi_{I_k}(\tau)\boldsymbol{\varphi}'_k(\tau)\Big)d\tau = \mathbf{x}_1,$$

and for $t = t_p$ we get

$$\begin{aligned}
\boldsymbol{\varphi}_{1\to p}(t_p) &= \mathbf{x}_1 + \int_{t_1}^{t_p} \Big(\sum_{k=1}^{p-1} \chi_{I_k}(\tau)\boldsymbol{\varphi}'_k(\tau)\Big)d\tau \\
&= \mathbf{x}_1 + \int_{t_1}^{t_p} \chi_{I_1}(\tau)\boldsymbol{\varphi}'_1(\tau)d\tau + \int_{t_1}^{t_p} \chi_{I_2}(\tau)\boldsymbol{\varphi}'_2(\tau)d\tau + \cdots + \\
&\int_{t_1}^{t_p} \chi_{I_{p-1}}(\tau)\boldsymbol{\varphi}'_{p-1}(\tau)d\tau \\
&= \underbrace{\underbrace{\mathbf{x}_1 + \underbrace{\int_{t_1}^{t_2} \boldsymbol{\varphi}'_1(\tau)d\tau}_{=\mathbf{x}_2}}_{=\mathbf{x}_3} + \int_{t_2}^{t_3} \boldsymbol{\varphi}'_2(\tau)d\tau + \cdots + \int_{t_{p-1}}^{t_p} \boldsymbol{\varphi}'_{p-1}(\tau)d\tau}_{=\mathbf{x}_p} = \mathbf{x}_p.
\end{aligned}$$

□

The result of Lemma 3.2.1 could be used to give a definition to spatio-temporal directed rooted trees and spatio-temporal exploring trees as in what follows.

Definition 3.2.1 (Spatio-temporal directed rooted tree)**.** Let $(X, \|\cdot\|)$ be a normed

vector space, $T = [t_0, t_f] \subset \mathbb{R}$ and $V \subset T \times X$ be a countable set. Moreover, let

$$E := \Big\{ \boldsymbol{\varphi}_{\mathbf{v}_1 \to \mathbf{v}_2} : \mathbf{v}_1 = (t_1, \mathbf{x}_1) \in V \text{ and } \mathbf{v}_2 = (t_2, \mathbf{x}_2) \in V,\, t_1 < t_2 \Big\},$$

be a set of trajectories between the points in V. Given $\mathbf{r} = (t_r, \mathbf{x}_r) \in V$, the triple $(V, E, \mathbf{r})$ is *spatio-temporal directed rooted tree* with *root* $\mathbf{r}$, if for every $\mathbf{v} \in V$, one can construct a unique trajectory

$$\boldsymbol{\varphi}_{\mathbf{r} \to \mathbf{v}}(t) := \mathbf{x}_r + \int_{t_r}^{t} \Big(\sum_k \chi_{I_k}(\tau) \boldsymbol{\varphi}'_k(\tau) \Big) d\tau,$$

as established in Lemma 3.2.1 such that $\boldsymbol{\varphi}_k \in E$ for every k. If such condition holds, it is easy to verify[7] that $|E| < |V|$.

Definition 3.2.2 (Spatio-temporal exploring tree)**.** Let $T = [t_0, t_f] \subset \mathbb{R}$ and $(X, \|\cdot\|)$ be a normed vector space. Moreover, let $Y \subset X$ be a given workable set, $\mathbf{r} = (t_r, \mathbf{x}_r) \in T \times Y$ be an assigned root and $N \in \mathbb{N}$ be a desired size. Then, a *Spatio-temporal exploring tree* of size N, donated by Υ_N, is a Spatio-temporal directed rooted tree $(V, E, \mathbf{r})$ with $|V| = N$ that is incrementally constructed based on the following algorithm:

$V := \{\mathbf{r}\}$, $E := \emptyset$;
while $|V| < N$ **do**
  $(\mathbf{w}, \boldsymbol{\varphi}_{\mathbf{v} \to \mathbf{w}}) := \text{Expand}(V, T \times Y)$;
  $V := V \cup \{\mathbf{w}\}$, $E := E \cup \{\boldsymbol{\varphi}_{\mathbf{v} \to \mathbf{w}}\}$;
end
return $\Upsilon_N := (V, E, \mathbf{r})$;

where $\text{Expand}(V, T \times Y)$ is a sub-algorithm that returns $\mathbf{w} \in T \times Y$ as a new vertex

[7]Similar to the discussion for Definition 3.1.2 we can show that uniqueness of trajectories between the root and tree vertices implies $|E| \geq |V|$.

and $\boldsymbol{\varphi}_{\mathbf{v}\to\mathbf{w}}$ as a trajectory connecting $\mathbf{v} \in V$ to $\mathbf{w}$.

In the following definition, we present an application of Spatio-temporal exploring trees as a tool to explore the state space of dynamic systems.

Definition 3.2.3 (Dynamics-based expanding tree)**.** Given a control system $\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y}(t), \mathbf{u}(t))$, as defined in Definition 2.1.1, Y as a connected subset of $\mathbb{R}^n$ with the usual topology, U as the set of feasible control functions, $T = [t_0, t_f] \subset \mathbb{R}$ and $\mathbf{y}_0 = \mathbf{y}(0) \in Y$. A *dynamics-based expanding tree* is a spatio-temporal exploring tree that is constructed based on the following Expand(V, Y) algorithm

$\mathbf{v}_0 = (t_0, \mathbf{y}_0) :=$ PickVertex(V);

$\delta t :=$ PickTime$(\mathbb{R}^+)$;

$\mathbf{u} :=$ PickControl(U);

$\boldsymbol{\varphi}(t) := \mathbf{y}_0 + \int_{t_0}^{t} \mathbf{f}(\tau, \mathbf{y}(\tau), \mathbf{u}(\tau))d\tau$;

if $\boldsymbol{\varphi}(t) \in Y$, *for all* $t \in [t_0, t_0 + \delta t]$ **then**

  return $\big((t_0 + \delta t, \boldsymbol{\varphi}(t_0 + \delta t)), \boldsymbol{\varphi}\big)$;

else

  return $(\emptyset, \emptyset)$;

end

where PickVertex, PickTime and PickControl are sub-algorithms to choose a vertex $\mathbf{v}$ from V, a positive time interval δt and a control function $\mathbf{u}$ from U, respectively. Depending on details of these sub-algorithms, we can tailor the procedure to generate trees suitable for different given problems.

In [23] we have demonstrated that using a set of dynamics-based expanding trees, $\{\Upsilon_n^j = (V_j, E_j, \mathbf{r}_j) : j \in \mathbb{N}, j = 1, \ldots, N < \infty\}$, we can solve different control problems with relatively complicated dynamics.

In what follows, we explore application of dynamics-based expanding tree in solving conditionally controllable problems.

3.3 Extension to conditionally controllable problems

Let $(\mathbf{f}, U, Y, \mathbf{y}_d, \widehat{Y}, t_0)$ be a conditionally controllable problem, then for every $\boldsymbol{\eta}_0 \in Y \setminus Y_N$, $\exists \mathbf{u} \in U_L$ such that $\lim_{t\to+\infty} \|\mathbf{y}_d(t) - \mathbf{y}_{\mathbf{u}}(t)\| = 0$. Let $t_1 \geq t_0$, if the cost of solving the initial value problem $\dot{\mathbf{y}} = \mathbf{f}(t, \mathbf{y}(t), \mathbf{u}(t))$; $\mathbf{y}(t_1) = \boldsymbol{\eta} \in Y$ is small enough, then we can check if the performance of a control function $\mathbf{u} \in U_L$ is satisfactory. This possibility of evaluating the control functions allows utilization of Ariadne's clew framework [24] where the planning problem is decomposed into two contemporaneous subtasks of Explore and Search that are:

i) Explore: build and enhance a representation of the accessible set;

ii) Search: check for possibility of reaching the target from accessible set based on a predefined criteria.

In the case of conditionally controllable problems, if the Explore subtask can identify a region intersecting with $Y \setminus Y_N$, then the Search subtask can utilize a function in U_L to guide the system trajectory toward $\mathbf{y}_d(t)$.

On a different note, complexities of the computations involved in synthesizing a planning-based controller demands utilization of a digital computer to evaluate and apply the control function. Thus, the resultant output will be a discrete signal that could be represented as a continuous input via zero-order hold[8]. Accordingly, the control signal attains a constant value during each sampling time, which needs to

[8]Let $\mathbf{x}[n] : \mathbb{N} \to \mathbb{R}^n$ represent a discreet signal and $t_s \in \mathbb{R}$ be the sampling time (also known as the sample interval), then zero-order hold representation of $\mathbf{x}$ is $\mathbf{x}_{\text{zoh}} : \mathbb{R} \to \mathbb{R}^n$ defined as

$$\mathbf{x}_{\text{zoh}}(t) = \sum_{n\in\mathbb{N}} \mathbf{x}[n]\, \chi_{[t_s n,\, t_s(n+1))}(t).$$

be accounted for in design of the algorithm[9]. Based on the presented discussions, we propose the following algorithm to synthesize control functions for a class of conditionally controllable problems where the set of feasible controls is L^∞. The presented method utilizes spatio-temporal exploring trees to explore the state space of the associated error dynamics as the Explore subtask of Ariadne's clew framework.

Method 3.3.1. Given $T = [t_0, t_f]$, $t_s, t_h \in \mathbb{R}^+$, $0 < t_s << (t_f - t_0)$, $1 \leq n_s << (t_f - t_0)/ts \in \mathbb{N}$, $t_h >> t_s$ and an error regularization problem $R_e = (\mathbf{f}_e, V, E, \mathbf{0}, \widehat{E}, t_0)$ with $V = L^\infty$ that is conditionally controllable when range of control functions in V is constrained to $D := \prod_{i=1}^{m}[w_{i_{min}}, w_{i_{max}}] \subset \mathbb{R}^m$. Let C be a given finite subset of ∂D. If there exists a control function

$$\mathbf{u}(t) := \sum_{k=0}^{N} \chi_{[t_k, t_k+\delta_k]}(t)\mathbf{w}_k,$$

for which $t_{k+1} = t_k + \delta_k$, $\delta_k \in \{t_s, n_s t_s\}$ for all k, $N \leq (t_f - t_0)/t_s$ and

$$\mathbf{w}_k \in \begin{cases} D, & \text{if } \delta_k = t_s, \\ C, & \text{if } \delta_k = n_s t_s, \end{cases}$$

that can regulate the error from a given $\mathbf{e}(t_0) \in E$ to zero. Then, $\mathbf{u}(t)$ could be constructed by determining the values of δ_k and $\mathbf{w}_k$ through the following algorithm:

[9]The effect of zero-order hold must also be considered for the functions in U_L, if they are meant to be implemented via a digital computer. In the following discussions, for the sake of brevity, we assume that the sampling time is small enough to guarantee the stability of the linearized model with a linear control signal passing through a zero-order hold filter. We will address this effect in the discussions presented on Section 3.4

Data: t_k, $\mathbf{e}(t_k)$, and H

Result: δ_k, $\mathbf{w}_k$ and H

$\tau_0 := t_k$; $\boldsymbol{\xi}_0 := \mathbf{e}(t_k)$;

$P := \{(\tau_0, \boldsymbol{\xi}_0)\}$; $Q := \emptyset$; $i := 0$;

while $\exists t \in [t_f,\, t_f + t_h]$ *such that* $\|\boldsymbol{\varphi}(t, \mathbf{p}_i, \mathbf{v})\| > \epsilon$ **do**

  $(\bar{\mathbf{p}}, \mathbf{w}) :=$ PickVertexAndControl(P, C, H);

  if $\bar{\mathbf{p}} = \emptyset \vee i > N$ **then**

    return $\big(t_s,$ SampleControl$(D),\, \emptyset\big)$;

  else if $\boldsymbol{\varphi}(t, \bar{\mathbf{p}}, \mathbf{w}) \in E(t) \,\wedge\, \dot{\boldsymbol{\varphi}}(t, \bar{\mathbf{p}}, \mathbf{w}) \in \widehat{E}(t),\ \forall t \in [t_{\bar{\mathbf{p}}}, t_{\bar{\mathbf{p}}} + n_s t_s]$ **then**

    $\mathbf{p}_{i+1} := (t_{\bar{\mathbf{p}}} + n_s t_s,\ \boldsymbol{\varphi}(t_{\bar{\mathbf{p}}} + n_s t_s, \bar{\mathbf{p}}, \mathbf{w}))$;

    $P := P \cup \big\{\mathbf{p}_{i+1}\big\}$; $Q := Q \cup \{\mathbf{w}\}$;

    $i := i + 1$

  else

    Mask$(\bar{\mathbf{p}}, \mathbf{w})$;

  end

end

if $|P| = 1$ **then**

  return $\big(t_s,\, \mathbf{v}_{\boldsymbol{\xi}_i}(t_0),\, \emptyset\big)$;

else

  return $\big(n_s t_s, \mathbf{w}_{k\to 1},\, \mathbf{w}_{k\to i}\big)$;

end

where $\mathbf{v} := \mathrm{Proj}_D\, \mathbf{K}\mathbf{e}(t)$ is projected linear control function that regulates the linear approximation of the error dynamics at $\mathbf{e}(t) = \mathbf{0}$, H is the results of planning obtained in the previous step, that is $\mathbf{u}_{\mathbf{e}(t_{k-1})\to\mathbf{0}}$; $H = \emptyset$ indicates unavailability of previous planning solution. $\epsilon > 0$ is an acceptable distance from $\mathbf{0}$ that suggest convergence of $\mathbf{e}(t)$ to zero. $N \in \mathbb{N}$ is a given maximum size of the tree ($|P|$). For

$\mathbf{p} = (t_p, \boldsymbol{\xi}_p) \in \mathbb{R} \times \mathbb{R}^n$ and control function $\mathbf{u}$, $\boldsymbol{\varphi}(t, \mathbf{p}, \mathbf{u})$ is defined as:

$$\boldsymbol{\varphi}(t, \mathbf{p}, \mathbf{u}) := \boldsymbol{\xi}_{\mathbf{p}} + \int_{t_p}^{t} \mathbf{f}_e\big(\tau, \mathbf{e}(\tau), \mathbf{u}(\tau)\big)\, d\tau.$$

Sub-algorithm PickVertexAndControl returns a vertex $\bar{\mathbf{p}} = (t_{\bar{\mathbf{p}}}, \boldsymbol{\xi}_{\bar{\mathbf{p}}}) \in P$ such that $t_{\bar{\mathbf{p}}} \leq t_f - n_s t_s$ and a control $\mathbf{u}(t) \equiv \mathbf{w} \in C$ based on the current vertices in P and history H. PickVertexAndControl returns $\bar{\mathbf{p}} = \emptyset$ if all the vertex and input combinations are used (to eliminate exploring repeated vertex-control pairs). SampleControl(D) sub-algorithm returns $\mathbf{u}(t) \equiv \mathbf{w} \in D$ based on a defined probability density function. The Mask sub-algorithm masks specific vertex and control tuple so that PickVertexAndControl does not pick the same combination again.

The procedure presented in Method 3.3.1 constructs the control function $\mathbf{u}(t)$ gradually from t_0 to t_f. At each step, the algorithm attempts to find a control function $\mathbf{u}_k(t)$ from the current point $(t_k, \mathbf{e}(t_k))$ to a point in the goal region $\{\boldsymbol{\xi} \in \mathbb{R}^n : \|\boldsymbol{\xi}\| < \epsilon\}$ satisfying $\|\mathbf{e}_{\mathbf{u}_k}(t)\| \leq \epsilon$ for all $t \in [t_f, t_f + t_h]$. If the algorithm terminates successfully, it returns $\mathbf{w}_k = \mathbf{u}_k(t_k)$ and $\delta_k \in \{t_s, n_s t_s\}$ along with $H \equiv \mathbf{v}_k$ which could be used as a reference for the next step planning. However, if the process terminates unsuccessfully, which can happen by completely exhausting all the inputs in C for all the vertices in P satisfying $t_{\mathbf{p}} \leq t_f - n_s t_s$ or reaching the maximum allowable tree size, then it will return a random control input from D to alter the current state $\mathbf{e}(t_k)$ and repeat the procedure.

To explore performance of Method 3.3.1 in synthesizing control functions we proceed with the following case studies of a simple pendulum, a point-mass double pendulum and a point-mass cart-pole system. In all the given examples, the PickVertexAndControl sub-algorithm is set to alternate between history and current explored vertices until it exhausts history H.

3.3.1 Simple pendulum

As our first demonstration, we continue with the simplified point-mass pendulum example presented in Section 2.2 with

$$\mathbf{f}(t, \mathbf{y}(t), u(t)) = \begin{bmatrix} y_2(t) \\ \sin\big(y_1(t)\big) + u(t) \end{bmatrix}, \tag{3.1}$$

and the error regularization problem $R_e = (\mathbf{f}_e, L^\infty, \mathbb{R}^2, \mathbf{0}, \mathbb{R}^2, 0)$ with the error dynamics $\mathbf{f}_e = -\mathbf{f}(t, -\mathbf{e}(t), -\mathbf{v}(t))$. Let the interval $D := [-w_{max}, w_{max}]$, $V_L = \{v \in L^\infty : v(t) = \text{sat}_D\langle \mathbf{K}, \mathbf{e}(t)\rangle, \mathbf{K} \text{ and } \mathbf{e}(t) \in \mathbb{R}^2\}$. Based on Theorem 2.2.3, R_e is a conditionally controllable if $w_{max} \in (0, 1)$. To evaluate the performance of Method 3.3.1 to synthesize a control function for R_e, we set $w_{max} = 0.5$ and picked two initial conditions: (i) $\mathbf{e}(0) = [-2,\, 0]^T \in E_N$ and (ii) $\mathbf{e}(0) = [-4,\, 2]^T \in E \setminus E_N$. The corresponding synthesized control function using Method 3.3.1, denoted by u, and the corresponding response of the system to u are illustrated in Figure 3.1. In addition, Figure 3.1 includes graphs of saturated linear function $v(t) =: \text{sat}_D\langle \mathbf{K}, \mathbf{e}(t)\rangle$ with[10] $\mathbf{K} = [-10, -3]^T$, and corresponding response of the system to v. As seen in the figure, even when $\mathbf{e}(0) \in E_N$, the control function u can meet the objective of the regularization problem. However, v saturates at w_{max} and results in an undamped oscillatory response of the system. A more interesting behavior is observed for $\mathbf{e}(0) = [-4,\, 2]^T \in E \setminus E_N$. In this case, although $\mathbf{e}_v(t)$ eventually reduces to zero in time, the exploring nature of the proposed method allows finding a shortcut in that results in a faster convergence of $\mathbf{e}_u(t)$. Other parameters used in this case study are: $t_f = 25$, $T = 10$, $\epsilon = \pi/10$, $n_s = 10$ and $t_s = 0.1$.

[10]The negative sign in $\mathbf{K}$ appears by defining the error dynamics with $\mathbf{e}(t) := \mathbf{y}_d(t) - \mathbf{y}(t) = -\mathbf{y}(t)$ and $v(t) := u_d(t) - u(t)$. This mismatch in the sign of $\mathbf{K}$ with the analysis presented in Section 2.2 is due to a sign simplification that is applied in Section 2.2 by replacing $v(t)$ with $-u(t)$.

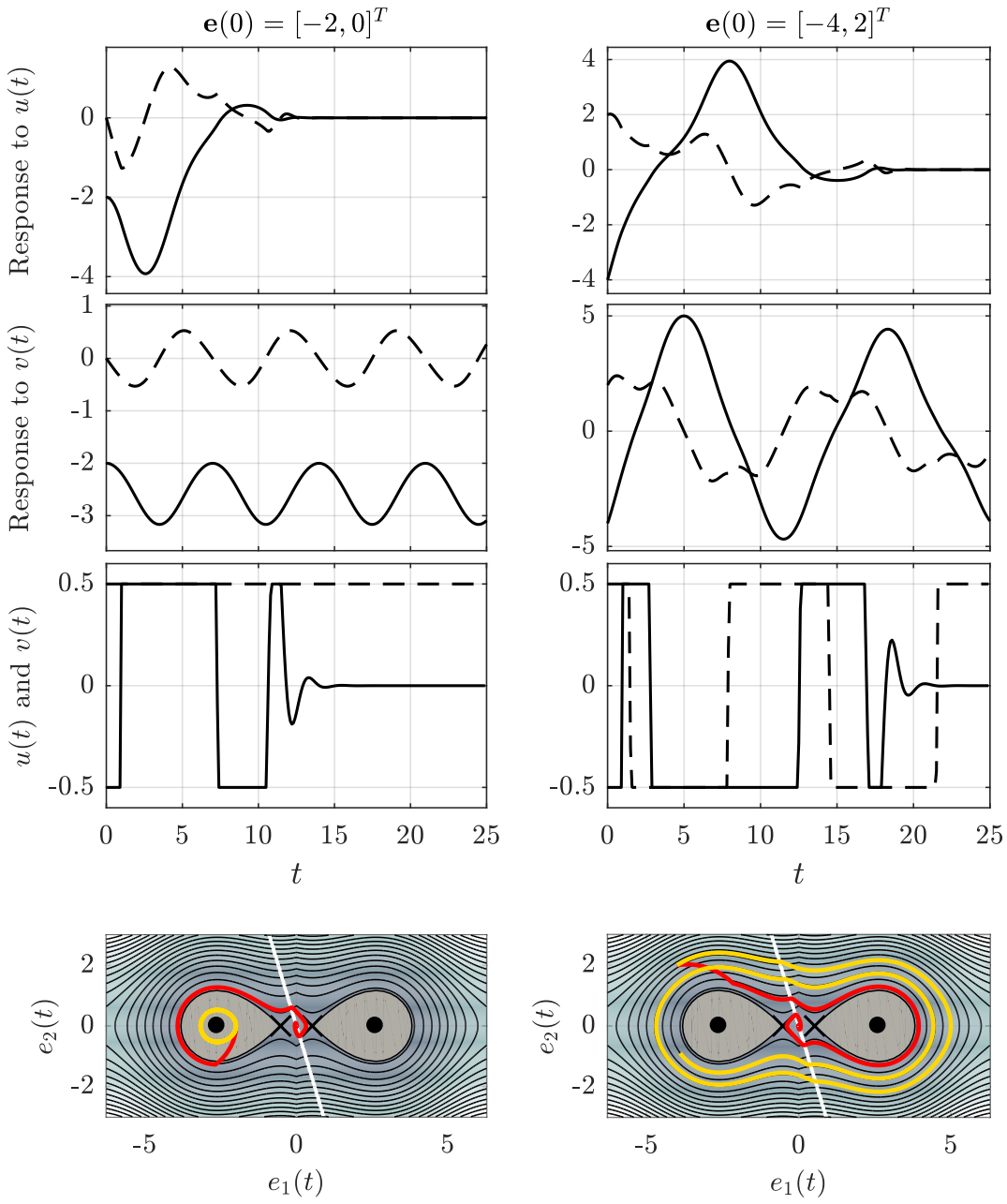

Figure 3.1: Response of the pendulum system to the synthesized control function based on Method 3.3.1, u and the projected linear control function $v(t) =: \text{Proj}_D\langle \mathbf{K}, \mathbf{e}(t)\rangle$. Solid and dashed lines in the first two rows indicate $e_1(t)$ and $e_2(t)$, respectively. In contrast, solid and dashed lines in the third row depict $u(t)$ and $v(t)$, respectively. The phase portrait of the system is depicted in the last row where red and yellow lines are used to illustrate system response to $u(t)$ and $v(t)$, respectively. The contour lines indicate Hamiltonian isoclines on Σ_L and Σ_R. The $\times$ and o markers are used to illustrate the singular points. Please refer to discussions in Section 2.2 and Figure 2.1 for more detailed description of the depicted phase planes.

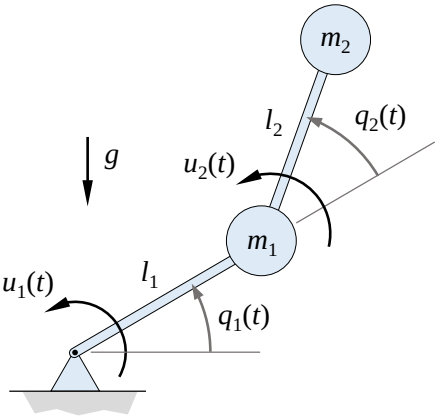

Figure 3.2: The point-mass double pendulum system and the corresponding parameters.

3.3.2 Double pendulum

As an example of a more complicated system, we can study performance of Method 3.3.1 in synthesizing control functions for a point-mass double pendulum system depicted in Figure 3.2. The detailed derivation of the differential equation of motion for the system with arbitrary values of m_1, m_2, l_1 and l_2 is presented in the Appendix B. In order to simplify the expressions, we take $m_1 = m_2 = l_1 = l_2 = g = 1$ (with appropriate units). Moreover, by defining $\mathbf{y}(t) := [q_1(t),\, q_2(t),\, \dot{q}_1(t),\, \dot{q}_2(t)]^T$ as the state vector and $\mathbf{u}(t) = [u_1(t),\, u_2(t)]^T$ as the input vector. we can obtain a first order representation for the system as

$$\dot{\mathbf{y}}(t) = \begin{bmatrix} y_3(t) \\ y_4(t) \\ \mathbf{M}^{-1}\big(\mathbf{y}(t)\big)\Big(\mathbf{u}(t) - \mathbf{\Phi}\big(\mathbf{y}(t)\big)\Big) \end{bmatrix}, \tag{3.2}$$

where

$$\mathbf{M}(\mathbf{y}(t)) = \begin{bmatrix} 2\cos(y_2(t)) + 3 & \cos(y_2(t)) + 1 \\ \cos(y_2(t)) + 1 & 1 \end{bmatrix}, \tag{3.3}$$

and

$$\mathbf{\Phi}(\mathbf{y}(t)) = \begin{bmatrix} \cos(y_1(t) + y_2(t)) + 2\cos(y_1(t)) - y_4(t)\big(2y_3(t) + y_4(t)\big)\sin(y_2(t)) \\ y_3^2(t)\sin(y_2(t)) + \cos(y_1(t) + y_2(t)) \end{bmatrix}. \tag{3.4}$$

Similar to the point-mass pendulum problem, we chose the objective of the control problem as swinging up and holding the system in the upward configuration. Accordingly, $\mathbf{y}_d(t) = [\pi/2,\, 0,\, 0,\, 0]^T \equiv \mathbf{y}_d$. Moreover, we assume additional constraints on the range of $q_1(t)$ and $q_2(t)$ and set $Y(t) \equiv Y := \{\boldsymbol{\eta} \in \mathbb{R}^4 : -2\pi \leq \eta_1 \leq 2\pi, -\pi \leq \eta_2 \leq \pi\}$. Since $\dot{\mathbf{y}}_d = \mathbf{0}$, we can set $\mathbf{u}_d$ to be the input required to make $\mathbf{y}_d$ an equilibrium point of (3.2). Thus,

$$\mathbf{u}_d - \underbrace{\mathbf{\Phi}(\mathbf{y}_d)}_{=\mathbf{0}} = \mathbf{0} \implies \mathbf{u}_d = \mathbf{0}. \tag{3.5}$$

Having defined the values of $\mathbf{y}_d$ and $\mathbf{u}_d$, we can proceed with defining the error dynamics. Following the formulation presented in (2.4), we get

$$\dot{\mathbf{e}}(t) = \begin{bmatrix} e_3(t) \\ e_4(t) \\ \mathbf{M}_e^{-1}\big(\mathbf{e}(t)\big)\Big(\mathbf{v}(t) + \mathbf{\Phi}_e\big(\mathbf{e}(t)\big)\Big) \end{bmatrix}, \tag{3.6}$$

where $\mathbf{e}(t) := \mathbf{y}_d - \mathbf{y}(t)$, $\mathbf{v}(t) := \mathbf{u}_d - \mathbf{u}(t)$,

$$\mathbf{M}_e(\mathbf{e}(t)) = \mathbf{M}(\mathbf{y}_d - \mathbf{e}(t)) = \begin{bmatrix} 2\cos(e_2(t)) + 3 & \cos(e_2(t)) + 1 \\ \cos(e_2(t)) + 1 & 1 \end{bmatrix}, \tag{3.7}$$

and

$$\begin{aligned}\boldsymbol{\Phi}_e(\mathbf{e}(t)) &= \boldsymbol{\Phi}(\mathbf{y}_d - \mathbf{e}(t)) \\ &= \begin{bmatrix} \sin(e_1(t)+e_2(t)) + 2\sin(e_1(t)) + e_4(t)\big(2e_3(t)+e_4(t)\big)\sin(e_2(t)) \\ \sin(e_1(t)+e_2(t)) - e_3^2(t)\sin(e_2(t)) \end{bmatrix}. \end{aligned} \tag{3.8}$$

Based on (2.6), we can obtain a linear approximation of (3.6) around $\mathbf{e}(t) = \mathbf{0}$ and $\mathbf{v}(t) = \mathbf{0}$ as

$$\dot{\mathbf{e}}(t) \approx \mathbf{A}\mathbf{e}(t) + \mathbf{B}\mathbf{v}(t), \tag{3.9}$$

where

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & -1 & 0 & 0 \\ -1 & 3 & 0 & 0 \end{bmatrix}, \text{ and } \mathbf{B} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & -2 \\ -2 & 5 \end{bmatrix}. \tag{3.10}$$

In order to find a gain matrix $\mathbf{K}$ for the linear control function $\mathbf{K}\mathbf{e}(t)$ that regulates the linear approximation of the system presented in (3.9), we follow the pole placement technique discussed in [6], which is known as Ackermann's formula. Based on this technique, the eigenvalues of the closed loop linear approximation of the system, that is $\dot{\mathbf{e}}(t) = \mathbf{A}_c\mathbf{e}(t)$ with $\mathbf{A}_c$ defined as

$$\mathbf{A}_c := \mathbf{A} + \mathbf{B}\mathbf{K}, \tag{3.11}$$

could be arbitrarily assigned as the roots of a desired characteristic polynomial, if the controllability matrix $\mathbf{C} := [\mathbf{B}, \mathbf{A}\mathbf{B}, \cdots, \mathbf{A}^{n-1}\mathbf{B}]$ is of rank n. For the double pendulum example rank controllability is 4 that allows us to proceed with Acker-

mann's formula. Choosing desired polynomial $\lambda(s) = (s+1)^2(s+2)^2$ leads to

$$\mathbf{K}^T = -[0,\, 0,\, \cdots,\, 1]\mathbf{C}^{-1}\lambda(\mathbf{A}) = -\begin{bmatrix} 13 & 5 & 15 & 6 \\ 5 & 3 & 6 & 3 \end{bmatrix}. \tag{3.12}$$

Finally, to construct the projected linear control function $\mathbf{v}(t) := \mathrm{Proj}_D \mathbf{Ke}(t)$, we pick $D = \{\mathbf{w} \in \mathbb{R}^2 : \|\mathbf{w}\|_\infty \leq 0.5\}$.

Figure 3.3 shows response of the system from two different initial conditions to both $\mathbf{v}$ and $\mathbf{u}$, which is the control function constructed by Method 3.3.1. As depicted in the figure, $\mathbf{u}$ can successfully reduce norm of error to zero in time. For the initial condition $\mathbf{e}(0) = [\pi,\, 0,\, 0,\, 0]^T$, the projected linear control function $\mathbf{v}$ saturates and remains at the boundary of D as $\|\mathbf{e_v}(t)\|$ demonstrates a periodic response. On the other hand, for the same initial condition, $\mathbf{u}(t)$ can successfully reduce $\|\mathbf{e_u}(t)\|$ to zero in time. This behavior is in line with the definition of conditionally controllable problems and suggests that the given regularization problem for the double pendulum system is conditional controllable and $[\pi,\, 0,\, 0,\, 0]^T \in E_N$ (the set E_N corresponds to set Y_N when the regularization problem is written in terms of the error dynamics as discussed in Theorem 2.1.3). Other parameters used in this case study are: $t_f = 20$, $T = 20$, $\epsilon = 0.5$, $n_s = 20$ and $t_s = 0.1$.

3.3.3 Cart-pole

To further investigate the performance of Method 3.3.1, we continue with a point-mass cart-pole system[11], which is a simple example of an under actuated system[12].

[11] The cart-pole system is also referred to as the inverted pendulum in the literature.

[12] Based on Newton's laws of motion, dynamics of mechanical systems are inherently second order [25]. Consequently, we can assume $\ddot{\mathbf{q}} = \mathbf{f}(t, \mathbf{q}, \dot{\mathbf{q}}, \mathbf{u})$ as a generic form of the system accelerations where $\mathbf{q}$ is the vector of generalized coordinates and $\mathbf{u}$ is the vector of input forces and moments. However, in many systems, namely control affine, $\ddot{\mathbf{q}}$ is an affine function of $\mathbf{u}$. As a result, we can write $\ddot{\mathbf{q}}$ for control affine systems as $\ddot{\mathbf{q}} = \mathbf{f}_1(t, \mathbf{q}, \dot{\mathbf{q}}) + \mathbf{f}_2(t, \mathbf{q}, \dot{\mathbf{q}})\mathbf{u}$. An under actuated system is a

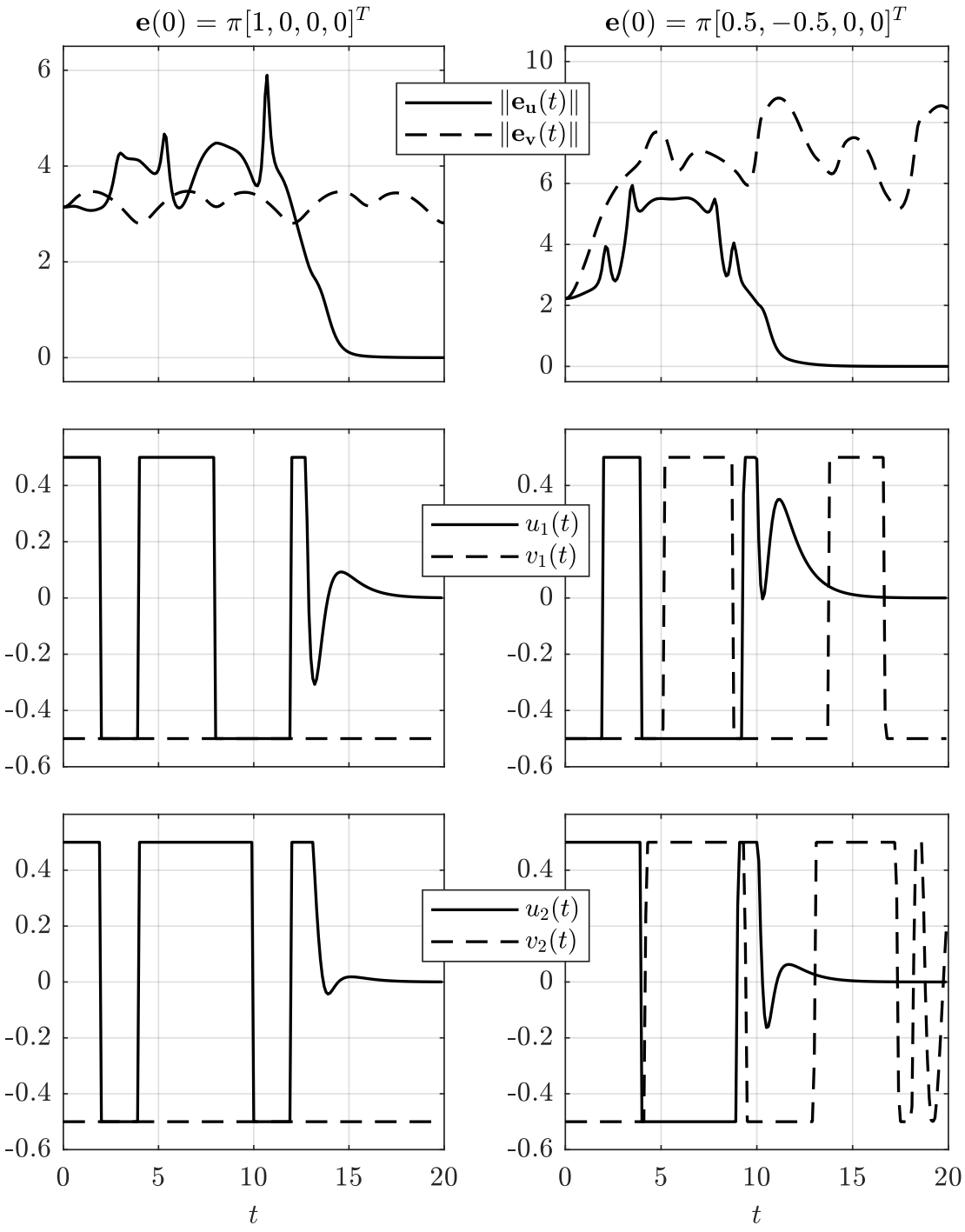

Figure 3.3: Response of the double pendulum system from two different initial conditions to both the synthesized control function based on Method 3.3.1, $\mathbf{u}$ and the projected linear control function $\mathbf{v}$.

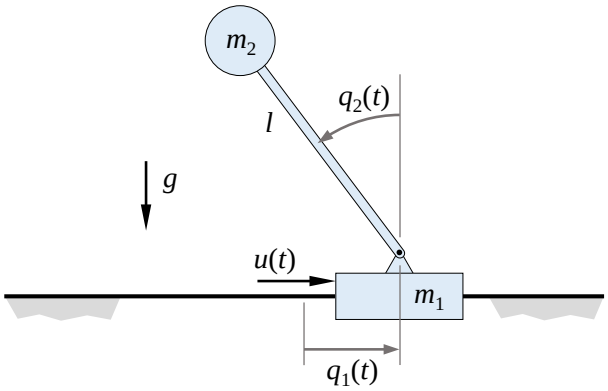

Figure 3.4: The point-mass cart-pole system and the corresponding parameters.

A point-mass cart-pole system is illustrated in Figure 3.4. The detailed derivation of the differential equation of motion for the system with arbitrary values of m_1, m_2 and l is presented in the Appendix B. In order to simplify the expressions, here we take $m_1 = m_2 = l_1 = g = 1$ (with appropriate units). Moreover, by assigning $\mathbf{y}(t) := [q_1(t),\, q_2(t),\, \dot{q}_1(t),\, \dot{q}_2(t)]^T$ as the state vector and $u(t)$ as the input force, we can obtain a first order representation for the system as

$$\dot{\mathbf{y}}(t) = \begin{bmatrix} y_3(t) \\ y_4(t) \\ \mathbf{M}^{-1}\big(\mathbf{y}(t)\big)\Big([u(t),\, 0]^T - \mathbf{\Phi}\big(\mathbf{y}(t)\big)\Big) \end{bmatrix}, \tag{3.13}$$

where

$$\mathbf{M}(\mathbf{y}(t)) = \begin{bmatrix} 2 & -\cos(y_2(t)) \\ -\cos(y_2(t)) & 1 \end{bmatrix}, \text{ and } \mathbf{\Phi}(\mathbf{y}(t)) = \begin{bmatrix} \sin(y_2(t))y_4^2(t) \\ -\sin(y_2(t)) \end{bmatrix}. \tag{3.14}$$

The objective of the cart-pole system is to balance the pole in an upward configuration as the cart stands at a specific position. Accordingly, we chose the desired

control affine system for which $\text{Rank}\big(\mathbf{f}_2(t, \mathbf{q}, \dot{\mathbf{q}})\big) < \text{Dim}(\mathbf{q})$.

state for the control problem as $\mathbf{y}_d(t) = \mathbf{0} \equiv \mathbf{y}_d$. Since $\dot{\mathbf{y}}_d = \mathbf{0}$, we can set u_d to be the input that makes $\mathbf{y}_d$ an equilibrium point of (3.13). Thus,

$$[u_d,\, 0]^T - \underbrace{\mathbf{\Phi}(\mathbf{0})}_{=\mathbf{0}} = \mathbf{0} \implies u_d = 0. \tag{3.15}$$

Having defined values of $\mathbf{y}_d$ and u_d, we can proceed with defining the error dynamics based on the formulation presented in (2.4) that leads to

$$\dot{\mathbf{e}}(t) = \begin{bmatrix} e_3(t) \\ e_4(t) \\ \mathbf{M}_e^{-1}\big(\mathbf{e}(t)\big)\Big([v(t),\, 0]^T + \mathbf{\Phi}_e\big(\mathbf{e}(t)\big)\Big) \end{bmatrix}, \tag{3.16}$$

where $\mathbf{e}(t) := \mathbf{y}_d - \mathbf{y}(t) = -\mathbf{y}(t)$, $v(t) := u_d - u(t) = -u(t)$,

$$\mathbf{M}_e(\mathbf{e}(t)) = \mathbf{M}(\mathbf{y}_d - \mathbf{e}(t)) = \begin{bmatrix} -2 & \cos(e_2(t)) \\ \cos(e_2(t)) & -1 \end{bmatrix}, \tag{3.17}$$

and

$$\mathbf{\Phi}_e(\mathbf{e}(t)) = \mathbf{\Phi}(\mathbf{y}_d - \mathbf{e}(t)) = \begin{bmatrix} \sin(e_2(t))e_4^2(t) \\ -\sin(e_2(t)) \end{bmatrix}. \tag{3.18}$$

Based on (2.6), we can obtain a linear approximation of (3.16) around $\mathbf{e}(t) = \mathbf{0}$ and $u(t) = 0$ as

$$\dot{\mathbf{e}}(t) \approx \mathbf{A}\mathbf{e}(t) + \mathbf{B}\mathbf{v}(t), \tag{3.19}$$

where

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \end{bmatrix}, \text{ and } \mathbf{B} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}. \tag{3.20}$$

Although, similar to the double pendulum example, we can use the pole placement techniques to find an appropriate gain matrix $\mathbf{K}$, here we follow LQR derivation. That is, we seek to find $\mathbf{K}$ such that the linear controller $a(t) := \mathbf{K}\mathbf{e}(t)$ minimizes the cost functional

$$J(\mathbf{e}, a) = \int_0^\infty \mathbf{e}(t)^T \mathbf{Q}\mathbf{e}(t) + Ra^2(t)\, dt, \tag{3.21}$$

subjected to $\dot{\mathbf{e}}(t) = \mathbf{A}\mathbf{e}(t) + \mathbf{B}a(t)$. As explained in Chapter 4, the optimal gain matrix $\mathbf{K}^*$ could be obtained by as

$$\mathbf{K}^* = R^{-1}\mathbf{B}^T\mathbf{P}, \tag{3.22}$$

where $\mathbf{P}$ is found by solving the continuous time algebraic Riccati equation

$$\mathbf{A}^T\mathbf{P} + \mathbf{P}\mathbf{A} - \mathbf{P}\mathbf{B}R^{-1}\mathbf{B}^T\mathbf{P} + \mathbf{Q} = 0. \tag{3.23}$$

Substituting $\mathbf{Q} = \mathbf{I}_4$ and $R = 20$ in (3.22) and (3.23) we obtain $\mathbf{K}^*$ as

$$\mathbf{K}^* \approx [0.22,\ -5.77,\ 1.31,\ -4.66]. \tag{3.24}$$

Finally, to construct the projected linear control function, we set $D = [-0.5, 0.5]$

that leads to

$$v(t) := \underset{D}{\text{Proj}}\, \mathbf{K}^* \mathbf{e}(t) = \underset{[-0.5,\, 0.5]}{\text{sat}}\, \mathbf{K}^* \mathbf{e}(t), \tag{3.25}$$

Figure 3.5 shows response of the system from four different initial conditions to both v and u (function u is the control synthesized by Method 3.3.1). As depicted in the figure, for all cases u can successfully reduce norm of the error to zero while for

$$\mathbf{e}(0) \in H := \left\{ [0,\, 1,\, 0,\, 0]^T,\, [0,\, 0,\, 1,\, 0]^T,\, [0,\, 0,\, 0,\, 1]^T \right\}, \tag{3.26}$$

the saturated linear control function, v, drives system to instability and causes norm of the error to increase in time. Similar to the double pendulum example, for these initial condition, v saturates and remains at the boundary of D while $\|\mathbf{e}_v(t)\|$ increases in time. However, for the same initial condition, the synthesized control function u can successfully reduce $\|\mathbf{e}_u(t)\|$ to zero in time. This set of conditions suggest conditional controllability of the problem and shows that $H \subset E_N$ (the set E_N corresponds to set Y_N when the regularization problem is written in terms of the error dynamics as discussed in Theorem 2.1.3). Other parameters used in this case study are: $t_f = 20$, $T = 20$, $\epsilon = 0.5$, $n_s = 10$ and $t_s = 0.1$.

3.4 Remarks and conclusions

In this chapter, we covered some preliminaries on spatial exploring trees and discussed their limitations in synthesizing control functions as solutions to planning problems. To address this limitations, we extended the idea of spatial exploring trees to include time and presented a formal setting for spatio-temporal exploring trees. Finally, the presented idea is used to as tool for synthesizing control functions. In particular, we proposed Method 3.3.1 that utilizes spatio-temporal exploring trees

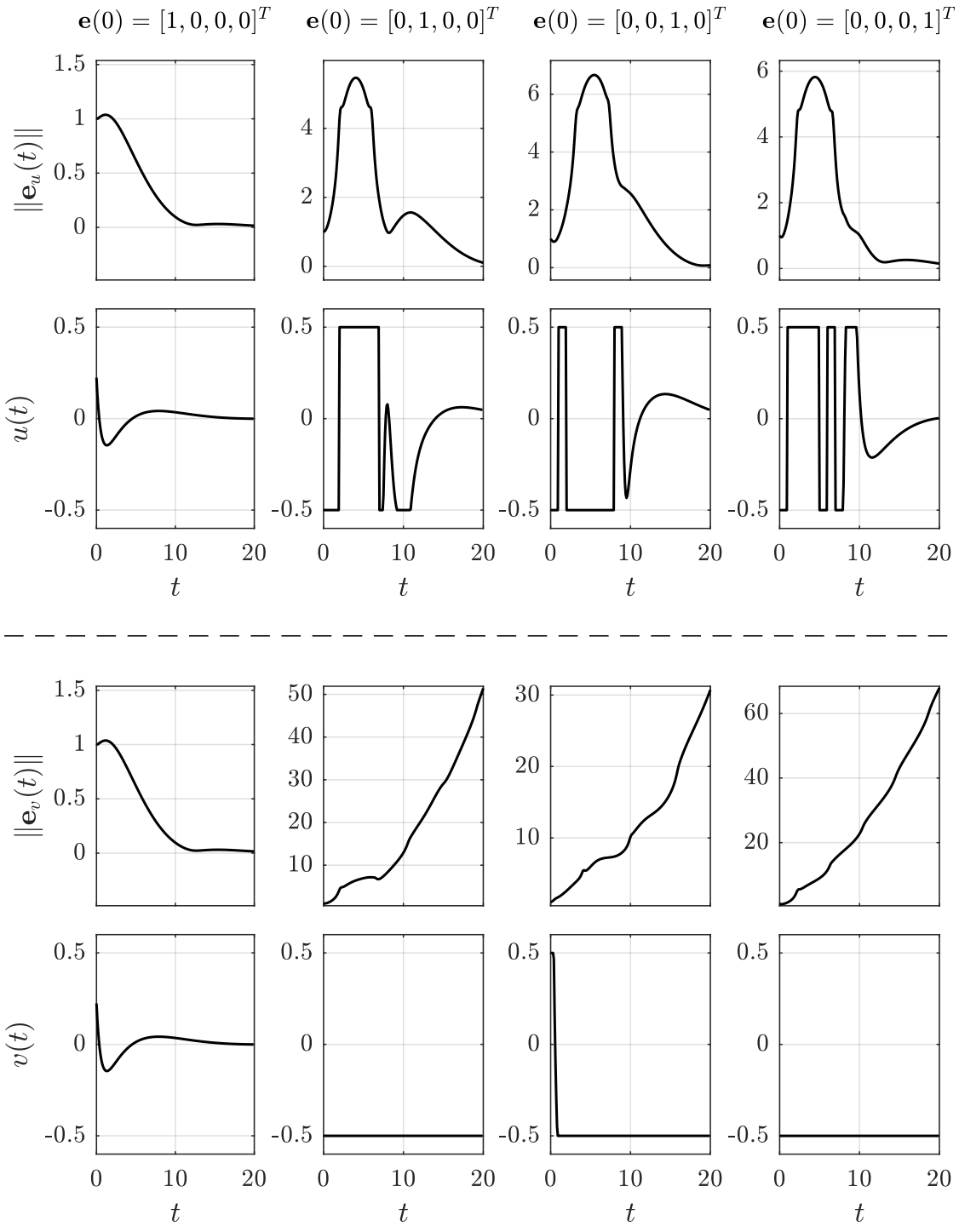

Figure 3.5: Response of the point-mass cart-pole system from four different initial conditions to both the synthesized control function based on Method 3.3.1, u, and the projected linear control function v.

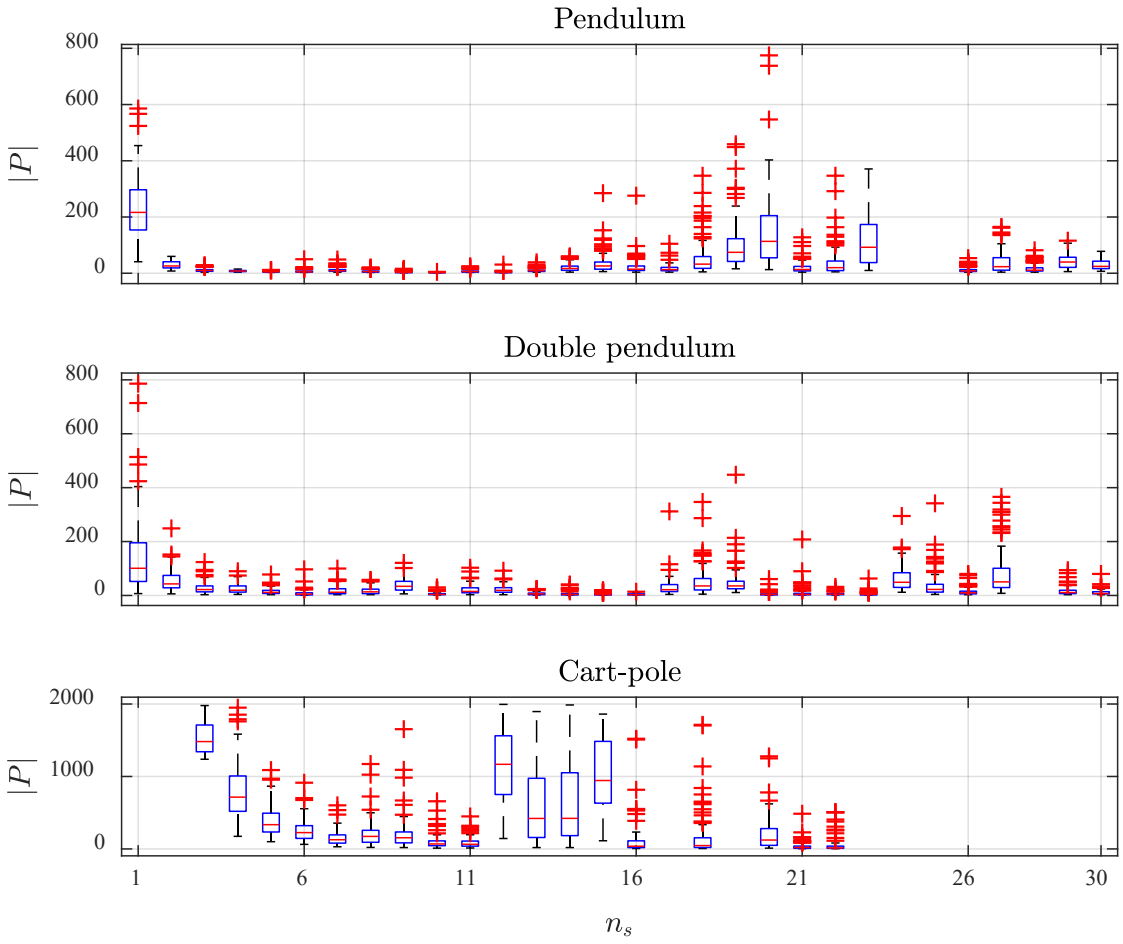

Figure 3.6: $|P|$ versus n_s when evaluating $\mathbf{w}_1$ and δ_1 for the three case studies. The missing boxes indicate inability of the method in finding a trajectory from the given initial condition satisfying $\|\mathbf{e}_v(t)\| \leq \epsilon$ for all $t \in [t_f, t_f + T]$.

in an Ariadne's clew framework to gradually construct a control function to regulate the error.

Effectiveness of the proposed method is explored by synthesizing control functions for three case studies from different initial conditions. The conducted simulations revealed effectiveness of the method in constructing control functions for each problem. However, through numerous experiments with the parameters associated with Method 3.3.1, we observed the importance of t_s, n_s and the linear control function gain, $\mathbf{K}$, in the success of the algorithm. In general, reducing t_s

increases the computation time and increasing t_s affects convergence of the error close to $\mathbf{e}(t) = \mathbf{0}$, where the linear approximation of the system is valid. In the implementation of the method, special care must be given to the sampling time t_s. If the effect of discretizing the saturated linear control function, $\mathbf{v}(t)$, which is a consequence of using digital computers in implementing Method 3.3.1, is not considered for in the Search subtask of Ariadne's clew framework, t_s should be small enough to reduce the divergence of $\mathbf{e}_{\mathbf{v}}(t)$ from $\mathbf{e}_{zoh(\mathbf{v})}(t)$. Based on the results presented in this chapter, if such divergence is small, the algorithm can successfully account for it by solving the planning problem from the diverged point in some future time step. Such action also suggest the noise rejection capability of Method 3.3.1. Similarly, smaller values of n_s result in shorter search horizons which increases the number of vertices required to solve the planning problem. Thus, reducing n_s, increases oscillations in the synthesized control function. On the other hand, increasing n_s leads to more scattered tree vertices that may reduce the chance of finding a solution. Figure 3.6 illustrates the effect of n_s on $|P|$ at $t = t_0$ for the three case studies. The corresponding results are obtained by running the algorithm with 100 different random number seeds for each n_s value. The parameters used in the algorithm are the same as for the case studies with the maximum allowable size of the search tree is set to 2000. The presented results suggest existence of n_s optima for which the planning problem is solved with fewer iterations, that corresponds to smaller $|P|$ at t_0. Another important observation is the effect of linear controller gain matrix $\mathbf{K}$ on the performance of the algorithm. Accordingly, some experimentation with a given system is required to find a gain matrix that results in better performance of the Search subroutine in Ariadne's clew framework. Finally, the time horizon t_h should be long enough to correctly reflect convergence of the projected linear controller. Choosing small t_h may deceive the search subroutine and result in an ineffectual

control function.

Chapter 4

Optimal control

This chapter focuses on utilizing methods of optimal control in synthesizing control functions for conditionally controllable problems. We start with a brief overview and preliminaries of optimal control theory and continue our discussions with the details of the proposed method to synthesize control functions for conditionally controllable problems. In particular, we present an algorithm to construct control functions through composition of a piecewise constant function with a linear controller. The coefficients of the piecewise function are determined by solving an optimization problem for which the cost is defined as the norm of state vector at a given finite time.

4.1 A brief introduction to optimal control theory

As an informal definition, we can state an optimal control problem as the following. Let $\dot{\mathbf{y}}(t) = \mathbf{f}(t, \mathbf{y}(t), \mathbf{u}(t))$ be a given control system, U be the set of admissible control functions, $J(\mathbf{y}, \mathbf{u})$ be a given performance measure and $T \subset R$ be a time

interval. An optimal control problem is to find $\mathbf{u}^* \in U$ such that

$$\mathbf{u}^* = \arg\min_{\mathbf{u}\in U} J(\mathbf{y}_\mathbf{u}, \mathbf{u}), \tag{4.1}$$

satisfying

$$\mathbf{\Psi}(t, \mathbf{y}(t), \mathbf{u}(t)) \geq \mathbf{0}, \tag{4.2}$$

$$\mathbf{E}(t, \mathbf{y}(t), \mathbf{u}(t)) = \mathbf{0}, \tag{4.3}$$

for all $t \in T$. Functions $\mathbf{\Psi}$ and $\mathbf{E}$ are defined based on the physical requirements of the system.

In what follows, we present a short introduction to optimal control theory by discussing common techniques used to solve an optimal control problem analytically and numerically. An interested reader may refer to [9], [26] and [27] for more detailed derivations and explanations.

4.1.1 The variational approach to optimal control problems

From historic perspective, contributions to calculus of variations by Edward J. McShane lead to major developments of optimal control theories, largely due to the work of Lev Pontryagin [28] and Richard Bellman in the 1950s. In what follows, we summarize Pontryagin's maximum principle, as stated in [9], as the following theorem. For the sake of brevity, we omit proof of the theorem and refer interested readers to [9] or [28] for detailed discussions and proofs.

Theorem 4.1.1 (Pontryagin's maximum principle)**.** *Given a control system* $\dot{\mathbf{y}}(t) = \mathbf{f}(t, \mathbf{y}(t), \mathbf{u}(t))$*, a time interval* $T = [t_0, t_f]$ *and a set of feasible control functions* U*.*

Let

$$J(\mathbf{u}) := h(t_f, \mathbf{y}(t_f)) + \int_{t_0}^{t_f} g(t, \mathbf{y}(t), \mathbf{u}(t))dt,$$

be a given performance measure. Then, the control function $\mathbf{u}^*$ *that minimizes* J *must satisfy the following necessary conditions:*

i) $\mathbf{y}^*(t) = \dfrac{\partial}{\partial \boldsymbol{\lambda}}\mathbf{H}(t, \mathbf{y}^*(t), \boldsymbol{\lambda}^*(t), \mathbf{u}^*(t))$, *for all* $t \in T$;

ii) $\dot{\boldsymbol{\lambda}}^*(t) = -\dfrac{\partial}{\partial \mathbf{y}}\mathbf{H}(t, \mathbf{y}^*(t), \boldsymbol{\lambda}^*(t), \mathbf{u}^*(t))$, *for all* $t \in T$;

iii) $\mathbf{H}(t, \mathbf{y}^*(t), \boldsymbol{\lambda}^*(t), \mathbf{u}^*(t)) \leq \mathbf{H}(t, \mathbf{y}^*(t), \boldsymbol{\lambda}^*(t), \mathbf{u}(t))$, *for all* $\mathbf{u} \in U$ *and all* $t \in T$;

iv) $\Big[\dfrac{\partial}{\partial \mathbf{y}}h(t_f, \mathbf{y}^*(t_f)) - \boldsymbol{\lambda}^*(t_f)\Big]^T \delta\mathbf{y}_f + \Big[\dfrac{\partial}{\partial t}h(t_f, \mathbf{y}^*(t_f))$
$+\mathbf{H}(t_f, \mathbf{y}^*(t_f), \boldsymbol{\lambda}^*(t_f), \mathbf{u}^*(t_f))\Big]\delta t_f = 0$;

where

$$\mathbf{H}(t, \mathbf{y}(t), \boldsymbol{\lambda}(t), \mathbf{u}(t)) := g(t, \mathbf{y}(t), \mathbf{u}(t)) + \boldsymbol{\lambda}^T(t)\mathbf{f}(t, \mathbf{y}(t), \mathbf{u}(t)),$$

is the Hamiltonian function. $\delta\mathbf{y}_f$ *and* δt_f *are variations of the final state and time, respectively.*

Moreover, if there is no constraint on the range of functions in U*, that is for all* $\mathbf{u} \in U$, $\mathbf{u} : \mathbb{R} \to \mathbb{R}^m$*, then condition (iii) simplifies to*

iii) $\dfrac{\partial}{\partial \mathbf{u}}\mathbf{H}(t, \mathbf{y}^*(t), \boldsymbol{\lambda}^*(t), \mathbf{u}^*(t)) = \mathbf{0}$ *for all* $t \in T$.

The results of Theorem 4.1.1 could be used to find an optimal control function for linear regulator problem as presented in the next subsection.

4.1.2 Linear Quadratic Regulator (LQR)

In this subsection, we apply Theorem 4.1.1 to find a linear control function of the form $\mathbf{v}(t) := \mathbf{K}(t)\mathbf{e}(t)$ for linearized system $\dot{\mathbf{e}} = \mathbf{A}(t)\mathbf{e}(t) + \mathbf{B}(t)\mathbf{v}(t)$. The result

presented here follows the discussions in [9] and are primarily due to the work of R. E. Kalman.

Let the performance measure be defined as

$$J := \frac{1}{2}\langle \mathbf{C}\mathbf{e}(t_f), \mathbf{e}(t_f)\rangle + \frac{1}{2}\int_{t_0}^{t_f} \langle \mathbf{Q}(t)\mathbf{e}(t),\, \mathbf{e}(t)\rangle + \langle \mathbf{R}(t)\mathbf{v}(t),\, \mathbf{v}(t)\rangle dt, \tag{4.4}$$

where $t_f < \infty$ is fixed, $\mathbf{C}, \mathbf{Q}(t) \in \mathbb{R}^{n\times n}$ are symmetric positive semi-definite matrices and $\mathbf{R} \in \mathbb{R}^{m\times m}$ is a symmetric positive definite matrix. It is also assumed that $\mathbf{e}(t) \in \mathbb{R}^n$ and $\mathbf{v}(t) \in \mathbb{R}^m$ are not bounded. Moreover, $\mathbf{e}(t_f)$ is not constrained. The corresponding Hamiltonian for the problem is

$$\begin{aligned}\mathbf{H}(t, \mathbf{e}(t), \boldsymbol{\lambda}(t), \mathbf{v}(t)) = \frac{1}{2}\langle \mathbf{Q}(t)\mathbf{e}(t),\, \mathbf{e}(t)\rangle + \frac{1}{2}\langle \mathbf{R}(t)\mathbf{v}(t),\, \mathbf{v}(t)\rangle \\ + \langle \boldsymbol{\lambda}(t), \mathbf{A}(t)\mathbf{e}(t) + \mathbf{B}(t)\mathbf{v}(t)\rangle,\end{aligned} \tag{4.5}$$

and the necessary conditions are

$$\mathbf{e}^*(t) = \mathbf{A}(t)\mathbf{e}^*(t) + \mathbf{B}(t)\mathbf{v}^*(t) \tag{4.6}$$

$$\dot{\boldsymbol{\lambda}}^*(t) = -\mathbf{Q}(t)\mathbf{e}^*(t) - \mathbf{A}^T(t)\boldsymbol{\lambda}^*(t) \tag{4.7}$$

$$\mathbf{0} = \frac{\partial \mathbf{H}}{\partial \mathbf{v}} = \mathbf{R}(t)\mathbf{v}^*(t) + \mathbf{B}^T(t)\boldsymbol{\lambda}^*(t). \tag{4.8}$$

Solving (4.8) for $\mathbf{v}^*(t)$ and substituting in condition (4.6) leads to

$$\dot{\mathbf{e}}^*(t) = \mathbf{A}(t)\mathbf{e}^*(t) - \mathbf{B}(t)\mathbf{R}^{-1}(t)\mathbf{B}^T(t)\boldsymbol{\lambda}^*(t). \tag{4.9}$$

Note that the existence of $\mathbf{R}^{-1}(t)$ is assured due to its positive definiteness. Equations (4.7) and (4.9) form a set of linear homogeneous differential equations of the

form

$$\begin{bmatrix} \dot{\mathbf{e}}^*(t) \\ \dot{\boldsymbol{\lambda}}^*(t) \end{bmatrix} = \begin{bmatrix} \mathbf{A}(t) & -\mathbf{B}(t)\mathbf{R}^{-1}(t)\mathbf{B}^T(t) \\ -\mathbf{Q}(t) & -\mathbf{A}^T(t) \end{bmatrix} \begin{bmatrix} \mathbf{e}^*(t) \\ \boldsymbol{\lambda}^*(t) \end{bmatrix}, \tag{4.10}$$

with the solution

$$\begin{bmatrix} \mathbf{e}^*(t_f) \\ \boldsymbol{\lambda}^*(t_f) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\phi}_{11}(t,t_f) & \boldsymbol{\phi}_{12}(t,t_f) \\ \boldsymbol{\phi}_{21}(t,t_f) & \boldsymbol{\phi}_{22}(t,t_f) \end{bmatrix} \begin{bmatrix} \mathbf{e}^*(t) \\ \boldsymbol{\lambda}^*(t) \end{bmatrix}. \tag{4.11}$$

From condition (iv) of Theorem 4.1.1 by setting $\delta t_f = 0$ (since the final time is fixed) we can obtain a boundary condition for $\boldsymbol{\lambda}^*(t)$ as

$$\boldsymbol{\lambda}^*(t_f) = \mathbf{C}\mathbf{e}^*(t_f). \tag{4.12}$$

In [9], by borrowing the idea presented in [29], it is shown that the above formulation simplifies to

$$\mathbf{v}^*(t) = -\mathbf{R}^{-1}(t)\mathbf{B}^T(t)\mathbf{G}_{t_f}(t)\mathbf{e}(t), \tag{4.13}$$

where

$$\mathbf{G}_{t_f}(t) = \Big(\boldsymbol{\phi}_{22}(t,t_f) - \mathbf{C}\boldsymbol{\phi}_{12}(t,t_f)\Big)^{-1}\Big(\mathbf{C}\boldsymbol{\phi}_{11}(t,t_f) - \boldsymbol{\phi}_{21}(t,t_f)\Big). \tag{4.14}$$

Generally, in order to implement $\mathbf{v}^*$, we need to resort to numerical procedures to evaluate $\boldsymbol{\phi}_{ij}(t,t_f)$. Alternatively, by substituting (4.12) in (4.11) and solving the results for $\boldsymbol{\lambda}^*(t)$ as function of $\mathbf{e}^*(t)$, which leads to $\boldsymbol{\lambda}^*(t) = \mathbf{G}_{t_f}(t)\mathbf{e}*(t)$, and finally taking derivative of $\boldsymbol{\lambda}^*(t) = \mathbf{G}_{t_f}(t)\mathbf{e}*(t)$ with respect to time we can show that

$\mathbf{G}_{t_f}(t)$ satisfies the Riccati type final value problem

$$\begin{cases} \dot{\mathbf{G}}_{t_f}(t) = -\mathbf{G}_{t_f}(t)\mathbf{A}(t) - \mathbf{A}^T(t)\mathbf{G}_{t_f}(t) - \mathbf{Q}(t) + \mathbf{G}_{t_f}(t)\mathbf{B}(t)\mathbf{R}^{-1}(t)\mathbf{B}^T(t)\mathbf{G}_{t_f}(t), \\ \mathbf{G}_{t_f}(t_f) = \mathbf{C}. \end{cases} \tag{4.15}$$

In addition, in [29] it is shown that if (i) $t_f = \infty$, (ii) $\mathbf{C} = \mathbf{0}$ and (iii) $\mathbf{A}$, $\mathbf{B}$, $\mathbf{Q}$ and $\mathbf{R}$ are constant matrices, then $\lim_{t\to\infty} \mathbf{G}_{t_f}(t) = \mathbf{G}$, that is a constant matrix. Consequently, the linear optimal control strategy is a constant linear map applied to $\mathbf{e}(t)$. In this case, $\mathbf{G}$ could be obtained by solving

$$\mathbf{G}\mathbf{A} + \mathbf{A}^T\mathbf{G} + \mathbf{Q} - \mathbf{G}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{G} = \mathbf{0}, \tag{4.16}$$

which is obtained by setting $\dot{\mathbf{G}}_{t_f}(t) = \mathbf{0}$ in (4.15).

Although we have used Pontryagin's maximum principle to derive the expression for optimal linear control function $\mathbf{v}$, one can obtain similar results by using the principles of dynamic programming, in particular Hamilton-Jacobi-Bellman equation, as discussed in [9]. For more information on dynamic programming, and in particular its application to control problems, please refer to [30]. As the presented discussion suggest, deriving an expression for the optimal control function is rather challenging and depending on nonlinearities of a given problem, it may not be possible to find an analytic solution. In the following subsections, we look into two common approaches of solving an optimal control problem numerically.

4.1.3 Numerical approaches

In general, optimal control problems are nonlinear, mostly due to nonlinearities of system dynamics, and therefore, it is not possible to derive an analytic expression

for the control function. To address this issue, we can explore numerical approaches to solve optimal control problems, which are often classified as direct and indirect methods [26].

Indirect methods

An indirect method attempts to solve the optimal control problem by finding a solution that satisfies the necessary optimality conditions, such as the ones stated in Theorem 4.1.1 that leads to a nonlinear two-point boundary value problem. The beauty of using indirect methods is that we can solve for both state ($\mathbf{y}$) and adjoint ($\boldsymbol{\lambda}$) equations and the obtained result is readily verified to be an extremal trajectory. The main disadvantage of indirect methods is that the obtained boundary-value problem is often extremely difficult to solve, specifically when Range($\mathbf{u}$) $\neq \mathbb{R}^m$, that is when the output of admissible control functions is constrained to a proper subset of $\mathbb{R}^m$. Moreover, control engineer or specialist needs to derive the expressions for the Hamiltonian, $\mathbf{H}$, and adjoint, $\boldsymbol{\lambda}$, and corresponding partial derivatives that can be complicated for specific systems. An introductory discussion on some basic indirect methods is available in [9].

Direct methods

The core idea behind direct methods is to convert state and control functions from infinite dimensional objects to finite dimensions through quantization of the functions in time. That is, state and control functions are approximated on a finite number of subintervals in time using functions such as piecewise constant, piecewise linear or polynomials. Accordingly, the cost functional is approximated as a cost function. Finally, the coefficients of the approximated functions for states and controls are treated as optimization variables and the problem is transcribed to a

nonlinear optimization problem. As stated in [26], such transcription is commonly achieved in three steps of

i) Converting control function and/or the dynamic system into a problem with a finite set of variables;

ii) Solving the obtained finite dimensional problem via a parameter optimization algorithm;

iii) Estimating the accuracy of the finite dimensional solution and repeating the process if necessary.

4.2 Application to solve conditionally controllable problems

The approach that has risen to prominence in numerical optimal control over the past two decades (i.e., from the 1980s to the present) is that of so-called direct methods.

In this section we explore the possibility of utilizing methods of optimal control in synthesizing control functions for conditionally controllable problems. It must be noted that our main objective is not to find an optimal control and trajectory functions but to find a relatively fast and computationally inexpensive algorithm that can solve conditionally controllable problems.

Having in mind that for every $\mathbf{e}_0 \in E \setminus E_N$, there exists $\mathbf{v} \in V_L$ that can regulate the problem, our objective is to find a feasible function that can take system trajectory to $E \setminus E_N$. In this regard, we can follow a procedure similar to Method , presented in Chapter 3, but rather than using a planner to find the function, we can utilize direct methods to construct the control function.

Prior to presenting the algorithm, we introduce the following objects. Let $N \in \mathbb{N}$, $t_0, t_f \in \mathbb{R}$ such that $t_0 < t_f$ and $0 < \delta \in \mathbb{R}$, then

$$\delta_N := \frac{t_f - t_0}{N+1}, \tag{4.17}$$

$$\tau_k^N := t_0 + \delta_N \cdot k,\ k \in \{0\} \cup \mathbb{N}, \tag{4.18}$$

$$t_j^N := t_0 + \frac{\delta_N}{\lceil \frac{t_f - t_0}{N+1} \rceil} \cdot j,\ j \in \{0\} \cup \mathbb{N}, \tag{4.19}$$

$$Z_N := \Big\{ \mathbf{z} : \mathbb{R} \to D : \mathbf{z}(t) := \sum_{k=0}^{N-1} \chi_{[\tau_k^N, \tau_{k+1}^N)}(t) \cdot \mathbf{c}_k,\ \mathbf{c}_k \in D\, \forall k \in \{0\} \cup \mathbb{N} \Big\}, \tag{4.20}$$

where for $x \in \mathbb{R}$, $\lceil x \rceil := \min n \in \mathbb{Z} : n \geq x$.

We propose the following algorithm to construct $\mathbf{u} \in U$ that can satisfy (i) and (ii).

Method 4.2.1. Given $T = [t_0, t_f]$, $\delta \in \mathbb{R}^+$, $N_0, \Delta N \in \mathbb{N}$, and an error regularization problem $R_e = (\mathbf{f}_e, V, \mathbb{R}^n, \mathbf{0}, \mathbb{R}^n, t_0)$ with $V = L^\infty$ that is conditionally controllable when range of control functions in V is constrained to $D := \prod_{i=1}^m [w_{i_{min}}, w_{i_{max}}] \subset \mathbb{R}^m$. The following algorithm can be employed to construct control function $\mathbf{u}$ by composing a piecewise constant function in Z_N with a projected linear controller in V_L.

$N := N_0$;

$\mathbf{z}^*(t) := \mathbf{0}$;

while $\exists\, t \in [t_f, t_f + t_h]$ *such that* $\|\mathbf{e}_{(\mathbf{z}^* \oplus \mathbf{v}, \mathbf{e}_0)}(t)\| > \epsilon$ **do**

  $N \longleftarrow N + \Delta N$;

  if $\delta_N < \delta$ **then**

    return $\mathbf{u} := \emptyset$;

  end

  $\mathbf{z}^* := \arg\min_{\mathbf{z} \in Z_N} \|\mathbf{e}_{(\mathbf{z} \oplus \mathbf{v}, \mathbf{e}_0)}(t_f)\|$;

end

return $\mathbf{u} := \mathbf{z}^* \oplus \mathbf{v}$;

where N_0 indicates the initial number of intervals in time, ΔN defines the increment in number of intervals at each iteration. $\mathbf{v} \in V_L$ is a projected linear control function that regulates the linearized system model around $\mathbf{e}(t) = \mathbf{0}$, and for $\mathbf{z} \in Z_N$, $\mathbf{z} \oplus \mathbf{v}$ is defined as

$$(\mathbf{z} \oplus \mathbf{v})(t) := \mathbf{z}(t) + \chi_{[\tau_N^N, \infty)}(t) \cdot \mathbf{v}(t). \tag{4.21}$$

In words, the procedure presented in Method 4.2.1 first checks if the error trajectory resulted by employing the projected linear controller from the given initial error value can satisfy the convergence criteria

$$\|\mathbf{e}_{(\mathbf{v}, \mathbf{e}_0)}\| < \epsilon,\ \forall t \in [t_f,\ t_f + t_h]. \tag{4.22}$$

If so, the algorithm terminates by returning $\mathbf{u} := \mathbf{v}$. Otherwise, it divides the time interval $[t_0,\ t_f]$ into $N = N_0 + \Delta N + 1$ intervals and proceeds with finding a function in Z_N that minimizes norm of the error at t_f, that is $\|\mathbf{e}_{(\mathbf{z} \oplus \mathbf{v}, \mathbf{e}_0)}(t_f)\|$. Note, based on the composition rule, the piecewise function will have support on all but the last subdivided interval of $[t_0,\ t_f]$ and the projected linear controller will only have a

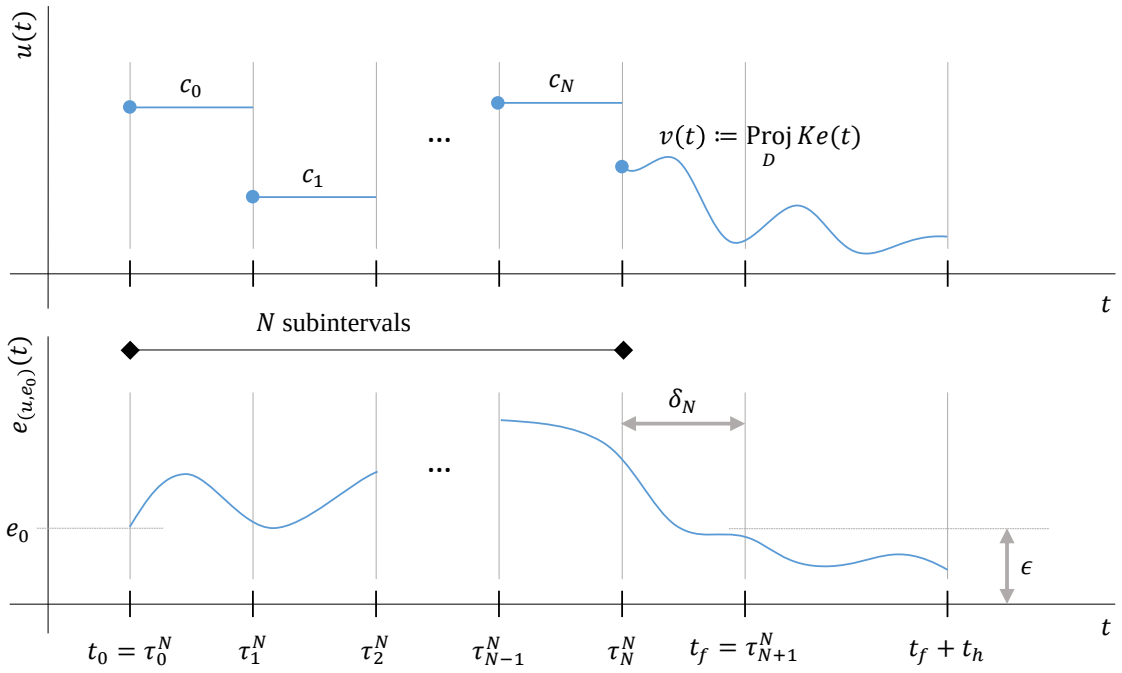

Figure 4.1: text

nonzero value on the very last subdivided interval. An example of the composed function $\mathbf{u} = (\mathbf{z} \oplus \mathbf{v})(t)$ and the corresponding subdivided intervals for a given N are illustrated in Figure 4.1. Note that if the algorithm can not find a suitable function in Z_N for the given δ, then the process will terminate by returning $\mathbf{u} := \emptyset$.

In what follows, we explore the effectiveness of Method 4.2.1 in synthesizing functions for the same case studies as used in Chapter 3.

4.2.1 Case studies

To evaluate the effectiveness of Method 4.2.1 in synthesizing control functions for conditionally controllable problems, we proceed with the three case studies that are presented in Chapter 3 that are: (i) a simple point-mass pendulum, (ii) point-mass double pendulum and (iii) a simple cart-pole system. For the sake of brevity, we do not discuss derivations of projected linear controllers for the case studies

and will refer to the related discussions in Chapter 3 for detailed derivations of the related subjects. The associated parameters of Method 4.2.1 used in all the following examples are: $N_0 = 0$, $\Delta N = 1$ and $\delta = 0.1$.

4.2.2 Simple pendulum

As our first system, we proceed with the simple point mass pendulum of Chapters 2. Considering the upward configuration of the pendulum, $\mathbf{y}_d \equiv \mathbf{0}$, as the desired point, we get $\mathbf{e}(t) = -\mathbf{y}(t)$. As discussed in Section 3.3.1, setting $\mathbf{y}_d$ as an equilibrium point leads to $u_d \equiv 0$ and consequently $v(t) = -u(t)$. Thus, the corresponding error dynamic of the system simplifies to

$$\begin{bmatrix} \dot{e}_1(t) \\ \dot{e}_2(t) \end{bmatrix} = \begin{bmatrix} e_2(t) \\ \sin(e_1(t)) + v(t) \end{bmatrix} =: \mathbf{f}_e\big(t, \mathbf{e}(t), v(t)\big). \tag{4.23}$$

Similar to the example in Section3.3.1, we set $D := [-w_{max}, w_{max}]$ that leads to $V_L = \{v \in L^\infty : v(t) = \text{Proj}_D\, \mathbf{K}\mathbf{e}(t),\, \mathbf{K} \,\text{and}\, \mathbf{e}(t) \in \mathbb{R}^2\}$. We pick $w_{max} = 0.5$ and two initial conditions: (i) $\mathbf{e}(0) = [-2,\, 0]^T \in E_N$ and (ii) $\mathbf{e}(0) = [-4,\, 2]^T \in E \setminus E_N$, which are the exact values used for the example in Section 3.3.1[1]. The results of utilizing Method 4.2.1 in synthesizing control functions for the pendulum example are illustrated on Figure 4.2. In this figure, $u(t)$ represents the function constructed by Method 4.2.1 and $v(t) := \text{Proj}_D\, \mathbf{K}\mathbf{e}(t)$, $\mathbf{K} = [-10, -3]$, is the projected linear controller on set D. The response of the system from both initial conditions to both u and v are depicted in the figure.

Similar to the results obtained by using Method 3.3.1, even when $\mathbf{e}(0) \in E_N$, the control function u can meet the objective of the regularization problem, while

[1]Recall that, based on Theorem 2.2.3, the regularization problem $R_e = (\mathbf{f}_e, L^\infty, \mathbb{R}^2, \mathbf{0}, \mathbb{R}^2, 0)$ is a conditionally controllable if $w_{max} \in (0, 1)$.

v saturates at w_{max} and results in an undamped oscillatory response of the system. And similarly, u leads to a faster convergence of $\mathbf{e}$ to zero for $\mathbf{e}(0) = [-4,\ 2]^T \in E \setminus E_N$. The final values of N for which the `while` loop is terminated are $N_f =$ and $N_f =$ for $\mathbf{e}(0) = [-2,\ 0]^T \in E_N$ and $\mathbf{e}(0) = [-4,\ 2]^T \in E \setminus E_N$, respectively.

4.2.3 Double pendulum

In continuation of our case studies, we proceed with the double pendulum system as discussed in Section 3.3.2. We set the objective of the corresponding regularization problem as to move and keep the system in the upward configuration where $q_1(t) = \pi/2$ and $q_2(t) = 0$ (for definition of the generalized coordinates q_1 and q_2 please refer to Figure 3.2) and use the same system parameters, constraint set D and projected linear controller $\mathbf{v}$ as defined in Section 3.3.2. The results of employing Method 4.2.1 are depicted in Figure 3.2. In this figure $\mathbf{u}$ denotes the control function constructed by Method 4.2.1. As depicted in the figure, $\mathbf{u}$ can successfully reduce norm of error to zero in time. As previously observed in Figure 3.3, for $\mathbf{e}(0) = [\pi,\ 0,\ 0,\ 0]^T$, the projected linear control function $\mathbf{v}$ saturates and remains at the boundary of D as $\|\mathbf{e}_{(\mathbf{v},\mathbf{e}_0)}(t)\|$ shows a periodic response. However, for the same initial condition, $\mathbf{u}(t)$ can successfully reduce $\|\mathbf{e}_{\mathbf{u}}(t)\|$ to zero in time. As also noted in Section 3.3.2, this behavior is in line with the definition of conditionally controllable problems and suggests that the given regularization problem for the double pendulum system is conditional controllable and $[\pi,\ 0,\ 0,\ 0]^T \in E_N$. In this case study, the final values of N for which the `while` loop is terminated are $N_f = 5$ for $\mathbf{e}(0) = \pi[1,\ 0,\ 0,\ 0]^T$ and $N_f = 1$ for $\mathbf{e}(0) = \pi[0.5,\ -0.5,\ 0,\ 0]^T$.

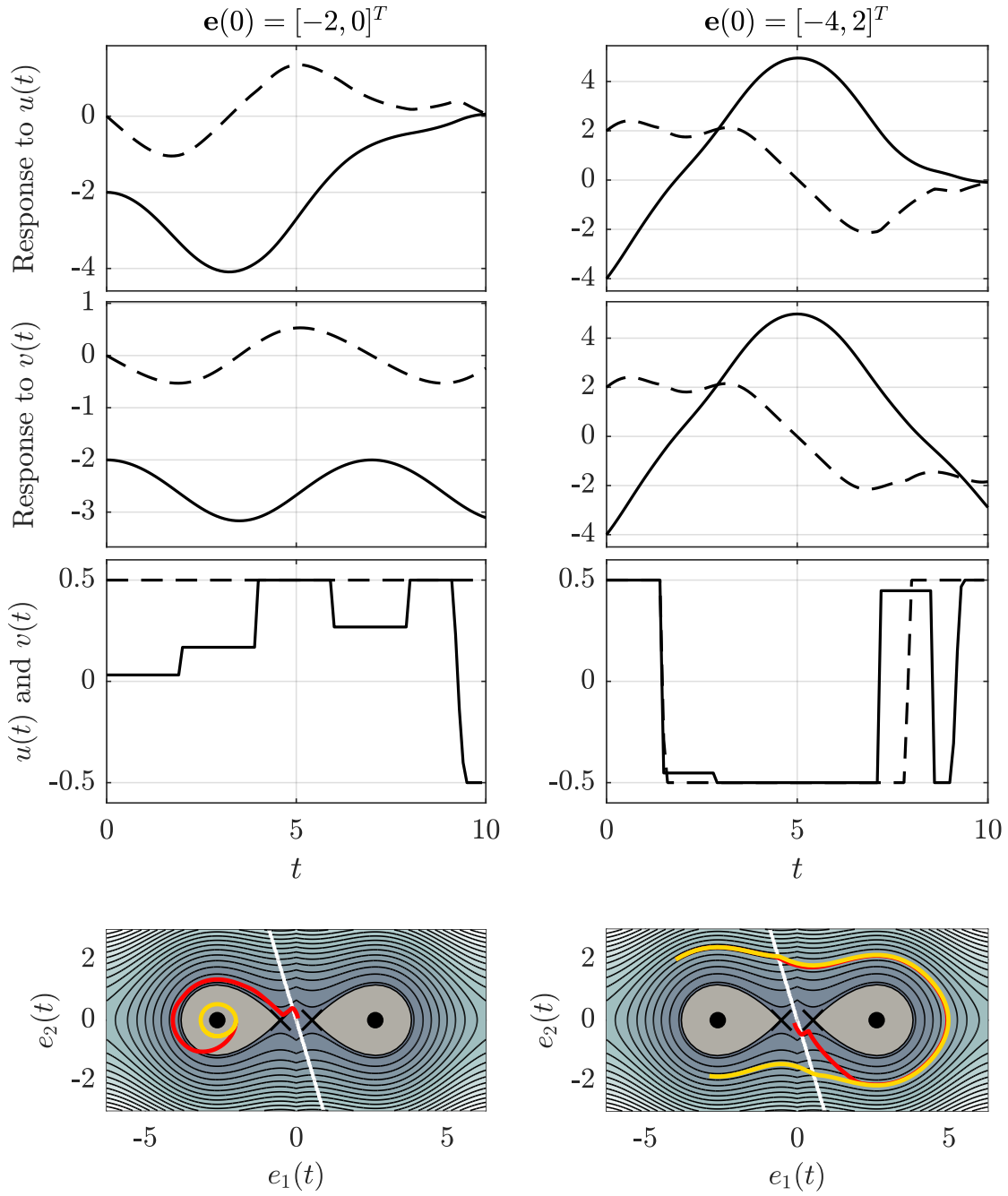

Figure 4.2: Response of the pendulum system to the synthesized control function based on Method 4.2.1, u and the projected linear control function $v(t) =: \text{Proj}_D\langle \mathbf{K}, \mathbf{e}(t)\rangle$. Solid and dashed lines in the first two rows indicate $e_1(t)$ and $e_2(t)$, respectively. In contrast, solid and dashed lines in the third row depict $u(t)$ and $v(t)$, respectively. The phase portrait of the system is depicted in the last row where red and yellow lines are used to illustrate system response to $u(t)$ and $v(t)$, respectively. The contour lines indicate Hamiltonian isoclines on Σ_L and Σ_R. The $\times$ and o markers are used to illustrate the singular points. Please refer to discussions in Section 2.2 and Figure 2.1 for more detailed description of the depicted phase planes.

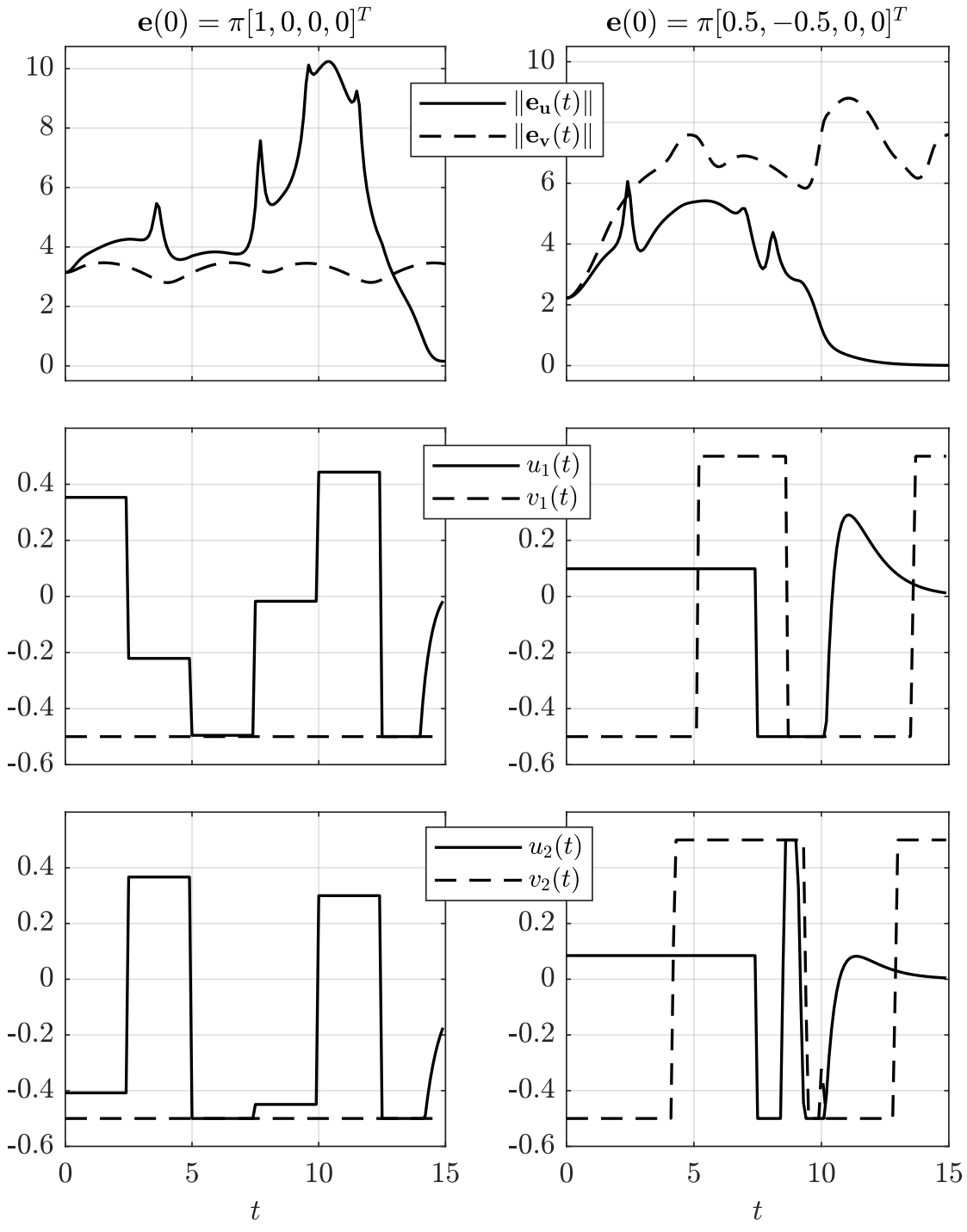

Figure 4.3: Response of the double pendulum system from two different initial conditions to both the synthesized control function based on Method 4.2.1, $\mathbf{u}$ and the projected linear control function $\mathbf{v}$.

4.2.4 Cart-pole

As our last example, we use Method 4.2.1 to construct control functions for the cart-pole system illustrated in Figure 3.4. Similar to the previous examples, we will refer to the discussion in Chapter 3 for problem setup and design of linear controllers. Here, we use the same system parameters, constraint set D and projected linear controller v as defined in Section 3.3.3.

The result of employing Method 4.2.1 to construct control functions for the cart-pole system for four different initial conditions are depicted in Figure 4.4. As shown in the figure, Method 4.2.1 can successfully synthesize control functions that regulates $\mathbf{e}_{(u,\mathbf{e}_0)}$ for all considered initial condition. However, similar to the results presented in Figure 3.5, for $\mathbf{e}_0 \in H$ where

$$H := \left\{[0,\, 1,\, 0,\, 0]^T,\, [0,\, 0,\, 1,\, 0]^T,\, [0,\, 0,\, 0,\, 1]^T\right\}, \tag{4.24}$$

the saturated linear control function, v, drives system to instability and causes norm of the error to increase in time. Noting that $v(t)$ saturates at ∂D for $\mathbf{e}_0 \in H$ while $\|\mathbf{e}_(v, \mathbf{e}_0 \in H)(t)\|$ increases in time, and since $\|\mathbf{e}_(u, \mathbf{e}_0 \in H)(t)\| \to 0$ in time, suggest that $H \subset E_N$ for the corresponding regularization problem. For the cart-pole example, the final values of N for which the `while` loop is terminated are $N_f = 0$ for $\mathbf{e}(0) = [1,\, 0,\, 0,\, 0]^T$ (which indicates that v can regulate the problem from this initial state), $N_f = 3$ for $\mathbf{e}(0) = [0,\, 1,\, 0,\, 0]^T$, $N_f = 10$ for $\mathbf{e}(0) = [0,\, 0,\, 1,\, 0]^T$ and $N_f = 3$ for $\mathbf{e}(0) = [0,\, 0,\, 0,\, 1]^T$.

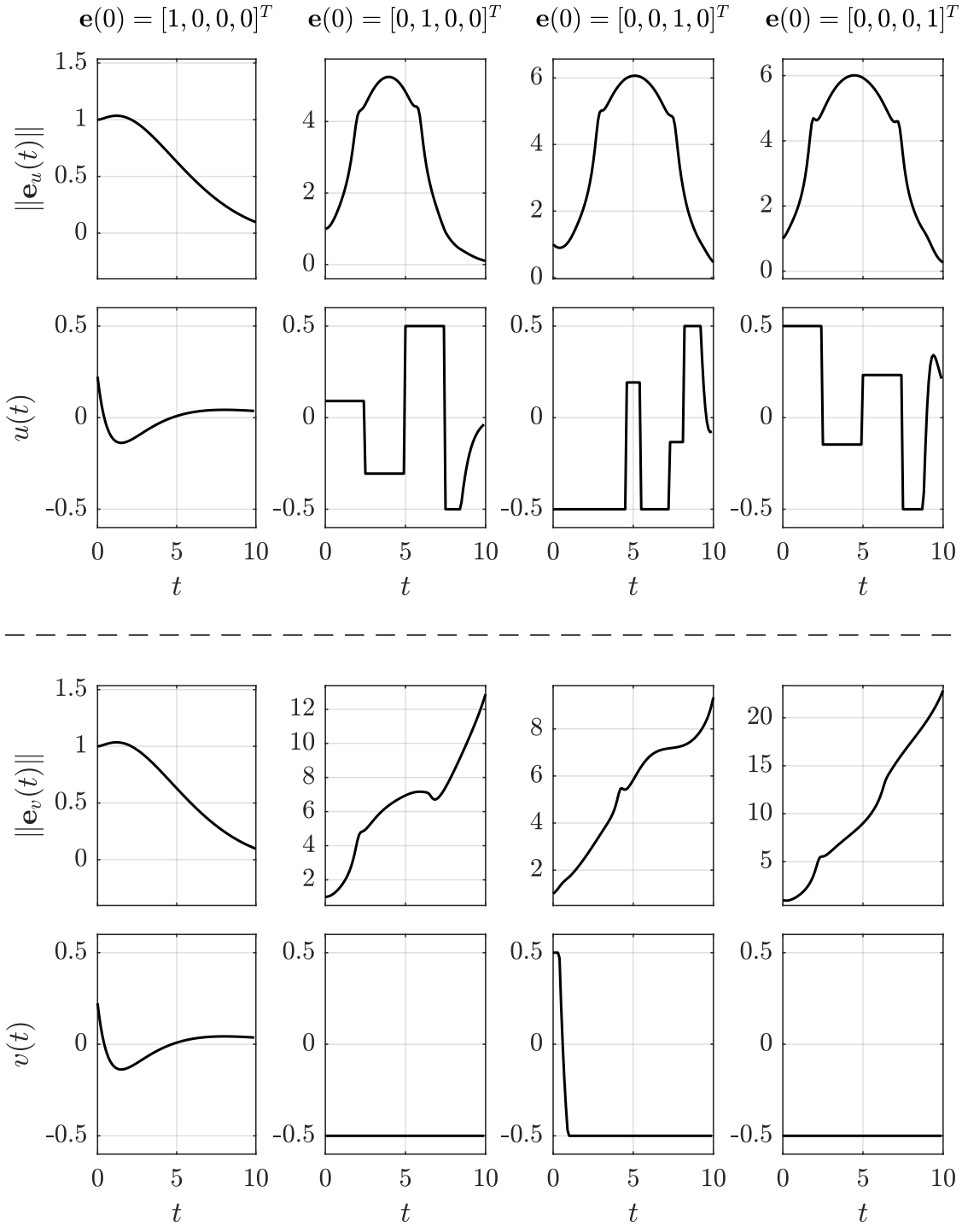

Figure 4.4: Response of the point-mass cart-pole system from four different initial conditions to both the synthesized control function based on Method 4.2.1, u, and the projected linear control function v.

4.3 Remarks and conclusions

In this chapter we studied possibility of synthesizing control functions for conditionally controllable problems by cascading a projected linear controller with a piecewise constant function. The parameters of the piecewise function are obtained by minimizing norm of the error vector at a specified finite time. The discretization used to define the piecewise function is incrementally refined until the solution satisfies a convergence criterion. In particular we set the convergence criterion as keeping the norm of the error vector below a specified threshold for a predefined time interval. The preliminary results obtained for the examined case studies shows the effectiveness of our proposed method in synthesizing control functions for systems with relatively complex dynamics. It must be noted that, utilizing projected linear controllers and model predictive nature of the algorithm can lead to more robust solutions in comparison to classical open-loop controllers derived in generic model-driven approaches. From the extermination with different time horizons it is also observed that the choice of final time, t_f, can affect the maximum number of intervals at the while loop termination, N_f. In all the case studies presented, we have used $N_0 = 0$ and $\Delta N = 1$. However, these values could be assigned based on the characteristics of a given system to reduce the number of iterations of the algorithm. Clearly, based on the definition of the objective function, the answer to the optimization problem that is solved at each iteration is not unique. Thus, utilizing a different objective function and convergence criterion may lead to more effective solutions, which could serve as a possible extension of this study.

Chapter 5

Conclusions

In this research we have explored two methods based on planning-based and optimal control theories to construct algorithms that can synthesize control functions for conditionally controllable problems. However, details in the proof of Theorem 2.2.3 suggests a strong relation between conditional controllability of a problem and existence of local extrema in the energy function. In particular, for the pendulum cases study presented in Section 2.2, the regions for which the projected linear controller fails to satisfy the regularization problem objective coincides with the domains around local minima of the system's Hamiltonian. In this regard, prior to finalizing the discussions of this manuscript, we present a brief introduction to energy-based control, where the objective of the control function is to regulate the energy of the system rather than the states. Moreover, we present an application of the energy based control in finding a control function for the pendulum example of Section 2.2. The concluding remarks of this research are presented in the last section of this chapter.

5.1 Energy-based control

Owing to their direct relation with physical systems, energy-based control approaches are relatively popular in synthesis of nonlinear or hybrid controllers for systems with reach dynamics such as control of cart-pole [31, 32] and multi-link pendulums [33, 34]. An extension of the method is used to synthesize control functions for aircraft automatic landing problem [35]. In what follows we present a brief introduction to energy-based control formulation and proceed with applying the method to solve the pendulum example of Section 2.2. Noting that the general approach in energy-based control is to regulate the energy of the system rather than directly controlling the state vector. Thus, to be more specific in the derivations, we will focus our discussion only on energy-based control application in rigid-body dynamics.

5.1.1 Energy dynamics in Lagrangian systems

Consider a generic form of equations of motion for an unconstrained Lagrangian system [8] with $d \in \mathbb{N}$ degrees of freedom

$$\mathbf{M}\big(\mathbf{q}(t), \dot{\mathbf{q}}(t)\big)\ddot{\mathbf{q}}(t) + \mathbf{C}\big(\mathbf{q}(t), \dot{\mathbf{q}}(t)\big)\dot{\mathbf{q}}(t) + \frac{\partial E_p\big(\mathbf{q}(t)\big)}{\partial \mathbf{q}(t)} = \mathbf{T}\big(\mathbf{q}(t)\big)\mathbf{u}(t), \tag{5.1}$$

where $\mathbf{q} : \mathbb{R} \to \mathbb{R}^d$ is the trajectory of generalized coordinates in time and respectively, $\dot{\mathbf{q}}(t) := \frac{d}{dt}\mathbf{q}(t)$ and $\ddot{\mathbf{q}}(t) := \frac{d}{dt}\dot{\mathbf{q}}(t)$ denote the trajectories of generalized velocities and accelerations. $E_p : \mathbb{R}^d \to \mathbb{R}$ maps every $\mathbf{q}(t)$ to the potential energy of the system. $\mathbf{M}\big(\mathbf{q}(t), \dot{\mathbf{q}}(t)\big) \in \mathbb{R}^{d\times d}$ is a symmetric positive definite matrix, which is commonly referred to as the mass matrix. $\mathbf{C}\big(\mathbf{q}(t), \dot{\mathbf{q}}(t)\big) \in \mathbb{R}^{d\times d}$ encloses Coriolis and centrifugal terms[1]. $\mathbf{T}\big(\mathbf{q}(t)\big) \in \mathbb{R}^{d\times m}$ maps input vector $\mathbf{u}(t) \in \mathbb{R}^m$ to generalized

[1]Since $\mathbf{C}\big(\mathbf{q}(t), \dot{\mathbf{q}}(t)\big)$ is derived from the Lagrangian of the system, it does not include frictional terms. The effect of friction could be captured by the terms in $\mathbf{T}\big(\mathbf{q}(t)\big)$ and as some added terms to the input function $\mathbf{u}(t)$.

forces applied to the system. In order to convert (5.1) to a set of first-order differential equations, we can define $\mathbf{y}(t) = [\mathbf{y}_1(t),\, \mathbf{y}_2(t)]^T := [\mathbf{q}(t),\, \dot{\mathbf{q}}(t)]^T$ that allows us to write (5.1) as

$$\begin{bmatrix} \dot{\mathbf{y}}_1(t) \\ \dot{\mathbf{y}}_2(t) \end{bmatrix} = \begin{bmatrix} \mathbf{y}_2(t) \\ \mathbf{\Phi}_1\big(\mathbf{y}(t)\big) + \mathbf{\Phi}_2\big(\mathbf{y}(t)\big)\mathbf{u}(t) \end{bmatrix}, \tag{5.2}$$

where

$$\mathbf{\Phi}_1\big(\mathbf{y}(t)\big) := -\mathbf{M}^{-1}\big(\mathbf{y}_1(t), \mathbf{y}_2(t)\big)\left(\mathbf{C}\big(\mathbf{y}_1(t), \mathbf{y}_2(t)\big)\mathbf{y}_2(t) + \frac{\partial E_p\big(\mathbf{y}_1(t)\big)}{\partial \mathbf{y}_1(t)}\right), \tag{5.3}$$

$$\mathbf{\Phi}_2\big(\mathbf{y}(t), \mathbf{u}(t)\big) := \mathbf{M}^{-1}\big(\mathbf{y}_1(t), \mathbf{y}_2(t)\big)\mathbf{T}\big(\mathbf{y}_1(t)\big). \tag{5.4}$$

The total energy of a system, that is the sum of kinetic and potential energies, is

$$E\big(\mathbf{y}(t)\big) = E_k\big(\mathbf{y}(t)\big) + E_p\big(\mathbf{y}_1(t)\big) = \frac{1}{2}\big\langle \mathbf{M}\big(\mathbf{y}_1(t), \mathbf{y}_2(t)\big)\mathbf{y}_2(t), \mathbf{y}_2(t)\big\rangle + E_p(\mathbf{y}_1(t)). \tag{5.5}$$

In the above equation, $E_k : \mathbb{R}^{2d} \to \mathbb{R}$ and $E_p : \mathbb{R}^{2d} \to \mathbb{R}$ denote the total kinetic and potential energies of the system, respectively. Taking derivative of $E(\mathbf{y}(t))$ with respect to time leads to

$$\begin{aligned} \dot{E}(\mathbf{y}(t)) &= \big\langle \mathbf{M}\big(\mathbf{y}_1(t), \mathbf{y}_2(t)\big)\dot{\mathbf{y}}_2(t), \mathbf{y}_2(t)\big\rangle + \frac{1}{2}\big\langle \dot{\mathbf{M}}\big(\mathbf{y}_1(t), \mathbf{y}_2(t)\big)\mathbf{y}_2(t),\, \mathbf{y}_2(t)\big\rangle \\ &\quad + \big\langle \frac{\partial E_p\big(\mathbf{y}_1(t)\big)}{\partial \mathbf{y}_1(t)},\, \mathbf{y}_2(t)\big\rangle \\ &= \big\langle \mathbf{T}\big(\mathbf{y}_1(t)\big)\mathbf{u}(t) - \mathbf{C}\big(\mathbf{y}_1(t), \mathbf{y}_2(t)\big)\mathbf{y}_2(t) - \frac{\partial E_p\big(\mathbf{y}_1(t)\big)}{\partial \mathbf{y}_1(t)},\, \mathbf{y}_2(t)\big\rangle \\ &\quad + \frac{1}{2}\big\langle \dot{\mathbf{M}}\big(\mathbf{y}_1(t), \mathbf{y}_2(t)\big)\mathbf{y}_2(t),\, \mathbf{y}_2(t)\big\rangle + \big\langle \frac{\partial E_p\big(\mathbf{y}_1(t)\big)}{\partial \mathbf{y}_1(t)},\, \mathbf{y}_2(t)\big\rangle \\ &= \big\langle \mathbf{T}\big(\mathbf{y}_1(t)\big)\mathbf{u}(t), \mathbf{y}_2(t)\big\rangle \\ &\quad + \frac{1}{2}\Big\langle \Big(\dot{\mathbf{M}}\big(\mathbf{y}_1(t), \mathbf{y}_2(t)\big) - 2\mathbf{C}\big(\mathbf{y}_1(t), \mathbf{y}_2(t)\big)\Big)\mathbf{y}_2(t),\, \mathbf{y}_2(t)\Big\rangle. \end{aligned} \tag{5.6}$$

In [36], it is shown that $\dot{\mathbf{M}}\big(\mathbf{y}_1(t), \mathbf{y}_2(t)\big) - 2\mathbf{C}\big(\mathbf{y}_1(t), \mathbf{y}_2(t)\big)$ is a skew-symmetric matrix. Moreover, since all the right hand terms of (5.6) are in $\mathbb{R}$, we have

$$\left\langle \Big(\dot{\mathbf{M}}\big(\mathbf{y}_1(t), \mathbf{y}_2(t)\big) - 2\mathbf{C}\big(\mathbf{y}_1(t), \mathbf{y}_2(t)\big)\Big)\mathbf{y}_2(t),\, \mathbf{y}_2(t) \right\rangle = 0 \tag{5.7}$$

Consequently, the time derivative of the total energy of the system simplifies to

$$\dot{E}(\mathbf{y}(t)) = \langle \mathbf{T}\big(\mathbf{y}_1(t)\big)\mathbf{u}(t), \mathbf{y}_2(t) \rangle. \tag{5.8}$$

5.1.2 Derivation of an energy-based controller

The results obtained in previous subsection indicates $\dot{E}\big(\mathbf{y}(t)\big)$ is a linear function of $\mathbf{T}(\mathbf{y}_1(t))\mathbf{u}(t)$ with a time varying gain $\mathbf{y}_2(t)$. Accordingly, a control strategy for $E\big(\mathbf{y}(t)\big)$ could be obtained using Lyapunov method [37] with a Lyapunov function candidate

$$V\big(\mathbf{y}(t)\big) := \frac{1}{2}\Big(E\big(\mathbf{y}(t)\big) - E_d\Big)^2, \tag{5.9}$$

for some desired energy $E_d \in \mathbb{R}$. Based on the definition, $V\big(\mathbf{y}(t)\big)$ is positive for every $E\big(\mathbf{y}(t)\big) \in \mathbb{R}$ and is zero when $E\big(\mathbf{y}(t)\big) = E_d$. Taking the derivative of $V\big(\mathbf{y}(t)\big)$ with respect to t yields

$$\dot{V}\big(\mathbf{y}(t)\big) = \Big(E\big(\mathbf{y}(t)\big) - E_d\Big)\dot{E}\big(\mathbf{y}(t)\big) = \Big(E\big(\mathbf{y}(t)\big) - E_d\Big)\Big\langle \mathbf{T}\big(\mathbf{y}_1(t)\big)\mathbf{u}(t), \mathbf{y}_2(t) \Big\rangle. \tag{5.10}$$

If we can chose $\mathbf{u}$ such that it satisfies:

i) $V\big(\mathbf{y}(t)\big) \geq 0$, for all $\mathbf{y}(t) \in \mathbb{R}^n$,

ii) $V\big(\mathbf{y}(t)\big) = 0 \implies \mathbf{y}(t) = \mathbf{0}$,

iii) $\dot{V}\big(\mathbf{y}(t)\big) < 0$, for all $\mathbf{y}(t) \in \mathbb{R}^n$,

then, based on Lyapunov's second method for stability [38], $\lim_{t\to\infty} V(\mathbf{y}(t)) = 0$. Moreover, based on the definition of V, $V(\mathbf{y}(t)) \to 0$ implies $|E(\mathbf{y}(t)) - E_d| \to 0$.

We must also point out that, as seen in (5.2), at every equilibrium point of the system $\mathbf{y}_2(t)$ must be zero. Consequently, regardless of choice of $\mathbf{u}$, we loose controllability of $\dot{E}$ at an equilibrium point. Such limitation of energy-based control demands a special attention when it is used as an state regulator. For further discussions on energy-based control and related derivations please see [31] and [39].

5.1.3 Application to the pendulum example

To implement an energy-based controller on the pendulum example of Section 2.2, we first need to define the mechanical energy of the pendulum as a summation of its kinetic and potential energies. Since the upward configuration of the pendulum corresponds to $y_1(t) = 0$, to simplify the equations, we can define the potential energy of the system as $E_p(t) = \cos(y_1(t)) - 1$, which is zero when $y_1(t) = 0$. Consequently, the energy function $E(t)$ and its time derivative $\dot{E}(t)$ are

$$E(t) = \frac{1}{2}y_2^2(t) + \cos(y_1(t)) - 1, \tag{5.11}$$

$$\dot{E}(t) = u(t)y_2(t). \tag{5.12}$$

Since $\mathbf{y}_d = \mathbf{0}$, then $\mathbf{e}(t) = -\mathbf{y}(t)$. Substituting $y_1(t) = -e_1(t)$ and $y_2(t) = -e_2(t)$ in (5.11) leads to

$$E(t) = \frac{1}{2}e_2^2(t) + \cos(e_1(t)) - 1, \tag{5.13}$$

$$\dot{E}(t) = -u(t)e_2(t). \tag{5.14}$$

Noting that $E(t) = 0$ if $\mathbf{e} = \mathbf{0}$, we let the desired energy value $E_d := 0$. Based on the discussion presented on Section 5.1.2, we proceed as the following

$$V(t) := \frac{1}{2}\big(E(t) - E_d\big)^2 \implies \dot{V}(t) = -\big(E(t) - E_d\big)e_2(t)u(t). \tag{5.15}$$

To satisfy the necessary conditions for Lyapunov's second method for stability, we define

$$u(t) := k\big(E(t) - E_d\big)\,\mathrm{sign}\,\big(e_2(t)\big), \tag{5.16}$$

for some $k > 0$. Consequently $\dot{V}(t)$ simplifies to

$$\dot{V}(t) = -k\big(E(t) - E_d\big)^2 e_2(t)\,\mathrm{sign}\,\big(e_2(t)\big) \leq 0,\ \forall t \in \mathbb{R}. \tag{5.17}$$

It is clear that the range of u as defined in (5.16) is $\mathbb{R}$. Thus, to limit the range of u to $D = [-w_{max},\, w_{max}] \subset \mathbb{R}$, we define $z(t)$ as

$$z(t) := \mathrm{sign}\,\big(e_2(t)\big) \underset{D}{\mathrm{sat}}\, k\big(E(t) - E_d\big). \tag{5.18}$$

Since for any $x \in \mathbb{R}$, $\mathrm{sign}(\mathrm{sat}_D(x)) = \mathrm{sign}(x)$, then by using the control function $z(t)$ we can still satisfy the sign requirement of $\dot{V}(t)$. Consequently, the energy of the pendulum system endowed with $z(t)$ must converge to zero in time. However, since the mapping from $\mathbf{e}(t)$ to $E(t)$ is not injective, there exists a subset in $\mathbb{R}^2$ such that $E(t) = 0$. As a result, simply regulating $E(t)$ will not be sufficient to regulate the error. To address this issue, we can combine $z(t)$ with projected linear controller $v(t) := \mathrm{Proj}_D\, \mathbf{K}\mathbf{e}(t)$, with $K \in \Omega$ as defined in (2.19) to obtain a new

control function $\hat{z}(t)$ defined as

$$\hat{z}(t) := \begin{cases} v(t), & \text{if } \|\mathbf{e}(t)\| < \epsilon, \\ z(t), & \text{otherwise,} \end{cases} \tag{5.19}$$

where $\epsilon > 0$ is defined based on a desired performance and it depends on w_{max}. The simulation results of the pendulum system endowed with control function $\hat{z}$ with $k = 1$ and $\mathbf{K} = [5,\, 2]^T$ is illustrated in Figure 5.1. As depicted in the figure, even with the imposed constraint on the range of the control function, $\hat{z}$ is capable of achieving the control objective from both initial conditions $\mathbf{e}_0 = [-2,\, 0]^T \in E_N$ and $\mathbf{e}_0 = [-4,\, 2]^T \in E \setminus E_N$.

5.2 Concluding remarks

In pursuit of finding an automatic approach to construct control functions capable of satisfying control objectives through utilization of inherent system nonlinearities, we explored performance of linear functions when used to solve nonlinear control problems with limited input range. In this regard, we started with developing a formal setting to identify a sub class of control problems that are solvable via fusing linear and nonlinear controllers.

We dedicated the discussions in Chapter 2 to this matter and proposed our definition of conditionally controllable problem as a subclass of regularization problems where the objective of control is to ensure convergence of state vector to a given trajectory. Furthermore, we explored the effect of system nonlinearities and limited range of control functions on convergence of the error for a simple point-mass pendulum system. In addition, we formally proved that for a specific allowable torque range the problem is conditionally controllable.

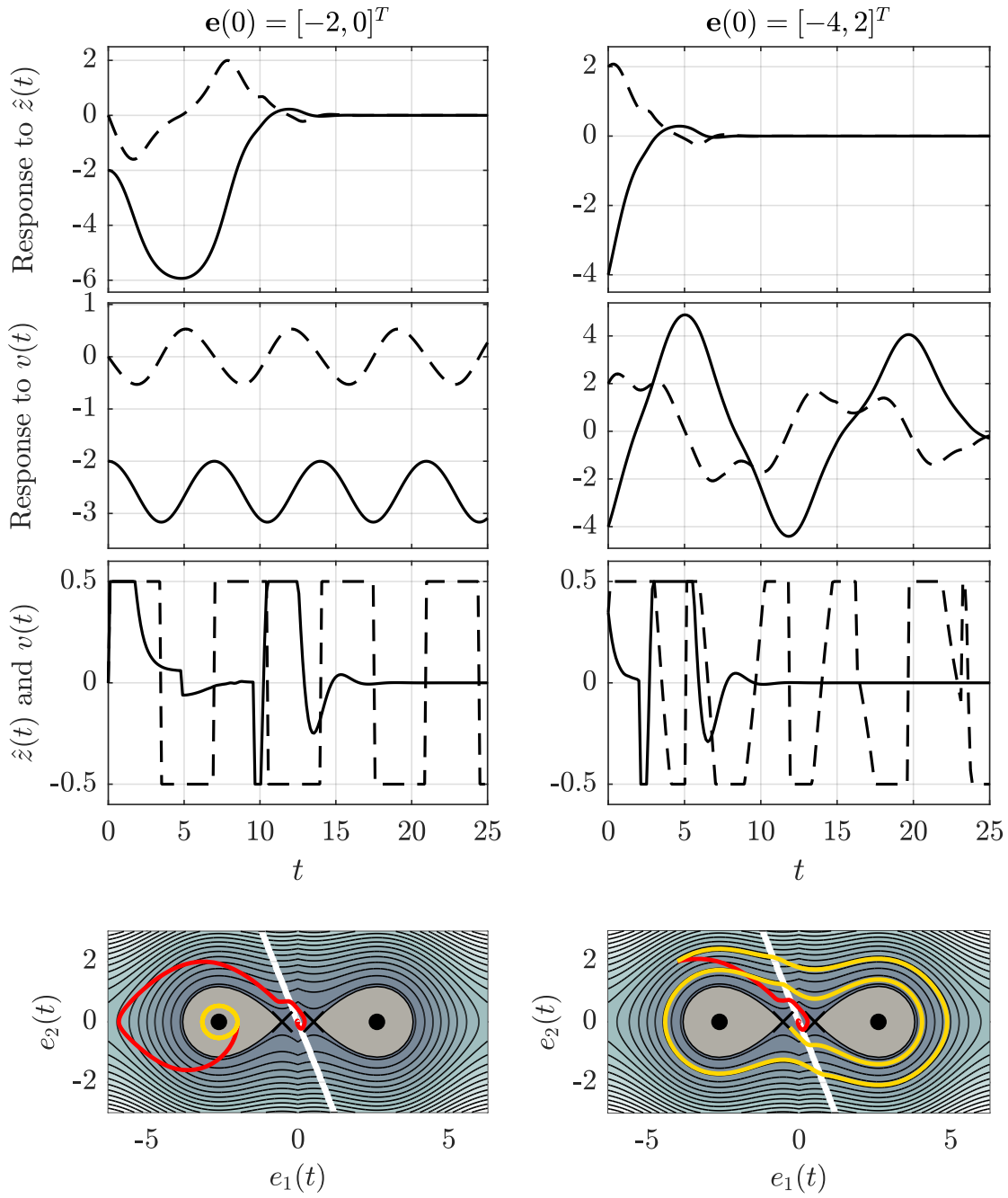

Figure 5.1: The pendulum system endowed with control function $\hat{z}$ as defined in (5.19). Solid and dashed lines in the first two rows indicate $e_1(t)$ and $e_2(t)$, respectively. In contrast, solid and dashed lines in the third row depict $\hat{z}(t)$ and $v(t)$, respectively. The phase portrait of the system is depicted in the last row where red and yellow lines are used to illustrate system response to $\hat{z}(t)$ and $v(t)$, respectively. The contour lines indicate Hamiltonian isoclines on Σ_L and Σ_R. The $\times$ and o markers are used to illustrate the singular points. Please refer to discussions in Section 2.2 and Figure 2.1 for more detailed description of the depicted phase planes.

Given the proposed definition and the corresponding sub class of control problems, we followed our discussions with two possible algorithms to synthesize control functions for conditionally controllable problems. The first algorithm, presented in Chapter 3, utilizes spatio-temporal exploring trees, as an extension of spatial exploring trees, to synthesize a control function for a given problem from a defined initial state. We explored the effectiveness of our proposed method in synthesizing control functions for three case study systems: pendulum, double pendulum and cart-pole. These systems are particularly used as case studies in control literature due to their relatively simple yet rich dynamics that resemble the behavior of many practical systems. Moreover, in order to build a foundation for our proposed algorithm and possible future developments, Chapter 3 includes a formal setting for exploring trees in normed vector spaces, their extension to include time, and their application in solving control problems.

In Chapter 4 we focused our discussion on the theory of optimal control, and following a short introduction on the subject, we proposed an algorithms to construct control function for conditionally controllable problems as an application of direct methods. It must be noted that our objective is to find a solution to conditionally controllable problems and not an optimal trajectory for the given problem. Accordingly, we have used a discretization in time to combine a piecewise function with a projected linear controller to synthesize a solution. The proposed time grid is refined at each iteration until the obtained control function satisfies a given convergence criterion. We have tested the effectiveness of the method with the same case studies that are used in Chapter 3. As a short comparison between the planning-based method presented in Chapter 3 and optimization based method of Chapter 4 we can state that the extra parameters associated with Method 3.3.1 increases the complexities of controller design and the problem needs to be examined to find pa-

rameter values that increases the efficiency of the solver. In contrast, Method 4.2.1 has fewer free parameters to be assigned. On a different note, the optimization procedure of Method 4.2.1 must be adjusted based on the topology of the set of feasible control inputs, namely the set D. However, Method 3.3.1 could be directly used for various systems with different structures of the D. It must be noted that, by using a different convergence criterion and objective function, one can adjust Method 4.2.1 to also solve for an optimal solution (that depends on the refinement used on the time grid) in parallel to solving the original problem. Such characteristic could be favorable in specific application.

Future work and research directions

As details of the proof for conditional controllability of the pendulum problem suggests, there appears to be a strong relation between existence of local minima in the Hamiltonian function and conditional controllability of a mechanical system. Exploring such dependence and deriving analytical expressions for necessary conditions of conditional controllability via energy functions could serve as a possible future work of this research. Such analysis can significantly simplify the process of synthesizing control functions for the problems and allow design of more elegant controllers. Accordingly, to explore more with the utilizing energy in control of conditionally controllable problem, we took a minor detour in Chapter 5 to discuss energy-based control and its application to the pendulum example before concluding the manuscript. However, we must note that the energy function is a mapping from the space of generalized coordinates and velocities to the set of real numbers and, consequently, it is not a injective map. Thus, implementation of energy based controller, in the form presented here, for systems with multiple degrees of freedom may not necessarily lead to an appropriate solution for conditionally controllable

problems. This is due to the fact that there may exist multiple regions in the set of feasible state vectors with the same energy level. If one of such regions lays within the set of states for which the projected linear controller fails to regulate the system, it may not be possible to direct trajectories toward the goal. Thus, the possibilities of using energy-based control in synthesizing controllers for conditionally controllable problems remains as an open question to be explored.

Appendix A

Intermediate lemmas used to prove conditional controllability of the pendulum example

Lemma A.0.1. *Let* $\mathbf{k} \in \Omega$. *If* $\mathbf{e_k} \in \Gamma$ *for all* $t \geq t_0$, *then* $\lim_{t\to\infty} \|\mathbf{e_k}(t)\| = 0$.

Proof. Since for every $\mathbf{k} \in \Omega_1$, $k_1 > 1$, then $E_\Gamma(\mathbf{e}(t))$ is convex and its minimum is located at $\mathbf{e}(t) = \mathbf{0}$. Thus, $E_\Gamma(\mathbf{e}(t))$ satisfies the conditions required for a Lyapunov function candidate. Accordingly let Lyapunov function $V(t) = E_\Gamma(\mathbf{e}(t))$, then

$$\dot{V}(t) = \dot{E}_\Gamma(\mathbf{e}(t)) = -k_2 e_2^2(t).$$

Since $\dot{V}(t) < 0$ for all $e_1(t)$ and $e_2(t) \neq 0$ and $\dot{V}(t) = 0$ when $e_2(t) = 0$, then $V(t)$ will decrease for all $\mathbf{e}(t)$ such that $e_2(t) \neq 0$. In addition, since $\mathbf{e}(t) = \mathbf{0}$ is the only equilibrium point of the system in Γ, then for $e_2(t) = 0$, $e_1(t) \neq 0 \implies \dot{e}_2(t) \neq 0$. Thus, $\lim_{t\to\infty} V(t) = \lim_{t\to\infty} E(\mathbf{e}(t)) = 0$ that implies $\lim_{t\to\infty} \|\mathbf{e}(t)\| = 0$. □

Lemma A.0.2. *Let* $\mathbf{k} \in \Omega$, $w_{\max} > 0$ *and* $\mathbf{e_k}(t_0) \in \Sigma_L \cap \Gamma$. *If* $\exists t_1 > t_0$, *such that*

$\mathbf{e_k}(t) \in \Gamma$ *for* $t \in [t_0, t_1]$ *and* $\mathbf{e_k}(t_1) \in \Gamma \cap \Sigma_R$*, then* $H_{\Sigma_L}(\mathbf{e_k}(t_0)) > H_{\Sigma_R}(\mathbf{e_k}(t_1))$.

Proof. Let $\mathbf{a} := \mathbf{e_k}(t_0)$ and $\mathbf{b} := \mathbf{e_k}(t_1)$. First we show that if $\exists t_1 > t_0$, such that $\mathbf{b} \in \Gamma \cap \Sigma_R$, then $b_2 \geq 0$ and $|a_1| > |b_1|$. Our first claim is that if $\mathbf{a} \in \Sigma_L \cap \Gamma$ and $\mathbf{e_k}(t) \in \Gamma$ for $t \in [t_0, t_1]$, then $a_2 \geq 0$. Since $a_2 < 0 \implies \dot{a}_1 < 0 \implies a_1(t_0^+) < a_1(t_0)$ and consequently $a_1(t_0^+) \in \Sigma_L \setminus \Gamma$, which is a contradiction. To show that $b_2 \geq 0$, by contradiction, assume that $b_2 < 0$. For $w_{max} > 0$, $\Sigma_L \cap \Sigma_R = \emptyset$ and for every $\mathbf{a} = [a_1, 0]^T \in \Gamma \cap \Sigma_L$ with $\mathbf{k} \in \Omega$, $\dot{a}_2 > 0$. Thus for every $a_2 \geq 0$, there must be a time $t_c \in (t_0, t_1)$ at which $(e_2)_{\mathbf{k}}(t_c) = 0$ and $(e_2)_{\mathbf{k}}(t) < 0$ for $t \in (t_c, t_1]$. Accordingly, there are three possible cases for $(e_1)_{\mathbf{k}}(t_c)$; that are:

(i) $(e_1)_{\mathbf{k}}(t_c) > \frac{w_{max}}{k_1}$: that implies $\mathbf{e_k}(t_c) \notin \Gamma$, which is a contradiction with $\mathbf{e_k}(t) \in \Gamma$ for $t \in [t_0, t_1]$.

(ii) $-\frac{w_{max}}{k_1} < (e_1)_{\mathbf{k}}(t_c) < \frac{w_{max}}{k_1}$: then for every $t \in (t_c, t_1]$, $(e_2)_{\mathbf{k}}(t) = (\dot{e}_1)_{\mathbf{k}}(t) < 0$, that implies $(e_1)_{\mathbf{k}}(t) < (e_1)_{\mathbf{k}}(t_c)$ for all $t \in (t_c, t_1]$. Since for every $\mathbf{b} \in \Gamma \cap \Sigma_R$ with $b_2 < 0$, $b_1 > \frac{w_{max}}{k_1}$; thus, $\mathbf{e_k}(t_1) \notin \Gamma \cap \Sigma_R$, which is contradiction with the main assumption.

(iii) $(e_1)_{\mathbf{k}}(t_c) = \frac{w_{max}}{k_1}$: then $(e_1)_{\mathbf{k}}(t_c) \in \Gamma \cap \Sigma_R$ and $t_1 = t_c$ which is a contradiction with $t_c \in (t_0, t_1)$.

Consequently, $b_2 \geq 0$. To show that $|a_1| > |b_1|$ we will proceed as the following. Since $(e_2)_{\mathbf{k}}(t) = (\dot{e}_1)_{\mathbf{k}}(t) \geq 0$ for $t \in [t_0, t_1]$, then $b_1 \geq a_1$. In addition, $a_2 \geq 0 \implies a_1 \leq -\frac{w_{max}}{k_1} < 0$ and $b_2 \geq 0 \implies b_1 \leq \frac{w_{max}}{k_1}$. Since $\dot{E}_\Gamma(\mathbf{e_k}(t)) < 0$ for all $t \in (t_0, t_1)$, then $\mathbf{a} = [-\frac{w_{max}}{k_1}, 0]^T \implies \mathbf{b} \neq [\frac{w_{max}}{k_1}, 0]^T$ (otherwise for $\mathbf{a} \neq \mathbf{b}$, $E_\Gamma(\mathbf{a}) = E_\Gamma(\mathbf{b})$ which is a contradiction with $\dot{E}_\Gamma(\mathbf{e_k}(t)) < 0$ for all $t \in (t_0, t_1)$). Thus, $b_1 < \frac{w_{max}}{k_1}$. Based on the above conditions, $b_1 \in (a_1, \frac{w_{max}}{k_1})$ and since $a_1 \in (-\infty, -\frac{w_{max}}{k_1}]$, then $|a_1| > |b_1|$.

Now we can use the obtained conditions on $\mathbf{a}$ and $\mathbf{b}$ to prove the claim of the lemma. Note that to show $H_{\Sigma_L}(\mathbf{a}) > H_{\Sigma_L}(\mathbf{b})$, it suffices to show that the difference

$$H_{\Sigma_L}(\mathbf{a}) - H_{\Sigma_L}(\mathbf{b}) = \frac{1}{2}a_2^2 + \cos(a_1) - \frac{1}{2}b_2^2 - \cos(b_1) - w_{max}(a_1 + b_1),$$

is positive. Let $H_{\Sigma_L}(\mathbf{a}) - H_{\Sigma_L}(\mathbf{b}) = h_1(\mathbf{a}, \mathbf{b}) + h_2(\mathbf{a}, \mathbf{b})$, where

$$\begin{aligned} h_1(\mathbf{a}, \mathbf{b}) &:= \frac{1}{2}a_2^2 + \cos(a_1) - \frac{1}{2}b_2^2 - \cos(b_1), \\ h_2(\mathbf{a}, \mathbf{b}) &:= -w_{max}(a_1 + b_1). \end{aligned}$$

Since the trajectory $\mathbf{e_k}(t) \in \Gamma$ for $t \in [t_0, t_1]$, then for every $t \in [t_0, t_1]$, we have $\dot{E}_\Gamma(t) \leq 0$ that implies $E_\Gamma(t) \geq E_\Gamma(t_0)$. Thus

$$E_\Gamma(\mathbf{a}) \geq E_\Gamma(\mathbf{b}) \implies \cos(b_1) - \cos(a_1) \leq \frac{1}{2}\Big(a_2^2 + k_1 a_1^2 - b_2^2 - k_1 b_1^2\Big). \qquad (*)$$

To show that $h_1(\mathbf{a}, \mathbf{b}) > 0$, we need to have

$$h_1(\mathbf{a}, \mathbf{b}) > 0 \implies \frac{a_2^2}{2} + \cos(a_1) > \frac{b_2^2}{2} + \cos(b_1) \implies \cos(b_1) - \cos(a_1) < \frac{a_2^2 - b_2^2}{2}.$$

Based on inequality $(*)$, $h_1(\mathbf{a}, \mathbf{b}) > 0$ if $a_2^2 - b_2^2 < a_2^2 + k_1 a_1^2 - b_2^2 - k_1 b_1^2$, that implies

$$0 < k_1(a_1^2 - b_1^2).$$

Since $k_1 > 1$ and $|a_1| > |b_1|$, then the above inequality is satisfied and consequently, $h_1(\mathbf{a}, \mathbf{b}) > 0$. To show that $h_2(\mathbf{a}, \mathbf{b}) > 0$, we know that $a_1 < 0$ and $|a_1| > |b_1|$ that implies $a_1 + b_1 < 0$. Thus, $h_2(\mathbf{a}, \mathbf{b}) = -w_{max}(a_1 + b_1) > 0$. Since both $h_1(\mathbf{a}, \mathbf{b})$ and $h_2(\mathbf{a}, \mathbf{b})$ are positive, then $H_{\Sigma_L}(\mathbf{a}) - H_{\Sigma_L}(\mathbf{b}) = h_1(\mathbf{a}, \mathbf{b}) + h_2(\mathbf{a}, \mathbf{b}) > 0$. □

Lemma A.0.3. *Let* $\mathbf{k} \in \Omega$, $w_{max} > 0$ *and* $\mathbf{e_k}(t_0) \in \Sigma_R \cap \Gamma$. *If* $\exists t_1 > t_0$, *such that* $\mathbf{e_k}(t) \in \Gamma$ *for* $t \in [t_0, t_1]$ *and* $\mathbf{e_k}(t_1) \in \Gamma \cap \Sigma_L$, *then* $H_{\Sigma_R}(\mathbf{e_k}(t_0)) > H_{\Sigma_L}(\mathbf{e_k}(t_1))$.

Proof. Let $\mathbf{a} := \mathbf{e_k}(t_0)$ and $\mathbf{b} := \mathbf{e_k}(t_1)$. Similar to the proof presented for Lemma A.0.2, it is possible to show that if $\exists t_1 > t_0$, such that $\mathbf{b} \in \Gamma \cap \Sigma_L$, then a_2 and b_2 are both negative, $0 < \frac{w_{max}}{k_1} \leq a_1$ and $|a_1| > |b_1|$. Consequently $H_{\Sigma_L}(\mathbf{a}) - H_{\Sigma_L}(\mathbf{b}) = \frac{1}{2}a_2^2 + \cos(a_1) - \frac{1}{2}b_2^2 - \cos(b_1) + w_{max}(a_1 + b_1) > 0$. □

Lemma A.0.4. *Let* $\mathbf{k} \in \Omega$, $w_{max} \in (0, 1)$ *and* $j \in \mathbb{N} \cup \{0\}$. *If* $\mathbf{e_k}(t_0) \in Q_L^j$, *then* $\mathbf{e_k}(t) \in Q_L^j$ *for all* $t \geq t_0$.

Proof. In order to prove the claim of the lemma, it suffices to show that ∂Q_L^j forms a closed path in $\mathbb{R}^2$. This is trivial since H_{Σ_L} only increases in e_2 direction. Moreover, $H_{\Sigma_L}(\mathbf{e}(t))$ has a minimum at $\mathbf{e}^*(t) = [-(2j+1)\pi + \sin^{-1}(w_{max}),\, 0]^T$. Also since

$$-2(j+1)\pi - \sin^{-1}(w_{max}) < e_1^*(t) < -2j\pi - \sin^{-1}(w_{max}), \tag{A.1}$$

and

$$H_{\Sigma_L}\Big(\big[-2(j+1)\pi - \sin^{-1}(w_{max}), 0\big]^T\Big) - H_{\Sigma_L}\Big(\big[-2j\pi - \sin^{-1}(w_{max}), 0\big]^T\Big) = 2\pi w_{max} > 0 \tag{A.2}$$

we get that the value of H_{Σ_L} increases up to $H_{\Sigma_L}\Big(\big[-2j\pi - \sin^{-1}(w_{max}), 0\big]^T\Big)$ which is located on the boundary of Q_L^j. Finally, based on the definition of Q_L^j, we have that ∂Q_L^j forms a closed path, since it coincides with a level set of H_{Σ_L}.

Noting that every trajectory in Σ_L coincides with a level set of H_{Σ_L}, we get that for $\mathbf{e_k}(t_0) \in Q_L^j$, then $\mathbf{e_k}(t) \in Q_L^j$ for all $t \geq t_0$.

□

Lemma A.0.5. *Let* $\mathbf{k} \in \Omega$, $w_{max} \in (0, 1)$ *and* $j \in \mathbb{N} \cup \{0\}$. *If* $\mathbf{e_k}(t_0) \in Q_R^j$, *then* $\mathbf{e_k}(t) \in Q_R^j$ *for all* $t \geq t_0$.

Proof. This lemma could be proved with similar steps as presented in the proof of Lemma A.0.5. □

Lemma A.0.6. *Let* $\mathbf{k} \in \Omega$ *and* $w_{max} \in (0,1)$. *If* $\mathbf{e_k}(t_0) \in \Sigma_L \setminus \bigcup_{j=0}^{\infty} Q_L^j$, *then* $\exists t_1 \geq t_0$ *such that* $\mathbf{e_k}(t_1) \in \Gamma$.

Proof. Throughout this proof, we will use the short notation $\mathbf{e}(t) = \mathbf{e_k}(t)$. By recalling the error dynamics in Σ_L, that is

$$\begin{bmatrix} \dot{e}_1(t) \\ \dot{e}_2(t) \end{bmatrix} = \begin{bmatrix} e_2(t) \\ \sin(e_1(t)) + w_{max} \end{bmatrix},$$

we can construct two sets a and b defined as

$$a := \{\boldsymbol{\xi} = [\xi_1, 0]^T \in \mathbb{R}^2 : \dot{\boldsymbol{\xi}}_2 = \sin(\xi_1) + w_{max} > 0\},$$
$$b := \{\boldsymbol{\xi} = [\xi_1, 0]^T \in \mathbb{R}^2 : \dot{\boldsymbol{\xi}}_2 = \sin(\xi_1) + w_{max} < 0\}.$$

Note that $a \cup b \cup S_L = \mathbb{R}^-$ and, as stated in (2.37), the set S_L contains all the stationary points of H_{Σ_L} and $\sin(e_1(t)) + w_{max} = 0$ if $\mathbf{e}(t) \in S_L$. Moreover, based on the definition of Q_L^j, we have $b \subset \bigcup_{j=0}^{\infty} Q_L^j$. Pick $\mathbf{e}(0) \in \{\boldsymbol{\xi} \in \Sigma_L \setminus \bigcup_{j=0}^{\infty} Q_L^j : \xi_2 > 0\}$, we can show that $e_2(t_0) > 0$ for $t \leq t_1$. Assume by contradiction that $\exists t_0 < t' < t_1$ such that $e_2(t') = 0$. Based on the sign of $\dot{e}_2(t)$, this point can only belong to $b \subset \bigcup_{j=0}^{\infty} Q_L^j$. However, since for every j, ∂Q_L^j forms a closed orbit, then there must be $t'' \in (t_0, t')$ such that $\mathbf{e}(t'') \in \partial Q_L^j$, which is a contradiction with the fact that in Σ_L, $\mathbf{e}(t)$ coincides with a level set of H_{Σ_L}. Consequently, for $\mathbf{e}(0) \in \{\boldsymbol{\xi} \in \Sigma_L \setminus \bigcup_{j=0}^{\infty} Q_L^j : \xi_2 > 0\}$ we have $\dot{e_1}(t) = e_2(t) > 0$ and since for such $\mathbf{e}(0)$, $e_1(0) < 0$, the value of $e_1(t)$ increases until it reaches to $\partial\Gamma$. Similarly, we can show that for every $\mathbf{e}(0) \in \{\boldsymbol{\xi} \in \Sigma_L \setminus \bigcup_{j=0}^{\infty} Q_L^j : \xi_2 < 0\}$, there must exist $t' > t_0$ such that $e_2(t') \in a \setminus \bigcup_{j=0}^{\infty} Q_L^j \subset \Sigma_L \setminus \bigcup_{j=0}^{\infty} Q_L^j$. Since $\dot{e}_2(t) > 0$ for $\mathbf{e}(t) \in a$, the

above argument applies. □

Lemma A.0.7. *Let* $\mathbf{k} \in \Omega$ *and* $w_{max} \in (0,1)$. *If* $\mathbf{e_k}(t_0) \in \Sigma_R \setminus \bigcup_{j=0}^{\infty} Q_R^j$, *then* $\exists t_1 \geq t_0$ *such that* $\mathbf{e_k}(t_1) \in \Gamma$.

Proof. The proof of the lemma exactly follows the proof of Lemma A.0.6 by changing the sign of w_max in the error dynamics and redefining the sets a and b. □

Appendix B

Detailed derivations of the differential equations of motion for the case study systems

B.1 Double pendulum system

The point-mass double pendulum system and the associated parameters are depicted in Figure 3.2. In this section, we follow Lagrange's approach to derive the differential equations of motion for the point-mass double pendulum system.

Let $\mathbf{q}(t) = [q_1(t), q_2(t)]^T$. For the sake of brevity, we drop the dependence of time from $\mathbf{q}(t)$ and $\mathbf{u}(t)$ vectors and simply denote them by $\mathbf{q}$ and $\mathbf{u}$. Correspondingly, $q_1(t) \equiv q_1$, $q_2(t) \equiv q_2$, $u_1(t) \equiv u_1$ and $u_2(t) \equiv u_2$. The kinetic and potential energies of the system are

$$E_k(\mathbf{q}) = \frac{1}{2} m_1 \langle \dot{\mathbf{r}}_1, \dot{\mathbf{r}}_1 \rangle + \frac{1}{2} m_2 \langle \dot{\mathbf{r}}_2, \dot{\mathbf{r}}_2 \rangle, \tag{B.1}$$

$$E_p(\mathbf{q}) = m_1\, g\, \langle \mathbf{r}_1, \hat{j} \rangle + m_2\, g\, \langle \mathbf{r}_2, \hat{j} \rangle, \tag{B.2}$$

where $\hat{i}$ and $\hat{j}$ are the unit vectors along the horizontal and vertical (parallel to the gravitational acceleration) directions. Moreover, $\mathbf{r}_1$ and $\mathbf{r}_2$ represent positions of m_1 and m_2 on the plane and are equal to

$$\mathbf{r}_1 = l_1\big(\cos(q_1)\hat{i} + \sin(q_1)\hat{j}\big), \tag{B.3}$$

$$\mathbf{r}_2 = \mathbf{r}_1 + l_2\big(\cos(q_1+q_2)\hat{i} + \sin(q_1+q_2)\hat{j}\big). \tag{B.4}$$

Thus, the Lagrangian of the system is

$$\mathcal{L}(\mathbf{q}) = E_k(\mathbf{q}) - E_p(\mathbf{q}) \tag{B.5}$$

$$= m_1 l_1\Big(\frac{l_1\dot{q}_1}{2} - g\sin(q_1)\Big) + \frac{m_2}{2}\Big(l_1\dot{q}_1^2 + l_2(\dot{q}_1+\dot{q}_2)^2\Big) \tag{B.6}$$

$$+ m_2 l_1 l_2\cos(q_2)(\dot{q}_1^2 + \dot{q}_1\dot{q}_2) - m_2 g\Big(l_1\sin(q_1) + l_2\sin(q_1+q_2)\Big). \tag{B.7}$$

Noting that the generalized forces associated with generalized coordinates q_1 and q_2 are u_1 and u_2, respectively, the set of differential equations of motion for the system could be obtained as the following.

$$\frac{d}{dt}\Big(\frac{\partial\mathcal{L}(\mathbf{q})}{\partial\dot{\mathbf{q}}_1}\Big) - \frac{\partial\mathcal{L}(\mathbf{q})}{\partial\mathbf{q}_1} = u_1, \tag{B.8}$$

$$\frac{d}{dt}\Big(\frac{\partial\mathcal{L}(\mathbf{q})}{\partial\dot{\mathbf{q}}_2}\Big) - \frac{\partial\mathcal{L}(\mathbf{q})}{\partial\mathbf{q}_2} = u_2. \tag{B.9}$$

Substituting $\mathcal{L}(\mathbf{q})$ from (B.5) in (B.8) and (B.9) and factoring $\ddot{\mathbf{q}}$ and $\mathbf{u}$ vectors leads to the differential equations of motion of the point-mass double pendulum system, that is

$$\mathbf{M}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{\Phi}(\mathbf{q},\dot{\mathbf{q}}) = \mathbf{u}, \tag{B.10}$$

where

$$\mathbf{M}(\mathbf{q}) = \begin{bmatrix} (m_1+m_2)l_1^2 + m_2 l_2^2 + 2l_1 l_2 m_2 \cos(q_2) & m_2 l_2 (l_1 \cos(q_2) + l_2) \\ m_2 l_2 (l_1 \cos(q_2) + l_2) & m_2 l_2^2 \end{bmatrix}, \quad \text{(B.11)}$$

and $\mathbf{\Phi}(\mathbf{q}, \dot{\mathbf{q}}) = [\Phi_1(\mathbf{q}, \dot{\mathbf{q}}),\ \Phi_2(\mathbf{q}, \dot{\mathbf{q}})]^T$ with Φ_1 and Φ_2 equal to

$$\Phi_1(\mathbf{q}, \dot{\mathbf{q}}) = (m_1+m_2) g l_1 \cos(q_1) + m_2 l_2 \Big(g \cos(q_1+q_2) - l_1 \dot{q}_2 (2\dot{q}_1 + \dot{q}_2) \sin(q_2) \Big), \quad \text{(B.12)}$$

$$\Phi_2(\mathbf{q}, \dot{\mathbf{q}}) = m_2 l_2 \Big(l_1 \dot{q}_1^2 \sin(q_2) + g \cos(q_1+q_2) \Big). \quad \text{(B.13)}$$

B.2 Cart-pole system

The point-mass cart-pole system and the associated parameters are depicted in Figure 3.4. In this section we derive the differential equations of motion of the point-mass cart-pole system using Lagrange's approach.

Let $\mathbf{q}(t) = [q_1(t), q_2(t)]^T$. For the sake of brevity, we drop the dependence of time from $\mathbf{q}(t)$ and $u(t)$ vectors and simply denote them by $\mathbf{q}$ and u. Correspondingly, $q_1(t) \equiv q_1$, $q_2(t) \equiv q_2$, $u(t) \equiv u$. The kinetic and potential energies of the system are

$$E_k(\mathbf{q}) = \frac{1}{2} m_1 \langle \dot{\mathbf{r}}_1, \dot{\mathbf{r}}_1 \rangle + \frac{1}{2} m_2 \langle \dot{\mathbf{r}}_2, \dot{\mathbf{r}}_2 \rangle, \quad \text{(B.14)}$$

$$E_p(\mathbf{q}) = m_2\, g\, \langle \mathbf{r}_2, \hat{j} \rangle, \quad \text{(B.15)}$$

where $\hat{i}$ and $\hat{j}$ are the unit vectors along the horizontal and vertical (parallel to the gravitational acceleration) directions. Moreover, $\mathbf{r}_1$ and $\mathbf{r}_2$ represent positions of m_1

and m_2 on the plane and are equal to

$$\mathbf{r}_1 = q_1\hat{i}, \tag{B.16}$$

$$\mathbf{r}_2 = \mathbf{r}_1 + l\big(\cos(q_2)\hat{j} - \sin(q_2)\hat{i}\big). \tag{B.17}$$

Thus, the Lagrangian of the system is

$$\mathcal{L}(\mathbf{q}) = E_k(\mathbf{q}) - E_p(\mathbf{q}) \tag{B.18}$$

$$= \frac{1}{2}(m_1 + m_2)\dot{q}_1^2 + m_2 l\dot{q}_2\big(\frac{l}{2}\dot{q}_2 - \dot{q}_1\cos(q_2)\big) - m_2 gl\cos(q_2). \tag{B.19}$$

Noting that u is the generalized force associated with generalized coordinates q_1, the set of differential equations of motion for the system could be obtained as the following.

$$\frac{d}{dt}\Big(\frac{\partial\mathcal{L}(\mathbf{q})}{\partial\dot{\mathbf{q}}_1}\Big) - \frac{\partial\mathcal{L}(\mathbf{q})}{\partial\mathbf{q}_1} = u, \tag{B.20}$$

$$\frac{d}{dt}\Big(\frac{\partial\mathcal{L}(\mathbf{q})}{\partial\dot{\mathbf{q}}_2}\Big) - \frac{\partial\mathcal{L}(\mathbf{q})}{\partial\mathbf{q}_2} = 0. \tag{B.21}$$

Substituting $\mathcal{L}(\mathbf{q})$ from (B.18) in (B.20) and (B.21) and factoring $\ddot{\mathbf{q}}$ and u terms leads to the differential equations of motion of the point-mass cart-pole system as

$$\mathbf{M}(\mathbf{q})\ddot{\mathbf{q}} + \mathbf{\Phi}(\mathbf{q}, \dot{\mathbf{q}}) = \mathbf{u}, \tag{B.22}$$

where

$$\mathbf{M}(\mathbf{q}) = \begin{bmatrix} m_1 + m_2 & -m_2 l\cos(q_2) \\ -m_2 l\cos(q_2) & m_2 l^2 \end{bmatrix}, \text{ and } \mathbf{\Phi}(\mathbf{q}, \dot{\mathbf{q}}) = \begin{bmatrix} m_2 l\dot{q}_2^2\sin(q_2) \\ -m_2 gl\sin(q_2) \end{bmatrix}. \tag{B.23}$$

Bibliography

[1] Franco Blanchini and Stefano Miani. Any domain of attraction for a linear constrained system is a tracking domain of attraction. *SIAM Journal on Control and Optimization*, 38(3):971–994, 2000.

[2] Aleksandar I Zečević and Dragoslav D Šiljak. Regions of attraction. In *Control of Complex Systems*, pages 111–141. Springer, 2010.

[3] Jean-Michel Coron. *Control and nonlinearity*. Number 136. American Mathematical Soc., 2007.

[4] A.D. Lewis. *Tautological Control Systems*. SpringerBriefs in Electrical and Computer Engineering. Springer International Publishing, 2014.

[5] Georgi V Smirnov. *Introduction to the theory of differential inclusions*, volume 41. American Mathematical Soc., 2002.

[6] Katsuhiko Ogata and Yanjuan Yang. *Modern control engineering*, volume 4. Prentice hall India, 2002.

[7] Peter Petersen and Springer Books. *Linear algebra*. Springer, New York, 2012.

[8] Jerry H Ginsberg. *Advanced engineering dynamics*. Cambridge University Press, 1998.

[9] Donald E Kirk. *Optimal control theory: an introduction*. Courier Corporation, 2012.

[10] Bo Lincoln and Bo Bernhardsson. Efficient pruning of search trees in lqr control of switched linear systems. In *Decision and Control, 2000. Proceedings of the 39th IEEE Conference on*, volume 2, pages 1828–1833. IEEE, 2000.

[11] Russ Tedrake, Ian R Manchester, Mark Tobenkin, and John W Roberts. Lqr-trees: Feedback motion planning via sums-of-squares verification. *The International Journal of Robotics Research*, 29(8):1038–1052, 2010.

[12] Steven M LaValle and James J Kuffner Jr. Randomized kinodynamic planning. *The international journal of robotics research*, 20(5):378–400, 2001.

[13] Joseph Moore and Russ Tedrake. Control synthesis and verification for a perching uav using lqr-trees. In *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pages 3707–3714. IEEE, 2012.

[14] Joseph Moore, Rick Cory, and Russ Tedrake. Robust post-stall perching with a simple fixed-wing glider using lqr-trees. *Bioinspiration & biomimetics*, 9(2):025013, 2014.

[15] Anirudha Majumdar, Amir Ali Ahmadi, and Russ Tedrake. Control design along trajectories with sums of squares programming. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 4054–4061. IEEE, 2013.

[16] Daniel Mellinger, Nathan Michael, and Vijay Kumar. Trajectory generation and control for precise aggressive maneuvers with quadrotors. *The International Journal of Robotics Research*, 31(5):664–674, 2012.

[17] Steven M LaValle and James J Kuffner Jr. Rapidly-exploring random trees: Progress and prospects. 2000.

[18] Jean-Claude Latombe. *Robot motion planning*, volume 124. Springer Science & Business Media, 2012.

[19] Jérôme Barraquand, Lydia Kavraki, Jean-Claude Latombe, Rajeev Motwani, Tsai-Yen Li, and Prabhakar Raghavan. A random sampling scheme for path planning. *The International Journal of Robotics Research*, 16(6):759–774, 1997.

[20] Jeong hwan Jeon, Sertac Karaman, and Emilio Frazzoli. Anytime computation of time-optimal off-road vehicle maneuvers using the rrt. In *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*, pages 3276–3282. IEEE, 2011.

[21] Sertac Karaman and Emilio Frazzoli. Optimal kinodynamic motion planning using incremental sampling-based methods. In *Decision and Control (CDC), 2010 49th IEEE Conference on*, pages 7681–7687. IEEE, 2010.

[22] Alexander Shkolnik and Russ Tedrake. Path planning in 1000+ dimensions using a task-space voronoi bias. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, pages 2061–2067. IEEE, 2009.

[23] Siamak G Faal and Cagdas D Onal. Regionally growing random trees: A synergistic motion planning and control algorithm for dynamic systems. In *Automation Science and Engineering (CASE), 2016 IEEE International Conference on*, pages 141–147. IEEE, 2016.

[24] Pierre Bessiere, J-M Ahuactzin, E-G Talbi, and Emmanuel Mazer. The" ariadne's clew" algorithm: Global planning with local methods. In *Intelligent*

Robots and Systems' 93, IROS'93. Proceedings of the 1993 IEEE/RSJ International Conference on, volume 2, pages 1373–1380. IEEE, 1993.

[25] James L Meriam and L Glenn Kraige. *Engineering mechanics: dynamics*, volume 2. John Wiley & Sons, 2012.

[26] John T Betts. *Practical methods for optimal control and estimation using nonlinear programming*, volume 19. Siam, 2010.

[27] Richard Vinter. *Optimal control*. Springer Science & Business Media, 2010.

[28] Lev Semenovich Pontryagin. *Mathematical theory of optimal processes*. Routledge, 2018.

[29] Rudolf Emil Kalman et al. Contributions to the theory of optimal control. *Bol. Soc. Mat. Mexicana*, 5(2):102–119, 1960.

[30] Rush D Robinett III, David G Wilson, G Richard Eisler, and John E Hurtado. *Applied dynamic programming for optimization of dynamical systems*, volume 9. Siam, 2005.

[31] Karl Johan Åström and Katsuhisa Furuta. Swinging up a pendulum by energy control. *Automatica*, 36(2):287–295, 2000.

[32] F Gordillo and J Aracil. A new controller for the inverted pendulum on a cart. *International Journal of Robust and Nonlinear Control*, 18(17):1607–1621, 2008.

[33] Isabelle Fantoni, Rogelio Lozano, and Mark W Spong. Energy based control of the pendubot. *IEEE Transactions on Automatic Control*, 45(4):725–729, 2000.

[34] Mark W Spong. Energy based control of a class of underactuated mechanical systems. *IFAC Proceedings Volumes*, 29(1):2828–2832, 1996.

[35] Rini Akmeliawati and Iven MY Mareels. Nonlinear energy-based control method for aircraft automatic landing systems. *IEEE Transactions on Control Systems Technology*, 18(4):871–884, 2010.

[36] Mark W Spong and Mathukumalli Vidyasagar. *Robot dynamics and control*. John Wiley & Sons, 2008.

[37] Jean-Jacques E Slotine, Weiping Li, et al. *Applied nonlinear control*, volume 199. Prentice hall Englewood Cliffs, NJ, 1991.

[38] Rudolf E Kalman and John E Bertram. Control system analysis and design via the "second method" of lyapunov: I-continuous-time systems. *Journal of Basic Engineering*, 82(2):371–393, 1960.

[39] Karl Johan Åström. Hybrid control of inverted pendulums. In *Learning, control and hybrid systems*, pages 150–163. Springer, 1999.