Bifurcation and Robust Control of Instabilities in the Presence of Uncertainties

by

Yiming Meng

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Applied Mathematics

Waterloo, Ontario, Canada, 2022

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Serdar Yüksel
Professor, Department of Mathematics and Statistics
Queen's University

Supervisors: Jun Liu
Associate Professor, Department of Applied Mathematics

N.Sri Namachchivaya
Professor, Department of Applied Mathematics

Internal Members: Sue Ann Campbell
Professor, Department of Applied Mathematics

Hans De Sterck
Professor, Department of Applied Mathematics

Internal-External Member: Stephen L. Smith
Professor, Department of Electrical and Computer Engineering

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In real-world applications, nominal mathematical models that are used to describe the state behaviors of dynamical systems are usually less robust to deal with environmental disruptions. Uncertainties, such as imprecision of signals, Gaussian-type white noise, and observation errors, may be injected into the systems and create substantial impacts on stability and safety etc. To better understand and hence robustly eliminate the potential negative impact, this thesis aims to develop novel control methods and bifurcation analysis for general nonlinear systems that are subjected to such types of perturbations.

The first main aspect of the research addresses the verification and control synthesis of more complex tasks with ω-regular linear-time properties besides stabilization problems for more general perturbed finite-dimensional nonlinear systems. Rigorous abstraction-based formal methods compute with guarantees a set of initial states from which the trajectories satisfy or a controller exists to realize the given specification, however, at the cost of heavy state-space discretization and potential difficulties of adjusting the speed of the dynamical flows. This thesis proposes discretization-free Lyapunov methods to handle verification and control synthesis for building-block specifications such as safety, stability, reachability, and reach-and-stay specifications. In the presence of non-stochastic and stochastic perturbations, respectively, rigorous analysis is conducted upon the fundamental mathematical guarantees of satisfying the above mentioned specifications using Lyapunov-like functions. A comparison between the proposed Lyapunov method and formal methods is illustrated via numerical simulations for the case with non-stochastic perturbations.

In terms of formal verification and control synthesis for stochastic systems, the current literature focuses on developing sound abstraction techniques for discrete-time stochastic dynamics without extra uncertain signals. However, soundness thus far has only been shown for preserving the satisfaction probability of certain types of temporal-logic specification. We focus on more general discrete-time nonlinear stochastic systems and present constructive finite-state abstractions for verifying or control synthesis of probabilistic satisfaction with respect to general ω-regular linear-time properties. Instead of imposing stability assumptions, we analyze the probabilistic properties from the topological view of metrizable space of probability measures. Such abstractions are both sound and approximately complete. That is, given a concrete discrete-time stochastic system and an arbitrarily small $\mathcal{L}^1$-perturbation of this system, there exists a family of finite-state Markov chains whose set of satisfaction probabilities contains that of the original system and meanwhile is contained by that of the slightly perturbed system. A direct consequence is that, given a probabilistic linear-time specification, initializing within the winning/losing region of the abstraction system can guarantee a satisfaction/dissatisfaction for

the original system. We make an interesting observation that, unlike the deterministic case, point-mass (Dirac) perturbations cannot fulfill the purpose of robust completeness.

The second aspect of the research addresses the bifurcation analysis in parabolic stochastic partial differential equations (SPDEs). We consider cases with small additive and multiplicative space-time noise, respectively, and conduct a local bifurcation analysis via a multiscale technique. In the presence of small additive noise, we make assumptions that the noise only acts on the stable fast-varying modes. We apply homogenization techniques based on recent advances for systems with one-dimensional critical mode directly to the perturbed Moore-Greitzer full model for the detection of modern jet engine compressor stall. We rigorously develop low-dimensional approximations using a multiscale analysis approach near the deterministic Hopf bifurcation point that occurs within the infinite-dimensional subspace. We also show that the reduced-dimension approximation model contains a multiplicative noise.

To better understand the long-term behavior of SPDEs near the deterministic Hopf bifurcation point and demonstrate the stochastic Hopf bifurcations under the impact of small multiplicative noise, we focus on the system with only cubic nonlinearities and use a different approach other than stochastic averaging/homogenization. We propose a simplified equation that has the same linearization as the original equation and prove the error bounds. It can be shown that the stable marginals do have a small impact on determining the stochastic bifurcation points. This approximation scheme does not reduce the stochastic effects from the stable modes to point-mass perturbations, and can be allied with the almost-sure exponential stability of the trivial solution to analyze the stochastic bifurcation diagram as the noise becomes smaller.

Acknowledgements

I would like to express my deepest gratitude to both of my supervisors, Prof. Jun Liu and Prof. N.Sri.Namachchivaya. I sincerely thank them for providing me this opportunity to study applied mathematics. I am extremely grateful for the past learning experiences with them, and their support and appreciation mean a great deal to me.

I thank Dr. Prof. Nicolas Perkowski and Prof. Alexander Schied for sharing their insightful knowledge on stochastic processes and analysis during my study.

I would also like to thank my examining committee members Prof. Sue Ann Campbell and Prof. Hans De Sterck, my internal/external examiner Prof. Stephen L. Smith, and my external examiner Prof. Serdar Yüksel for taking their time reading my thesis. I am very grateful for the valuable comments on improving this thesis.

I would like to thank my friends I have met at Waterloo. I enjoyed every moment that I spent with them. I sincerely thank my family for providing a great amount of support.

I acknowledge financial support for this work from Hybrid Systems Laboratory, National Sciences and Engineering Research Council (NSERC) Discovery grant 50503-10802, TECSIS /Fields-CQAM Laboratory for Inference and Prediction, NSERC-CRD grant 543433-19, as well as other forms of support provided by Department of Applied Mathematics and University of Waterloo.

Table of Contents

List of Figures

List of Abbreviations

BMDP Bounded-Parameter Markov Decision Process 3, 74, 143–147, 151, 153, 249

CBF Control Barrier Function 24–26, 72, 81

DA Deterministic Automaton 270, 271

DRA Deterministic Rabin Automata 271

HOCBF High-Order Control Barrier Function 25

IMC Interval-Valued Markov chain 3, 74, 121–123, 126–128, 130, 132, 135, 136, 141, 153, 249

LLN Law of Large Numbers 108, 273

LT Linear Time 3, 16, 115, 118, 135, 249, 268, 269

LTL Linear Temporal Logic 2, 16, 17, 115, 118, 127, 128, 134, 139, 141, 144, 145, 147, 150, 249, 250, 268–271

NA Non-deterministic Automaton 270, 271

NBA Non-deterministic Büchi Automata 271

ODE Ordinary Differential Equation 2, 19, 38, 45, 46, 154, 155, 247

PCTL Probabilistic Computation Tree Logic 3

PDE Partial Differential Equation 1, 4, 5, 154, 225, 245, 250, 275

Pr-UA Uniformly Attractive in Probability Law 88, 93, 95

Pr-UAS Uniformly Asymptotically Stable in Probability Law 88, 90, 94, 95, 97

Pr-US Uniformly Stable in Probability Law 88, 89, 91, 94–96

QP Quadratic Programming 19, 47, 48, 86, 248

R-SCBF Reciprocal Stochastic Control Barrier Function 82, 84, 87, 110

r.v. random variable 284

RBF Reciprocal Barrier Function 22, 24, 80

RKHS Reproducing Kernel Hilbert Space 185, 283–286

SCBF Stochastic Control Barrier Function 85–87, 104, 109, 110

SDE Stochastic Differential Equations 74–76, 99, 100, 134, 154, 156, 180, 209, 275

SLF Stochastic Lyapunov Function 94–98, 101, 110

SPDE Stochastic Partial Differential Equation 3, 4, 13, 14, 153, 154, 180, 181, 209, 225, 226, 247, 249, 250, 275, 282, 285, 286

UAS Uniformly Asymptotically Stable 27–31, 34–37, 40

UpAS Uniformly Pre-Asymptotically Stable 54, 55, 57, 60, 71

Z-SCBF Zeroing Stochastic Control Barrier Function 83–85, 87, 110

ZBF Zeroing Barrier Function 23, 25, 80, 81

List of Symbols

Notations in Part I (Chapter 2 to 4) and Part II (Chapter 5 to 8) are generally not the same. We make great effort to keep notations succinct and consistent. Due to the complexity, some of the notations are inevitably overloaded. To reduce confusion, frequently used notations and notations with different meanings are listed below. Notations that are used in local definitions are not provided in this list.

$\mathbb{R}^n$	Standard Euclidean space of n dimensions.
$\mathbb{R}_{\geq 0}$	The set of nonnegative real numbers.
$\mathbb{C}$	The space of complex numbers
$\lvert \cdot \rvert$	Euclidean norm (Chapter 2 and 3).
	∞-norm in $\mathbb{R}^n$ (Chapter 4).
	The modulus of a complex number.
$B\backslash A$	Set difference for sets B and A, i.e., $\{x \in B : x \notin A\}$.
A^c	The complement of set A.
$\bar{A}$	The closure of set A.
∂A	The boundary of set A.
$\mathrm{Int}(A)$	The interior of set A.
$\mathbb{B}$	The unit open ball centered at $\mathbf{0}$ w.r.t. $\lvert \cdot \rvert$.
$x + r\mathbb{B}$	The r-neighborhood of x w.r.t. $\lvert \cdot \rvert$, i.e., $\{y \in \mathbb{R}^n : \lvert y - x\rvert < r\}$.
$\mathbb{B}_r(x)$	Alternative notation for $x + r\mathbb{B}$.
$\lvert x\rvert_A$	The distance from $x \in \mathbb{R}^n$ to a closed set $A \subseteq \mathbb{R}^n$ w.r.t. $\lvert \cdot \rvert$, i.e., $\inf_{y\in A} \lvert x - y\rvert$.
$A + r\mathbb{B}$	The r-neighborhood of A, i.e., $\bigcup_{x\in A} x + r\mathbb{B}$.
$\mathbb{B}_r(A)$	Alternative notation for $A + r\mathbb{B}$.

$\mathbb{Z}$	The set of integer numbers.
$\mathbb{Z}_0$	$\mathbb{Z} \setminus \{0\}$.
$\mathscr{B}(E)$	The Borel σ-algebra of the set E.
$\bigvee$	The smallest σ-algebra generated by the union of collections of sets.
$\mathfrak{P}(E)$	The space of all probability measures on $\mathscr{B}(E)$.
$\mathcal{B}$	The unit open ball of random variables $X : \mathbf{\Omega} \to \mathbb{R}^n$ such that $\mathbf{E}\lvert X\rvert < 1$.
$r\mathcal{B}$	$\{X : \mathbf{E}\lvert X\rvert < r\}$.
$\mathscr{H}$	A Hilbert space in general.
$\mathcal{H}$	A separable infinite-dimensional Hilbert space.
$\mathcal{K}$	Class $\mathcal{K}$ functions (see Page 20).
$\mathcal{K}_\infty$	Extended class $\mathcal{K}$ functions (see Page 20).
$C(D)$	The space of real-valued continuous functions on D.
$C^k(D)$	The space of real-valued continuous functions with k^{th}-order continuous derivatives on D.
$C(D; E)$	The space of E-valued continuous functions on D.
$C_b(D)$	The space of real-valued bounded continuous functions on D.
$C_b^k(D)$	The space of real-valued bounded continuous functions with k^{th}-order continuous derivatives on D.
$C_b(D; E)$	The space of E-valued bounded continuous functions on D.
L^p	The family of functions $h : D \to \mathbb{R}$ such that $\int_D \lvert h(x)\rvert^p dx < \infty$.
$\mathcal{L}^p$	The family of random variables $X : \mathbf{\Omega} \to \mathbb{R}^n$ such that $\mathbf{E}\lvert X\rvert^p < \infty$.
$\mathcal{L}_2(E, K)$	The set of all Hilbert-Schmidt operators from E to K with operator norm $\lVert \cdot \rVert_{\mathcal{L}_2(E,K)}$.
$a \wedge b$	$\min(a, b)$.

id	Identity operator.
$\mathbb{1}_A$	Indicator function w.r.t. set A.
$\prod$	The product of ordinary sets, spaces, or function values.
$\otimes$	The product of collections of sets, or sigma algebras, or measures.

Chapter 1

Introduction

1.1 Motivation

The research is motivated to develop theory and methods to detect, control and mitigate dynamic instabilities in modern jet engines. Controlling compressor instability (surge and stall) is essential for increasing compressor efficiency, preventing damage or failure, and lengthening the life-span of the engine components. This thesis aims to develop novel methods for more general nonlinear models with uncertainties, and the results can be applied to controlling of compression system instabilities and enhancing performance.

The abstract evolution equation is of the form of parameter-dependent semi-linear differential equations with uncertainties

$$\partial_t u(t) = Au(t) + f(u(t);\gamma) + \varepsilon\xi(t), \tag{1.1}$$

where $u(t)$ takes value in a general (finite or infinite-dimensional) Hilbert space $\mathscr{H}$ for each t; the linear operator $A : \mathscr{H} \to \mathscr{H}$ generates an analytic compact C_0 semigroup, and hence equips point spectrum; the field $f(u(t);\gamma)$, also written as $f_\gamma(u(t))$, is a 3^+-order Fréchet differentiable nonhomogeneity that also depends on the parameter $\gamma \in \mathbb{R}$; ξ represents the effect of perturbations. In the absence of noise ($\varepsilon = 0$), linearization around each equilibrium point $u_e(\gamma)$ gives rise to the linear operator $A + Df_{u_e}(\gamma)$. As γ changes monotonically, the point spectrum of $A + Df_{u_e}(\gamma)$ satisfies (Hopf) bifurcation conditions and creates local instabilites.

The motivating application problem, detection and control of modern jet engine compressor instabilities, is based on the Moore-Greitzer full model [74] with perturbations. This commonly used mathematical model consists of a PDE, which describes the behavior of disturbances in the inlet region of compression systems, and two ODEs, which describe the coupling of the disturbances within the mean flow. The perturbed version of the Moore-Greitzer model can be abstracted by (1.1), where $\mathscr{H} = \mathcal{H} \times \mathbb{R} \times \mathbb{R}$ and $\mathcal{H}$ is an infinite-dimensional separable Hilbert space in this case. The subspace $\mathbb{R}^2$ is decoupled from $\mathcal{H}$ when the state variable in $\mathcal{H}$ reaches the steady state. This allows us to consider $\mathcal{H}$ and $\mathbb{R}^2$ separately under certain conditions.

Determined by compressor geometry, as the parameter (throttle coefficient) decreases, Hopf bifurcation occurs, the jet engine compressors may exhibit potentially three types of instabilities near their Hopf bifurcation points (optimal operating range), namely rotating stall, surge, and a combination of both. Rotating stall is an instability where the circumferential flow pattern is disturbed, it manifests itself as a region of severely reduced flow that rotates at a fraction of the rotor speed and causes a drop in performance; surge is a pumping oscillation that can cause flameout and engine damage. The occurance of such instabilities reduce performance and cause damage of the compressor blades. To increase compressor efficiency, prevent damage or failure, and lengthen the life-span of the engine components, controllers regarding stabilization and performance improvement are needed.

Motivated by the above problem, the first main aspect of the research addresses the development of controllers to fulfill more complex tasks and hence improve the performance for more general perturbed nonlinear systems with finite-dimensional state space, i.e., $\mathscr{H} = \mathbb{R}^n$. We do not confine ourselves to stabilization problems so as to enlarge the potential application areas. Classical applications are such as robotic motion planning and regulation of trajectories [57, 61, 63, 58]. We consider complex tasks such as pickup-delivery, parts assembly, surveillance and persistent monitoring [108], etc., as the potential application areas. Systems with deterministic (point-mass) perturbations and stochastic perturbations will be investigated separately.

To specify complex ω-regular tasks (see Appendix A for details), LTL is used as an expressive language. Amongst all the tools of verifying dynamical behaviors and synthesizing controllers w.r.t. LTL specifications, formal methods are rigorous mathematical techniques specifying and verifying hardware and software systems [14]. Formal synthesis is to design controllers from a temporal logic specification using formal analysis or model checking. One of the advantages of formal methods is that they compute with guarantees a set of initial states from which a controller exists to realize the given specification [21], however, at the cost of heavy state-space discretization.

So far, abstraction-based formal verification for deterministic systems has gained its maturity [21]. Whilst bisimilar (equivalent) symbolic models exist for linear and linear control systems [96, 158], sound and approximately complete finite abstractions can be achieved via stability assumptions [136, 69] or robustness (in terms of point-mass perturbations) [112, 110, 113]. An illustrative example is provided in Example 2.0.1 to describe how abstraction-based methods work. The recent investigation considered building blocks, such as safety, reachability, and reach-and-stay specifications, of temporal logics for specifying more complex task objectives and developed specified (robustly) complete algorithms by adaptive state-space partitions [108, 110]. Despite the improvement in terms of the computational complexities, it still remains a fundamental challenge to overcome the curse of dimensionality for verification and control synthesis. In addition, for systems with tunable parameters, e.g. the 2-dimensional Moore-Greitzer model that undergoes a surge-type Hopf bifurcation, the abstraction-based algorithms manifest difficulties of adjusting the speed of the dynamical flows [122].

In terms of formal verification and control synthesis for stochastic systems, a common theme is to construct IMC or BMDP, a family of finite-state Markov chains with uncertain transitions, as finite-state abstractions to approximate the probability of satisfaction in proper ways. The previous works [102] argued without proof that for every PCTL formula, the probability of (path) satisfaction of the IMC abstractions forms a compact interval, which contains the real probability of the original system. The algorithm provides a fundamental view of computing the bounds of satisfaction probability given IMCs/BMDPs. However, the intuitive reasoning for soundness seems inaccurate based on our observation. The most recent work in [51, 52] and [44] claimed the soundness of verifying general ω-regular[1] properties using IMC abstractions, but a proof is not provided. To the best of our knowledge, we currently lack a general framework, as the one presented in the thesis, for guaranteeing soundness of IMC abstractions for verifying ω-regular properties.

Under this background, in comparison with formal methods, we propose discretization-free Lyapunov methods to handle verification and control synthesis for safety, stability, reachability, and reach-and-stay specifications. We analyze the fundamental mathematical guarantees of satisfying the above mentioned specifications using Lyapunov-like functions for finite-dimensional nonlinear systems with non-stochastic and stochastic perturbations, respectively. In terms of formal methods for nonlinear stochastic systems, we for the first time propose the concept of completeness for the stochastic abstractions. We also present constructive finite-state abstractions for verification and control synthesis of probabilistic satisfaction of general ω-regular LT properties. We analyze the probabilistic properties from the topological view of metrizable space of probability measures. Such abstractions are both sound and approximately complete.

The second aspect of the research addresses the bifurcation analysis in SPDE in the form of (1.1) within the state space of $\mathcal{H}$. We consider (infinite-dimensional) systems driven by additive and multiplicative noise, respectively, and conduct a local bifurcation analysis via a multiscale technique.

[1] See Appendix A for details.

In contrast to the deterministic setting, the stochastic bifurcation is not that well understood, in particular for infinite-dimensional systems [11, 17]. Two concepts of bifurcation are defined within the framework of random dynamical systems [8, Section 9]. The first is a phenomenological bifurcation (P-bifurcation), where the density of a unique invariant measure changes its structure. On the other hand, as the stability of the trivial invariant measure changes, the corresponding random dynamical system undergoes a dynamical bifurcation (D-bifurcation) where the number of invariant measures and the structure of random attractors change. In general, there is not a strong connection between these two types of bifurcation. In particular, with the appearance of additive noise, i.e. a random process that uniformly perturbs the dynamics, the D-bifurcation pattern of $\mathbb{R}^n$ systems is destroyed [41]. A complete diagram of stochastic Hopf bifurcation is obtained in [11] for $\mathbb{R}^2$ systems with multiplicative noise under proper conditions.

In the vinicity of the bifurcation points, the study of center manifold for the deterministic PDE model makes it possible to approximate the local flow behavior by a reduced-dimensional set of equations [176]. The existance of center manifolds for several classes of PDEs has been well established, the technical tool of using projection method (e.g. Lyapunov-Schmidt reduction) avoids putting the linear part into a normal form and gains its popularity [101][78]. However, under the stochastic purturbations, the dominating dynamics in the critical modes may appear differently in the linear terms and hence may affect the stability structure.

It is worth mentioning that there are several results regarding the approximation of the transient dynamics within the finite-dimensional critical manifold on a sufficiently long timescale. For example, the work presented in [26, 27, 29], describe the stochastic bifurcations using amplitude equations and multiscale analysis techniques. More rigorous analysis on the approximation of the invariant measure was developed in [28]. Furthermore, in contrast to the cubic nonlinearities [26], bilinearities [29] tend to mix the dynamics of the slow and fast varing modes more strongly. For the analysis of stochastic Hopf bifurcations, both bilinearities and cubic terms are needed for the local analysis, which make it more complicated than that of simple bifurcations. We extend the above work and rigorously develop low-dimensional approximations for Moore-Greitzer PDE model with additive noise using a multiscale analysis approach near the deterministic stall bifurcation points.

With the appearance of multiplicative noise, stochastic averaging/homogenization techniques have been applied for dimensional reduction and obtaining amplitude equations [125, 150, 27] to analyze dynamical bifurcation. The results verify that as the noise becomes smaller, a lower dimensional Markov process characterizes the limiting behavior. The low-dimensional approximation near a deterministic Hopf bifurcation point performs well in terms of simulating the distribution, density, as well as the top Lyapunov exponent of the full system.

To investigate the long-term behavior of SPDEs near the deterministic Hopf bifurcation point and demonstrate the dynamical Hopf bifurcation under the impact of small multiplicative noise, we use a different approach other than stochastic averaging/homogenization. We propose a simplified equation that has the same linearization as the original equation and show that the error does not affect the topological stability. Compared to stochastic homogenization techniques, the proposed method does not eliminate the interactions between the critical and stable modes, which allows us to associate the approximation scheme with the almost-sure exponential stability of the trivial invariant measure to analyze the D-bifurcation diagram as the noise becomes smaller.

As the research is motivated by a solid application field, the thesis starts with reviews on the unperturbed Moore-Greitzer PDE model. We focus on the introduction of Moore-Greitzer model and existing control strategies dealing with the stabilization problem.

1.2 Unperturbed Moore-Greitzer Model

The structure of the compression system and the compressor geometry are given in Fig. 1.1 and Fig. 1.2.

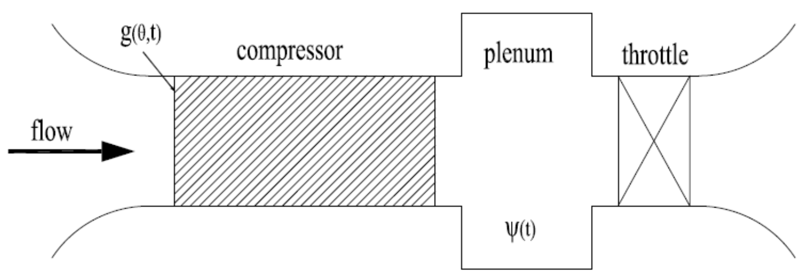


Figure 1.1: Compression system.

The compressor gives pressure rise to the upstream flow and sends it into the plenum through the downstream duct. The throttle controls the averaged mass flow through the system at the rear of the plenum. The sources of instability are twofold: (stall) the upstream non-uniform disturbance generates a locally higher angle of attack, and propagates along the blade row without mitigation; (surge) the average mean flow and pressure rise oscillate constantly and formulate standing waves [74]. The deterministic Moore-Greitzer model captures the dy-

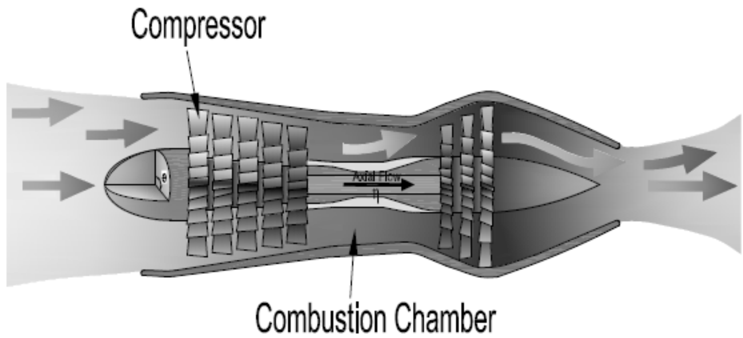


Figure 1.2: Compressor geometry.

namic evolution of the above states, and is given explicitly as [175]:

$$\frac{\partial}{\partial t}\begin{bmatrix} \mathrm{g} \\ \Phi \\ \Psi \end{bmatrix} = \begin{bmatrix} \mathrm{K}^{-1}(\frac{\mathfrak{v}}{2}\frac{\partial^2}{\partial\theta^2} - \frac{1}{2}\frac{\partial}{\partial\theta}) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} \mathrm{g} \\ \Phi \\ \Psi \end{bmatrix} + \begin{bmatrix} \mathfrak{a}\,\mathrm{K}^{-1}(\psi_c(\Phi+\mathrm{g}) - \overline{\psi}_c) \\ \frac{1}{\mathfrak{l}_c}(\overline{\psi}_c - \Psi) \\ \frac{1}{4\mathfrak{l}_c\mathfrak{B}^2}(\Phi - \gamma\sqrt{\Psi}) \end{bmatrix}, \tag{1.2}$$

where the states $[\mathrm{g}(t), \Phi(t), \Psi(t)]^T \in \mathscr{H} := \mathcal{H} \times \mathbb{R} \times \mathbb{R}$ are as introduced before. The physical meaning of the states is explained as follow: $\mathrm{g}(t,\theta)$ represents the velocity of upstream disturbance along the axial direction at the duct entrance, $\Phi(t)$ is the averaged mean flow rate, $\Psi(t)$ is the averaged pressure. To fulfill the physical meaning of g, we also require that $\mathrm{g}(t,0) = \mathrm{g}(t,2\pi)$, $\mathrm{g}_\theta(t,0) = \mathrm{g}_\theta(t,2\pi)$ and $\int_0^{2\pi}\mathrm{g}(t,\theta)d\theta = 0$. Thus, we can write g as a Fourier expansion as follows

$$\mathrm{g}(t,\theta) = \sum_{n\in\mathbb{Z}_0}\mathrm{g}_n(t)e^{in\theta}.$$

The operator K is defined as a Fourier multiplier,

$$\mathrm{K}(\mathrm{g}) = \sum_{n\in\mathbb{Z}_0}\left(1 + \frac{\mathfrak{a}\mathfrak{m}}{|n|}\right)\mathrm{g}_n(t)e^{in\theta},$$

where $\mathfrak{a}$ (that also appears in (1.2)) is the internal compressor lag and $\mathfrak{m}$ is the duct parameter. The compressor characteristic ψ_c is given in a cubic form,

$$\psi_c(\Phi) = \psi_{c_0} + \iota\left[1 + \frac{3}{2}\left(\frac{\Phi}{\mathfrak{M}} - 1\right) - \frac{1}{2}\left(\frac{\Phi}{\mathfrak{M}} - 1\right)^3\right], \tag{1.3}$$

where ψ_{c_0}, ι and $\mathfrak{M}$ are real-valued parameters that are defined by the compressor configuration. We also define

$$\overline{\psi}_c := \frac{1}{2\pi}\int_0^{2\pi} \psi_c(\Phi + \mathrm{g})d\theta.$$

The meaning of the other parameters in (1.2) is as follows: $\mathfrak{l}_c > 0$ is the compressor length, $\mathfrak{B} > 0$ is the plenum-to-compressor volume ratio, $\mathfrak{v} > 0$ is the viscous coefficient. The parameter $\gamma \in \mathbb{R}$ represents the throttle coefficient, the decrease of which will cause the stability change.

Remark 1.2.1. *The solution of* $\mathrm{g}(t)$ *lies in an infinite-dimensional Hilbert space* $\mathcal{H} := \{f \in \mathrm{L}^2 : \int_0^{2\pi} f(\theta)d\theta = 0\}$ *equipped with the inner product*

$$\langle\,\cdot,\ \cdot\rangle_{\mathcal{H}} := \langle\,\cdot,\ \mathrm{K}\,\cdot\rangle \tag{1.4}$$

as well as the induced norm $\|\cdot\|_{\mathcal{H}}$. *Note that the Fourier multiplier* K *is bounded from above and from below. In general, due to the spatial periodicity and the zero-average property of* $\mathrm{g}(t)$, *we can expect the solution to be at least in a Sobolev space* $\mathbf{H}^2_{\mathrm{per}} \subseteq \mathcal{H}$,

$$\mathbf{H}^2_{\mathrm{per}} := \left\{\mathrm{g} \in \mathbf{H}^2 : \mathrm{g}(0) = \mathrm{g}(2\pi),\ \mathrm{g}_\theta(0) = \mathrm{g}_\theta(2\pi), \int_0^{2\pi} \mathrm{g}(\tau,\theta)d\theta = 0\right\},$$

where $\mathbf{H}^2 \subseteq \mathcal{H}$ *is the Sobolev Hilbert space of* L^2 *functions with weak derivatives of order up to* 2 *in* L^2, *and* $\mathbf{H}^2_{\mathrm{per}} \subseteq \mathbf{H}^2$ *is the subspace of periodic functions. Note that the condition* $\mathrm{g}(0) = \mathrm{g}(2\pi),\ \mathrm{g}_\theta(0) = \mathrm{g}_\theta(2\pi)$ *in the definition of* $\mathbf{H}^2_{\mathrm{per}}$ *makes sense because, by the Sobolev embedding theorem [56], we have* $\mathbf{H}^2 \subseteq C^1$. *The space* $\mathscr{H} = \mathcal{H} \times \mathbb{R} \times \mathbb{R}$ *is then a product Hilbert space with inner product defined by*

$$\begin{aligned}\langle u^{(1)}, u^{(2)}\rangle_{\mathscr{H}} &= \langle(\mathrm{g}^{(1)}, \Phi^{(1)}, \Psi^{(1)}), (\mathrm{g}^{(2)}, \Phi^{(2)}, \Psi^{(2)})\rangle_{\mathscr{H}} \\ &:= \langle \mathrm{g}^{(1)}, \mathrm{g}^{(2)}\rangle_{\mathcal{H}} + \mathfrak{l}_c\Phi^{(1)}\Phi^{(2)} + (4\mathfrak{l}_c\mathfrak{B}^2)\Psi^{(1)}\Psi^{(2)}.\end{aligned} \tag{1.5}$$

1.2.1 Abstract Form

We can write (1.2) as an abstract form

$$\partial_t u = Au + f_\gamma(u), \tag{1.6}$$

where $u = [\mathrm{g}, \Phi, \Psi]^T \in \mathscr{H}$, A is the operator matrix

$$A = \begin{bmatrix} \mathrm{K}^{-1}(\frac{\mathfrak{v}}{2}\frac{\partial^2}{\partial\theta^2} - \frac{1}{2}\frac{\partial}{\partial\theta}) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

and

$$f_\gamma(u) = \begin{bmatrix} \mathfrak{a}\,\mathrm{K}^{-1}(\psi_c(\Phi+\mathrm{g}) - \overline{\psi}_c) \\ \frac{1}{\mathfrak{l}_c}(\overline{\psi}_c - \Psi) \\ \frac{1}{4\mathfrak{l}_c\mathfrak{B}^2}(\Phi - \gamma\sqrt{\Psi}) \end{bmatrix}.$$

We consider a fixed point of the form $u_e(\gamma) = [0, \Phi_e(\gamma), \Psi_e(\gamma)]^T$ at which $f_\gamma(u_e(\gamma)) = \mathbf{0}$ for each γ. In particular $(\Phi_e(\gamma), \Psi_e(\gamma))$ is determined by the intersection of the compressor characteristic $\Psi = \psi_c(\Phi)$ and the throttle characteristic $\Phi = \gamma\sqrt{\Psi}$ (see [175, Figure 4] for details).

Note that by definition, we have the following expansion:

$$\psi_c(\Phi+\mathrm{g}) = \psi_c(\Phi) + \iota\left[\frac{3}{2}\left(\frac{\mathrm{g}}{\mathfrak{M}}\right) - \frac{1}{2}\left(\frac{\mathrm{g}}{\mathfrak{M}}\right)^3 - \frac{3}{2}\left(\frac{\Phi}{\mathfrak{M}} - 1\right)^2\frac{\mathrm{g}}{\mathfrak{M}} - \frac{3}{2}\left(\frac{\Phi}{\mathfrak{M}} - 1\right)\left(\frac{\mathrm{g}}{\mathfrak{M}}\right)^2\right].$$

Since $\mathrm{g} = \sum_{n\in\mathbb{Z}_0} \mathrm{g}_n\, e^{in\theta}$,

$$\begin{aligned} \bar{\psi}_c &= \frac{1}{2\pi}\int_0^{2\pi} \psi_c(\Phi+\mathrm{g})d\theta = \frac{1}{2\pi}\int_0^{2\pi}\psi_c(\Phi)d\theta \\ &+ \frac{1}{2\pi}\int_0^{2\pi}\iota\left[\frac{3}{2}\left(\frac{\mathrm{g}}{\mathfrak{M}}\right) - \frac{1}{2}\left(\frac{\mathrm{g}}{\mathfrak{M}}\right)^3 - \frac{3}{2}\left(\frac{\Phi}{\mathfrak{M}} - 1\right)^2\frac{\mathrm{g}}{\mathfrak{M}} - \frac{3}{2}\left(\frac{\Phi}{\mathfrak{M}} - 1\right)\left(\frac{\mathrm{g}}{\mathfrak{M}}\right)^2\right]d\theta \\ &= \psi_c(\Phi) - \frac{3\iota}{2\mathfrak{M}^2}\left(\frac{\Phi}{\mathfrak{M}} - 1\right)\sum_{\substack{j,k\in\mathbb{Z}_0\\ k+j=0}} \mathrm{g}_j\,\mathrm{g}_k - \frac{\iota}{6\mathfrak{M}^3}\sum_{\substack{j,k,l\in\mathbb{Z}_0\\ k+j+l=0}}\mathrm{g}_j\,\mathrm{g}_k\,\mathrm{g}_l \\ &= \psi_c(\Phi) + \frac{\psi_c''(\Phi)}{2}\Pi^{(2)}\,\mathrm{g}^2 + \frac{\psi_c'''(\Phi)}{6}\Pi^{(3)}\,\mathrm{g}^3, \end{aligned} \tag{1.7}$$

where we have used notations $\Pi^{(2)}uv = \sum_{\substack{j,k\in\mathbb{Z}_0\\ k+j=0}} u_j v_k$ and $\Pi^{(3)}uvw = \sum_{\substack{j,k,l\in\mathbb{Z}_0\\ k+j+l=0}} u_j v_k w_l$ for any arbitrary $u, v, w \in \mathbf{H}^2_{\mathrm{per}}$. Therefore, the noisy perturbation of g that will be added later in the thesis enters the flow equations via $\Pi^{(2)}\,\mathrm{g}^2$ and $\Pi^{(3)}\,\mathrm{g}^3$. However, the operation points of the compressor, a family of stable fixed points $(\Phi_e(\gamma), \Psi_e(\gamma))$, are not influenced by g.

Linearization of (1.6) about $u_e(\gamma)$ for each γ results in the following linear operator

$$\mathcal{A}(\gamma) := A + Df_{u_e}(\gamma).$$

The Fréchet derivative at $u_e(\gamma)$ is given as

$$Df_{u_e}(\gamma) = \begin{bmatrix} \mathfrak{a}(\psi_c'(\Phi_e(\gamma)) - \overline{\psi_c}')\,\mathrm{K}^{-1} & 0 & 0 \\ 0 & \frac{1}{\mathfrak{l}_c}\overline{\psi_c}' & -\frac{1}{\mathfrak{l}_c} \\ 0 & \frac{1}{4\mathfrak{B}^2\mathfrak{l}_c} & \frac{1}{4\mathfrak{B}^2\mathfrak{l}_c}\frac{\gamma}{2\sqrt{\Psi_e(\gamma)}} \end{bmatrix}, \tag{1.8}$$

where $\psi_c'(\Phi_e(\gamma))$ and $\overline{\psi_c}'$ are the Fréchet derivatives of ψ_c and $\overline{\psi_c}$ at u.

1.2.2 Linear Operator Properties and Hopf Bifurcations

In this subsection, we list the properties of the linear operator $\mathcal{A}(\gamma)$.

1. $\mathcal{A}(\gamma)$ generates an analytic compact C_0 semigroup $S(t) := e^{\mathcal{A}(\gamma)t}$ on $\mathscr{H}$ [175].

2. For each γ, there exist constants $\omega \geq 0$ and $M \geq 1$ such that

$$\|S(t)\|_{\mathscr{H}} \leq Me^{\omega t},\ \forall t > 0.$$

   For the stable projection, there exists $\omega > 0$ and $M > 0$ such that

$$\|P_s S(t)\|_{\mathscr{H}} \leq Me^{-\omega t},\ \forall t > 0.$$

3. $\mathcal{A}(\gamma)$ can be represented as

$$\mathcal{A}(\gamma) = \begin{bmatrix} \mathcal{A}|_{\mathcal{H}}(\gamma) & 0 \\ 0 & \mathcal{A}|_{\mathbb{R}^2}(\gamma) \end{bmatrix}, \tag{1.9}$$

   where $\mathcal{A}|_{\mathcal{H}} : \mathcal{H} \to \mathcal{H}$ is the restriction of $\mathcal{A}$ onto $\mathcal{H}$, whilst $\mathcal{A}$ restricted to $\mathbb{R}^2$ is a 2×2 matrix $\mathcal{A}|_{\mathbb{R}^2}$. The decoupling of the eigenspace makes the linearized flow of g and (Φ, Ψ) invariant respectively under the semigroups $e^{\mathcal{A}|_{\mathcal{H}}(\gamma)t}$ and $e^{\mathcal{A}|_{\mathbb{R}^2}(\gamma)t}$.

4. We can expect the solution u to be in the domain of $\mathcal{A}(\gamma)$ (a subspace of $\mathscr{H}$), which is

$$\mathrm{dom}(\mathcal{A}(\gamma)) = \mathbf{H}^2_{per} \times \mathbb{R} \times \mathbb{R}.$$

5. The spectrum of $\mathcal{A}(\gamma)$ [175] is

$$\{\rho_{\pm n}(\gamma), \mathfrak{r}_{\pm 1}(\gamma)\} \text{ for } n \in \mathbb{Z}^+,$$

where $\rho_{\pm n}(\gamma) = \frac{\mathfrak{a}|n|}{|n|+\mathfrak{am}}\left(\psi_c'(\Phi_e(\gamma)) - \frac{\mathfrak{v}n^2}{2\mathfrak{a}} \pm \frac{|n|}{2\mathfrak{a}}i\right) \in \mathbb{C}$ for $n \in \mathbb{Z}^+$ are the eigenvalues of $\mathcal{A}|_{\mathcal{H}}$ corresponding to the eigenvectors $\upsilon_{\pm n}= [e^{\pm in\theta}, 0, 0]^{\mathrm{T}}$. The eigenvalues of $\mathcal{A}|_{\mathbb{R}^2}$ are

$$\mathfrak{r}_{\pm 1}(\gamma) = \frac{\mathfrak{N}(\gamma) - \mathfrak{A}(\gamma)}{2} \pm i\frac{\sqrt{\frac{1}{\mathfrak{B}^2} - \left(\psi_c'(\Psi_e(\gamma)) + \frac{\gamma}{8\mathfrak{B}^2\sqrt{\Psi_e(\gamma)}}\right)^2}}{2\mathfrak{l}_c}$$

where $\mathfrak{N}(\gamma) = \frac{1}{\mathfrak{l}_c}(\psi_c'(\Phi_e(\gamma))$ and $\mathfrak{A}(\gamma) = -\frac{1}{4\mathfrak{B}^2\mathfrak{l}_c}\frac{\gamma}{2\sqrt{\Psi_e(\gamma)}}$. The eigenvectors associated to $\mathfrak{r}_{\pm 1}(\gamma)$ are given by $\upsilon_{\mathfrak{r}_j}(\gamma) = \left[0, 1, \nu_{\psi_j}(\gamma)\right]^{\mathrm{T}}$ for $j \in \{\pm 1\}$, where

$$\upsilon_{\psi_j}(\gamma) = \frac{\mathfrak{l}_c(\mathfrak{N}(\gamma) + \mathfrak{A}(\gamma))}{2} - ij\frac{\sqrt{\frac{1}{\mathfrak{B}^2} - \left(\psi_c'(\Psi_e(\gamma)) + \frac{\gamma}{8\mathfrak{B}^2\sqrt{\Psi_e(\gamma)}}\right)^2}}{2}.$$

6. We verify the type of Hopf bifurcation by the sign of the indicator[175]

$$\Delta := \frac{\psi_{c_0} + \iota\left[1 + \frac{3}{2}\sqrt{1 - \frac{\mathfrak{vM}}{3\mathfrak{a}\iota}} - \frac{1}{2}\left(\sqrt{1 - \frac{\mathfrak{vM}}{3\mathfrak{a}\iota}}\right)^3\right]}{\mathfrak{M}\left(1 + \sqrt{1 - \frac{\mathfrak{vM}}{3\mathfrak{a}\iota}}\right)} - \frac{\mathfrak{a}}{4\mathfrak{B}^2\mathfrak{v}}. \tag{1.10}$$

More specifically, surge bifurcation happens when $\Delta > 0$, stall bifurcation happens when $\Delta < 0$. When $\Delta = 0$, then the bifurcations occur in both of the $\mathcal{H}$ and $\mathbb{R}^2$ subspaces.

Remark 1.2.2. *The spectral decomposition of $\mathcal{A}(\gamma)$ (as in 3 of the above) results in*

$$\mathscr{H} = \mathscr{H}_1 \oplus \mathscr{H}_2,$$

where $\mathscr{H}_1$ is an infinite-dimensional Hilbert space isomorphic to $\mathcal{H}$ and $\mathscr{H}_2$ is isomorphic to $\mathbb{R}^2$. This decomposition provides local coordinates in terms of which the dynamics near the equilibrium have a convenient form. Hence, the corresponding eigenvalues $\rho_{\pm 1}$ and $\mathfrak{r}_{\pm 1}$ change independently. Depending on which pair of eigenvalues crosses the imaginary axis first as the bifurcation parameter γ is varied, there are three possible types of Hopf bifurcations: If $\rho_{\pm 1}$ crosses the imaginary axis first, the physical oscillations are dominated by stall effects; if $\mathfrak{r}_{\pm 1}$ satisfies the Hopf bifurcation condition, then surge effects dominate; if $\rho_{\pm 1}$ and $\mathfrak{r}_{\pm 1}$ cross the imaginary axis simultaneously, we see a mixture of both effects. The critical subspaces are given respectively as $\mathscr{H}_1^c = [\mathrm{span}\{e^{\pm i\theta}\}, 0, 0]^T$, $\mathscr{H}_2^c = \mathscr{H}_2$, and $\mathscr{H}_1^c \oplus \mathscr{H}_2^c$, where the superscript c represents 'critical'. The indicator (1.10) verifies that the oscillation type is only determined by the fluid's viscosity and the geometric structure of the compressors.

Example 1.2.3. *Figure 1.3 shows an example of the regions of stable/unstable equilibrium points under the settings of* $\mathrm{g} \equiv \mathbf{0}$ *and*

$$\mathfrak{l}_c = 8,\ \iota = 0.18,\ \mathfrak{M} = 0.25,\ \psi_{c_0} = 1.67\iota, \mathfrak{a} = \frac{1}{3.5}, \mathfrak{v} = 0.1, \mathfrak{B} = 2, \tag{1.11}$$

as well as examples of phase portraits given different values of the parameter γ *that are away from the surge bifurcation point* $\gamma_c \approx 0.6123$*. The convergent/divergent rates given the values of* γ *in Figure 1.3 w.r.t. the equilibrium points are exponentially fast. However, when* γ *is in a small neighborhood of the surge bifurcation point, the dynamics as shown in Figure 1.4 appear to be slowly-varying.*

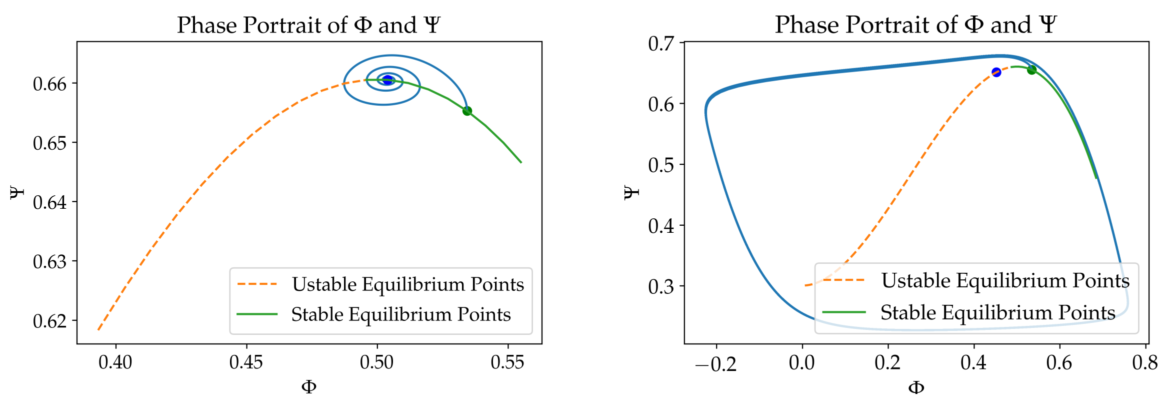


Figure 1.3: Regions of stable (green) / unstable (red) equilibrium points in the $\mathbb{R}^2$ subspace of the Moore-Greitzer model. Given the initial condition of $(\Phi_0, \Psi_0) = (0.5343, 0.6553)$, the parameter settings in (1.11), and $\mathrm{g}(0) = \mathbf{0}$, the phase portraits in the $\mathbb{R}^2$ subspace are generated with $\gamma = 0.62$ (left) and $\gamma = 0.56$ (right).

□

Example 1.2.4. *Let the configuration parameters be*

$$\mathfrak{l}_c = 8,\ \iota = 0.18,\ \mathfrak{M} = 0.25,\ \psi_{c_0} = 1.67\iota, \mathfrak{a} = \frac{1}{3.5}, \mathfrak{v} = 0.1, \mathfrak{B} = 0.72061. \tag{1.12}$$

Set $\mathrm{g}(0, \theta) = 0.005 \sin(\theta)$ *and* $(\Phi_0, \Psi_0) = (0.51, 0.66)$*. Then the indicator as in (1.10) is* $\Delta \approx 0$*, which implies the surge and stall bifurcations happen simultaneously. Now let* $\gamma = 0.56$*, under which the instabilities occur in both of the subspaces. The phase portrait in the* $\mathbb{R}^2$ *subspace is shown in Figure 1.5. The state evolution of* $\mathrm{g}(t)$ *at* $t = 40$ *and* $t = 50$ *in the* $\mathcal{H}$ *subspace is shown in Figure 1.6.* □

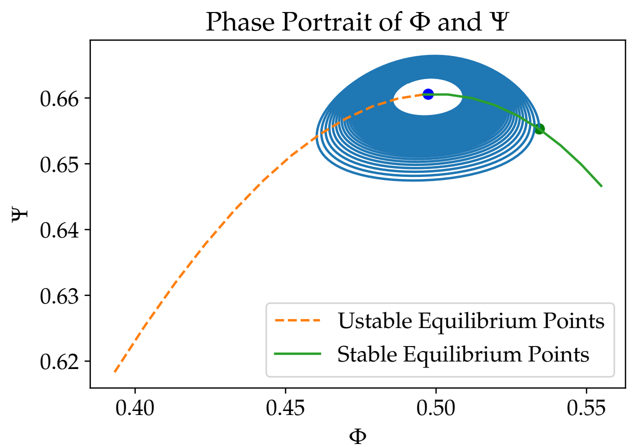


Figure 1.4: Phase portrait in the $\mathbb{R}^2$ subspace of the Moore-Greitzer model given the initial condition of $(\Phi_0, \Psi_0) = (0.5343, 0.6553)$, the parameter settings in (1.11) and $\gamma = 0.612$, and $\mathrm{g}(0) = \mathbf{0}$.

More details about bifurcation analysis in $\mathcal{H}$ can be found in the introduction part of Chapter 5.

1.2.3 Existing Results on Control of Instabilities of Jet Engine Compressors

Surge/stall avoidance control was invented to prevent the compressor from operating in a region near and beyond the surge line. The recycling of the flow lowers the efficiency of the system and may limit the transient performance of the compressor. However, active surge/stall control strategies were introduced to overcome the drawbacks of avoidance control. The approach is to stabilize some part of the unstable area using feedback controllers.

The methods of controller design were mainly regarding stabilization feedback. Feedback linearization was considered for surge control in [13], backstepping as well as the robustness backstepping was used in [98] and [74]. Passivity based surge control was also introduced to render robust L^2-stable [75].

As for the stall (infinite-dimensional) control, Banaszuk et al. in [15] designed a back stepping control, Birnir and Hauksson in [25] went one step further to construct a control strategy using the knowledge of asymptotic dynamics. Wen et al. in [170] studied the local feedback stabilization of Hopf bifurcation for infinite-dimensional nonlinear systems where only one pair of

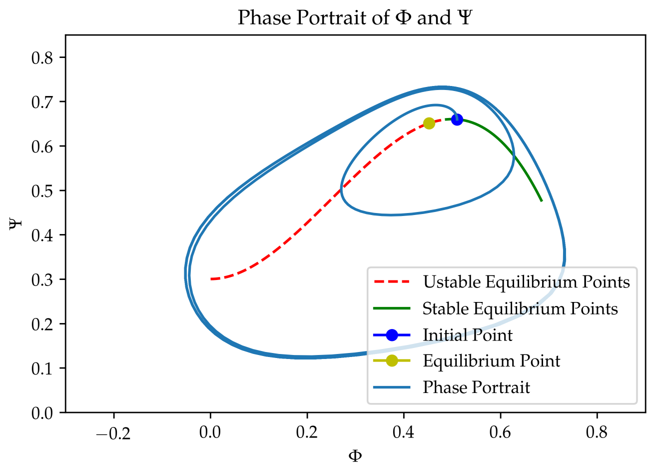


Figure 1.5: Phase portrait in the $\mathbb{R}^2$ subspace of the Moore-Greitzer model given the initial condition of $(\mathrm{g}(0,\theta), \Phi_0, \Psi_0) = (0.005\sin(\theta), 0.51, 0.66)$, the parameter settings in (1.12) and $\gamma = 0.56$.

the modes are at the risk of being unstable. Sufficient and necessary conditions were obtained for the controllability.

However, the above investigations are under the assumption that the parameter γ is time invariant. Apart from stabilization, reactive control strategies are needed to adapt to more complex time-varying objectives with less restriction on the parameter. To bridge such a practical gap, real-time controllers with optimal action at each current time should be discovered before stepping into the stochastic control problem. Other formal methods regarding complicated specifications with bounded noise as well as uncertainties will be helpful for understanding the regions of attraction for the system.

1.3 Organization of the Thesis

The goal of the thesis was stated at the end of Section 1.1. As mentioned earlier, the scope of the thesis is separated into two parts: robust control for finite-dimensional nonlinear systems with non-stochastic and stochastic uncertainties, and stochastic Hopf bifurcation analysis in the presence of small additive and multiplicative noise for SPDEs.

Part I:

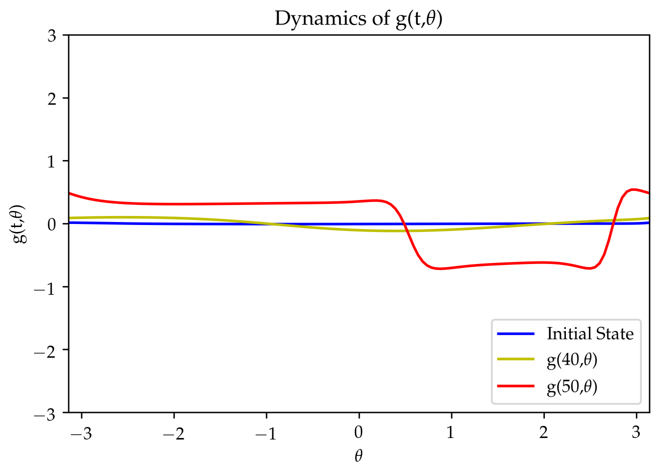


Figure 1.6: State snapshots of $\mathrm{g}(t)$ in the $\mathcal{H}$ subspace of the Moore-Greitzer model at $t = 40$ and $t = 50$ given the initial condition of $(\mathrm{g}(0,\theta), \Phi_0, \Psi_0) = (0.005\sin(\theta), 0.51, 0.66)$, the parameter settings in (1.12) and $\gamma = 0.56$.

Chapter 2 considers Lyapunov-barrier function approaches for general nonlinear systems with point-mass or deterministic uncertainties. Theories are developed upon verification and control synthesis for stability and safety related specifications. Comparisons between the proposed method and formal methods are made via numerical examples.

Chapter 3 extends the Lyapunov-barrier function approaches to the stochastic systems with extra uncertain signals.

Chapter 4 develops mathematical foundations of stochastic abstractions for verification and control synthesis of probabilistic specifications. The concept of robust completeness for stochastic abstractions is proposed. The philosophy of abstractions is discussed.

Part II can be read independent of Part I.

Chapter 5 develops the Hopf bifurcation analysis for a stochastic version of the Moore and Greitzer PDE model with additive noise based on recent advances in stochastic PDEs. Rigorous analysis is conducted to show the homogenized critical dynamics with relatively small error in proper sense.

Chapter 6 to 8 consider parabolic SPDEs with cubic nonlinearities and multiplicative noise. An approximation scheme, the almost-sure asymptotic stability, as well as the stochastic bifurcation structure based on the approximation scheme and the almost-sure asymptotic stability, are developed respectively in Chapter 6, 7, and 8.

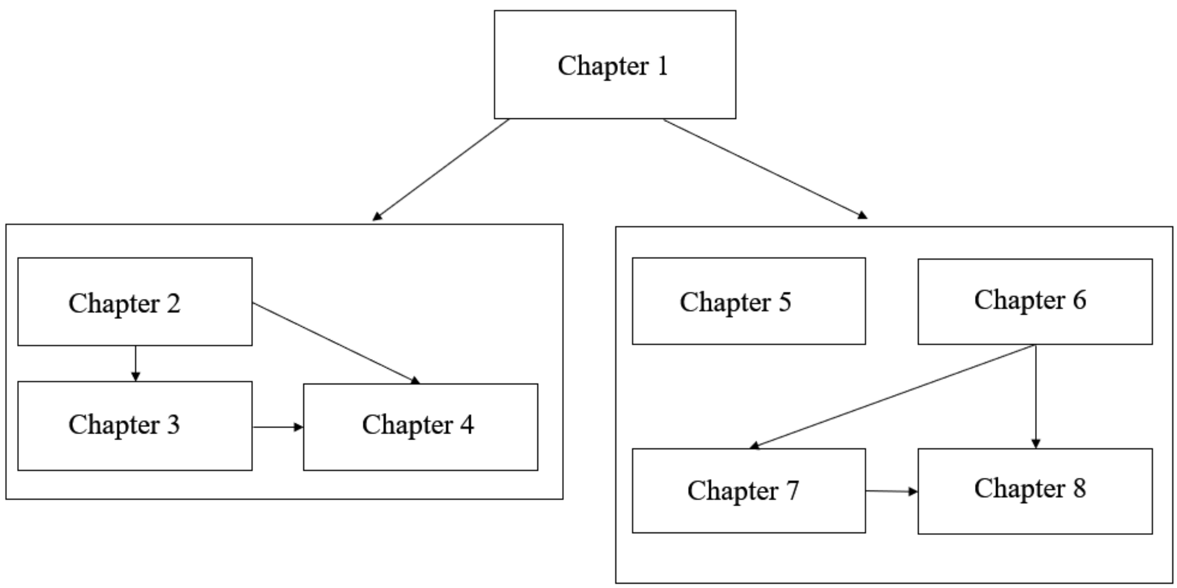


Figure 1.7: Interconnections between chapters.

Chapter 2

Lyapunov-Barrier Approaches for Verification and Control of Deterministic Systems

Stability, reachability and safety are important aspects in safety-critical control of dynamical systems. Beyond these, reach-avoid-stay specifications is such that the trajectory always avoids an unsafe set whilst the state reaches a target set within a finite time and stays inside it afterwards. In recent years, safety and reachability related properties for dynamical systems received considerable attention, primarily motivated by safety-critical control applications, such as in autonomous cyber-physical systems and robotics [2, 38, 86, 127, 4, 180, 63, 58, 129].

In view of temporal logic specifications (see details in Appendix A), safety and reachability related dynamical behaviors[1] are regarded as building blocks for specifying more complex task objectives. The advent of various temporal logic languages and the corresponding model checking algorithms [14] made it possible to verify and synthesize controllers w.r.t. LT properties for finite-transition systems [129, 96, 158]. Formal methods for nonlinear systems rely on a finite abstraction (or symbolic model) of the original systems, based on which computational methods are developed [69, 112]. Apart from the complicated abstraction analysis and the computational complexity caused by state space discretization, formal methods in control synthesis compute with guarantees a set of initial states from which a controller exists to realize the given specification [21].

[1]In view of LTL formula (see Appendix A), given an unsafe set U and a target set Γ, 'safety' specification can be regarded as $\Box(\neg U)$, 'reachability' is $\Diamond\Gamma$, and 'reach-avoid-stay' is $(\Diamond\Box\Gamma) \wedge (\Box(\neg U))$.

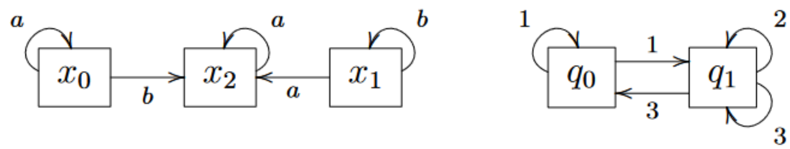


Figure 2.1: Transition systems T_1 (left) and T_2 (right).

Example 2.0.1 (Finite abstraction)**.** *We insert a simple example here, which is rephrased from [112, Example 1], to illustrate the idea of finite abstraction. Consider two transition systems (see Definition A.0.1)* $T_i = (Q_i, A_i, R_i, \mathrm{AP}, L_i)$*,* $i = 1, 2$*, where* $Q_1 = \{x_0, x_1, x_2\}$*,* $Q_2 = \{q_0, q_1\}$*,* $\mathrm{AP} = \{\textit{Initial}, \textit{Goal}\}$*,* $L_1(x_0) = L_1(x_1) = L_2(q_0) = \{\textit{Initial}\}$ *and* $L_1(x_2) = L_2(q_1) = \{\textit{Goal}\}$*. The actions and transition relations are given in Figure 2.1. We regard* T_1 *as the original system and treat* T_2 *as the abstraction of* T_1 *in the following (nonrigorous) sense: (1) for all states in* Q_1*, there exists a state in* Q_2 *with the same label; (2) for all transitions in* T_1*, there exists a transition in* T_2 *with the same starting and ending labels as in* T_1*, respectively.*

It can be seen that the number of states in T_2 *is less than that of* T_1*, whereas the transition from* q_1 *(Goal) to* q_0 *(Initial) in* T_2 *does not have a counterpart in* T_1*. The abstraction* T_2 *also contains non-deterministic transitions given the same actions, e.g., under action number* 3 *on* q_1*, the post-transition state could be either* q_0 *or* q_1*. This somewhat heuristically illustrates the principle of constructing finite abstractions. We abstract states in the original system into finite and less states at a cost of potentially enlarging the transition relations. We then use the abstracted non-deterministic transition graph to connect with some automation generated by an LTL formula for model checking (see Remark A.0.8). Once the abstract model is verified to satisfy or be controllable w.r.t. some LTL specification, the original system should have the same property.*

In practice, T_1 *is usually given by a dynamical system driven by nonlinear vector fields with uncountably many states. Rigorous finite abstraction analysis and algorithms should be developed based on state-space discretization [112, 110, 108].* □

In [110], a fixed-point algorithm was developed for reach-and-stay specification regarding the computation complexity issue by adaptively partitioning the state space. However, for systems that also depend on tunable parameters and undergo bifurcations, such as the Moore-Greitzer model for jet engine compressors studied in this thesis, two challenges are raised for using formal methods: (1) the sampling time for constructing abstractions is highly related to the parameters since the system state evolves with different rates for different parameters, and (2) there are constraints on the change rates of the tunable parameters that can be treated as control inputs. Formal control synthesis tools such as [145, 109] cannot be used readily to design control strategies for such systems because of these challenges. While the multiscale abstractions [68, 85] can be used to deal with the first challenge, the existing formal control methods is incapable to deal with constraints on the change rates of the control inputs.

In contrast to formal methods, Lyapunov-like functions are able to provide feedback stabilizing and feedback set-invariance related controllers without state-space discretization and considering local dynamics [2, 5, 86, 127]. It has been a well established fact in control theory that stability properties can be characterized by Lyapunov functions. Reachability properties can also be naturally captured by Lyapunov functions for finite-time stability. On the other hand, barrier functions [138] are used to certify that solutions of a given dynamical system can stay within a prescribed safe set, along with their control variants, called control barrier functions [171, 5], to provide feedback controls that render the system safety. The barrier function approach can be further combined with Lyapunov method to satisfy stability and safety requirements simultaneously [159, 5, 174, 144, 87, 130]. Such formulations are amenable to optimization-based solutions enabled by quadratic programming [5, 179] or model predictive control [174], provided that the control system is in a control-affine form.

Another important characteristic of a dynamical system is whether or not its solutions can reach a certain target set from a given initial set with or without control. This is defined as reachability, which plays a key role in analysis and in particular control of dynamical systems [23]. Reachability analysis and control can also be viewed as an important special case of verification and control of dynamical systems with respect to more general formal specifications [110]. Since asymptotic stability entails asymptotic attraction, reachability can be naturally captured by asymptotic stability and Lyapunov conditions.

The stability/reachability and safety objectives, however, are sometimes conflicting. For example, while a system can reach a target set from a given initial set, it may have to traverse an unsafe region to do so. For this reason, when formulating the problem as an optimization problem, some authors defined safety as a hard constraint, and reachability/stability as a soft (performance) requirements [5].

The main objective of this chapter is to provide a theoretical perspective on uniting Lyapunov and barrier functions. The level sets of Lyapunov functions naturally define invariant sets that can be used to certify safety. The work in [159] used the notion of "barrier Lyapunov function" to ensure stability under state constraints is achieved. As pointed out by [5], such conditions sometimes are overly strong and conservative. The more recent work [144] proposed the notion of (control) Lyapunov-barrier function (the lower-bounded function W in [144, Proposition 1 and Definition 2]), and derived sufficient conditions for stability and stabilization with guaranteed safety. Despite the potential practical value of the control design framework presented in [144], the type of Lyapunov-barrier functions considered in [144] (defined on $\mathbb{R}^n$ and radially unbounded) implicitly imposes strong conditions on the unsafe set (e.g., it has to be unbounded [31, Theorem 11]). The authors of [31] then proposed sufficient conditions for safe stabilization using non-smooth control Lyapunov functions (see also [33]). The same authors also pointed out a technical inconsistency (see [32] for details) of the control Lyapunov-barrier conditions proposed in [144]. All these indicate that unifying Lyapunov and barrier functions is a non-trivial task.

Different from the aforementioned work, we aim to formulate sufficient and necessary Lyapunov conditions for asymptotic stability under state constraints. We show that, if we restrict the domain of the Lyapunov function to the set of initial conditions from which solutions can simultaneously satisfy the conditions of asymptotic stability and safety, then a smooth Lyapunov function can be found, building upon earlier results on converse Lyapunov functions [99, 160]. In particular, the results from [160] play a key role in inspiring us to formulate a Lyapunov function that is defined on the entire set of initial conditions from which the stability with safety specification is satisfied. We further extend the converse theorems to reach-avoid-stay type specifications. Since reachability (similar to asymptotic attraction) does not impose any stability conditions (see Vinograd's example [120, p. 120]), we in general cannot expect to find a Lyapunov function that is defined in a neighborhood of the target set. We use a robustness argument [114] to obtain a slightly weaker statement in the sense that if a reach-avoid-stay specification is satisfied robustly, then there exists a robust Lyapunov-barrier function that is robust under perturbations arbitrarily close to that of the original system.

The main results of this chapter are summarized as follows.

(1) We formulate the problems of stability with safety and reach-avoid-stay specifications and establish connections between them.

(2) We prove a smooth converse Lyapunov-barrier function theorem that is defined on the entire set of initial conditions from which the stability with safety property is satisfied.

(3) We extend the converse Lyapunov-barrier function theorem to reach-avoid-stay type specifications using a robustness argument. We show by example that such statements are the strongest one can obtain.

(4) We extend the converse Lyapunov-barrier functions to converse control Lyapunov-barrier functions w.r.t. reach-avoid-stay specifications, provided that there exists a Lipschitz continuous feedback law. A comparison between the proposed Lyapunov method and formal methods based on a fixed-point algorithm is illustrated by an application to enhancing the performance of jet engine compressors, which is based on a reduced Moore-Greitzer nonlinear ODE model. We apply a QP framework to reactively synthesize controllers in the case study.

(5) We prove the connections between stability with safety and reach-avoid-stay specifications for more general hybrid systems relying on the concept of solutions. We show that the Lyapunov-barrier approach can be extended to verification of reach-avoid-stay specifications for hybrid systems with differential and difference inclusions.

Conventions for Notation:

A continuous and strictly increasing function $\alpha : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is said to belong to class $\mathcal{K}$ if $\alpha(0) = 0$. It is said to belong to $\mathcal{K}_\infty$ if it belongs to class $\mathcal{K}$ and is unbounded. We denote by $|\cdot|$ the Euclidean norm, and by $\mathbb{B}$ the open unit ball centered at $\mathbf{0}$ w.r.t. $|\cdot|$.

2.1 System Description

Consider a control-free continuous-time dynamical system

$$\dot{x} = f(x), \tag{2.1}$$

where $x \in \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}^n$ is assumed to be locally Lipschitz. For each $x_0 \in \mathbb{R}^n$, we denote the unique solution starting from x_0 and defined on the maximal interval of existence by $\phi(t; x_0)$. For simplicity of notation, we may also write the solution as $\phi(t)$ if x_0 is not emphasized or as ϕ if the argument t is not emphasized.

Given a scalar $\vartheta \geq 0$, a ϑ-*perturbation* of the dynamical system (2.1) is described by the differential inclusion

$$\dot{x} \in F_\vartheta(x), \tag{2.2}$$

where $F_\vartheta(x) = f(x) + \vartheta\overline{\mathbb{B}}$. An equivalent description of the ϑ-perturbation of system (2.1) can be given by

$$\dot{x} = f(x) + \vartheta\xi, \tag{2.3}$$

where $\xi : \mathbb{R} \to \overline{\mathbb{B}}$ is any measurable signal. We denote system (2.1) by $\mathcal{S}$ and its ϑ-perturbation by $\mathcal{S}_\vartheta$. Note that $\mathcal{S}_\vartheta$ reduces to $\mathcal{S}$ when $\vartheta = 0$. A solution of $\mathcal{S}_\vartheta$ starting from x_0 can be denoted by $\phi(t; x_0, \xi)$, where ξ is a given disturbance signal. We may also write the solution simply as $\phi(t)$ or ϕ.

Remark 2.1.1. *We avoid using ε, ϵ and δ to denote the intensity of perturbations, since ϵ is frequently used from Chapter 2 to 4 for convergence analysis, ε causes confusion with ϵ, and δ is commonly used to denote a Dirac measure in the stochastic context. To keep the notations consistent in the first part of the thesis (Chapter 2 to 4), we use ϑ instead.*

We introduce some notation for reachable sets of $\mathcal{S}_\vartheta$. Denote the set of all solutions for $\mathcal{S}_\vartheta$ starting from x_0 by $\mathfrak{S}^{\vartheta}_{\mathcal{S}}(x_0)$. Let $\mathcal{R}^t_\vartheta(x_0)$ denote the set reached by solutions of $\mathcal{S}_\vartheta$ at time t starting from x_0, i.e.,

$$\mathcal{R}^t_\vartheta(x_0) = \left\{\phi(t) :\, \phi \in \mathfrak{S}^{\vartheta}_{\mathcal{S}}(x_0))\right\}.$$

For $T \geq 0$, we define

$$\mathcal{R}^{t\geq T}_\vartheta(x_0) = \bigcup_{t\geq T} \mathcal{R}^t_\vartheta(x_0), \quad \mathcal{R}^{0\leq t\leq T}_\vartheta(x_0) = \bigcup_{0\leq t\leq T} \mathcal{R}^t_\vartheta(x_0),$$

and write $\mathcal{R}_\vartheta(x_0) = \mathcal{R}^{t\geq 0}_\vartheta(x_0)$. For a set $\mathcal{X}_0 \subseteq \mathbb{R}^n$, we further define

- ✧ $\mathcal{R}^t_\vartheta(\mathcal{X}_0) = \bigcup_{x_0\in\mathcal{X}_0} \mathcal{R}^t_\vartheta(x_0)$;
- ✧ $\mathcal{R}^{t\geq T}_\vartheta(\mathcal{X}_0) = \bigcup_{x_0\in\mathcal{X}_0} \mathcal{R}^{t\geq T}_\vartheta(x_0)$;
- ✧ $\mathcal{R}^{0\leq t\leq T}_\vartheta(\mathcal{X}_0) = \bigcup_{x_0\in\mathcal{X}_0} \mathcal{R}^{0\leq t\leq T}_\vartheta(x_0)$;
- ✧ $\mathcal{R}_\vartheta(\mathcal{X}_0) = \bigcup_{x_0\in\mathcal{X}_0} \mathcal{R}_\vartheta(x_0)$.

Given a nonempty compact convex set of control inputs $\mathcal{U} \subseteq \mathbb{R}^p$, consider a nonlinear system of the form

$$\dot{x} = f(x) + g(x)\mathfrak{u} + \vartheta\xi, \tag{2.4}$$

where the mapping $g : \mathbb{R}^n \to \mathbb{R}^{n\times p}$ is smooth; $\mathfrak{u} : \mathbb{R}_{\geq 0} \to \mathcal{U}$ is a locally bounded measurable control signal, whilst the other notation remains the same.

Definition 2.1.2 (Control strategy)**.** *A control policy is a function*

$$\kappa : \mathbb{R}^n \to \mathcal{U}. \tag{2.5}$$

We further denote $\mathcal{S}^\kappa_\vartheta$ by the control system driven by (2.4) that is comprised by $\mathfrak{u} = \kappa(x)$.

Definition 2.1.3. *Let $h : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function. We introduce the following notations for (2.4):*

(1) $L_{f,d}h(x) = \nabla h(x) \cdot (f(x) + d)$; if $d = 0$, we simply use $L_f h(x) = \nabla h(x) \cdot f(x)$.

(2) $L_g h(x) = \nabla h(x) \cdot g(x)$.

2.2 Review of Barrier Conditions for Invariance Specifications

Before proceeding, we take a review on barrier functions and barrier conditions that ensure set invariance for nonlinear systems.

Definition 2.2.1 (Forward invariance)**.** *A set $\mathcal{C} \subset \mathbb{R}^n$ is said to be forward invariant for $\mathcal{S}_\vartheta$ (or ϑ-robustly forward invariant for $\mathcal{S}$), if solutions from $\mathcal{C}$ are forward complete (i.e., defined for all positive time) and $\mathcal{R}_\vartheta(\mathcal{C}) \subseteq \mathcal{C}$.*

Assumption 2.2.2. *We assume that $\mathcal{C}$ is defined as*

$$\mathcal{C} = \{x \in \mathbb{R}^n : h(x) \geq 0\}, \tag{2.6}$$

where $h : \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable function. Consequently,

$$\partial\mathcal{C} = \{x \in \mathbb{R}^n : h(x) = 0\}, \tag{2.7}$$

$$\text{Int}(\mathcal{C}) = \{x \in \mathbb{R}^n : h(x) > 0\}. \tag{2.8}$$

2.2.1 Reciprocal and Zeroing Barrier Functions

We start with two notions of barrier functions that are frequently used to guarantee forward invariance of a set. To illustrate the methodology of barrier functions, we consider control-free nonlinear systems $\mathcal{S}$ instead of a more general $\mathcal{S}_\vartheta$.

Definition 2.2.3 (Reciprocal barrier function)**.** *For $\mathcal{S}$, given a continuously differentiable function h as in (2.6), a continuously differentiable function $B : \text{Int}(\mathcal{C}) \to R$ is said to be a RBF for the set $\mathcal{C}$ if there exist functions $\alpha_1, \alpha_2, \alpha_3 \in \mathcal{K}$ such that, for all $x \in \text{Int}(\mathcal{C})$,*

$$\frac{1}{\alpha_1(h(x))} \leq B(x) \leq \frac{1}{\alpha_2(h(x))}, \tag{2.9}$$

$$\nabla B(x) \cdot f(x) \leq \alpha_3(h(x)). \tag{2.10}$$

Proposition 2.2.4. *[5, Theorem 1] Given a set $\mathcal{C} \subset \mathbb{R}^n$ defined in (2.6) for a continuously differentiable function h, if there exists an RBF $B : \text{Int}(\mathcal{C}) \to \mathbb{R}$, then $\text{Int}(\mathcal{C})$ is forward invariant.*

One drawback of RBFs is that unbounded function values occur as the argument of RBFs approaches $\partial\mathcal{C}$, which may be 'undesirable when real-time/embedded implementations are considered' [5]. We relax the conditions and consider the following notion of barrier functions.

Definition 2.2.5 (Zeroing barrier function)**.** *For $\mathcal{S}$, a continuously differentiable function h is said to be a ZBF for the set $\mathcal{C}$ if there exist an $\alpha \in \mathcal{K}$ and a set D with $\mathcal{C} \subseteq D \subset \mathbb{R}^n$ such that, for all $x \in D$,*

$$\nabla h(x) \cdot f(x) \geq -\alpha(h(x)). \tag{2.11}$$

In [5, Proposition 1], the authors verified that the existence of a ZBF for $\mathcal{S}$ implies the forward invariance of $\mathcal{C}$. However, a counterexample is provided in [166, Remark 4], which we rephrase as below.

Example 2.2.6. *Consider*

$$\dot{x} = -1, \ \ x(0) = x_0$$

and

$$h(x) = \begin{cases} \frac{2\sqrt{2}}{3\sqrt{3}} x^{3/2}, & x \geq 0, \\ -\frac{2\sqrt{2}}{3\sqrt{3}} x^{3/2}, & x < 0. \end{cases}$$

Then h is continuously differentiable on $\mathbb{R}$ and

$$\mathcal{C} = \{x \in \mathbb{R}^n : h(x) \geq 0\} = [0, \infty).$$

Now let $D = \mathcal{C}$ as required in Definition 2.2.5, then for all $x \in D$, we have

$$\nabla h(x) \cdot f(x) = -1 \cdot \frac{\sqrt{2}}{\sqrt{3}} x^{1/2} = -h^{1/3}(x).$$

Let $\alpha(\cdot) = (\cdot)^{1/3}$ on $[0, \infty)$, then it is clear that $\alpha \in \mathcal{K}$ and therefore h satisfies (2.11). However, the point 0 loses asymptotic behavior (due to the non-Lipschitz property of h) and $h(x)$ will reach 0 within finite time for any $x_0 > 0$.

We fix this problem by adding local Lipshitz condition on α. The modified version is given in Proposition 2.2.9.

Remark 2.2.7. *An alternative modification in Definition 2.2.5 to fix the problem is to impose D to be an open set strictly containing $\mathcal{C}$. We refer readers to [114, Remark 21] for more details.*

Lemma 2.2.8. *[70] Let $z : [t_0, t_f) \to \mathbb{R}$ be a continuously differentiable function satisfying the differential inequality*

$$\dot{z}(t) \geq -\alpha(z(t)), \ \forall t \in [t_0, t_f), \tag{2.12}$$

where $\alpha : \mathbb{R} \to \mathbb{R}$ is a locally Lipschitz extended class $\mathcal{K}$ function. Then there exists a class $\mathcal{KL}$ function $\beta : [0, \infty) \times [0, \infty) \to [0, \infty)$ (only depending on α) such that

$$z(t) \geq \beta(z(t_0), t - t_0), \quad \forall t \in [t_0, t_f).$$

Proof. Let

$$\dot{\tilde{z}}(t) = -\alpha(\tilde{z}(t)),\ \tilde{z}(t_0) = z(t_0). \tag{2.13}$$

Thus, the following estimates [103] are valid for any solution of (2.12):

$$z(t) \geq \tilde{z}(t),\ \forall t \in [t_0, t_f] \tag{2.14}$$

However, the solution to (2.13) is uniquely defined by $\tilde{z}(t) = \eta^{-1}(\eta(\tilde{z}(t_0)) + t - t_0)$ when $\tilde{z}(t_0) > 0$ [89], where $\eta(\tilde{z}) = -\int_b^{\tilde{z}} \frac{dx}{\alpha(x)}$ and $b > 0$. It can be verified that: (i) η is strictly decreasing on $(0, \infty)$; (ii) $\lim_{\tilde{z}\to 0} \eta(\tilde{z}) = \infty$; (iii) $\tilde{z}(t) \to 0$ as $t \to t_f$. When $\tilde{z}(t_0) = 0$, $\tilde{z}(t) = 0$ for all $t \in [t_0, t_f]$. Therefore, $\tilde{z}(t) \geq 0$ whenever $\tilde{z}(t_0) \geq 0$, and $z(t) \geq \tilde{z}(t) \geq 0$ for all $t \in [t_0, t_f]$ whenever $z(t_0) \geq 0$. □

Proposition 2.2.9. *Given a continuously differentiable function $h : \mathbb{R}^n \to \mathbb{R}$ and dynamics on $\mathbb{R}^n$*

$$\dot{x} = f(x) \tag{2.15}$$

such that $f : \mathbb{R}^n \to \mathbb{R}$ is locally Lipschitz. Let $\mathcal{C} = \{x :\ h(x) \geq 0\}$, and $\mathrm{Int}(\mathcal{C}) := \{x :\ h(x) > 0\}$. If the Lie derivative of h along the trajectories of x satisfies

$$\nabla h(x) \cdot f(x) \geq -\alpha(h(x)),\ \forall x \in \mathcal{C} \tag{2.16}$$

where α is a locally Lipschitz extended class $\mathcal{K}$ function, then the set $\mathrm{Int}(\mathcal{C})$ is forward invariant.

Proof. If $\mathrm{Int}(\mathcal{C}) = \emptyset$, then it is invariant. Otherwise, we apply Lemma 2.2.8, it follows that if $x(t_0) \in \mathrm{Int}(\mathcal{C})$, then we have $h(x(t)) > 0$ for all $t \in [t_0, t_f)$, where $[t_0, t_f)$ is the maximal interval of existence for $x(t)$ starting from $x(t_0)$. □

Remark 2.2.10. *Note that the result cannot be extended to the invariance of the entire set $\mathcal{C}$, despite that it is widely stated so in the literature. A simple counterexample is that we let $h(x) = -x^2$, and therefore $\mathcal{C} = \{0\}$. Then, for $\dot{x} = c \neq 0$, even though we have a satisfaction of (2.16) on $\mathcal{C} = \{0\}$, it is not invariant under the flow.*

2.2.2 Control Barrier Functions

We extend the notion of RBF to CBF for systems with controls. To illustrate the methodology, we consider system (2.4) with $\vartheta \equiv 0$.

Definition 2.2.11. *Let $h : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function and $\mathcal{C} \subseteq D \subset \mathbb{R}^n$ be as in (2.6). Then h is a CBF if*

(1) $L_g h(x) \neq 0$ *for all* $x \in D$,

(2) and there exists an $\alpha \in \mathcal{K}_\infty$ *such that for the control system* (2.1),

$$\sup_{u \in \mathcal{U}} [L_f h(x) + L_g h(x)u] \geq -\alpha(h(x)), \quad \forall x \in D.$$

We further define a set of state-dependent control strategy

$$\mathfrak{K}(x) := \{u \in \mathcal{U} : L_f B(x) + L_g B(x)u + \alpha(h(x)) \geq 0 \; \forall x \in D\}.$$

By a similar argument as in Proposition 2.2.9, we suppose α is also locally Lipschitz continuous. Then, if $\kappa(x(t)) \in \mathfrak{K}(x(t))$ and Lipschitz continuous for all $t \geq 0$, the set $\text{Int}(\mathcal{C})$ is controlled invariant for $\mathcal{S}^\kappa$.

The notion of CBF can be extended to high relative-degree control systems where $L_g h(x) = 0$. We can correspondingly define the relative degree of the continuously differentiable function h on a set with respect to a system as in (2.1), which is the number of times we need to differentiate h along the dynamics of the system before the control input $\mathfrak{u}$ explicitly appears. The set invariance given the existence of HOCBF is guaranteed [166]. We omit the details due to the less relevance to the main topic of the thesis.

We have seen ZBFs and the associated CBFs with the relaxed barrier conditions. However, this type of relaxed barrier conditions fail to guarantee a set invariance with high probability in the stochastic context (see details in Chapter 3). In particular, for control systems with high relative degree, the relaxed conditions perform even worse. Different concepts of stochastic control barrier functions are compared in Section 3.2 in purpose of guaranteeing a set invariance with a high probability.

2.2.3 Robustness and Converse Barrier Functions

Now that the mechanism of barrier functions and conditions are understood via two special cases, we involve robustness and define robust barrier function aiming to characterize safety.

Definition 2.2.12 (Robustly safe set)**.** *[114] Given an unsafe set* $U \subseteq \mathbb{R}^n$, *a set* $\mathcal{C} \subseteq \mathbb{R}^n$ *is said to be* ϑ*-robustly safe w.r.t. to U if all solutions of* $\mathcal{S}_\vartheta$ *starting from* $\mathcal{C}$ *will not enter U.*

It has been verified in [114, Proposition 4] that : If there exists a ϑ-robustly invariant set D such that $\mathcal{C} \subseteq D$ and $U \cap D = \emptyset$, then $\mathcal{C}$ is ϑ-robustly safe w.r.t. to U.

We generalize RBF and ZBF for $\mathcal{S}_\vartheta$ in the following sense [114, Definition 5].

Definition 2.2.13 (Robust barrier function)**.** *Given sets $\mathcal{C}, U \subseteq \mathbb{R}^n$, a continuously differentiable function $B : \mathbb{R}^n \to \mathbb{R}$ is said to be a ϑ-robust barrier function for $\mathcal{C}$ and U if the following conditions are satisfied:*

(1) $B(x) \geq 0$ *for all* $x \in \mathcal{C}$*;*

(2) $B(x) < 0$ *for all* $x \in U$*;*

(3) $\nabla B(x) \cdot (f(x) + \vartheta d) > 0$ *for all* x *such that* $B(x) = 0$ *and* $d \in \overline{\mathbb{B}}$*.*

Robust control barrier functions can be defined in a similar way as in Definition 2.2.11. Sufficiency of barrier functions in terms of ensuring a set safety is also guaranteed, the proof of which falls in standard Lyapunov-type arguments.

Remark 2.2.14. *Note that a barrier function B works in a way to separate two disjoint sets $\mathcal{C}$ and U, and meanwhile guarantees the invariance of $\mathcal{C}$. However, the sets $\mathcal{C}$ and U are free to choose as long as they are disjoint, e.g. it is not necessary that $\overline{\mathcal{C}} \cap \overline{U} \neq \emptyset$ or $\mathcal{C} \cup U = \mathbb{R}^n$. Furthermore, to deal with safety control problems, it is natural to arbitrarily look for a set $\mathcal{C}$ containing the initial set of states $\mathcal{X}_0$ and conversely find a CBF for $(\mathcal{C}, U)$.*

The necessity of the existence of B given the satisfaction of a safety specification w.r.t. $(\mathcal{C}, U)$ can be verified based on the compactness of reachable set from the set of initial conditions $\mathcal{X}_0$, which does not intersect with U by the *a priori* safety assumption [114]. This compactness in turn makes it easier to construct B via the Lyapunov function V.

The latest converse barrier function theorem by [64], however, does not depend on the boundedness of $\mathbb{R}^n \setminus U$, and hence does not require the compactness of $\mathcal{X}_0$. The construction of barrier function given the safety of robust systems relies on the 'time-to-impact' function $B_{\mathcal{R}}$ w.r.t. the reachable set $\mathcal{R}$ of robust systems. This $B_{\mathcal{R}}$ turns out to satisfy condition (1) and (2) of Definition 2.2.13 as well as a weaker certificate[2]

$$\nabla B(x) \cdot f(x) > 0, \;\; \forall x \in \partial S, \;\; f \in F(x) + \vartheta \overline{\mathbb{B}}.$$

The result in [64] is of great theoretical interest in the sense of constructing barrier functions with relaxed topological requirement and barrier conditions. Their construction relies on the seemingly less intuitive time-to-impact functions, whereas our approach aims to unify Lyapunov and barrier functions. It would be interesting to investigate in future work whether the Lyapunov approach can be extended to handle unbounded reachable sets.

[2]Note that we have adapted the notion to be consistent with the notation used in this section.

2.3 Lyapunov-Barrier Theorems for Asymptotic Stability with Safety Constraints and Reach-Avoid-Stay Specifications

In this section, we formally define two common types of properties for solutions of $\mathcal{S}_\vartheta$ and highlight the connections between them. We then derive converse Lyapunov-barrier function theorems for $\mathcal{S}_\vartheta$ satisfying such specifications, respectively.

The first one is on reaching a target set in finite time and remaining there after, while avoiding an unsafe set. This is often called a reach-avoid-stay type specification.

Definition 2.3.1 (Reach-avoid-stay specification)**.** *We say that $\mathcal{S}_\vartheta$ satisfies a reach-avoid-stay specification $(\mathcal{X}_0, U, \Gamma)$, where $\mathcal{X}_0, U, \Gamma \subseteq \mathbb{R}^n$, if the following conditions hold:*

(1) (reach and stay w.r.t. Γ) Solutions of $\mathcal{S}_\vartheta$ starting from $\mathcal{X}_0$ are defined for all positive time (i.e., forward complete) and there exists some $T \geq 0$ such that $\mathcal{R}_\vartheta^{t\geq T}(\mathcal{X}_0) \subseteq \Gamma$.

(2) (safe w.r.t. U) $\mathcal{R}_\vartheta(\mathcal{X}_0) \cap U = \emptyset$.

If these conditions hold, we also say that $\mathcal{S}$ ϑ-robustly satisfies the reach-avoid-stay specification $(\mathcal{X}_0, U, \Gamma)$.

A closely related property for solutions of $\mathcal{S}_\vartheta$ is stability with safety guarantees. We first define stability for solutions of $\mathcal{S}_\vartheta$ w.r.t. a closed set.

Definition 2.3.2 (Set stability)**.** *A closed set $A \subset \mathbb{R}^n$ is said to be UAS for $\mathcal{S}_\vartheta$ if the following two conditions are met:*

(1) (uniform stability) For every $\epsilon > 0$, there exists an $\eta_\epsilon > 0$ such that $|\phi(0)|_A < \eta_\epsilon$ implies that $\phi(t)$ is defined for $t \geq 0$ and $|\phi(t)|_A < \epsilon$ for any solution ϕ of $\mathcal{S}_\vartheta$ for all $t \geq 0$; and

(2) (uniform attractivity) There exists some $\varrho > 0$ such that, for every $\epsilon > 0$, there exists some $T > 0$ such that $\phi(t)$ is defined for $t \geq 0$ and $|\phi(t)|_A < \epsilon$ for any solution ϕ of $\mathcal{S}_\vartheta$ whenever $|\phi(0)|_A < \varrho$ and $t \geq T$.

If these conditions hold, we also say that A is ϑ-robustly UAS for $\mathcal{S}$.

Definition 2.3.3 (Domain of attraction)**.** *If a closed set $A \subseteq \mathbb{R}^n$ is ϑ-robustly UAS for $\mathcal{S}$, we further define the domain of attraction of A for $\mathcal{S}_\vartheta$, denoted by $\mathcal{G}_\vartheta(A)$, as the set of all initial states $x \in \mathbb{R}^n$ such that any solution $\phi \in \mathfrak{S}_{\mathcal{S}}^{\vartheta}(x)$ is defined for all positive time and converges to the set A, i.e.,*

$$\mathcal{G}_\vartheta(A) = \left\{x \in \mathbb{R}^n : \forall \phi \in \mathfrak{S}_{\mathcal{S}}^{\vartheta}(x),\ \lim_{t\to\infty} |\phi(t)|_A = 0\right\}.$$

Definition 2.3.4 (Stability with safety guarantee)**.** *We say that $\mathcal{S}_\vartheta$ satisfies a stability with safety guarantee specification $(\mathcal{X}_0, U, A)$, where $\mathcal{X}_0, U, A \subseteq \mathbb{R}^n$ and A is closed, if the following conditions hold:*

(1) (asymptotic stability w.r.t. A) The set A is UAS for $\mathcal{S}_\vartheta$ and the domain of attraction of A contains $\mathcal{X}_0$, i.e. $\mathcal{X}_0 \subseteq \mathcal{G}_\vartheta(A)$.

(2) (safe w.r.t. U) $\mathcal{R}_\vartheta(W) \cap U = \emptyset$.

If these conditions hold, we also say that $\mathcal{S}$ ϑ-robustly satisfies the stability with safety guarantee specification $(\mathcal{X}_0, U, A)$.

2.3.1 Lyapunov-Barrier Function for Stability with Safety Guarantees

In this subsection, we derive a converse Lyapunov-barrier function theorem for $\mathcal{S}_\vartheta$ satisfying a stability with safety guarantee specification $(\mathcal{X}_0, U, A)$.

Definition 2.3.5. *[160] Let $A \subseteq \mathbb{R}^n$ be a compact set contained in an open set $D \subseteq \mathbb{R}^n$. A continuous function $\mathfrak{w} : D \to \mathbb{R}_{\geq 0}$ is said to be a proper indicator for A on D if the following two conditions hold: (1) $\mathfrak{w}(x) = 0$ if and only if $x \in A$; (2) $\lim_{m\to\infty} \mathfrak{w}(x_m) = \infty$ for any sequence $\{x_m\}$ in D such that either $x_m \to p \in \partial D$ or $|x_m| \to \infty$ as $m \to \infty$.*

Intuitively, a proper indicator for a compact set $A \subseteq D$, where $D \subseteq \mathbb{R}^n$ is open, is a continuous function whose value equals zero if and only if on A and approaches infinity at the boundary of D or at infinity. It generalizes the idea of a radially unbounded function.

Remark 2.3.6. *Let $A \subseteq \mathbb{R}^n$ be a compact set contained in an open set $D \subseteq \mathbb{R}^n$. There is always a proper indicator for A on D defined by [160, Remark 2]*

$$\mathfrak{w}(x) = \max\left\{|x|_A, \frac{1}{|x|_{\mathbb{R}^n\setminus D}} - \frac{2}{dist(A, \mathbb{R}^n \setminus D)}\right\},$$

where $dist(A, \mathbb{R}^n \setminus D) = \inf_{x\in A} |x|_{\mathbb{R}^n\setminus D}$. Indeed, $\mathfrak{w}$ is clearly continuous. If $x \in A$, we have $\mathfrak{w}(x) = |x|_A = 0$. If $x \in D \setminus A$, we have $\mathfrak{w}(x) \geq |x|_A > 0$. For any $\{x_m\}$ in D such that either $x_m \to p \in \partial D$ or $|x_m| \to \infty$ as $n \to \infty$, we either have $|x_m|_A \to \infty$ or $\frac{1}{|x_m|_{\mathbb{R}^n\setminus D}} \to \infty$.

Theorem 2.3.7. *Suppose that A is compact, U is closed, and $A \cap U = \emptyset$. Then the following two statements are equivalent:*

(1) $\mathcal{S}_\vartheta$ satisfies the stability with safety guarantee specification $(\mathcal{X}_0, U, A)$.

(2) There exists an open set D such that $(A \cup \mathcal{X}_0) \subseteq D$ and $D \cap U = \emptyset$, a smooth function $V : D \to \mathbb{R}_{\geq 0}$ and class $\mathcal{K}_\infty$ functions α_1 and α_2 such that, for all $x \in D$ and $d \in \vartheta\overline{\mathbb{B}}$,

$$\alpha_1(\mathfrak{w}(x)) \leq V(x) \leq \alpha_2(\mathfrak{w}(x)), \tag{2.17}$$

and

$$\nabla V(x) \cdot (f(x) + d) \leq -V(x), \tag{2.18}$$

where $\mathfrak{w}$ be any proper indicator for A on D,

Moreover, the set D can be taken as the following set

$$\mathcal{W}_\vartheta = \left\{ x \in \mathbb{R}^n : \forall \phi \in \mathfrak{S}^{\vartheta}_{\mathcal{S}}(x), \lim_{t\to\infty} |\phi(t)|_A = 0 \text{ and } \phi(t) \notin U, \forall t \geq 0 \right\}.$$

Clearly, the set $\mathcal{W}_\vartheta$ defined above includes all initial states from which solutions of $\mathcal{S}_\vartheta$ will approach A and avoid the unsafe set U. The following lemma establishes some basic properties of the set $\mathcal{W}_\vartheta$. The proof is completed in Section 2.3.3.

Lemma 2.3.8. *Suppose that A is compact, U is closed, and $A \cap U = \emptyset$. If $\mathcal{S}_\vartheta$ satisfies a stability with safety guarantee specification $(\mathcal{X}_0, U, A)$, then the set $\mathcal{W}_\vartheta$ is open, forward invariant, and satisfies $\mathcal{X}_0 \subseteq \mathcal{W}_\vartheta \subseteq \mathcal{G}_\vartheta(A)$.*

The proof of Theorem 2.3.7 relies on the following result, which states that, on any forward invariant open subset D of $\mathcal{G}_\vartheta(A)$, we can find a "global" Lyapunov function relative to D.

Proposition 2.3.9. *Let $A \subseteq \mathbb{R}^n$ be a compact set that is UAS for $\mathcal{S}_\vartheta$. Let $D \subseteq \mathbb{R}^n$ be an open set such that $A \subseteq D \subseteq \mathcal{G}_\vartheta(A)$ and D is forward invariant for $\mathcal{S}_\vartheta$, where $\mathcal{G}_\vartheta(A)$ is the domain of attraction of A for $\mathcal{S}_\vartheta$. Let $\mathfrak{w}$ be any proper indicator for A on D. Then there exists a smooth function $V : D \to \mathbb{R}_{\geq 0}$ and class $\mathcal{K}_\infty$ functions α_1 and α_2 such that conditions (2.17) and (2.18) hold for all $x \in D$ and $d \in \vartheta\overline{\mathbb{B}}$.*

This proposition can be proved by combining the proof for Proposition 3 and the statements of Theorem 2 and Theorem 1 in [160]. The main difference being that the results in [160] are stated for more general differential inclusions and Proposition 3 in [160] is proved on the *whole* *domain of attraction* of A, whereas the above results are for specific ϑ-perturbations of a Lipschitz ordinary differential equation and for any open forward invariant set containing the set A. Due to this subtlety, Proposition 3 of [160] is not directly applicable for our purpose. For completeness, we provide a more direct proof of this result in Section 2.3.3.

Proof of Theorem 2.3.7 We first prove (2) $\Longrightarrow$ (1). The fact that V is a smooth Lyapunov function, i.e., satisfying conditions (2.17) and (2.18), on an open neighborhood D containing A shows that A is UAS for $\mathcal{S}_\vartheta$. We show that the set D is forward invariant. Let $x_0 \in D$. Then for any $\phi \in \mathfrak{S}^{\vartheta}_{\mathcal{S}}(x_0)$, we have

$$\frac{dV(\phi(t))}{dt} = \nabla V(\phi(t)) \cdot (f(\phi(t)) + d(t)) \leq 0$$

holds for almost all $t \geq 0$. It follows that $V(\phi(t)) \leq V(x_0) < \infty$. Hence $\phi(t)$ is bounded, defined, and satisfies $\phi(t) \in D$ for all $t \geq 0$. By forward invariance of D and $\mathcal{X}_0 \subseteq D$, we have $\mathcal{R}_\vartheta(W) \subseteq D$ and $\mathcal{R}_\vartheta(W) \cap U = \emptyset$. It remains to show that $\mathcal{X}_0 \subseteq \mathcal{G}_\vartheta(A)$. For any $x_0 \in \mathcal{X}_0$ and any $\phi \in \mathfrak{S}^{\vartheta}_{\mathcal{S}}(x_0)$, we have $\phi(t) \in D$ for all $t \geq 0$. Hence

$$\frac{dV(\phi(t))}{dt} = \nabla V(\phi(t)) \cdot (f(\phi(t)) + d(t)) \leq -V(\phi(t)) < 0$$

as long as $\phi(t) \notin A$. A standard Lyapunov argument shows that $|\phi(t)|_A \to 0$ as $t \to \infty$. Hence $x_0 \in \mathcal{G}_\vartheta(A)$ and $\mathcal{X}_0 \subseteq \mathcal{G}_\vartheta(A)$. We have verified that $\mathcal{S}_\vartheta$ satisfies a stability with safety guarantee specification $(\mathcal{X}_0, U, A)$.

We then prove (1) $\Longrightarrow$ (2). By Lemma 2.3.8, we can let $D = \mathcal{W}_\vartheta$. Then $(A \cup \mathcal{X}_0) \subseteq D \subseteq \mathcal{G}_\vartheta(A)$. Furthermore, D is open and forward invariant. The conclusion follows from that of Proposition 2.3.9. $\square$

Remark 2.3.10. *Compared with related results on sufficient Lyapunov conditions for stability with safety guarantees (e.g., [144, 31, 33]), to the best knowledge of the authors, Theorem 2.3.7 provides the first converse Lyapunov-barrier theorem and we show that the converse Lyapunov function is defined on whole set of initial conditions from which asymptotic stability with safety guarantees is satisfied. In other words, we provide a Lyapunov characterization of the problem of asymptotic stability with safety guarantees. We also note that several converse barrier functions have been reported in the literature [172, 140, 114]. In particular, the recent work [114] makes a connection between converse Lyapunov function and converse barrier function via a robustness argument, which, to some extent, inspired our work in this section to unify converse Lyapunov and barrier functions. The results of this section significantly differ from that in [114], because converse results are established for both stability with safety guarantees and reach-avoid-stay specifications,* *whereas the results in [114] only concern safety. We achieved this non-trivial extension by adapting converse Lyapunov theorems (e.g., [160]), as in Proposition 2.3.9, to work with safety requirements, enabled by characterizing all initial states from which solutions will satisfy stability with safety guarantees, as in Lemma 2.3.8.*

While Theorem 2.3.7 gives a single smooth Lyapunov function satisfying the strong set of conditions (2.17) and (2.18), we propose the following set of sufficient conditions for two reasons. First, they appear to be weaker (although in fact theoretically equivalent in view of Theorem 2.3.7) and perhaps easier to verify in practice [122]. Second, they agree with the notions of Lyapunov and barrier functions commonly seen in the literature.

Proposition 2.3.11. *Suppose that A is compact, U is closed, and $A \cap U = \emptyset$. If there exists an open set D such that $(A \cup \mathcal{X}_0) \subseteq D$ and smooth functions $V : D \to \mathbb{R}_{\geq 0}$ and $B : \mathbb{R}^n \to \mathbb{R}$ such that*

(1) V is positive definite on D w.r.t. A, i.e., $V(x) = 0$ if and only if $x \in A$;

(2) $\nabla V(x) \cdot (f(x) + d) < 0$ for all $x \in D \setminus A$ and $d \in \vartheta\overline{\mathbb{B}}$;

(3) $\mathcal{X}_0 \subseteq C = \{x \in \mathbb{R}^n : B(x) \geq 0\} \subseteq D$ and $B(x) < 0$ for all $x \in U$;

(4) $\nabla B(x) \cdot (f(x) + d) \geq 0$ for all $x \in D$ and $d \in \vartheta\overline{\mathbb{B}}$,

then $\mathcal{S}_\vartheta$ satisfies the stability with safety guarantee specification $(\mathcal{X}_0, U, A)$. Furthermore, if $\mathcal{X}_0$ is compact, then conditions (1)–(4) are also necessary for $\mathcal{S}_\vartheta$ to satisfy the stability with safety guarantee specification $(\mathcal{X}_0, U, A)$.

Proof. We first prove the sufficiency part. Conditions (1)–(2) state that V is a local Lyapunov function for $\mathcal{S}_\vartheta$ w.r.t. A. Hence A is UAS for $\mathcal{S}_\vartheta$. Conditions (3)–(4) state that B is a barrier function for $\mathcal{S}_\vartheta$ w.r.t. $(\mathcal{X}_0, U)$.

We can easily show that the set $\mathcal{C} = \{x \in \mathbb{R}^n : B(x) \geq 0\}$ is forward invariant. Indeed, if $\mathcal{C}$ is not forward invariant, then there exists some $x_0 \in \mathcal{C}$, a solution $\phi \in \mathfrak{S}^\vartheta_{\mathcal{S}}(x_0)$, and some $\tau > 0$ such that $B(\phi(\tau)) < 0$. Define

$$\bar{t} = \sup\{t \geq 0 : \phi(t) \in \mathcal{C}\}.$$

Then $\bar{t}$ is well defined and finite. By continuity of $B(\phi(t))$, we have $B(\phi(\bar{t})) = 0$. Since $\phi(\bar{t}) \in D$ and D is open, for $\epsilon > 0$ sufficiently small, we have $\phi(t) \in D$ for almost all $t \in [\bar{t}, \bar{t} + \epsilon]$. This implies that, for all $t \in [\bar{t}, \bar{t} + \epsilon]$,

$$\frac{dB(\phi(t))}{dt} = \nabla B(\phi(t)) \cdot (f(\phi(t)) + d(t)) \geq 0.$$

Hence we have $B(\phi(t)) \geq B(\phi(\bar{t})) = 0$ for all $t \in [\bar{t}, \bar{t} + \epsilon]$. This contradicts the definition of $\bar{t}$. Hence $\mathcal{C}$ must be forward invariant. Since $\mathcal{X}_0 \subseteq \mathcal{C}$ and $\mathcal{C} \cap U = \emptyset$, we have $\mathcal{R}_\vartheta(\mathcal{X}_0) \subseteq \mathcal{C}$ and $\mathcal{R}_\vartheta(\mathcal{X}_0) \cap U = \emptyset$.

It remains to show that $\mathcal{X}_0 \subseteq \mathcal{G}_\vartheta(A)$. For any $x_0 \in \mathcal{X}_0$ and any $\phi \in \mathfrak{S}^\vartheta_{\mathcal{S}}(x_0)$, we have $\phi(t) \in \mathcal{C} \subseteq D$ for all $t \geq 0$. Hence

$$\frac{dV(\phi(t))}{dt} = \nabla V(\phi(t)) \cdot (f(\phi(t)) + d(t)) < 0$$

as long as $\phi(t) \notin A$. A standard Lyapunov argument shows that $|\phi(t)|_A \to 0$ as $t \to \infty$.

We then prove the necessity part. This follows from Theorem 2.3.7. Let V and D be given by Theorem 2.3.7(2). Choose any $c > 0$ such that $\mathcal{X}_0 \subseteq \mathcal{C} = \{x \in D : V(x) \leq c\}$. This is always possible, because $\mathcal{X}_0$ is compact and we can simply take $c = \sup_{x \in \mathcal{X}_0} V(x)$. The idea is to construct the function B in the form of $B(x) = c - V(x)$. To make this precise, choose $c_2 > c_1 > c$ and let $D_i = \{x \in D : V(x) < c_i\}$ ($i = 1, 2$). Construct an open cover $\{O_1, O_2, O_3\}$ of $\mathbb{R}^n$ by $O_1 = D_1$, $O_2 = D_2 \setminus \mathcal{C}$, and $O_3 = \mathbb{R}^n \setminus \overline{D_1}$. Define $B_1(x) = c - V(x)$ for $x \in O_1$, $B_2(x) = c - V(x)$ for $x \in O_2$, and $B_3(x) = c - c_1$ for $x \in O_3$; elsewhere, they are defined to be zero. Then $B_i(x)$ ($i = 1, 2, 3$) are locally smooth. The rest of the proof follows the idea of the extension lemma for smooth functions [106, Lemma 2.26], which justifies the existence of a globally smooth function that coincides with the locally smooth function when it is restricted to the subdomain. Now, let $\{\psi_1, \psi_2, \psi_3\}$ be a smooth partition of unity [106, p. 43] subordinate to $\{O_1, O_2, O_3\}$. Construct $B(x) = \sum_{i=1}^3 \psi_i(x) B_i(x)$, for $x \in \mathbb{R}^n$. It is easy to verify that $B(x)$ satisfies conditions (3)–(4) with $D = D_1$. First, for $x \in O_1 = D_1$, we have $\psi_3(x) = 0$ and $B(x) = \sum_{i=1}^2 \psi_i(x) B_i(x) = c - V(x)$. Hence, by the construction of V and (2.18), $\nabla B(x) \cdot (f(x) + d) = -\nabla V(x) \cdot (f(x) + d) \geq V(x) \geq 0$, for all $x \in D_1$ and $d \in \vartheta\overline{\mathbb{B}}$. This verifies condition (4). Furthermore, if $x \notin D_1$, then either $x \notin D$ or $x \in D$ but $V(x) \geq c_1$. It follows that $B(x) \leq \sum_{i=1}^3 \psi_i(x)(c - c_1) = c - c_1 < 0$. Hence, we have $\mathcal{C} = \{x \in \mathbb{R}^n : B(x) \geq 0\} \subseteq D_1$ and we already have $\mathcal{X}_0 \subseteq \mathcal{C}$. Since $D \cap U = \emptyset$, for $x \in U$, we have $\psi_1(x) = \psi_2(x) = 0$ and $B(x) = B_3(x) = c - c_1 < 0$. This verifies condition (3). Clearly, by the construction of V, it verifies conditions (1)–(2) with $D = D_1$. □

Remark 2.3.12. *We can see from the proof of Proposition 2.3.11 that it is without loss of generality to construct barrier functions from Lyapunov functions [114]. Indeed, the invariant set $\mathcal{C}$, which separates the safe initial conditions from the unsafe set, is given by a level set of the Lyapunov function:*

$$\mathcal{C} = \{x \in \mathbb{R}^n : B(x) \geq 0\} = \{x \in D : V(x) \leq c\}.$$

Remark 2.3.13. *We compare the Lyapunov-barrier conditions with that in [144], which provided a novel control framework for stabilization with guaranteed safety for nonlinear systems. Nonetheless, we restrict the formulation to autonomous systems (cf. Proposition 1 in [144]). This is without* *loss of generality, because the control framework in [144] is fundamentally built upon the conditions for autonomous systems, as clearly indicated in [144] (see, e.g., the remark before and proof of [144, Theorem 3]). We also change the notion slightly to be consistent with the notation used in this thesis. In [144], a set of sufficient conditions for a smooth function $V : \mathbb{R}^n \to \mathbb{R}$ to be called a Lyapunov-barrier function for the system (2.1) with respect to the origin and an unsafe set U were formulated as follows:*

(1) V is lower-bounded and radially unbounded;

(2) $V(x) > 0$ for all $x \in U$;

(3) $\nabla V(x) \cdot f(x) < 0$ for all $x \in \mathbb{R}^n \setminus (U \cup \{0\})$; and

(4) $\overline{\mathbb{R}^n \setminus (U \cup \mathcal{C})} \cap \overline{U} = \emptyset$, where the set $\mathcal{C}$ is given by $\mathcal{C} = \{x \in \mathbb{R}^n : V(x) \le 0\}$.

In [31], it is shown that the above conditions imply the set U is necessarily unbounded. Here we show another property that indicates the restrictive nature of condition (4); that is,

$$x \in \partial U \text{ implies } V(x) = 0. \tag{2.19}$$

In fact, suppose that this is not the case, then $V(x) > 0$. There exists a sequence $\{x_n\} \to x \in \partial D$ such that $V(x_n) > 0$ (and hence $\{x_n\} \cap \mathcal{C} = \emptyset$) and $\{x_n\} \cap U = \emptyset$ (this is possibly because $x \in \partial U$). Hence $\{x_n\} \subseteq \mathbb{R}^n \setminus (U \cup \mathcal{C})$. It follows that $x \in \overline{\mathbb{R}^n \setminus (U \cup \mathcal{C})}$. By condition (4) above, $x \notin \overline{U}$, which contradicts $x \in \partial U$. In view of (2.19), condition (4) above is somewhat restrictive, because it implies that the boundary of the unsafe set U lies entirely on a level curve of V.

Remark 2.3.14. *Figure 2.2 provides an illustration of the sets defined for proving Theorem 2.3.7.*

2.3.2 Lyapunov-Barrier Function for Reach-Avoid-Stay Specifications

The converse results proved in the previous section can be extended to reach-avoid-stay specifications under some mild modifications.

Suppose that $\mathcal{S}_\vartheta$ satisfies a reach-avoid-stay specification $(\mathcal{X}_0, U, \Gamma)$.

Lemma 2.3.15. *Suppose that Γ is compact and $\mathcal{X}_0$ is nonempty. If $\mathcal{S}_\vartheta$ satisfies a reach-avoid-stay specification $(\mathcal{X}_0, U, \Gamma)$, then the set*

$$A = \left\{x \in \Omega : \forall \phi \in \mathfrak{S}^{\vartheta}_{\mathcal{S}}(x),\ \phi(t) \in \Omega, \forall t \ge 0\right\}. \tag{2.20}$$

is a nonempty compact invariant set for $\mathcal{S}_\vartheta$.

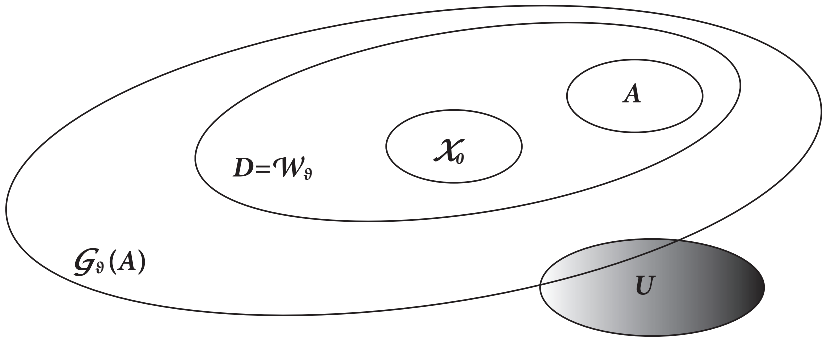


Figure 2.2: An illustration of the sets involved in Theorem 2.3.7, Lemma 2.3.8, and Proposition 2.3.9. While the domain of attraction $\mathcal{G}_\vartheta(A)$ can potentially intersect with the unsafe set U, the winning set $\mathcal{W}_\vartheta$ defined in (2.19) characterizes the set of initial conditions from which the stability with safety constraints is satisfied. Clearly, the system $\mathcal{S}_\vartheta$ satisfies a stability with safety specification $(\mathcal{X}_0, U, A)$ if and only if $\mathcal{X}_0 \subset \mathcal{W}_\vartheta$. Theorem 2.3.7 (together with Lemma 2.3.8 and Proposition 2.3.9) states that a smooth Lyapunov function can be found on the set $D = \mathcal{W}_\vartheta$ to verify the specification $(\mathcal{X}_0, U, A)$. [121]

Proof. We first show that A is nonempty. By the definition of reach-avoid-stay specification $(\mathcal{X}_0, U, \Gamma)$, solutions of $\mathcal{S}_\vartheta$ starting from $\mathcal{X}_0$ are forward complete and there exists some $T \geq 0$ such that $\mathcal{R}_\vartheta^{t\geq T}(\mathcal{X}_0) \subseteq \Gamma$. It is easy to verify that the set $\mathcal{R}_\vartheta^{t\geq T}(\mathcal{X}_0)$ is forward invariant for $\mathcal{S}_\vartheta$. Clearly, $\mathcal{R}_\vartheta^{t\geq T}(\mathcal{X}_0) \subseteq A$ and A is nonempty.

We next show that A is compact. Since $A \subseteq \Gamma$ and Γ is compact, we only need to show that A is closed. Note that A is forward invariant by definition. Let $\{x_m\}$ be a sequence in A that converges to x. Since Γ is compact, we have $x \in \Gamma$. Suppose that $x \notin A$. Then there exists some $\phi \in \mathfrak{S}_\mathcal{S}^\vartheta(x)$ and some $\tau > 0$ such that $\phi(\tau) \notin \Gamma$. By continuous dependence of solutions of $\mathcal{S}_\vartheta$ on initial conditions, there exists a sequence of solutions $\phi_m \in \mathfrak{S}_\mathcal{S}^\vartheta(x_m)$ that converges to ϕ uniformly on $[0, \tau]$. We have $\phi_m(\tau) \to \phi(\tau) \notin \Gamma$ as $m \to \infty$. Since $\mathbb{R}^n \setminus \Gamma$ is open, this implies that for m sufficiently large, $\phi_m(\tau) \notin \Gamma$. This contradicts the definition of A (recall that $x_m \in A$ and $\phi_m \in \mathfrak{S}_\mathcal{S}^\vartheta(x_m)$). Hence $x \in A$ and A is compact. □

The following proposition states that any compact robustly invariant set of $\mathcal{S}_\vartheta$ is UAS for $\mathcal{S}_{\vartheta'}$, where ϑ' can be taken to be arbitrarily close to ϑ. This fact was essentially proved in [114] in a slightly different context. The conclusion does not hold for $\vartheta' = \vartheta$ (see Example 2.3.20).

Proposition 2.3.16. *Any nonempty compact invariant set A for $\mathcal{S}_\vartheta$ is UAS for $\mathcal{S}_{\vartheta'}$ whenever $\vartheta' \in [0, \vartheta)$.*

The proof relies on the following technical lemma from [114].

Lemma 2.3.17. *[114] Fix any $\vartheta' \in (0, \vartheta)$ and $\tau > 0$. Let $K \subseteq \mathbb{R}^n$ be a compact set. Then there exists some $r = r(K, \tau, \vartheta', \vartheta) > 0$ such that the following holds: if there is a solution ϕ of $\mathcal{S}_{\vartheta'}$ such that $\phi(s) \in K$ for all $s \in [0, T]$, where $T \geq \tau$, then for any $y_0 \in \phi(0) + r\overline{\mathbb{B}}$ and any $y_1 \in \phi(T) + r\overline{\mathbb{B}}$, we have $y_1 \in \mathcal{R}_{\vartheta}^{T}(y_0)$, i.e., y_1 is reachable at T from y_0 by a solution of $\mathcal{S}_{\vartheta}$.*

Proof. We verify conditions (1) uniform stability and (2) uniform attractivity as required by Definition 2.3.2.

(1) For any $\epsilon > 0$, let $\tau > 0$ be the minimal time that is required for solutions of $\mathcal{S}_{\vartheta'}$ to travels from the interior of $A + \frac{\epsilon}{2}\overline{\mathbb{B}}$ to $\mathbb{R} \setminus (A + \epsilon\overline{\mathbb{B}})$. The existence of such a τ follows from that f is locally Lipschitz and an argument using Gronwall's inequality. Pick $\eta_\epsilon < \min(r, \frac{\epsilon}{2})$, where r is from Lemma 2.3.17, applied to the set $A + \epsilon\overline{\mathbb{B}}$ and scalars τ, ϑ', and ϑ. Let ϕ be any solution of $\mathcal{S}_{\vartheta'}$ such that $|\phi(0)|_A < \eta_\epsilon$. We show that $|\phi(t)|_A < \epsilon$ for all $t \geq 0$. Suppose that this is not the case. Then $|\phi(t_1)|_A \geq \epsilon$ for some $t_1 \geq \tau > 0$. Since $\eta_\epsilon < r$ and A is compact, we can always pick $y_0 \in A$ such that $y_0 \in \phi(0) + r\overline{\mathbb{B}}$. By Lemma 2.3.17, there exists a solution of $\mathcal{S}_{\vartheta}$ from $y_0 \in A$ to $y_1 = \phi(t_1) \notin A$. This contradicts that A is forward invariant for $\mathcal{S}_{\vartheta}$.

(2) Fix any $\epsilon_0 > 0$. Following part (1), choose η_{ϵ_0} such that $|\phi(0)|_A < \eta_{\epsilon_0}$ implies $|\phi(t)|_A < \epsilon_0$ for any solution $\phi(t)$ of $\mathcal{S}_{\vartheta'}$. Let r be chosen according to Lemma 2.3.17 with the set $A + \epsilon_0\overline{\mathbb{B}}$ and scalars $\tau = 1$, ϑ', and ϑ. Choose $\varrho \in (0, r)$. Let ϕ be any solution of $\mathcal{S}_{\vartheta'}$. We show that $|\phi(0)|_A < \varrho$ implies that $\phi(t) \in A$ for all $t \geq 1$. Suppose that this is not the case. Then there exists some $t_1 \geq 1$ such that $\phi(t_1) \notin A$. Since $\varrho < r$, we can pick y_0 such that $y_0 \in \phi(0) + r\overline{\mathbb{B}}$ and $y_0 \in A$. By Lemma 2.3.17, there exists a solution of $\mathcal{S}_{\vartheta}$ from $y_0 \in A$ to $y_1 = \phi(t_1) \notin A$. This contradicts that A is forward invariant for $\mathcal{S}_{\vartheta}$. Hence $\phi(t) \in A$ for all $t \geq 1$. This clearly implies (2). □

Proposition 2.3.16 establishes a link between robust invariance and asymptotic stability. By combining Lemma 2.3.15, Proposition 2.3.16, and Theorem 2.3.7, we can obtain the following converse theorem for a reach-avoid-stay specification.

Theorem 2.3.18. *Suppose that Γ is compact, U is closed, and $\Omega \cap U = \emptyset$, and $\mathcal{S}_{\vartheta}$ satisfies the reach-avoid-stay specification $(\mathcal{X}_0, U, \Gamma)$. Then there exists a compact set $A \subseteq \Gamma$ such that, for any $\vartheta' \in [0, \vartheta)$ and any proper indicator Γ for A on D, there exists an open set D such that $(A \cup \mathcal{X}_0) \subseteq D$ and $D \cap U = \emptyset$, a smooth function $V : D \to \mathbb{R}_{\geq 0}$ and class $\mathcal{K}_\infty$ functions α_1 and α_2 such that conditions (2.17) and (2.18) hold for all $x \in D$ and $d \in \vartheta'\overline{\mathbb{B}}$.*

Proof. By Lemma 2.3.15, there exists a compact set $A \subseteq \Gamma$ that is ϑ'-UAS for any $\vartheta' \in [0, \vartheta)$ by Proposition 2.3.16. Furthermore, as shown in the proof of Lemma 2.3.15, $\mathcal{R}_{\vartheta}^{t \geq T}(\mathcal{X}_0) \subseteq A$. This implies that, for any $\vartheta' \in [0, \vartheta)$, the domain of attraction of A for $\mathcal{S}_{\vartheta'}$ includes $\mathcal{X}_0$. Hence $\mathcal{S}_{\vartheta'}$ satisfy the stability with safety guarantee specification $(\mathcal{X}_0, U, A)$. The conclusion follows from that of Theorem 2.3.7. $\square$

Remark 2.3.19. *Figure 2.3 provides an illustration of the sets defined for proving Theorem 2.3.18.*

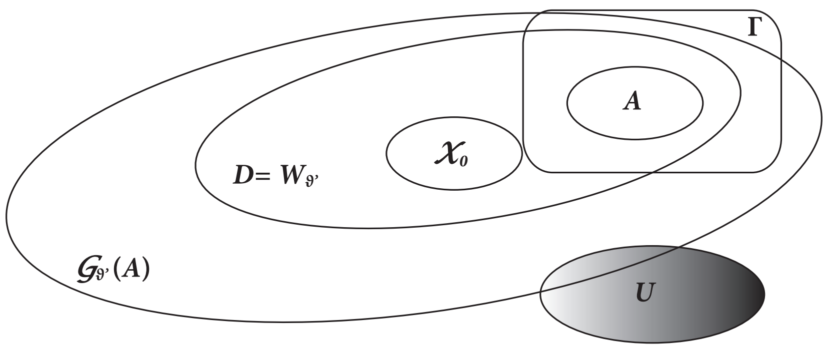


Figure 2.3: An illustration of the sets involved in Theorem 2.3.18. If a reach-avoid-stay specification $(\mathcal{X}_0, U, \Gamma)$ is satisfied, then for each $\vartheta' \in [0, \vartheta)$, we can find a set A such that $\mathcal{S}_{\vartheta'}$ satisfies the stability with safety guarantee specification $(\mathcal{X}_0, U, A)$. Consequently, a set D and a Lyapunov function V defined on D can be found such that the Lyapunov conditions (2.17) and (2.18) hold for $\mathcal{S}_{\vartheta'}$. The conclusion of Theorem 2.3.18 follows from that of Theorem 2.3.7. [121]

It would be tempting to draw a stronger conclusion than the one in Theorem 2.3.18 by allowing $\vartheta' = \vartheta$. The following example shows that the conclusion of Theorem 2.3.18 cannot be strengthened in this regard: Under the current assumptions of Theorem 2.3.18, there may not exist a converse Lyapunov-barrier function satisfying conditions (2.17) and (2.18) for $\mathcal{S}_\vartheta$, even if $\mathcal{S}_\vartheta$ satisfies a reach-avoid-stay specification $(\mathcal{X}_0, U, \Gamma)$.

Example 2.3.20. *Consider $\mathcal{S}$ defined by $\dot{x} = -x + x^2$. Let $\mathcal{X}_0 = [-1, -0.9]$, $U = [0.6, \infty)$, $\Gamma = [-0.25, 0.5]$, and $\vartheta = 0.25$. It is easy to verify that $\mathcal{S}_\vartheta$ satisfies the reach-avoid-stay specification $(\mathcal{X}_0, U, \Gamma)$. However, solutions of $\mathcal{S}_\vartheta$ starting from $x_0 = 0.5 + \epsilon$, where $\epsilon > 0$, with $d(t) = \vartheta$ will tend to infinity. Furthermore, for any $x_0 \in \Gamma$, there exists a solution of $\mathcal{S}_\vartheta$ that approaches 0.5. Hence, there does not exist an open set D as in Theorem 2.3.18 and a converse Lyapunov-barrier function defined on D that satisfies conditions (2.17) and (2.18) for all $x \in D$ and $d \in \vartheta\overline{\mathbb{B}}$. The reason for this is that the conclusion of Proposition 2.3.16 does not hold for $\mathcal{S}_\vartheta$, i.e., the set A defined by (2.20) may not be UAS for $\mathcal{S}_\vartheta$, even though it is UAS for $\mathcal{S}_{\vartheta'}$ whenever $\vartheta' \in [0, \vartheta)$. Indeed, it is* *not difficult to verify that the set* $A = [\frac{1}{2} - \frac{\sqrt{2}}{2}, 0.5]$ *and, by the observation above, the set* A *is not UAS for* $\mathcal{S}_\vartheta$. ☐

Similarly, Proposition 2.3.11 can be adapted to give the following version of converse theorem for reach-avoid-stay specifications.

Proposition 2.3.21. *Suppose that* Γ *and* $\mathcal{X}_0$ *are compact,* U *is closed, and* $\Gamma \cap U = \emptyset$*, and* $\mathcal{S}_\vartheta$ *satisfies the reach-avoid-stay specification* $(\mathcal{X}_0, U, \Gamma)$*. Then for any* $\vartheta' \in [0, \vartheta)$*, there exists a compact* $A \subseteq \Gamma$*, an open set* D *such that* $(A \cup \mathcal{X}_0) \subseteq D$*, and smooth functions* $V : D \to \mathbb{R}_{\geq 0}$ *and* $B : \mathbb{R}^n \to \mathbb{R}$ *such that*

(1) V *is positive definite on* D *w.r.t.* A*, i.e.,* $V(x) = 0$ *if and only if* $x \in A$*;*

(2) $\nabla V(x) \cdot (f(x) + d) < 0$ *for all* $x \in D \setminus A$ *and* $d \in \vartheta'\overline{\mathbb{B}}$*;*

(3) $\mathcal{X}_0 \subseteq C = \{x \in D : B(x) \geq 0\} \subseteq D$ *and* $B(x) < 0$ *for all* $x \in U$*;*

(4) $\nabla B(x) \cdot (f(x) + d) \geq 0$ *for all* $x \in D$ *and* $d \in \vartheta'\overline{\mathbb{B}}$*.*

Proof. Similar to that of Proposition 2.3.11. ☐

The above converse results (Theorem 2.3.18 and Proposition 2.3.21) reveal that the verification and design for reach-avoid-stay specifications can indeed be centered around the problem of stability/stabilization with safety guarantees. This is *without loss of generality* at least from a robustness point of view. In this regard, Lemma 2.3.15 and Proposition 2.3.16 connect robust reach-avoid-stay specification with stability with safety guarantees. We can also prove a result in the converse direction. These statements are summarized in the following proposition.

Proposition 2.3.22. *(Connections):*

(1) If $\mathcal{S}_\vartheta$ *satisfies a stability with safety guarantee specification* $(\mathcal{X}_0, U, A)$ *and* $\mathcal{X}_0$ *is compact, then for every* $\epsilon > 0$*,* $\mathcal{S}_\vartheta$ *satisfies the reach-avoid-stay specification* $(\mathcal{X}_0, U, A + \epsilon\overline{\mathbb{B}})$*.*

(2) If $\mathcal{S}_\vartheta$ *satisfies a reach-avoid-stay specification* $(\mathcal{X}_0, U, \Gamma)$*, then there exists a compact set* $A \subseteq \Gamma$ *such that, for any* $\vartheta' \in [0, \vartheta)$*,* $\mathcal{S}_{\vartheta'}$ *satisfies the stability with safety guarantee specification* $(\mathcal{X}_0, U, A)$*.*

Proof. (1) The conclusion follows from the uniform attractivity property for solutions for $\mathcal{S}_\vartheta$ under the stability assumption. The proof is provided in Section 2.3.3. (2) It follows from Lemma 2.3.15, Proposition 2.3.16, and the definitions of the specifications. ☐

2.3.3 Proofs of Results

Proof of Lemma 2.3.8

We first state two lemmas on the properties of the solutions of $\mathcal{S}_\vartheta$.

The first one is well known from the basic theory of ODEs (see, e.g, [153, Theorem 55, Appendix C]).

Lemma 2.3.23 (Continuous dependence). *Suppose that for some $x_0 \in \mathbb{R}^n$ there exists some $T > 0$ such that solutions for $\mathcal{S}_\vartheta$ starting from x_0 are defined on $[0, T]$. Then there exists some $\vartheta > 0$ such that solutions starting from $x_0 + \vartheta\overline{\mathbb{B}}$ are also defined on $[0, T]$ and there exists a constant C (depending on T and x_0) such that*

$$|\phi(t; x, d) - \phi(t; x_0, d)| \leq C\,|x - x_0|$$

for all $x \in x_0 + \vartheta\overline{\mathbb{B}}$ and $d : [0, T] \to \vartheta\overline{\mathbb{B}}$.

The next result is on topological properties of solutions of differential inclusions satisfying some basic conditions. It can be found, e.g., in [59, Theorem 3, Section 7]. Note that the differential inclusion we consider $\mathcal{S}_\vartheta : x' \in F_\vartheta(x) := f(x)+\vartheta\overline{\mathbb{B}}$ straightforwardly satisfies the basic conditions there (i.e., F_ϑ is upper semicontinuous and takes nonempty, compact, and convex values).

Lemma 2.3.24 (Compactness of reachable sets). *Let $K \subseteq \mathbb{R}^n$ be a compact set. Suppose that there exists some $\tau > 0$ such that solutions of $\mathcal{S}_\vartheta$ starting from K are always defined on $[0, \tau)$. Then, for any $T \in [0, \tau)$, $\mathcal{R}_\vartheta^{0\leq t\leq T}(K)$ is a compact set. Furthermore, solutions of $\mathcal{S}_\vartheta$ on $[0, T]$ form a compact set under the uniform convergence topology.*

The following result shows that under the uniform stability assumption (i.e., condition (1) in Definition 2.3.2), attraction of solutions starting from any compact set within the domain of attraction is always uniform. The proof of the following result is modeled after the proof for Proposition 3 in [160, cf. Claim 4].

Proposition 2.3.25 (Uniformity of attraction). *Suppose that a closed set $A \subseteq \mathbb{R}^n$ is uniformly stable for $\mathcal{S}_\vartheta$, i.e., condition (1) of Definition 2.3.2 holds. Let K be a compact set. Then the following two statements are equivalent:*

(1) For any $x_0 \in K$ and any $\phi \in \mathfrak{S}_{\mathcal{S}}^{\vartheta}(x_0)$, ϕ is defined for all $t \geq 0$ and

$$\lim_{t\to\infty} |\phi(t)|_A = 0.$$

(2) For every $\epsilon > 0$, there exists $T = T(\epsilon) > 0$ such that

$$|\phi(t)|_A < \epsilon$$

holds for any $x_0 \in K$, $\phi \in \mathfrak{S}_{\mathcal{S}}^{\vartheta}(x_0)$, and $t \geq T$.

Proof. Clearly, (2) implies (1). We prove that (1) also implies (2) under the uniform stability assumption. Suppose that (2) does not hold. Then there exists some $\epsilon_0 > 0$ such that for all $n > 0$ there exists $x_n \in K$, $\phi_n \in \mathfrak{S}_{\mathcal{S}}^{\vartheta}(x_n)$, and $t_n \geq n$ such that

$$|\phi_n(t_n)|_A \geq \epsilon_0. \tag{2.21}$$

Let $\eta_0 = \eta_{\epsilon_0}$ be given by condition (1) of Definition 2.3.2. For every $n > 0$, we must have

$$|\phi_n(t)|_A \geq \eta_0, \quad \forall t \in [0, n]. \tag{2.22}$$

Claim 2.3.26. *There exist subsequences $\{x_n\}$ and $\phi_n \in \mathfrak{S}_{\mathcal{S}}^{\vartheta}(x_n)$ such that x_n converges to x and ϕ_n converges to a solution $\phi \in \mathfrak{S}_{\mathcal{S}}^{\vartheta}(x)$. The latter convergence is uniform on every compact interval of $\mathbb{R}_{\geq 0}$.*

Proof of Claim 2.3.26 From (1), we know that solutions starting from K are always forward complete. Since K is compact, we can assume without loss of generality that $\{x_n\}$ converges to $x \in K$ (otherwise we can pick a subsequence). By Lemma 2.3.24, there exists a subsequence of $\{\phi_n\}$, denoted by $\{\phi_{1m}\}$, that converges uniformly on $[0, 1]$ to a solution $\phi_1 \in \mathfrak{S}_{\mathcal{S}}^{\vartheta}(x)$. By the same argument, $\{\phi_{1m}\}$ has a subsequence, denoted by $\{\phi_{2m}\}$, that converges uniformly on $[0, 2]$ to a solution $\phi_2 \in \mathfrak{S}_{\mathcal{S}}^{\vartheta}(x)$. Repeat this argument and pick the diagonal $\{\phi_{mm}\}$. Then $\{\phi_{mm}\}$ has the claimed property. □

Let $\phi \in \mathfrak{S}_{\mathcal{S}}^{\vartheta}(x)$ be given by the claim. By statement (1), there exists $T > 0$ such that

$$|\phi(t)|_A < \frac{\eta_0}{2}, \quad \forall t \geq T. \tag{2.23}$$

However, since $\{\phi_{mm}\}$ converges to ϕ uniformly on $[0, T]$, there exists some $n \geq T$ such that

$$|\phi_n(t) - \phi(t)| < \frac{\eta_0}{2}, \quad \forall t \in [0, T]. \tag{2.24}$$

The equations (2.23) and (2.24) give $|\phi_n(T)|_A < \eta_0$, which contradicts (2.21). □

Proof of Lemma 2.3.8 We can easily verify that $\mathcal{W}_\vartheta$ is forward invariant and $W \subseteq \mathcal{W}_\vartheta \subseteq \mathcal{G}_\vartheta(A)$ by its definition. We show that $\mathcal{W}_\vartheta$ is open.

Let $x_0 \in \mathcal{W}_\vartheta$. Let $\varrho > 0$ be given by condition (2) from Definition 2.3.2 for UAS of A. Choose $\epsilon_0 < \varrho$ such that $(A + \epsilon_0\overline{\mathbb{B}}) \cap U = \emptyset$. Choose $\eta_0 = \eta_{\epsilon_0}$ according to condition (1) in Definition 2.3.2 for UAS of A. Clearly, $\eta_0 \leq \epsilon_0 < \varrho$.

Then, by Proposition 2.3.25 in the Appendix, there exists some $T = T(\eta_0) > 0$ such that

$$|\phi(t)|_A < \frac{\eta_0}{2}$$

for any solution $\phi \in \mathfrak{S}^\vartheta_{\mathcal{S}}(x_0)$ and all $t \geq T$. By Lemma 2.3.24 in the Appendix, the set $K = \mathcal{R}^{0\leq t\leq T}_\vartheta(x_0)$ is compact. Let $\epsilon_1 < \frac{\epsilon_0}{2}$ be chosen such that $(K + \epsilon_1\overline{\mathbb{B}}) \cap U = \emptyset$.

By continuous dependence of solutions of $\mathcal{S}_\vartheta$ with respect to initial conditions, there exists some $r > 0$ such that, for all $x \in x_0 + r\overline{\mathbb{B}}$ and any $\psi \in \mathfrak{S}^\vartheta_{\mathcal{S}}(x)$, there exists a solution $\phi \in \mathfrak{S}^\vartheta_{\mathcal{S}}(x_0)$ such that

$$|\phi(t) - \psi(t)| < \epsilon_1, \quad \forall t \in [0, T].$$

It follows that

$$\mathcal{R}^{\leq T}_\vartheta(x_0 + \vartheta\overline{\mathbb{B}}) \subseteq K + \epsilon_1\overline{\mathbb{B}}. \tag{2.25}$$

Furthermore, at $t = T$, we have $|\psi(T)|_A \leq |\phi(T)|_A + \epsilon_1 < \frac{\eta_0}{2} + \frac{\eta_0}{2} = \eta_0$. It follows from condition (1) in Definition 2.3.2 that

$$\psi(t) \in A + \epsilon_0\overline{\mathbb{B}} \subseteq A + \varrho\overline{\mathbb{B}} \tag{2.26}$$

for all $\psi \in \mathfrak{S}^\vartheta_{\mathcal{S}}(x)$, $x \in x_0 + r\overline{\mathbb{B}}$, and $t \geq T$. By condition (2) in Definition 2.3.2, $\lim_{t\to\infty} |\psi(t)|_A = 0$. In view of (2.25) and (2.26), $\psi(t) \notin U$ for all $t \geq 0$. We have shown that $x \in \mathcal{W}_\vartheta$ for all $x \in \overline{\mathbb{B}}_r(x_0)$. Hence $\mathcal{W}_\vartheta$ is open. □

Proof of Proposition 2.3.9

The existence of a Lyapunov function can be proved based on the $\mathcal{KL}$-stability (i.e. given in Definition 2.3.27), following the techniques developed in [160] on converse Lyapunov functions for $\mathcal{KL}$-stability. The $\mathcal{KL}$-stability considered here is in fact a special case of that in [160], because we do not need to consider stability with respect to two different measures as in [160]. We provide a definition of $\mathcal{KL}$-stability below, adapted for a proper indicator of a compact set.

Definition 2.3.27. *Let $A \subseteq \mathbb{R}^n$ be a compact set contained in an open set $D \subseteq \mathbb{R}^n$. Let $\mathfrak{w}$ be any proper indicator for A on D. The system $\mathcal{S}_\vartheta$ is said to be $\mathcal{KL}$-stable on D w.r.t. $\mathfrak{w}$ if any solution $\phi \in \mathfrak{S}_{\mathcal{S}}^{\vartheta}(x)$ with $x \in D$ is defined and remain in D for all $t \geq 0$ and there exists a $\mathcal{KL}$-function β such that*

$$\mathfrak{w}(\phi(t;x)) \leq \beta(\mathfrak{w}(x), t), \quad \forall t \geq 0, \tag{2.27}$$

for all $x \in D$ and $\phi \in \mathfrak{S}_{\mathcal{S}}^{\vartheta}(x)$.

The key step in proving Proposition 2.3.9 is the following lemma.

Lemma 2.3.28. *Assume that the assumptions of Proposition 2.3.9 hold. Then the system $\mathcal{S}_\vartheta$ is $\mathcal{KL}$-stable on D w.r.t. $\mathfrak{w}$.*

Proof of Lemma 2.3.28 Let $\mathcal{C}_r := \{x \in D : \mathfrak{w}(x) \leq r\}$. Then by the assumptions, since $\mathfrak{w}$ is a proper indicator $\mathfrak{w}$ for A on D, $\mathcal{C}_r$ is compact subset of D for each $r \geq 0$. Fix $\varrho > 0$ such that $A + \varrho\overline{\mathbb{B}} \subseteq D$. We can find a $\mathcal{K}_\infty$-class function satisfying $\alpha(s) \geq \sup_{x \in D, \|x\|_A \leq \min(\varrho, s)} \mathfrak{w}(x)$. Therefore, for all $\|x\|_A \leq \varrho$, we have $\mathfrak{w}(x) \leq \alpha(\|x\|_A)$.

Claim 2.3.29. *There exists a $\mathcal{K}_\infty$ function φ such that, for each $x \in D$, $\mathfrak{w}(\phi(t;x)) \leq \varphi(\mathfrak{w}(x))$ for all $t \geq 0$ and $\phi \in \mathfrak{S}_{\mathcal{S}}^{\vartheta}(x)$.*

Proof of Claim 2.3.29 Indeed, for each $x \in D$, we can find an $r \geq 0$ such that $x \in \mathcal{C}_r$. By Proposition 2.3.25, for any $\varrho > 0$ chosen above, we can find a T such that $\|\phi(t;x)\|_A \leq \varrho$ for all $x \in \mathcal{C}_r$ and $\phi \in \mathfrak{S}_{\mathcal{S}}^{\vartheta}(x)$. By forward invariance of D, it follows that $\mathcal{R}_\vartheta(\mathcal{C}_r) \subseteq \mathcal{R}_\vartheta^{0 \leq t \leq T}(\mathcal{C}_r) \cup (A + \varrho\overline{\mathbb{B}}) \subseteq D$. Since $\mathcal{C}_r$ is compact, by Lemma 2.3.24, for any finite T, $\mathcal{R}_\vartheta^{0 \leq t \leq T}(\mathcal{C}_r)$ is also compact. The boundedness of $\mathcal{R}_\vartheta(\mathcal{C}_r)$ implies that $\overline{\mathcal{R}_\vartheta(\mathcal{C}_r)}$ is a compact subset of D. Let $M(r) = \max_{x \in \overline{\mathcal{R}_\vartheta(\mathcal{C}_r)}} \mathfrak{w}(x)$. Then $\mathfrak{w}(\phi(t;x)) \leq M(\mathfrak{w}(x))$ for all $x \in D$, $\phi \in \mathfrak{S}_{\mathcal{S}}^{\vartheta}(x)$, and $t \geq 0$. Clearly, $M(r)$ is nondecreasing (due to the inclusion relation of reachable sets from $\mathcal{C}_r$ with different r) and $\lim_{r \to 0} M(r) = 0$ (due to the uniform stability property). The $\varphi \in \mathcal{K}_\infty$ in the claim can be chosen such that $M(r) \leq \varphi(r)$ for all $r \geq 0$. □

Claim 2.3.30. *For each $r > 0$, there exists a strictly decreasing function $\psi_r : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ with $\lim_{t \to \infty} \psi_r^{-1}(t) = 0$ such that $\mathfrak{w}(\phi(t;x)) \leq \psi_r^{-1}(t)$ for all $t > 0$ whenever $\mathfrak{w}(x) \leq r$ and $\phi \in \mathfrak{S}_{\mathcal{S}}^{\vartheta}(x)$.*

Proof of Claim 2.3.30 For each $0 < \epsilon \leq \varphi(r)$, by Proposition 2.3.25, we can find a $T_r(\epsilon) = T(\min(\alpha^{-1}(\epsilon), \varrho)) > 0$ such that for all $x \in \mathcal{C}_r$, $\phi \in \mathfrak{S}_{\mathcal{S}}^{\vartheta}(x)$, and $t \geq T_r(\epsilon)$, we have

$$\|\phi(t;x)\|_A < \min(\alpha^{-1}(\epsilon), \varrho) \leq \varrho. \tag{2.28}$$

Equation (2.28) also implies

$$\mathfrak{w}(\phi(t;x)) \leq \alpha(\alpha^{-1}(\epsilon)) = \epsilon, \tag{2.29}$$

for all $x \in \mathcal{C}_r$, $\phi \in \mathfrak{S}_{\mathcal{S}}^{\vartheta}(x)$, and $t \geq T_r(\epsilon)$. For $\epsilon > \varphi(r)$, we set $T_r(\epsilon) = 0$ and (2.29) still holds because $\mathfrak{w}(\phi(t;x)) \leq \varphi(r) < \epsilon$ for all $t \geq 0$ by Claim 2.3.29. Note that for each fixed r, the function $T_r(\epsilon)$ can be chosen to be nonincreasing in ϵ and by definition $\lim_{\epsilon \to +\infty} T_r(\epsilon) = 0$; for each fixed $\epsilon > 0$, $T_r(\epsilon)$ can be chosen to be nondecreasing in r. Based on $T_r(\epsilon)$, we can find $\psi_r : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ such that $\psi_r(\epsilon) \geq T_r(\epsilon)$ for all $\epsilon > 0$. The function ψ_r can be constructed as strictly decreasing to zero (hence its inverse is defined on $\mathbb{R}_{>0}$ and also strictly decreasing) and satisfying $\lim_{t\to\infty} \psi_r^{-1}(t) = 0$. For each $t > 0$, let $\epsilon = \psi_r^{-1}(t)$. We have $t = \psi_r(\epsilon) \geq T_r(\epsilon)$. Hence $x \in \mathcal{C}_r$ implies that $w(\phi(t;x)) \leq \epsilon = \psi_r^{-1}(t)$. □

Now we force $\psi_r^{-1}(0) = \infty$ defined in Claim 2.3.30 and let

$$\beta(s,t) := \min\{\varphi(s), \inf_{r \in (s,\infty)} \psi_r^{-1}(t)\}$$

with φ defined in Claim 2.3.29. Then $\beta \in \mathcal{KL}$[3] and (2.27) holds. □

Once Lemma 2.3.28 is proved, the proof of Proposition 2.3.9 follows from a standard converse Lyapunov argument (see [160, proof of Theorem 1]). We provide an outline of the proof as follows.

Lemma 2.3.31 (Sontag [152])**.** *For each $\beta \in \mathcal{KL}$ and each $\lambda > 0$, there exist functions $\alpha_1, \alpha_2 \in \mathcal{K}_\infty$ such that α_1 is locally Lipschitz and*

$$\alpha_1(\beta(s,t)) \leq \alpha_2(s)e^{-\lambda t}, \quad \forall (s,t) \in \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0}. \tag{2.30}$$

Proof of Proposition 10 Based on the Lemma 2.3.28 and by Sontag's lemma (Lemma 2.3.31) on $\mathcal{KL}$-estimates, we can find α_1 and α_2 such that

$$\alpha_1(\mathfrak{w}(\phi(t;x))) \leq \alpha_1(\beta(\mathfrak{w}(x),t)) \leq \alpha_2(\mathfrak{w}(x))e^{-2t} \tag{2.31}$$

for any $x \in D$, $\phi \in \mathfrak{S}_{\mathcal{S}}^{\vartheta}(x)$, and $t \geq 0$. Now define

$$V(x) := \sup_{t\geq 0, \phi \in \mathfrak{S}_{\mathcal{S}}^{\vartheta}(x)} \alpha_1(\mathfrak{w}(\phi(t;x)))e^t. \tag{2.32}$$

Then $V(x) \geq \sup_{\phi \in \mathfrak{S}_{\mathcal{S}}^{\vartheta}(x)} \alpha_1(\mathfrak{w}(\phi(t;x))) = \alpha_1(\mathfrak{w}(x))$ for all $x \in D$, and it is straightforward from (2.31) that $V(x) \leq \sup_{t\geq 0} \alpha_2(\mathfrak{w}(x))e^{-t} \leq \alpha_2(\mathfrak{w}(x))$. Therefore condition (2.17) in Theorem 2.3.7 is satisfied.

[3] This construction of $\mathcal{KL}$ function does not impose continuity. Nonetheless, as pointed out in [160, Remark 3], any (potentially noncontinuous) $\mathcal{KL}$ function can be upper bounded by a continuous $\mathcal{KL}$ function.

To show the satisfaction of condition (2.18) in Theorem 2.3.7, we can show that

$$V(\phi(t;x)) \leq V(x)e^{-t}, \quad \forall \phi \in \mathfrak{S}_{\mathcal{S}}^{\vartheta}(x), \forall t \geq 0, \tag{2.33}$$

with a similar reasoning as the Claim 1 in [160]. The local Lipschitz continuity of V on $D \setminus A$ follows from the Claim 3 in [160]. Then we have

$$\begin{aligned} \nabla V(x) \cdot (f(x) + d) &\leq \liminf_{t \to 0^+} \frac{V(\phi(t;x,d)) - V(x)}{t} \\ &\leq \liminf_{t \to 0^+} V(x) \frac{e^{-t} - 1}{t} = -V(x). \end{aligned} \tag{2.34}$$

A smooth approximation for V exists given its local Lipschitz continuity [111]. The smoothness can also be extended from the local region $D \setminus A$ to the whole set D (by following the proof of Theorem 1 (step 3) in [160]). □

2.4 Application of Lyapunov-Barrier Approaches for Control of Reach-Avoid-Stay Specifications

In this section, we first take advantage of the results from Section 2.3.2 and make a straightforward derivation on a converse control Lyapunov-barrier function theorem for $\mathcal{S}_\vartheta$ satisfying a reach-avoid-stay specification $(\mathcal{X}_0, U, \Gamma)$ under controls. We then consider the separate form of control Lyapunov-barrier functions (as in Proposition 2.4.3) to guarantee reach-avoid-stay specifications for control systems. The effectiveness is numerically verified in a case study of jet engine compressor control problem. Throughout this section, we consider control systems (2.4) with the set of control inputs $\mathcal{U} \subseteq \mathbb{R}^p$.

Definition 2.4.1 (Reach-avoid-stay controllable). *A system $\mathcal{S}_\vartheta$ is called reach-avoid-stay controllable w.r.t. $(\mathcal{X}_0, U, \Gamma)$, where $\mathcal{X}_0, U, \Gamma \subseteq \mathbb{R}^n$, if there exists a Lipschitz continuous control strategy κ such that the system $\mathcal{S}_\vartheta^\kappa$ satisfies the reach-avoid-stay specification $(\mathcal{X}_0, U, \Gamma)$.*

2.4.1 Control Lyapunov-Barrier Functions for Reach-Avoid-Stay Specifications

We first show that reach-avoid-stay controllability implies the existence of a control Lyapunov-barrier function w.r.t. the reach-avoid-stay specification.

Theorem 2.4.2. *Suppose that Γ is compact, U is closed, and $\Gamma \cap U = \emptyset$, and $\mathcal{S}_\vartheta$ is reach-avoid-stay controllable w.r.t. $(\mathcal{X}_0, U, \Gamma)$. Then there exists a compact set $A \subseteq \Gamma$ such that, for any $\vartheta' \in [0, \vartheta)$ and any proper indicator Γ for A on D, there exists an open set D such that $(A \cup \mathcal{X}_0) \subseteq D$ and $D \cap U = \emptyset$, a smooth function $V : D \to \mathbb{R}_{\geq 0}$ and class $\mathcal{K}_\infty$ functions α_1 and α_2 such that, for all $x \in D$ and $d \in \vartheta'\overline{\mathbb{B}}$, Equation (2.17) is satisfied and*

$$\inf_{u \in \mathcal{U}} \sup_{x \in D} \sup_{d \in \vartheta \overline{\mathbb{B}}} [L_{f,d}V(x) + L_g V(x) u + V(x)] \leq 0. \tag{2.35}$$

Proof. By assumption, there exists a Lipschitz continuous κ that renders the solutions satisfy reach-avoid-stay specification $(\mathcal{X}_0, U, \Gamma)$. Then by Theorem 2.3.18, for any proper indicator w for A on D, there exists a function $V : D \to \mathbb{R}_{\geq 0}$ satisfying (2.17) and

$$\sup_{d \in \vartheta' \overline{\mathbb{B}}} [L_{f,d}V(x) + L_g V(x)\kappa(x) + V(x)] \leq 0$$

for all $x \in D$. Taking the supremum over all $x \in D$, we have

$$\sup_{x \in D} \sup_{d \in \vartheta' \overline{\mathbb{B}}} [L_{f,d}V(x) + L_g V(x)\kappa(x) + V(x)] \leq 0.$$

Since we have the control $\kappa(x) \in \mathcal{U}$, it follows that

$$\inf_{u \in \mathcal{U}} \sup_{x \in D} \sup_{d \in \vartheta' \overline{\mathbb{B}}} [L_{f,d}V(x) + L_g V(x) u + V(x)] \leq 0.$$

□

With a similar approach, Proposition 2.3.21 can be applied to give the following version of converse control Lyapunov-barrier functions theorem for reach-avoid-stay specifications.

Proposition 2.4.3. *Suppose that Γ and $\mathcal{X}_0$ are compact, U is closed, and $\Gamma \cap U = \emptyset$, and $\mathcal{S}_\vartheta$ is reach-avoid-stay controllable w.r.t. $(\mathcal{X}_0, U, \Gamma)$. Then for any $\vartheta' \in [0, \vartheta)$, there exists a compact $A \subseteq \Gamma$, an open set D such that $(A \cup \mathcal{X}_0) \subseteq D$, and smooth functions $V : D \to \mathbb{R}_{\geq 0}$ and $B : D \to \mathbb{R}$ such that*

(1) V is positive definite on D w.r.t. A, i.e., $V(x) = 0$ if and only if $x \in A$;

(2) $\inf_{u \in \mathcal{U}} \sup_{x \in D} \sup_{d \in \vartheta' \overline{\mathbb{B}}} [L_{f,d}V(x) + L_g V(x) u] < 0$;

(3) $\mathcal{X}_0 \subseteq C = \{x \in D : B(x) \geq 0\}$ and $B(x) < 0$ for all $x \in U$;

(4) $\sup_{u\in\mathcal{U}}[L_{f,d}B(x) + L_gB(x)u] \geq 0$ *for all* $x \in D$ *and* $d \in \vartheta'\overline{\mathbb{B}}$*.*

In real-world applications, our target is formulated as follows.

Problem 2.4.4 (Reach-avoid-stay control)**.** *Given a reach-avoid-stay specification* $(\mathcal{X}_0, U, \Gamma)$*, design a control policy* κ *such that the resulting solutions of* $\mathcal{S}_\vartheta^\kappa$ *satisfy* $(\mathcal{X}_0, U, \Gamma)$*.*

Note that, if the functions V and B (as in Proposition 2.4.3) exist, one can obtain a control strategy solving Problem 2.4.4, based on which a set of proper constraints on the state-dependent control signals can be obtained. The sufficient conditions on the state-dependent reach-avoid-stay control signals are derived correspondingly from the functions V and B with their certificates as in (1)-(4) of Proposition 2.4.3. For simplicity, we call the pair (V, B) the control Lyapunov-barrier functions for the reach-avoid-stay specification $(\mathcal{X}_0, U, \Gamma)$. We consider this separate form of control Lyapunov-barrier functions for potentially less efforts in construction.

The sufficient conditions can be verified based on Proposition 2.3.11 and 2.3.22. We formally provide the statement as below.

Theorem 2.4.5. *Given a closed set* A *and an unsafe set* U *such that* $(A + \epsilon\overline{\mathbb{B}}) \cap U = \emptyset$ *for some* $\epsilon > 0$*, suppose that there exists an open set* D *such that* $(A \cup \mathcal{X}_0) \subseteq D$*, and control Lyapunov-barrier functions* (V, B) *satisfying (1) to (4) of Proposition 2.4.3. Let*

$$\mathfrak{K}(x) = \{u \in \mathcal{U} :\ L_{f,d}V(x) + L_gV(x)u < 0 \text{ and } L_{f,d}B(x) + L_gB(x)u \geq 0,\ \forall x \in D,\ \forall d \in \vartheta\overline{\mathbb{B}}\}$$

be the associated set of control strategies generated by (V, B)*. Then, for any control strategy* κ *such that* $\kappa(x(t)) \in \mathfrak{K}(x(t))$ *and* $\mathfrak{K}(x) \neq \emptyset$ *for all* $x \in D$*, the system* $\mathcal{S}_\vartheta^\kappa$ *satisfies the reach-avoid-stay specification* $(\mathcal{X}_0, U, A + \epsilon\overline{\mathbb{B}})$*. In other words, if such a* κ *exists, the system* $\mathcal{S}_\vartheta$ *is reach-avoid-stay controllable w.r.t.* $(\mathcal{X}_0, U, A + \epsilon\overline{\mathbb{B}})$*.*

2.4.2 Case Study of Jet Engine Compressor Control

We have seen in Section 1.2 that the reduced Moore-Greitzer ODE model (restricted to the $\mathbb{R}^2$ subspace) is a commonly used nonlinear model for capturing average flow Φ and average pressure Ψ of axial-flow jet engine compressors. As the throttle coefficient γ decreases, surge instability occurs and generates a pumping oscillation (Hopf-bifurcation) that can cause flame-out and engine damage [13, 73]. In practice, to deal with the requests from downstream, the operation point needs to be switched during the process. However, without any controls, operation points are determined by γ and a smaller γ may result in unstable operation points [175].

To alleviate the oscillation and prevent substantial pressure loss during the switch, it is motivated to design controllers to lead the state (Φ, Ψ) reach-and-stay in a small region around an unstable operation point, and meanwhile avoid touching the region with low average pressure.

We consider the reduced Moore-Greitzer ODE model with an additive control input $[v, 0]^T$ and no extra perturbations:

$$\frac{d}{dt}\begin{bmatrix}\Phi(t)\\ \Psi(t)\end{bmatrix} = \begin{bmatrix}\frac{1}{\mathfrak{l}_c}(\psi_c - \Psi(t))\\ \frac{1}{16\mathfrak{l}_c}\left(\Phi(t) - \gamma\sqrt{\Psi(t)}\right)\end{bmatrix} + \begin{bmatrix}v(t)\\ 0\end{bmatrix} \tag{2.36}$$

with parameters as

$$\mathfrak{l}_c = 8,\ \iota = 0.18,\ \mathfrak{M} = 0.25,\ \psi_{c_0} = 1.67\iota, \mathfrak{a} = \frac{1}{3.5}, \mathfrak{v} = 0.1.$$

Problem 2.4.6. *We aim to manipulate the throttle coefficient γ and v simultaneously such that the state (Φ, Ψ) are regulated to satisfy reach-avoid-stay specification $(\mathcal{X}_0, U, \Gamma)$. We require that $\gamma : \mathbb{R}_{\geq 0} \to [0.5, 1]$ is time-varied with $\gamma(0) \in [0.62, 0.66]$ and $|\gamma(t + \tau) - \gamma(t)| \leq 0.01\tau$ for any $\tau > 0$. We define $\mathcal{X}_0 = \{(\Phi_e(\gamma(0)), \Psi_e(\gamma(0)))\}$ (i.e. a sub-region of stable equilibrium points, also see Example 1.2.3); Γ to be the ball that centered at $\zeta = (0.4519, 0.6513)$ with radius $r = 0.003$, i.e. $\Gamma = \zeta + r\overline{\mathbb{B}}$; $U = \{(x, y) : x \in (0.497, 0.503), y \in (0.650, 0.656)\}$. We set $v(t) \in \mathcal{U} = [-0.05, 0.05] \cap \mathbb{R}$ for all t.*

Addressing Problem 3.3.32, we apply the proposed Lyapunov method and compare the effectiveness with formal methods.

Remark 2.4.7. *For this special case, the purpose that we treat γ as a time-varied signal with a Lipschitz continuity restriction is to prevent unnecessary extra pumping. The system (2.36) can be transformed to fit in the general form of (2.4). Now we provide two ways of transformation.*

(a) The system (2.36) is equivalent as

$$\frac{d}{dt}\begin{bmatrix}\Phi(t)\\ \Psi(t)\\ \gamma(t)\end{bmatrix} = \begin{bmatrix}\frac{1}{\mathfrak{l}_c}(\psi_c - \Psi(t))\\ \frac{1}{16\mathfrak{l}_c}\left(\Phi(t) - \gamma(t)\sqrt{\Psi(t)}\right)\\ 0\end{bmatrix} + \begin{bmatrix}v(t)\\ 0\\ u_\gamma(t)\end{bmatrix}, \tag{2.37}$$

where $u_\gamma : \mathbb{R}_{\geq 0} \to [-0.01, 0.01]$ is an extra input signal on γ such that γ satisfies $|\gamma(t + \tau) - \gamma(t)| \leq 0.01\tau$. The γ becomes time invariant when $u_\gamma \equiv 0$. Equation (2.37) is in the *form of* (2.4)*, in particular,* $\mathfrak{u}(t) = [v(t), 0, u_\gamma(t)]^T$*, and*

$$f(x) = \begin{bmatrix} \frac{1}{\mathfrak{l}_c}(\psi_c - \Psi) \\ \frac{1}{16\mathfrak{l}_c}\left(\Phi - \gamma\sqrt{\Psi}\right) \\ 0 \end{bmatrix}, \; g(x) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The Problem 3.3.32 is converted to synthesizing v *and* u_γ *such that the trajectory satisfies the reach-avoid-stay specification* $(\tilde{\mathcal{X}}_0, \tilde{U}, \tilde{\Gamma})$*, where* $\tilde{\mathcal{X}}_0 = \{(\Phi_e(\gamma(0)), \Psi_e(\gamma(0)), \gamma(0)) : \gamma(0) \in [0.62, 0.66]\}$*;* $\tilde{U} = U \times (\mathbb{R} \setminus [0.5, 1])$*;* $\tilde{\Gamma} = \Gamma \times [0.5, 1]$*.*

(b) The system (2.36) *can be rewritten as*

$$\frac{d}{dt}\begin{bmatrix} \Phi(t) \\ \Psi(t) \end{bmatrix} = \begin{bmatrix} \frac{1}{\mathfrak{l}_c}(\psi_c - \Psi(t)) \\ \frac{1}{16\mathfrak{l}_c}\Phi(t) \end{bmatrix} + \begin{bmatrix} v(t) \\ -\gamma(t)\sqrt{\Psi(t)} \end{bmatrix}. \tag{2.38}$$

Let $\mathfrak{u}(t) = [v(t), \gamma(t)]^T$ *and*

$$g(x(t)) = \begin{bmatrix} 1 & 0 \\ 0 & -\sqrt{\Psi(t)} \end{bmatrix},$$

then (2.38) *fits in the general form of* (2.4)*, except that we require additional initial restriction and Lipschitz continuity restriction on* γ*. We make a little abuse of notation here to define* $\mathcal{U} = [-0.05, 0.05] \times [0.5, 0.1]$*.*

Lyapunov-barrier approach

The Lyapunov method can deal with both forms of the conversion (i.e. (2.37) and (2.38)), we only provide an example based on (2.38). We first derive the sufficient conditions on the state-dependent control signal $\mathfrak{u} := [v, \gamma]^T$ and then embed such conditions as constraints into a QP framework [5]. Meanwhile the cost function is selected in a sense that the control effort

$$|\mathfrak{u}(t)|^2 + \frac{2\mathfrak{u}(t)}{\mathfrak{l}_c}(\psi_c - \Psi(t)) + \left(\frac{1}{4\mathfrak{l}_c}(\Phi(t) - \gamma(t))\sqrt{\Psi(t)}\right)^2 \tag{2.39}$$

is minimized for every $t > 0$.

To derive the sufficient conditions, we need to first select a closed set A such that $A + \epsilon\overline{\mathbb{B}} \subseteq \Gamma$ and control Lyapunov-barrier functions (V, B) such that the reach-avoid-stay control problem can be reduced to the stabilization with safety guarantee control problem $(\mathcal{X}_0, U, A)$.

To demonstrate the sufficiency of the conditions, we simply choose $A = \{\zeta\}$. Now we let $h_1(x) = -|x-\zeta|_\infty + r$ and $h_2(x) = |x-(0.500, 0.653)|_\infty - r$, then the sets $\Gamma = \{x : h_1(x) \geq 0\}$, $U = \{x : h_2(x) < 0\}$. Therefore, $B(x) = 1/h_2(x)$ is a proper control barrier function as required. The set D can be considered as an open set $\{x : h_2(x) > 0\}$, and $V(x)$ can be chosen as $V(x) = |x - \zeta|^2$ for all $x \in D \setminus A$. The set of control strategies for the reach-avoid-stay problem based on (V, B) is then obtained as $\mathfrak{K}(x)$ (i.e. determined straightforwardly from Theorem 2.4.5). However, for Problem 3.3.32, we have additional restrictions on γ (i.e. the second entry of $\mathfrak{u}$), which is given as

$$\mathcal{M} := \{\gamma \in [0.5, 1] : \gamma(0) \in [0.62, 0.66],\ |\gamma(t+\tau) - \gamma(t)| \leq 0.01\tau,\ \forall \tau > 0\}. \tag{2.40}$$

Therefore, the final sufficient conditions on the state-dependent control signals are

$$\mathfrak{u}(t) \in \mathfrak{K}(x(t)) \cap \mathcal{M}. \tag{2.41}$$

To embed the conditions of (2.41) into the quadratic programming with the selected cost function, we choose sampling time as 0.1 and use numerical iteration method to obtain the discrete dynamics. The results justified that the sufficient conditions are effective for any $\gamma(0)$ and $(\Phi_0, \Psi_0) \in \mathcal{X}_0$, but we only show the case when $(\Phi_0, \Psi_0) = (0.5343, 0.6553)$ and $\gamma(0) = 0.66$ as an example.

As a result, the control signal v and γ are shown in Figure 2.4.2. The sufficient conditions on the signals generated by control Lyapunov-barrier functions are shown to be effectively embedded within the QP with the minimum input energy (2.39). In particular, the extra conditions on the changing rate of signals are reactively included.

The phase portraits of the resulting trajectory is shown in Figure 2.5. The local Lipschitz continuity of $\mathfrak{u}$ can also be guaranteed in this framework ([5, Theorem 3]). The synthesis of v and γ mainly depends on the sufficient conditions, as long as it is feasible ($\mathfrak{K}(x(t)) \cap \mathcal{M} \neq \emptyset$) for the current iteration, it will proceed to the next iteration. The chattering effect of v around time 20 is due to the relatively fast change of γ, which in turn affects the varying speed of the dynamics. The signal γ decided by the QP tends to converge to the γ under which equilibrium point is ζ, which is around 0.56.

In particular, when γ is around the Hopf-bifurcation point, the set control strategies $\mathfrak{K}(x)$ can force the trajectories to reach the set A with an exponential rate, the transient speed of the local dynamic will not affect the decision process of v and γ.

Discussions on comparisons with formal methods

To apply standard formal methods to Problem 3.3.32, we fix a sampling time 0.1 and use ROCS [109] to compute an inner approximation of the winning set w.r.t. the reach-avoid-stay specifi-

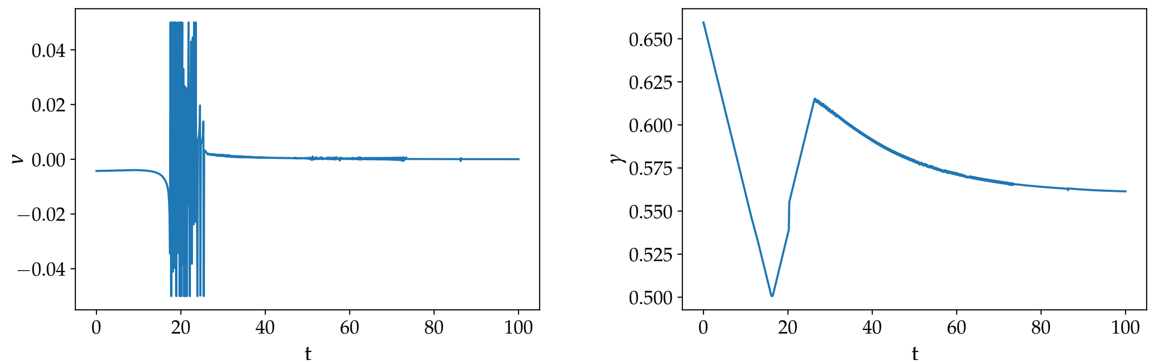


Figure 2.4: Control signal v and γ solved by the quadratic programming with condition (2.41) as constraints and (2.39) as the cost function.

cation as well as synthesize the control strategy. It turned out that the approximated winning set, which is shown in Figure 2.6, fails to cover the given initial condition.

The main reason is that, for γ around the bifurcation point, the sampling time 0.1 is too small to avoid spurious self-transitions in the abstraction. Hence, the reachable set computed on the abstraction is much smaller than the real one. The system state variable Ψ evolves very slowly when γ is around the bifurcation point, and this range of γ cannot be avoided due to the constraint on the change rate of γ.

To further see the effect of such a constraint, we remove the constraint $|\gamma(t+\tau) - \gamma(t)| \leq 0.01\tau$ and perform control synthesis with sampling time 0.01 by using ROCS for (2.36) directly with γ as a control variable. In this case, a winning set that can cover the initial condition is obtained (see Figure 2.6). From the closed-loop simulation result given in Figure 2.7, we notice that the bifurcation point (≈ 0.613) is skipped, leading to potentially unphysical control signals.

In summary, this case study poses challenges to formal methods for two reasons. First, the rate of change for the system state is sensitive to changes in the parameter. This would require the use of a parameter-dependent sampling time in constructing abstractions or computing reachable sets. Current tools [145, 109] and even multiscale methods [68, 85] cannot be readily used to handle such situations. Second, the system includes constraints on the change rates of control inputs. Even though (2.36) can be reformulated to (2.37), considering the throttle coefficient γ as a state variable, as opposed to a control variable, will lead to a more conservative control strategy, because of the curse of dimensionality and additional spurious transitions induced by a coarser abstraction/discretization scheme.

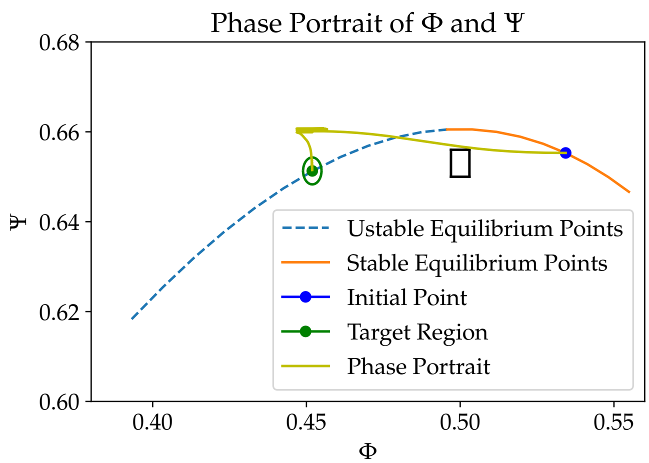


Figure 2.5: Phase portrait of (Φ, Ψ) generated based on signal v and γ.

2.5 Lyapunov-Barrier Characterization for Reach-Avoid-Stay Specifications of Hybrid Systems

We have seen in Section 2.3 that it is possible to construct united Lyapunov and barrier functions, or even a single Lyapunov function, that are defined on the entire set of initial conditions from which stabilization with safety guarantees is satisfied. The connection between stabilization with safety guarantees and reach-avoid-stay was also established via robustness.

In this section, we rely on the concept of solutions and show that the Lyapunov-barrier approach can be extended to verification of reach-avoid-stay specifications for hybrid systems with differential and difference inclusions. Unlike systems with diffusion (e.g. stochastic systems with Itô diffusion as will be shown in Chapter 3), reach-avoid-stay properties of solutions to hybrid systems can be converted to stabilization with safety guarantees under mild conditions. The approximated equivalence of these two types of specifications can be established via robustness. We show that smooth Lyapunov-barrier functions can be constructed given the compactness of target set and the set of initial states.

Note that even though the idea follows the previous work in Section 2.3, the underlying topology of the solutions to hybrid systems are different. Hence, the results from Section 2.3 are not directly applicable. We aim to leverage the rich results on stability theory of hybrid systems [71] to unify Lyapunov and barrier conditions in the context of converse Lyapunov-barrier theorems. We first provide the main results to show the similarity as in Section 2.3, then we would like to highlight the proofs in Section 2.5.5 based on the new topology of hybrid

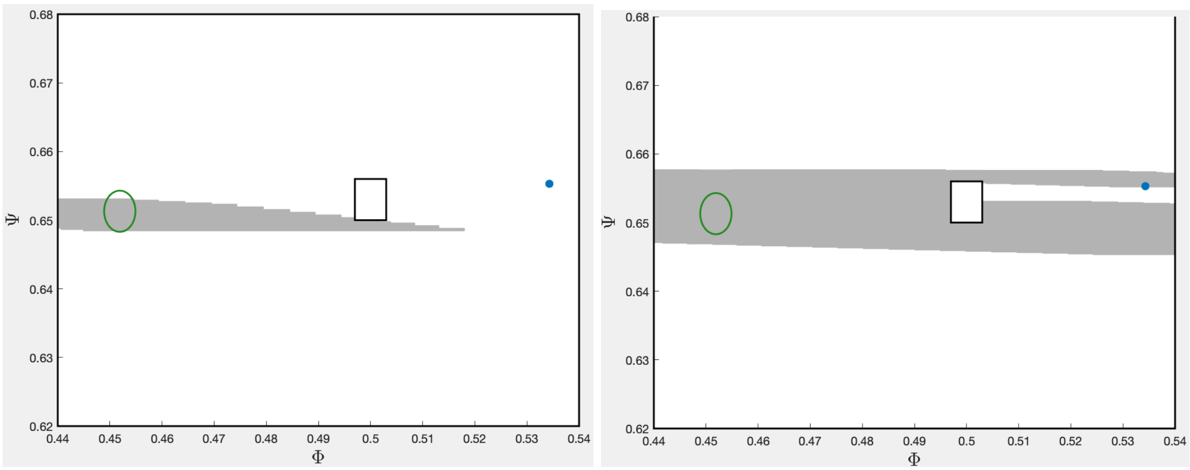


Figure 2.6: The approximated winning sets (shaded area) for system (2.37) (left) with sampling time 0.1 and system (2.38) (right) with sampling time 0.01, respectively. The green ball: the target set; the black box: the avoid area; the blue dot: the initial condition. [122]

systems.

We mention that the latest converse barrier theorems[4] for differential inclusions [64] provide a possibility to construct less smooth Lyapunov-barrier functions under less restricted topological requirement, e.g., unbounded reachable set. In this section, we only consider the case where the uniformly asymptotically stable set is compact, which seems already satisfactory in practice. For this reason, we stick with the compactness assumption on the target set and the set of initial states.

Before proceeding, we present the preliminaries for the hybrid systems, concepts of solutions, as well as other important definitions.

2.5.1 Preliminaries

Hybrid systems

Consider a hybrid system $\mathbf{H} = (C, F, D, G)$ with dynamics

$$\dot{x} \in F(x), \quad x \in C, \tag{2.42a}$$

$$x^+ \in G(x), \quad x \in D, \tag{2.42b}$$

[4]Recall that barrier functions are intended for merely safety properties.

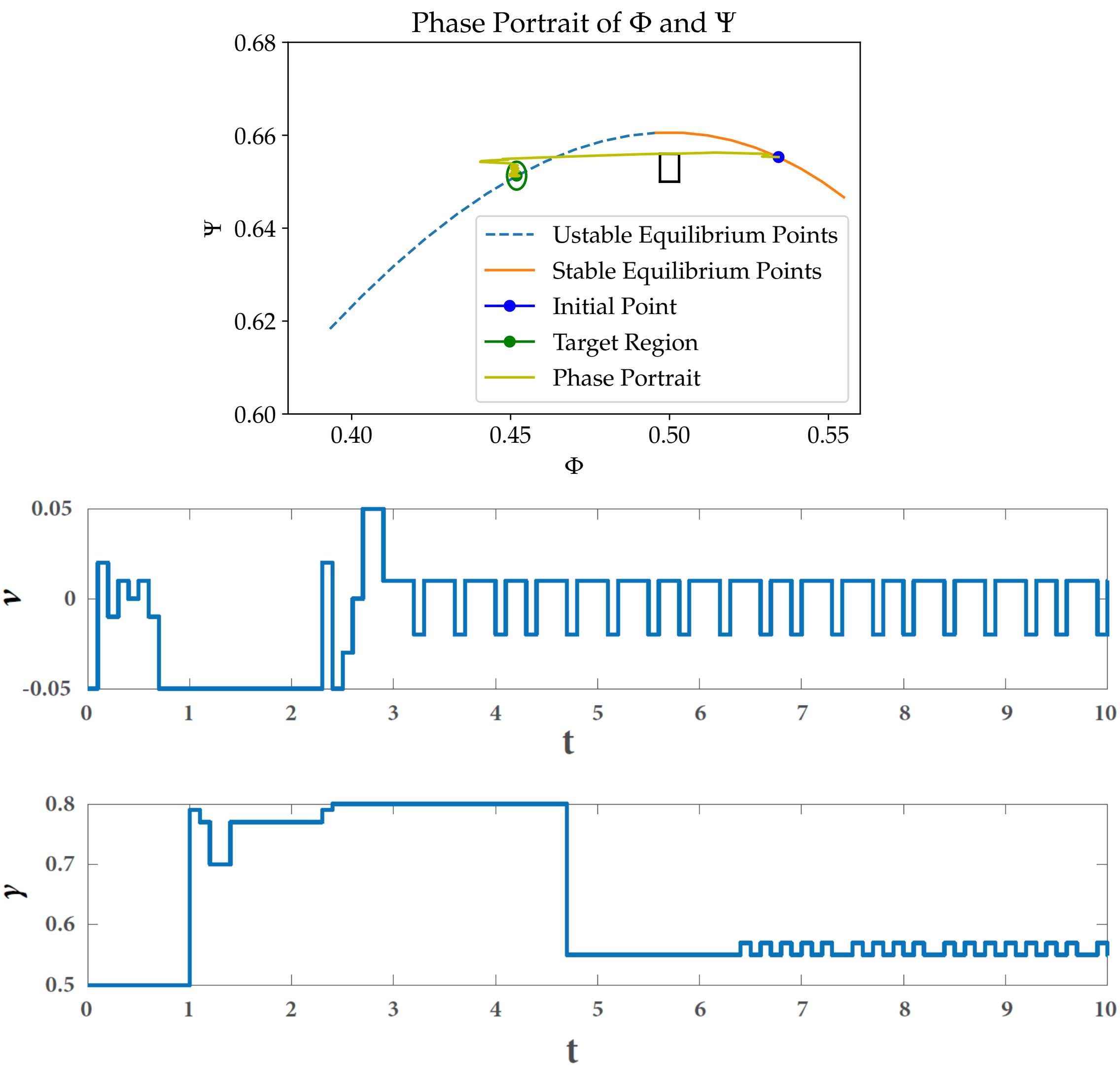


Figure 2.7: The closed-loop simulation with the control policy generated without the constraint on the change rate of γ. [122]

where $C, D \subseteq \mathbb{R}^n$ represent the flow set and the jump set respectively, and the set-valued maps $F : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, $G : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ represent flow and jump maps respectively.

Given a scalar $\vartheta \geq 0$, the (additive) ϑ-perturbation of $\mathbf{H}$, denoted by $\mathbf{H}_\vartheta$, is described as

$$\dot{x} \in F(x) + \vartheta\overline{\mathbb{B}}, \quad x \in C_\vartheta, \tag{2.43a}$$

$$x^+ \in G(x) + \vartheta\overline{\mathbb{B}}, \quad x \in D_\vartheta, \tag{2.43b}$$

where $C_\vartheta = C + \vartheta\overline{\mathbb{B}}$ and $D_\vartheta = D + \vartheta\overline{\mathbb{B}}$. A hybrid time domain is a subset $E \subseteq \mathbb{R}_{\geq 0} \times \mathbb{N}$ of the form $\cup_{j=0}^{J}[t_j, t_{j+1}] \times \{j\}$, where $J \in \mathbb{N} \cup \{\infty\}$ and $0 = t_0 \leq t_1 \leq t_2 \leq \dots$. Given $(t, j), (t', j') \in E$, the natural ordering is such that $t + j \leq t' + j'$ if $t \leq t'$ and $j \leq j'$. A hybrid arc is a function $\phi : E \to \mathbb{R}^n$ from a hybrid domain E to $\mathbb{R}^n$ and, for each fixed j, $t \mapsto \phi(t, j)$ is locally absolutely continuous on the interval $I^j = \{t : (t, j) \in \mathrm{dom}(\phi)\}$.

Definition 2.5.1 (Solution concept)**.** *A hybrid arc ϕ is a solution to* (2.42) *if*

(1) $\phi(0, 0) \in \overline{C} \cup D$;

(2) for all $j \in \mathbb{N}$ such that $\mathrm{Int}(I^j) \neq \emptyset$, we have

$$\phi(t, j) \in C \;\; \textit{for all } t \in \mathrm{Int}(I^j),$$
$$\dot{\phi}(t, j) \in F(\phi(t, j)) \;\; \textit{for almost all } t \in I^j;$$

(3) for all $(t, j) \in \mathrm{dom}(\phi)$ such that $(t, j + 1) \in \mathrm{dom}(\phi)$, we have $\phi(t, j) \in D$ and $\phi(t, j + 1) \in G(\phi(t, j))$.

We define the range of a solution arc ϕ as $\mathrm{rge}(\phi) = \{\phi(t, j) : (t, j) \in \mathrm{dom}(\phi)\}$ for convenience. For $\vartheta \geq 0$, we denote by $\mathfrak{S}^\vartheta_{\mathbf{H}}(x)$ the set of all maximal solutions[5] starting from $x \in \mathbb{R}^n$ for a hybrid system $\mathbf{H}_\vartheta$; we denote by $\mathfrak{S}^\vartheta_{\mathbf{H}}(K)$ the set of all maximal solutions starting from the set $K \subseteq \mathbb{R}^n$ for $\mathbf{H}_\vartheta$.

We introduce correspondingly some notations for reachable sets of $\mathbf{H}_\vartheta$ from some $K \subseteq \mathbb{R}^n$. For $T \geq 0$, we define

$$\mathcal{R}^{\leq T}_\vartheta(K) = \left\{\phi(t, j) : \phi \in \mathfrak{S}^\vartheta_{\mathbf{H}}(K), \phi(0, 0) \in K, t + j \leq T\right\}.$$

The reachable sets $\mathcal{R}^{<T}_\vartheta(K)$ and $\mathcal{R}^{\geq T}_\vartheta(K)$ are defined in a similar way (see examples in Section 2.1). The infinite-horizon reachable set from K for $\mathbf{H}_\vartheta$ is

$$\mathcal{R}_\vartheta(K) = \left\{\phi(t, j) : \phi \in \mathfrak{S}^\vartheta_{\mathbf{H}}(K),\; (t, j) \in \mathrm{dom}(\phi)\right\}.$$

[5]We omit this standard definition from this section. For a formal definition see e.g. [71, Definition 2.7]

Assumption 2.5.2. *We make the standing assumption that* $\mathbf{H}$ *should satisfy the basic conditions:*

(C1) $C, D \in \mathbb{R}^n$ *are closed.*

(C2) F *is outer semicontinuous, locally bounded, and convex for all* $x \in C$*.*

(C3) G *is outer semicontinuous and locally bounded for all* $x \in D$*.*

We also introduce the following concept of closeness of hybrid arcs that will be frequently used in proofs.

Definition 2.5.3. *[71, Definition 5.23] Given* $\tau, \epsilon > 0$*, two hybrid arcs* ϕ_1 *and* ϕ_2 *are* (τ, ϵ)*-close if*

(1) for all $(t, j) \in \operatorname{dom}(\phi_1)$ *with* $t + j \leq \tau$ *there exists* s *such that* $(s, j) \in \operatorname{dom}(\phi_2)$*,* $|t - s| < \epsilon$ *and* $|\phi_1(t, j) - \phi_2(s, j)| < \epsilon$*;*

(2) for all $(t, j) \in \operatorname{dom}(\phi_2)$ *with* $t + j \leq \tau$ *there exists* s *such that* $(s, j) \in \operatorname{dom}(\phi_1)$*,* $|t - s| < \epsilon$ *and* $|\phi_2(t, j) - \phi_1(s, j)| < \epsilon$*.*

Properties pertaining to stability and safety

We define properties that are related to stability and safety of solutions to (2.43) in a similar way as in Section 2.3. Particularly, we formally introduce the concepts of stability with safety guarantees and reach-avoid-stay type specifications.

Definition 2.5.4 (Forward (pre-) invariance)**.** *A set* $\mathfrak{I} \subseteq \mathbb{R}^n$ *is said to be forward pre-invariant for* $\mathbf{H}_\vartheta$ *if for every* $\phi \in \mathfrak{S}_{\mathbf{H}}^{\vartheta}(\mathfrak{I})$*,* $\operatorname{rge}(\phi) \subseteq \mathfrak{I}$*. The set* $\mathfrak{I}$ *is said to be forward invariant for* $\mathbf{H}_\vartheta$ *if for every forward complete* $\phi \in \mathfrak{S}_{\mathbf{H}}^{\vartheta}(\mathfrak{I})$*,* $\operatorname{rge}(\phi) \subseteq \mathfrak{I}$*, i.e.,* $\mathcal{R}_\vartheta(\mathfrak{I}) \subseteq \mathfrak{I}$*.*

Remark 2.5.5. *The term 'pre' is in the sense that non-complete maximal solutions are not excluded. This concept allows us to describe the completeness and dynamical behaviors separately for general maximal solutions. For future references, we only define the 'pre' properties of solutions in consideration of the space limitation.*

We next consider stability for solutions of $\mathbf{H}_\vartheta$ w.r.t. a closed set.

Definition 2.5.6 (UpAS)**.** *A closed set* $A \subseteq \mathbb{R}^n$ *is said to be UpAS for* $\mathbf{H}_\vartheta$ *if the following two conditions are met:*

(1) (uniform stability) for every $\epsilon > 0$, there exists a $\eta = \eta(\epsilon) > 0$ such that every solution to $\mathbf{H}_\vartheta$ with $|\phi(0,0)|_A \leq \eta$ satisfies $|\phi(t,j)|_A \leq \epsilon$ for all $(t,j) \in \mathrm{dom}(\phi)$;

(2) (uniform pre-attractivity) there exists some $r > 0$ such that, for every $\epsilon > 0$, there exist some $T > 0$ such that, for every solution ϕ to $\mathbf{H}_\vartheta$, $|\phi(t,j)|_A \leq \epsilon$ whenever $|\phi(0,0)|_A \leq r$, $(t,j) \in \mathrm{dom}(\phi)$, and $t + j \geq T$.

Definition 2.5.7 (Basin of pre-attraction)**.** *If a closed set $A \subseteq \mathbb{R}^n$ is UpAS for $\mathbf{H}_\vartheta$, we further define the basin of pre-attraction of A, denoted by $\mathfrak{B}_\vartheta(A)$, as the set of initial states $x \in \mathbb{R}^n$ such that every solution ϕ to $\mathbf{H}_\vartheta$ with $\phi(0,0) = x$ is bounded and, if it is complete, then also converges to the set A, i.e., $\lim_{t+j\to\infty} |\phi(t,j)|_A = 0$.*

We define stability with safety guarantee and reach-avoid-stay properties for the system $\mathbf{H}_\vartheta$ in the following definitions.

Definition 2.5.8 (Stability with safety guarantee)**.** *Let $\mathcal{X}_0, U, A \subseteq \mathbb{R}^n$ and A is a closed set. We say that $\mathbf{H}_\vartheta$ satisfies a stability with safety guarantee specification $(\mathcal{X}_0, U, A)$ if*

(1) (pre-asymptotic stability w.r.t. A) the set A is UpAS for $\mathbf{H}_\vartheta$ and $\mathcal{X}_0 \subseteq \mathfrak{B}_\vartheta(A)$;

(2) (safety w.r.t. U) $\mathrm{rge}(\phi) \cap U = \emptyset$ for all $\phi \in \mathfrak{S}_{\mathbf{H}}^{\vartheta}(\mathcal{X}_0)$.

Definition 2.5.9 (Reach-avoid-stay specification)**.** *Let $\mathcal{X}_0, U, \mathfrak{I} \subseteq \mathbb{R}^n$. We say that $\mathbf{H}_\vartheta$ satisfies a reach-avoid-stay specification $(\mathcal{X}_0, U, \mathfrak{I})$ if*

(1) (reach and stay w.r.t. $\mathfrak{I}$) there exists some $T \geq 0$ such that $\mathcal{R}_\vartheta^{\geq T}(\mathcal{X}_0) \subseteq \mathfrak{I}$;

(2) (safety w.r.t. U) $\mathrm{rge}(\phi) \cap U = \emptyset$ for all $\phi \in \mathfrak{S}_{\mathbf{H}}^{\vartheta}(\mathcal{X}_0)$.

2.5.2 Connection Between Stability with Safety Guarantees and Reach-Avoid-Stay Specifications

Stability with safety implies reach-avoid-stay

Throughout this section, we suppose that $\mathbf{H}_\vartheta$ satisfies a stability with safety guarantee specification $(\mathcal{X}_0, U, A)$ for any fixed $\vartheta \geq 0$.

We first show some basic properties of solutions to $\mathbf{H}_\vartheta$ with a fixed $\vartheta \geq 0$. The following proposition combines the result from [71, Proposition 7.4, Lemma 7.8].

Proposition 2.5.10. *Given a hybrid system $\mathbf{H}_\vartheta$, let $A \subseteq \mathbb{R}^n$ be a compact set that is uniform stable for $\mathbf{H}_\vartheta$ (condition (1) of Definition 2.5.6 holds), and every solution ϕ to $\mathbf{H}_\vartheta$ with proper initial condition be either bounded or complete such that $\lim_{t+j\to\infty} |\phi(t,j)|_A = 0$. Suppose $K \subseteq \mathfrak{B}_\vartheta(A)$ is compact. Then*

(1) $\mathfrak{B}_\vartheta(A)$ is open and $A \subseteq \mathfrak{B}_\vartheta(A)$.

(2) For every $\epsilon > 0$, there exists some $T = T(\epsilon) > 0$ such that

$$|\phi(t,j)|_A \leq \epsilon$$

for all $\phi \in \mathfrak{S}_\vartheta(K)$, $(t,j) \in \operatorname{dom}(\phi)$ with $t + j \geq T$.

(3) $A \cup \mathcal{R}(K)$ is a compact subset of $\mathfrak{B}_\vartheta(A)$.

By (2) of Proposition 2.5.10, which indicates the equivalence of uniform attraction and asymptotica attraction for solutions to $\mathbf{H}_\vartheta$ under the uniform stability condition, it can be shown straightforwardly the reach-avoid-stay property (see also Proposition 2.3.9 for comparison). The statement is given in the following proposition.

Proposition 2.5.11. *Let $A, \mathcal{X}_0 \subseteq \mathbb{R}^n$ be compact sets. Suppose that $\mathbf{H}_\vartheta$ satisfies a stability with safety guarantee specification $(\mathcal{X}_0, U, A)$. Then for every $\epsilon > 0$, the system $\mathbf{H}_\vartheta$ also satisfies the reach-avoid-stay specification $(\mathcal{X}_0, U, A + \epsilon\overline{\mathbb{B}})$.*

The converse side

Throughout this subsection, we suppose that the perturbed system $\mathbf{H}_\vartheta$ satisfies a reach-avoid-stay specification $(\mathcal{X}_0, U, \mathfrak{I})$. We make an additional assumption on the local Lipschitz property of F and G to prove the converse connection.

Assumption 2.5.12. *We assume in this section that F and G in (2.43) are locally Lipschitz on some open subset of C and D respectively.*

Definition 2.5.13 (Locally Lipschitz set-valued maps)**.** *Let $\mathcal{O} \subseteq \mathbb{R}^n$ be an open set. A set-valued flow map F is also locally Lipschitz on $\mathcal{O}$, that is, for each $x \in \mathcal{O}$, there exists a neighborhood $\mathcal{N} \subset \mathcal{O}$ of x and an $L > 0$ such that for any $x_1, x_2 \in \mathcal{N}$,*

$$F(x_1) \subseteq F(x_2) + L|x_2 - x_1|\overline{\mathbb{B}}.$$

The following fact is a direct consequence of being locally Lipschitz [160, Lemma 9].

Lemma 2.5.14. *If $F : \mathcal{O} \rightrightarrows \mathbb{R}^n$ is locally Lipschitz, then for any compact subset $K \subseteq \mathcal{O}$ there exists an $L_K > 0$ such that for any $x_1, x_2 \in K$, we have*

$$F(x_1) \subseteq F(x_2) + L_K|x_2 - x_1|\overline{\mathbb{B}}.$$

Lemma 2.5.15 (Perturbed solutions)**.** *For any $\vartheta > 0$, fix a $\vartheta' \in [0, \vartheta)$ and a $\tau > 0$. Let $T \geq \tau$ and $K \subseteq \mathbb{R}^n$ be a compact set such that $K \cap (C_\vartheta \cup D_\vartheta) \neq \emptyset$. Then there exists an $r = r(K, \tau, \vartheta', \vartheta) > 0$ such that for every solution $\phi \in \mathfrak{S}^{\vartheta'}_{\mathbf{H}}(K)$ with $\phi(t', j') \in K$ for all $(t', j') \in \mathrm{dom}(\phi)$ and $t' + j' \leq T$, and for all $x \in \phi(0,0) + r\overline{\mathbb{B}}$, there exists a $\psi \in \mathfrak{S}^{\vartheta}_{\mathbf{H}}(x)$ with $\mathrm{dom}(\psi) = \mathrm{dom}(\phi)$ and*

$$\psi(t, j) = \phi(t, j), \quad t + j = T. \tag{2.44}$$

Remark 2.5.16. *We provide the proof in Section 2.5.5. The above result shows that for any compact solution ϕ to $\mathbf{H}_{\vartheta'}$ (that exists till $t + j = \tau$), there exists a solution ψ to a slightly more perturbed system $\mathbf{H}_\vartheta$ such that the endpoints of ϕ and ψ are related within some time period. The construction and estimation rely on the local Lipschitz continuity of F and G. The proof states that no matter how conservative the estimation is, we are able to find the small neighborhood with radius r such that for any initial condition within $\phi(0,0) + r\overline{\mathbb{B}}$, there exists a ψ as a solution to $\mathbf{H}_\vartheta$ converging to ϕ within finite time. The requirement that ϑ' should be strictly less than ϑ is necessary. Unlike the robustness concept given in [71, Lemma 7.37], the perturbed systems $\mathbf{H}_\vartheta$ need inflation of C and D whose intensities increase as ϑ increases. This subtle difference is in consideration of when $\phi(0,0) \in \partial(C_{\vartheta'} \cup D_{\vartheta'})$ whilst the constructed ψ is still well posed.*

Applying the proceeding results, we show in the next proposition a hybrid-system version of Proposition 2.3.16.

Proposition 2.5.17. *Any nonempty, compact, forward pre-invariant set A for $\mathbf{H}_\vartheta$ is UpAS for $\mathbf{H}_{\vartheta'}$ with any $\vartheta' \in [0, \vartheta)$.*

Remark 2.5.18. *The proof is similar to that of Proposition 2.3.16. That the ϵ in Definition 2.5.6 can be arbitrarily given (at least in a small range) is guaranteed by the reach-avoid-stay assumption on $\mathbf{H}_\vartheta$. The idea is to suppose the opposite, i.e., the solution ϕ of $\mathbf{H}_{\vartheta'}$ starting within a sufficiently small neighborhood $A + r\overline{\mathbb{B}}$ can possibly not reach A or even flow/jump out of the compact set $K := |\phi(t, j)|_A \leq \epsilon$ for $t + j \geq \tau$, where r, τ are as in Lemma 2.5.15. However, Lemma 2.5.15 indicates that these are not the cases in that for any of those ϕ, there exists some ψ to $\mathbf{H}_\vartheta$ that will be identical to ϕ after $t + j \geq \tau$, which contradicts the pre-invariance assumption. We omit the details since there is no convergence concepts used in the proof that are special for hybrid systems.*

The following result is an analogue of Lemma 2.3.15. We complete the proof in Section 2.5.5.

Proposition 2.5.19 (Existence of compact invariant set)**.** *Suppose that $\mathfrak{I}$ is compact and $\mathcal{X}_0$ is nonempty. Suppose that $\mathbf{H}_\vartheta$ satisfies a reach-avoid-stay specification $(\mathcal{X}_0, U, \mathfrak{I})$. Then the set*

$$A = \left\{x \in \mathfrak{I} : \forall \phi \in \mathfrak{S}_{\mathbf{H}}^{\vartheta}(x), \mathrm{rge}(\phi) \subseteq \mathfrak{I}\right\} \tag{2.45}$$

is nonempty, compact, and forward pre-invariant for $\mathbf{H}_\vartheta$.

Combining Proposition 2.5.17 and Proposition 2.5.19, we immediately obtain the following connection between reach-avoid-stay specifications and stability with safety guarantees.

Proposition 2.5.20. *Let Assumption 2.5.12 be satisfied. Let $\mathcal{X}_0$ be a nonempty set and $\mathfrak{I}$ be a compact set. Suppose that $\mathbf{H}_\vartheta$ satisfies a reach-avoid-stay specification $(\mathcal{X}_0, U, \mathfrak{I})$. Then there exists a compact set $A \subseteq \mathfrak{I}$ such that any less perturbed system $\mathbf{H}_{\vartheta'}$ with $\vartheta' \in [0, \vartheta)$ satisfies the stability with safety guarantee specification $(\mathcal{X}_0, U, A)$.*

2.5.3 Lyapunov-Barrier Functions for Reach-Avoid-Stay Specifications

In this subsection, we provide Lyapunov-barrier characterizations for stability with safety guarantee specifications as well as reach-avoid-stay specification. In particular, such characterizations are (robustly) complete in the sense that Lyapunov-barrier functions exist based on the specified dynamical behavior of the solutions. The proofs of results from this subsection are provided in Section 2.5.5.

We focus on the Lyapunov-barrier functions for stability with safety guarantees, the result for reach-avoid-stay comes after the connection given in Proposition 2.5.20.

Definition 2.5.21 (Basin of pre-attraction with safety)**.** *The extracted basin of pre-attraction with safety is a subset of $\mathfrak{B}_\vartheta(A)$ defined by*

$$\widehat{\mathfrak{B}}_\vartheta(A) := \{x \in \mathfrak{B}_\vartheta(A) : \forall \phi \in \mathfrak{S}_{\mathbf{H}}^{\vartheta}(x), \ \ \mathrm{rge}(\phi) \cap U = \emptyset\}, \tag{2.46}$$

where A, U are given in the specification $(\mathcal{X}_0, U, A)$.

The following lemma verifies some basic properties of the set $\widehat{\mathfrak{B}}_\vartheta(A)$.

Lemma 2.5.22. *Suppose that A is compact, U is closed, and $A \cap U = \emptyset$. If $\mathbf{H}_\vartheta$ satisfies the stability with safety guarantee specification $(\mathcal{X}_0, U, A)$, then*

(1) $\mathcal{X}_0 \subseteq \widehat{\mathfrak{B}}_\vartheta(A) \subseteq \mathfrak{B}_\vartheta(A)$;

(2) $\widehat{\mathfrak{B}}_\vartheta(A)$ *is open and forward pre-invariant.*

The following result shows that based on the reachable region of solutions with stability and safety guarantees, a single Lypunov like function exists which is also effective as a barrier function to guarantee the safety.

Theorem 2.5.23. *Suppose that A is compact, U is closed, and $A \cap U = \emptyset$. Then the following two statements are equivalent:*

(1) $\mathbf{H}_\vartheta$ *satisfies the stability with safety guarantee specification* $(\mathcal{X}_0, U, A)$

(2) *There exists an open and forward pre-invariant set $\mathfrak{O}$ such that $(A \cup \mathcal{X}_0) \subseteq \mathfrak{O}$ and $\mathfrak{O} \cap U = \emptyset$, a smooth function $V : \mathfrak{O} \to \mathbb{R}_{\geq 0}$ and class $\mathcal{K}_\infty$ functions α_1, α_2 such that,*

$$\begin{aligned} \alpha_1(\mathfrak{w}(x)) \leq V(x) \leq \alpha_2(\mathfrak{w}(x)), \\ \forall x \in (C_\vartheta \cup D_\vartheta \cup G(D_\vartheta)) \cap \mathfrak{O}, \end{aligned} \tag{2.47}$$

$$\nabla V(x) \cdot f \leq -V(x), \;\; \forall x \in C_\vartheta \cap \mathfrak{O}, \;\; f \in F(x) + \vartheta\overline{\mathbb{B}}, \tag{2.48}$$

and

$$V(g) \leq V(x)/e, \;\; \forall x \in D_\vartheta \cap \mathfrak{O}, \;\; g \in G(x) + \vartheta\overline{\mathbb{B}}, \tag{2.49}$$

where $\mathfrak{w}$ is a proper indicator for A on $\mathfrak{O}$.

In particular, we can let $\mathfrak{O} = \widehat{\mathfrak{B}}_\vartheta(A)$.

Remark 2.5.24. *Note that to use [71, Corollary 7.33], we need to verify that the UpAS holds for the following system of the following modifications:*

$$\dot{x} \in F_\varrho(x), \qquad x \in (C_\vartheta)_\varrho, \tag{2.50a}$$

$$x^+ \in G_\varrho(x), \quad x \in (D_\vartheta)_\varrho, \tag{2.50b}$$

where

$$\begin{aligned} (C_\vartheta)_\varrho &= \left\{x \in \mathbb{R}^n : (x + \varrho(x)\overline{\mathbb{B}} \cap C_\vartheta \neq \emptyset\right\}, \\ F_\varrho(x) &= \overline{\text{con}} F((x + \varrho(x)\overline{\mathbb{B}}) \cap C_\vartheta) + (\varrho(x) + \vartheta)\overline{\mathbb{B}}, \\ (D_\vartheta)_\varrho &= \left\{x \in \mathbb{R}^n : (x + \varrho(x)\overline{\mathbb{B}} \cap D_\vartheta \neq \emptyset\right\}, \end{aligned}$$

$G_\varrho(x) = \left\{v \in \mathbb{R}^n : v \in g + (\varrho(g) + \vartheta)\overline{\mathbb{B}}\right\}$, *for $g \in G(x + \varrho(x)\overline{\mathbb{B}}) \cap D_\vartheta$. However, this is guaranteed by Assumption 2.5.2 for each $\vartheta \geq 0$.*

The proof for [71, Corollary 7.33] relies on the conversion from UpAS to the $\mathcal{KL}$ pre-asymptotic stability based on changing the basin of attraction. In contrast, for robust systems driven by differential equations with F being a single-valued function, Proposition 2.3.9 provides a more explicit way of constructing the $\mathcal{KL}$-function β such that $\mathfrak{w}(\phi(t)) \leq \beta(\mathfrak{w}(x), t)$ for all solutions with initial conditions $x \in \mathfrak{D}$. The argument is based on the reachable set properties as well as a connection between the solution-to-set distance and the proper indicator. The extension of the explicit construction from Proposition 2.3.9 to the general hybrid systems should be similar.

By the connection obtained in Section 2.5.2, the following results follow.

Corollary 2.5.25. *Suppose that $A, \mathcal{X}_0$ are compact sets and (2) of Theorem 2.5.23 holds. Then $\mathbf{H}_\vartheta$ satisfies the reach-avoid-stay specification $(\mathcal{X}_0, U, A + \epsilon\overline{\mathbb{B}})$ for any $\epsilon > 0$.*

Theorem 2.5.26. *Let Assumption 2.5.12 be satisfied. Suppose that $\mathfrak{I}$ is compact, U is closed and $\mathfrak{I} \cap U = \emptyset$. Suppose that $\mathbf{H}_{\vartheta'}$ satisfies the reach-avoid-stay specification $(\mathcal{X}_0, U, \mathfrak{I})$. Then there exists a compact set $A \subseteq \mathfrak{I}$ such that (2) of Theorem 2.5.23 holds for any $\mathbf{H}_\vartheta$ with $\vartheta \in [0, \vartheta')$.*

The above constructed Lyapunov functions for the stability with safety guarantees and reach-avoid-stay specifications play an implicit role as a barrier function. We mimic Proposition 2.3.21 and provide an equivalent characterization as in Theorem 2.5.23 with separate Lyapunov-barrier functions to complete this section. The separate Lyapunov-barrier functions for reach-avoid-stay specifications are omitted due to the similarity.

Theorem 2.5.27. *Suppose that A is compact, U is closed and $A \cap U = \emptyset$. If there exists an open set $\mathfrak{D}$ such that $A \cup \mathcal{X}_0 \subseteq \mathfrak{D}$, a smooth function $V : \mathfrak{D} \to \mathbb{R}_{\geq 0}$ satisfying*

(1) there exist $\alpha_1, \alpha_2 \in \mathcal{K}$ and a continuous positive function ϱ such that

$$\alpha_1(|x|_A) \leq V(x) \leq \alpha_2(|x|_A), \quad \forall x \in (C_\vartheta \cup D_\vartheta \cup G(D_\vartheta)) \cap \mathfrak{D}, \tag{2.51}$$

$$\nabla V(x) \cdot f \leq -\varrho(|x|_A), \quad \forall x \in C_\vartheta, \ f \in F(x) + \vartheta\overline{\mathbb{B}}, \tag{2.52}$$

$$V(g) - V(x) \leq -\varrho(|x|_A), \quad \forall x \in D_\vartheta, \ g \in G(x) + \vartheta\overline{\mathbb{B}}; \tag{2.53}$$

and $B : \mathbb{R}^n \to \mathbb{R}$ that is smooth in $C \cap \mathfrak{D}$ such that

(2) the set $S := \{x \in \mathbb{R}^n : B(x) \geq 0\} \subseteq \mathfrak{D}$ and $\mathcal{X}_0 \subseteq S$;

(3) $x \in U$ implies $B(x) < 0$;

(4)

$$\nabla B(x) \cdot f \geq 0, \;\; \forall x \in C_\vartheta \cap \mathfrak{O}, \; f \in F(x) + \vartheta\overline{\mathbb{B}}, \tag{2.54}$$

$$B(g) - B(x) \geq 0, \;\; \forall x \in D_\vartheta \cap \mathfrak{O}, \; g \in G(x) + \vartheta\overline{\mathbb{B}}, \tag{2.55}$$

then $\mathbf{H}_\vartheta$ *satisfies the stability with safety guarantee specification* $(\mathcal{X}_0, U, A)$.

If $\mathcal{X}_0$ *is also compact, then the converse side also holds, i.e., the existence of smooth* V *and* B *with conditions (1)-(4) is necessary for* $\mathbf{H}_\vartheta$ *satisfying the stability with safety guarantee specification* $(\mathcal{X}_0, U, A)$.

Remark 2.5.28. *For the sufficient part, condition (4) intends to regulate the safe direction of solutions on the entire open set* $\mathfrak{O}$ *rather than directly on its subset* $S := \{x \in \mathbb{R}^n : B(x) \geq 0\}$. *This condition seems stronger but also necessary. Imposing condition (4) on* S *can only guarantee the invariance of the interior of* S, *for counters examples see e.g. [167, Remark 4].*

2.5.4 Examples

We provide two examples in this subsection to validate our results.

Bouncing ball

As a classical mechanical system with impulse-momentum change, the model of bouncing balls is frequently used to illustrate dynamical behaviors of hybrid systems. We model a tennis ball as a point-mass, and consider dropping it from a fixed height with a constant horizontal speed. The vertical direction is forced by the gravity. As the ball hits the horizontal surface, the instantaneous vertical velocity is reversed with a small dissipation of energy.

To describe the hybrid dynamics, the state of the ball is given as

$$\mathfrak{x} = (x, y, z)^T \in \mathbb{R}^3,$$

where x is the horizontal position, y represents the height above the surface, and z is the vertical velocity. The flow set is given

$$C = \{\mathfrak{x} : y > 0 \text{ or } y = 0, z \geq 0\}.$$

We consider the flow[6] as

$$\dot{\mathfrak{x}} = f(\mathfrak{x}) := \begin{bmatrix} 1 \\ z \\ -a \end{bmatrix}, \;\; \mathfrak{x} \in C,$$

[6]As in [71], it is natural to set $f(\mathbf{0}) = \mathbf{0}$ regardless of the noncontinuity at $\mathbf{0}$.

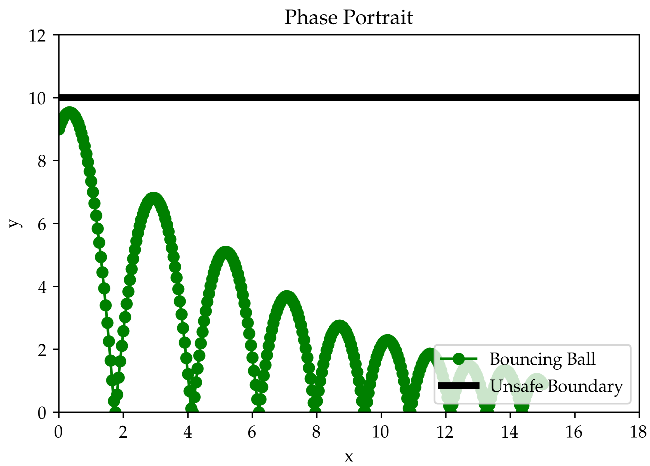


Figure 2.8: Snapshot of bouncing ball in xy-plane.

where $a = 9.8$ is the acceleration due to the gravity. The jump set captures the domain when the vertical velocity flips the sign, which is geiven as

$$D = \{\mathfrak{x} : y = 0 \text{ and } z < 0\}.$$

The jump dynamic is such that

$$\mathfrak{x}^+ = g(\mathfrak{x}) := \begin{bmatrix} x \\ 0 \\ -\varsigma z \end{bmatrix}, \quad \mathfrak{x} \in D,$$

for some dissipation coefficient $\varsigma \in (0, 1)$.

Now fix $\varsigma = 0.8$, let the initial condition to be $\mathfrak{x}(0) = (0, 9, 0.8)^T \in C$, and model a block (the unsafe set) by

$$U := \{\mathfrak{x} : y > 10\}.$$

We consider the target set as

$$\mathfrak{I} := \{\mathfrak{x} : y \in [0, 0.1]\}.$$

It can be shown analytically and numerically (see Figure 2.8) that the system satisfies the reach-avoid-stay specification $(\{\mathfrak{x}(0)\}, U, \mathfrak{I})$. We now validate the existence of Lyapunov-barrier functions (V, B).

Note that, the function

$$V(\mathfrak{x}) = \left(1 + \frac{1-\varsigma^2}{\pi(1+\varsigma^2)}\arctan(z)\right)\cdot\left(\frac{1}{2}z^2 + ay\right)$$

has been verified in [71, Example 3.19] to be a valid Lyapunov function w.r.t. the set $\{\mathfrak{x} : y = 0\}$. It suffices to find valid candidate of the barrier function B. Consider a sigmoid function

$$\sigma(x) = \frac{1}{1+\exp(-5x)}$$

and let

$$B(\mathfrak{x}) = \frac{1}{2}\sigma(x) - y - \frac{1}{2a}z^2 + 9.5.$$

It can be verified that $B(\mathfrak{x}) < 0$ for $\mathfrak{x} \in U$,

$$\nabla B(\mathfrak{x}) \cdot f(\mathfrak{x}) = \frac{1}{2}\sigma(x)(1-\sigma(x)) - z + z \geq 0, \;\; \mathfrak{x} \in C,$$

and

$$B(g(\mathfrak{x})) - B(\mathfrak{x}) = \frac{1}{2a}(z^2 - \varsigma^2 z^2) > 0, \;\; \mathfrak{x} \in D,$$

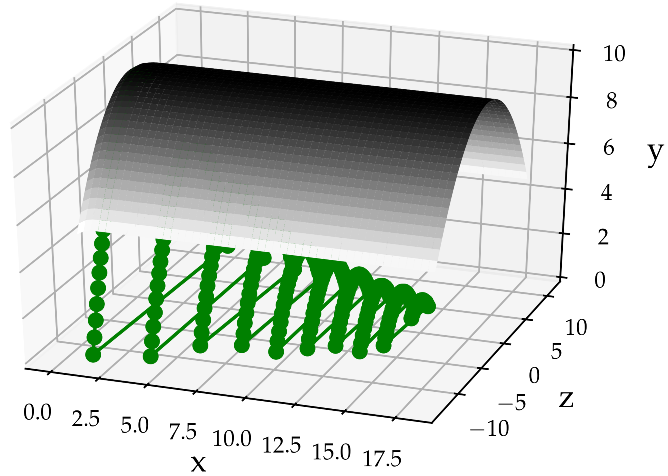
Therefore, the above $B(\mathfrak{x})$ is a valid barrier function. Let $S := \{\mathfrak{x} : B(\mathfrak{x}) \geq 0\}$ be the set as in Theorem 2.5.27, then it is also clear that $\mathfrak{x}(0) \in S$. The evolution of the state $\mathfrak{x} \in \mathbb{R}^3$ as well as the barrier function B are provided in Figure 2.9. It can be seen that for all $t \geq 0$, $\mathfrak{x}(t) \in S$ and therefore remain safety w.r.t. U.

Sample-and-hold control

In this example, we revisit the case study in Section 2.4.2 from the perspective of hybrid system. Due to the issues of less frequent and inaccurate state measurement, the errors inject into the closed-loop control system and may cause unsatisfactory performance. We are motivated by the above issues to convert the system (2.38), which is the equivalent form of (2.36), into a hybrid system. We then apply the conditions of the Lyapunov-barrier functions in Theorem 2.5.27 to synthesize valid sample-and-hold controllers aiming at fulfilling the task in Problem 2.4.6.

For simplicity, we define the state $\mathfrak{x} = (\Phi, \Psi)^T$ and the control input $\mathfrak{u} = (v, \gamma)$. We write

$$\tilde{f}(\mathfrak{x}, u) = \begin{bmatrix} \frac{1}{\mathfrak{l}_c}(\psi_c - \Psi(t)) \\ \frac{1}{16\mathfrak{l}_c}\Phi(t) \end{bmatrix} + \begin{bmatrix} v(t) \\ -\gamma(t)\sqrt{\Psi(t)} \end{bmatrix},$$



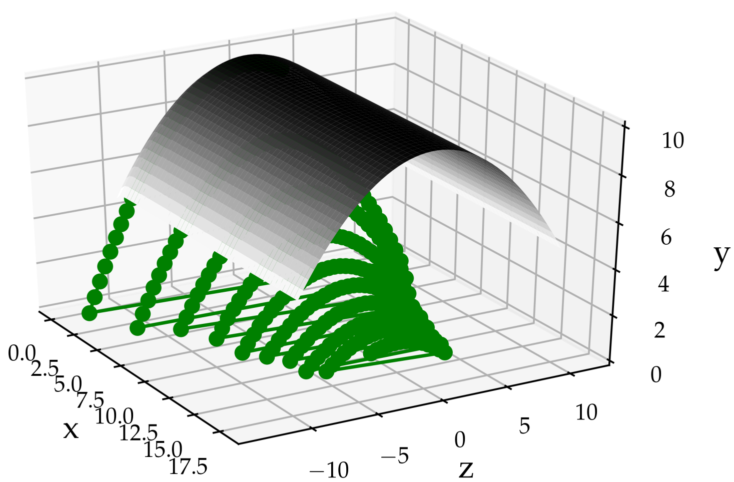


Figure 2.9: Solution of bouncing ball (the green dots) and barrier function B (the grey surface) in xyz-plane.

which is the r.h.s. of (2.38).

We then introduce $\tau \in [0, 0.5)$ as a timer variable and set

$$\mathfrak{z} = \begin{bmatrix} \mathfrak{x} \\ \mathfrak{u} \\ \tau \end{bmatrix} \in \mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}.$$

The flow set of the sample-and-hold hybrid system is given as $C = \mathbb{R}^2 \times \mathbb{R}^2 \times [0, 0.5)$, on which the flow dynamics are such that

$$\dot{\mathfrak{z}} = f(\mathfrak{z}) = \begin{bmatrix} \tilde{f}(\mathfrak{x}, \mathfrak{u}) \\ 0 \\ 1 \end{bmatrix}, \quad \mathfrak{z} \in C.$$

The jumps happen when the timer is up to 0.5 and decisions are made, i.e., $D = \mathbb{R}^2 \times \mathbb{R}^2 \times \{0.5\}$ and

$$\mathfrak{z}^+ = g(\mathfrak{z}) = \begin{bmatrix} \mathfrak{x} \\ \kappa(\mathfrak{x}) \\ 0 \end{bmatrix}, \quad \mathfrak{z} \in D. \tag{2.56}$$

We choose Lyapunov-barrier functions (V, B) as in Section 2.4.2 and define a set of high-gain robust control policy as follows:

$$\mathfrak{K}(x) = \{u \in \mathcal{U} : L_f V(x) + L_g V(x) u + V(x) \leq -\varsigma \text{ and } L_f B(x) + L_g B(x) u \geq \varsigma, \ \forall x \in D\}, \tag{2.57}$$

where $\varsigma > 0$ is intended to compensate the error that is generated when the same control input is imposed on the flow dynamics. Since $L_f V$, $L_g V$, $L_f B$ and $L_g B$ are locally Lipschitz continuous, and the quantity $|\mathfrak{x}(\tau) - \mathfrak{x}(0)|$ for any fixed $\mathfrak{x}(0) \in D$ is locally bounded given $\mathfrak{x}(\tau) = \mathfrak{x}(0) + \int_0^\tau \tilde{f}(\mathfrak{x}(s), u) ds$ and $\tau \in [0, 0.05)$. It is clear that both

$$|L_f V(\mathfrak{x}(\tau)) + L_g V(\mathfrak{x}(\tau)) u + V(\mathfrak{x}(\tau)) - L_f V(\mathfrak{x}(0)) + L_g V(\mathfrak{x}(0)) u + V(\mathfrak{x}(0))|$$

and

$$|L_f B(\mathfrak{x}(\tau)) + L_g B(\mathfrak{x}(\tau)) u - L_f B(\mathfrak{x}(0)) + L_g B(\mathfrak{x}(0)) u|$$

are locally bounded. We set the bound for both of the above quantities to be ς, such that for any $\mathfrak{u} = \kappa(\mathfrak{x}) \in \mathfrak{K}(\mathfrak{x})$ and $\mathfrak{x} \in D$, we have the Lyapunov-barrier conditions satisfied in the flow, i.e.,

$$L_f V(\mathfrak{x}) + L_g V(\mathfrak{x}) u + V(\mathfrak{x}) \leq 0 \text{ and } L_f B(\mathfrak{x}) + L_g B(\mathfrak{x}) \mathfrak{u} \geq 0, \ \forall \mathfrak{x} \in C.$$

This in turn guarantees the reach-avoid-stay property of the system.

In the numerical experiments, we empirically set $\varsigma = 0.07$ and impose an extra set of constraints for the control input as in (2.40). The $\mathfrak{u}$ is selected such that (2.39) is minimized for each $\mathfrak{z} \in D$. The results are shown in Figure 2.10 and 2.11. It can be seen that given the robust correction in the decision making, the high-gain controller can still fulfill the goal apart from the intensive chattering trajectory along the Φ direction in between $t = 20$ and $t = 80$.

2.5.5 Proofs of Results

Proof of Lemma 2.5.15

Proof. Note that, implicitly, every reference hybrid arc ϕ should exist for hybrid time $t+j \geq T$. We suppose the total number of jumps before T is N, thereby $N \in [0, T]$. Now, set

$$\psi(t', j') = \phi(t', j') + \left(1 - \frac{t' + j'}{T}\right) \cdot (\psi(0,0) - \phi(0,0))$$

for all $t' + j' \in [0, T]$. Then the constructed ψ satisfies (2.44) and

$$|\psi(t', j') - \phi(t', j')| \leq |\psi(0,0) - \phi(0,0)| \leq r.$$

In particular, for $N > 0$, the total distance of $|\psi(t', j') - \phi(t', j')|$ created by jumps should be bounded by $\frac{Nr}{T}$, and for each jump and at the end point $t' \in I^{j'}$, we have

$$|\psi(t', j'+1) - \phi(t', j'+1)| \leq \frac{r}{T}.$$

Similarly, suppose $N < T$ (i.e., flows exist), then for all $t' \in I^{j'}$, we have

$$\left|\dot{\psi}(t', j') - \dot{\phi}(t', j')\right| \leq \frac{T - N}{T(T-N)} |\psi(0,0) - \phi(0,0)| = \frac{r}{T}.$$

We show that this ψ is a solution to $\mathbf{H}_\vartheta$. On the flow set C_ϑ, let $\widehat{C} := C_\vartheta \cap (K + r\overline{\mathbb{B}})$. Then for $x \in \widehat{C}$, for all $t' \in I^{j'}$, we have

$$\begin{aligned}
\dot{\psi}(t', j') \subseteq & \dot{\phi}(t', j') + \frac{r}{T}\overline{\mathbb{B}} \subseteq F(\phi(t', j')) + \left(\vartheta' + \frac{r}{T}\right)\overline{\mathbb{B}} \\
\subseteq & F(\psi(t', j')) + \left(\vartheta' + \frac{r}{T}\right)\overline{\mathbb{B}} + L_{\widehat{C}}|\psi(t', j') - \phi(t', j')|\overline{\mathbb{B}} \\
\subseteq & F(\psi(t', j')) + \left(\vartheta' + \frac{r}{T} + rL_{\widehat{C}}\right)\overline{\mathbb{B}} \\
\subseteq & F(\psi(t', j')) + \left(\vartheta' + \frac{r}{\tau} + rL_{\widehat{C}}\right)\overline{\mathbb{B}},
\end{aligned}$$

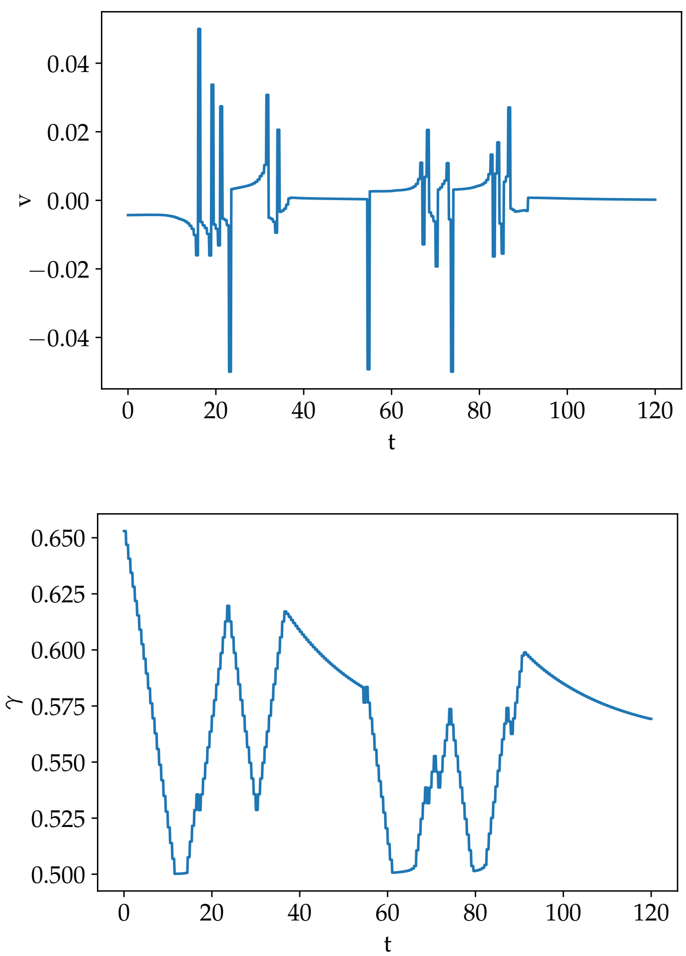


Figure 2.10: Control signal v and γ solved by the quadratic programming for the hybrid conversion of (2.38).

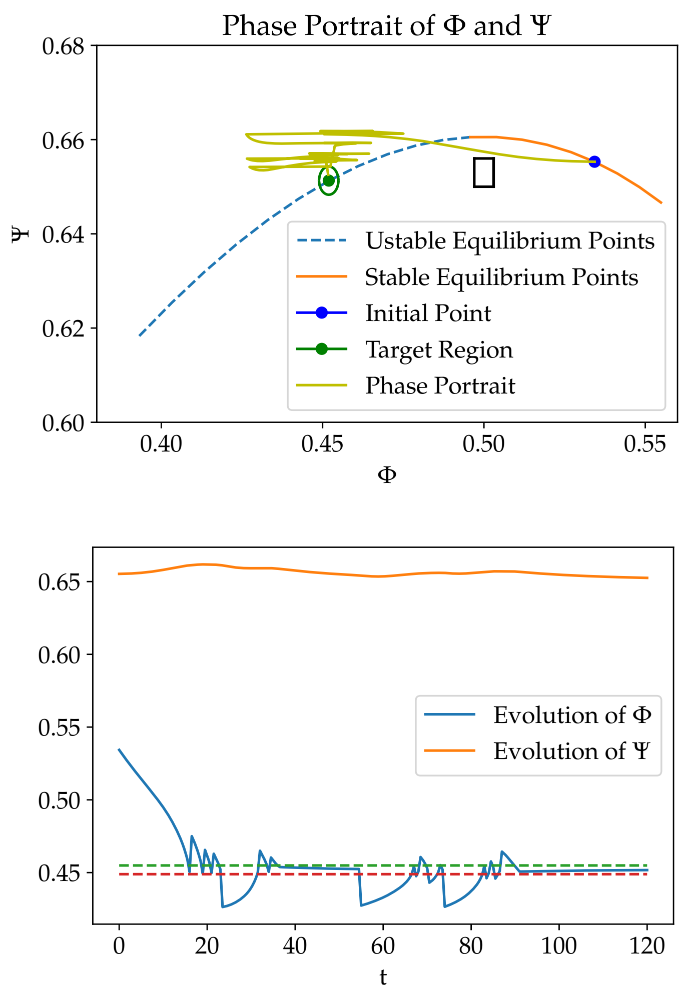


Figure 2.11: Phase portraits and evolution of Φ and Ψ given the sample-and-hold control signal υ and γ.

where the constant $L_{\widehat{C}}$ in line 4 of the above inclusions is obtained by Lemma 2.5.14.

Similarly, for each end point t' of some $I^{j'}$, the jump should satisfy

$$\psi(t', j'+1) \subseteq G(\psi(t', j')) + \left(\vartheta' + \frac{r}{\tau} + rL_{\widehat{D}}\right)\overline{\mathbb{B}},$$

where $\widehat{D} := D_\vartheta \cap (K + r\overline{\mathbb{B}})$ and $L_{\widehat{D}}$ is obtained based on Lemma 2.5.14 for G.

Now we are able to bound the above by ϑ such that ψ is a solution to $\mathbf{H}_\vartheta$. The r can be selected accordingly such that all of the followings are satisfied: $r \leq \vartheta - \vartheta'$, $\frac{2r}{\tau} + \vartheta' + 2rL_{\widehat{C}} \leq \vartheta$ and $\frac{2r}{\tau} + \vartheta' + 2rL_{\widehat{D}} \leq \vartheta$.

Note that the above choice of r should work for every possible case: the extreme cases when only flows ($N = 0$) or jumps ($N = T$) happen, as well as the mixed flow/jump case ($N \in (0, T)$). □

Proof of Proposition 2.5.19:

To prove Proposition 2.5.19, we need the following lemma.

Lemma 2.5.29. *Let the hypothesis in Proposition 2.5.19 be satisfied. Let $x_0 \in A \subseteq \mathfrak{I}$ be fixed. Then for every $\epsilon > 0$ and $\tau \geq 0$, there exists $\kappa > 0$ such that, for every solution $\phi \in \mathfrak{S}^\vartheta_{\mathbf{H}}(x_0 + \kappa\overline{\mathbb{B}})$ there exists a solution $\psi \in \mathfrak{S}^\vartheta_{\mathbf{H}}(x_0)$ such that ϕ and ψ are (τ, ϵ)-close.*

Proof. Without loss of generality, we can assume that ϵ and τ are arbitrarily small such that $x_0 + \kappa\overline{\mathbb{B}} \subseteq \mathfrak{I}$. Note that by the reach-avoid-stay property of $\mathbf{H}_\vartheta$, the solution starting from $\mathfrak{I}$ should be either eventually bounded or complete. Suppose the statement were to fail, then for some arbitrarily small ϵ and τ, for each $n \in \mathbb{N}$ and $\phi_n \in \mathfrak{S}^\vartheta_{\mathbf{H}}(x_0 + 1/n\overline{\mathbb{B}})$ with $x_0 + 1/n\overline{\mathbb{B}} \subseteq \mathfrak{I}$, there exists no solution $\psi \in \mathfrak{S}^\vartheta_{\mathbf{H}}(x_0)$ is (τ, ϵ)-close to ϕ_n. However, $\{\phi_n\}_n$ is (locally eventually) bounded. By [71, Theorem 6.1], we can extract a subsequence, still denoted by $\{\phi_n\}_n$, that is graphically convergent to some $\phi \in \mathfrak{S}^\vartheta_{\mathbf{H}}(x)$. This implies that there exists some sufficiently large n such that ϕ_n and ϕ are (τ, ϵ)-close as a consequence of graphical convergence of bounded sequences, which leads to a contradiction. □

Proof of Proposition 2.5.19: By Definition 2.5.9, there exists some $T \geq 0$ such that $\mathcal{R}^{\geq T}_\vartheta(\mathcal{X}_0) \subseteq \mathfrak{I}$. It is easy to verify that for any $x \in \mathcal{R}^{\geq T}_\vartheta(\mathcal{X}_0) \subseteq \mathfrak{I}$, $x \in \mathfrak{I}$, and for all $\phi \in \mathfrak{S}^\vartheta_{\mathbf{H}}(x)$, we have $\phi(t, j) \in \mathcal{R}^{\geq T}_\vartheta(\mathcal{X}_0)$ for all $(t, j) \in \mathrm{dom}(\phi)$. This shows that A is nonempty. The forward pre-invariance is verified by setting $T = 0$ and $\mathcal{X}_0 \subseteq \mathfrak{I}$.

It suffices to show the closedness of A, which will imply the compactness due to the boundedness assumption on $\mathfrak{I}$. Pick any sequence $\{x_n\}_n \subseteq A$ such that $x_n \to x$. Then for all $\phi_n \in \mathfrak{S}_{\mathbf{H}}^{\vartheta}(x_n)$, we have $\mathrm{rge}(\phi_n) \subseteq \mathfrak{I}$ and hence bounded. Suppose that $x \notin A$, then there exists a $\phi \in \mathfrak{S}_{\mathbf{H}}^{\vartheta}(x)$ such that $\phi(t,j) \notin \mathfrak{I}$ for some $(t,j) \in \mathrm{dom}(\phi)$. Select sufficiently small τ and ϵ as in Lemma 2.5.29, then there exists some $\kappa = \kappa(\tau, \epsilon) > 0$ and sufficiently large $n \in \mathbb{N}$ (with $x \in x_n + \kappa\overline{\mathbb{B}}$) such that $\phi \in \mathfrak{S}_{\mathbf{H}}^{\vartheta}(x_n + \kappa\overline{\mathbb{B}})$. However, by Lemma 2.5.29, there exists a solution $\phi_n \in \mathfrak{S}_{\mathbf{H}}^{\vartheta}(x_n)$ such that ϕ and ϕ_n are arbitrarily (τ, ϵ)-close. Since $\mathbb{R}^n \setminus \mathfrak{I}$ is open, we have $\phi_n(s,j) \in \mathbb{R}^n \setminus \mathfrak{I}$ for some $s + j \leq \tau$ with $|t - s| < \epsilon$. This contradicts the forward pre-invariance property of ϕ_n. Hence, $x \in A$ and A depicts compactness. □

Proof of Lemma 2.5.22

Proof. The first claim can be easily verified by Definition 2.5.8 and 2.5.21. The forward pre-invariance of (2) comes after Definition 2.5.4. It suffices to show that $\widehat{\mathfrak{B}}_{\vartheta}(A)$ is also open given U is closed.

Suppose the opposite, then pick any $x \in \widehat{\mathfrak{B}}_{\vartheta}(A)$, there exists a sequence of points $\{x_n\}_n \subseteq \mathfrak{B}_{\vartheta} \setminus \widehat{\mathfrak{B}}_{\vartheta}(A)$ with $x_n \to x$. Note that for each n, there exist $\phi_n \in \mathfrak{S}_{\mathbf{H}}^{\vartheta}(x_n)$ that are either bounded or complete with convergence to A. Either way, we have ϕ_n bounded for each n by Proposition 2.5.10. By [71, Theorem 6.1], we can extract a subsequence, still denoted by $\{\phi_n\}_n$, that is graphically convergent. By Assumption 2.5.2, the limit satisfies $\phi \in \mathfrak{S}_{\mathbf{H}}^{\vartheta}(x)$, and is again either bounded or complete with convergence to A.

Suppose the graphical limit ϕ is complete. Let $r > 0$ be given by condition (2) from Definition 2.5.6. Pick $\epsilon < r$ such that $(A + \epsilon\overline{\mathbb{B}}) \cap U = \emptyset$. Choose $\eta_\epsilon \leq \epsilon$ according to condition (1) of Definition 2.5.6 by the uniform stability. Then there exists a $T = T(\eta_\epsilon) > 0$ such that

$$|\widehat{\phi}(t,j)|_A \leq \eta_\epsilon$$

for all $\widehat{\phi} \in \mathfrak{S}_{\mathbf{H}}^{\vartheta}(x)$ and $t + j \geq \mathrm{dom}(\widehat{\phi})$ with $t + j \geq T$. Note that the reachable set $\mathcal{R}_{\vartheta}^{\leq T}(x)$ is compact and we can select some $\epsilon' < \epsilon/2$ such that $(\mathcal{R}_{\vartheta}^{\leq T}(x) + \epsilon'\overline{\mathbb{B}}) \cap U = \emptyset$. Let $\epsilon'' = \min\{\epsilon', \eta_\epsilon\}$. As the consequence of graphical convergence, there exists a sufficiently large $n \in \mathbb{N}$ such that for each $\tau > 0$, the solution ϕ_n and ϕ are (τ, ϵ'')-close. No matter $\tau \geq T$ or $\tau < T$, we can verify by the choice of ϵ'' and n that $\phi_n(t,j) \in (\mathcal{R}_{\vartheta}(x) + \epsilon'\overline{\mathbb{B}}) \cup (A + \epsilon\overline{\mathbb{B}})$ for all $(t,j) \in \mathrm{dom}(\phi_n)$, which contradicts the property of $\phi_n \in \mathfrak{S}_{\mathbf{H}}^{\vartheta}(x_n)$ with $x_n \in \mathfrak{B}_{\vartheta}(A) \setminus \widehat{\mathfrak{B}}_{\vartheta}(A)$. For the case that ϕ is bounded, we can proceed and show the contradiction by a similar way. Combining the above, we have $\widehat{\mathfrak{B}}_{\vartheta}(A)$ is open. □

Proof of Theorem 2.5.23

Proof. (1) $\implies$ (2): By [71, Corollary 7.33], the existence of smooth Lyapunov-like function holds on forward pre-invariant open subsets of $\mathfrak{B}_\vartheta(A)$. The result follows immediately by Lemma 2.5.22 for $\mathfrak{D} = \widehat{\mathfrak{B}}_\vartheta(A)$.

(2) $\implies$ (1): By a standard Lyapunov argument for hybrid systems, we are able to show that for any ϕ to $\mathbf{H}_\vartheta$ with $\phi(0,0) \in \mathfrak{D}$,

$$V(\phi(t,j)) \leq V(\phi(0,0))e^{-(t+j)/3},$$

which implies the forward pre-invariance of $\mathfrak{D}$ and the UpAS property of any ϕ starting within $\mathfrak{D}$. Now that $\mathfrak{D} \cap U = \emptyset$, for any $\phi \in \mathfrak{S}^\vartheta_{\mathbf{H}}(\mathcal{X}_0)$ with $\mathcal{X}_0 \subseteq \mathfrak{D}$, we have $\mathrm{rge}(\phi) \cap U = \emptyset$, which implies the safety. □

Proof of Theorem 2.5.27

Proof. We only show the sufficient part, the converse construction can be argued in a similar way to Proposition 2.3.11. We start with the stability. Note that by a standard Lyapunov argument for hybrid systems, the existence of V with condition (1) is sufficient to guarantee the UpAS property of A for $\mathbf{H}_\vartheta$. The condition (2)-(4) of B intends to separate the set S and U, such that all the solutions starting from S will stay within it.

We show formally the mechanism of B. Suppose the opposite, then there exists some $x \in S$, a solution $\phi \in \mathfrak{S}^\vartheta_{\mathbf{H}}(x)$, and some hybrid time such that $B(\phi(t,j)) < 0$. Then we are able to define a finite time $\tau := \sup\{t+j \geq 0 : \phi(t,j) \in S\}$. As a consequence, there exist t' and j' with $t'+j' = \tau$ such that $B(\phi(t',j')) = 0$. Note that since the ϕ exists till $B(\phi(t,j)) < 0$. For a small perturbation ϵ of t', ϕ is either a flow such that $t'+\epsilon \in I^{j'}$, or triggers a jump such that $t'+\epsilon \in I^{j'+1}$. For the first case, since $\mathfrak{D}$ is open, for arbitrary $\epsilon > 0$, we still have $\phi(t'',j') \in C_\vartheta \cap \mathfrak{D}$ for almost $t'' \in [t', t'+\epsilon]$. Thus, we have $\dot{B}(\phi(t'',j')) = \nabla B(\phi(t'',j)) \cdot f \geq 0$ for all $f \in F(x) + \vartheta\overline{\mathbb{B}}$ and almost all $t'' \in [t', t'+\epsilon]$, which means $t''+j' \geq \tau$ but $\phi(t'',j') \in S$. This contradicts the definition of τ. For the second case, by a similar argument, we have $B(\phi(t',j'+1)) \geq 0$, which also leads to a contradiction. Hence S is forward pre-invariant. This verifies condition (2) of Definition 2.5.8. A direct consequence of B is that, for any $\mathcal{X}_0 \subseteq \mathfrak{D}$, any solution $\phi \in \mathfrak{S}^\vartheta_{\mathbf{H}}(\mathcal{X}_0)$ stays within S and hence $\mathfrak{D}$. Condition (1) guarantees that $\phi \in \mathfrak{B}_\vartheta(A)$. This verifies condition (1) of Definition 2.5.8. □

2.6 Summary

In this chapter, we started with a quick review of barrier functions and the associated certificates to guarantee set invariance. A slight modification of barrier conditions were made in Section 2.2.1, based on which the CBF was proposed and verified to sufficiently guarantee the set safety of the controlled trajectories.

We then proved two converse Lyapunov-barrier function theorems for nonlinear systems satisfying either asymptotic stability with a safety constraints or a reach-avoid-stay type specification. In the former case, we show that a smooth Lyapunov-barrier function can be defined on the entire set of initial conditions from which asymptotic stability with a safety constraint can be satisfied. For the latter, we establish a converse theorem via a robustness argument. It is shown by example that the statement cannot be strengthened without additional assumptions. We further extend the results to establish converse control Lyapunov-barrier functions for systems with control inputs. The focus of the current chapter is on converse Lyapunov-barrier functions, applying which we make a quick extension to converse control Lyapunov-barrier function. However, we only considered an additive measurable disturbance in the right-hand side of the dynamical systems for the purpose of establishing converse Lyapunov-barrier results. In addition, similar to other converse Lyapunov theorems, the existence results are not constructive.

We investigated the effectiveness in a case study of jet engine compressor control problem. In the control problem, we concern the parameter as a time-varied signal to be decided along with the control signal v for the state variable Φ. It is shown that the sufficient conditions on v and γ generated by Lyapunov method can be flexibly embedded into a quadratic programming framework with a minimum energy cost. In contrast to formal methods, which fail to handle non-uniform speed of dynamics determined by $\gamma(t)$ using the existing tool boxes, Lyapunov methods analytically characterize the topological structure on the solutions w.r.t. the reach-avoid-stay specifications without considering local dynamics.

We finally showed that under mild conditions, the connection can be made, via a robustness argument, between stability with safety guarantees and reach-avoid-stay specifications for robust hybrid systems driven by differential and difference inclusions. We further extended the Lyapunov-barrier function theorems to robust hybrid systems that satisfy stability with safety guarantees and reach-avoid-stay specifications. Under the concept of solutions to hybrid systems, as well as natural requirements on the compactness of target set and the set of initial conditions, we showed that the existence of Lyapunov-barrier functions is necessary to the two specifications of our interests. It is interesting to compare with the latest converse theorems on barrier functions for systems driven by differential inclusions [64]. The mentioned reference provides a construction based on time-to-impact functions w.r.t. the robust reachable sets, which in turn to be qualified as barrier functions. In terms of safety, this method allows us to consider relaxations for the topological set-ups of target sets and the set of initial conditions, however, at the cost of not relying on the existing converse Lyapunov theorems and possibly sacrificing the smoothness of barrier functions.

Chapter 3

Lyapunov-Barrier Approaches for Verification and Control of Stochastic Systems

In this chapter, we extend the Lypanov-Barrier approaches to verification and control synthesis of stochastic systems. We focus on probabilistic invariance/safety and reachability related specifications for continuous-time stochastic systems modelled by SDE driven by Brownian motions.

Considering the appearance of noises with Itô diffusion, instead of direct requirements on trajectory behaviors in the state space *per se*, a proper specification is to specify a probability of sample paths satisfying certain state-space behaviors, namely probabilistic specifications. This turns out to fit the culture of stochastic dynamical systems: when it comes to observations in the state space, we usually concern how probability laws on the canonical space distribute the corresponding weak solutions.

As for verification and control synthesis of probabilistic stability-safety type problems, it appears more challenging. Authors in [102, 35, 117, 52, 51] applied abstract models, such as IMC and BMDP, on discrete-time continuous-state stochastic systems to compute an inclusion of the real satisfying probability and synthesize controllers for probabilistic specifications (including probabilistic reachability on an infinite horizon). Works in [143, 164, 88] characterized value functions for reachability/reach-avoid problems in discrete-time continuous-state stochastic systems and applied dynamic programming for synthesizing optimal controllers. Authors in [53] developed a weak dynamic programming principle for the value functions of probabilistic reach-avoid specifications in continuous-time continuous-state stochastic systems, which provides compatibility for non-almost-sure probabilistic requirements. It remains a fundamental challenge to overcome the curse of dimensionality in discretization-based approaches for verification and control synthesis.

Since small perturbations should necessarily be taken into account due to reasons such as modelling uncertainties and measurement errors of the state, robust analysis provides guarantees in a worst-case scenario. Despite the current theme of regarding 'inaccuracy' from the computation of probability measures as the 'uncertainty' [102, 35, 117, 52], to make a closer analogy of the deterministic case, we consider uncertainties as a result of perturbed stochastic systems which create an inclusion of solutions. A robust satisfaction of a probabilistic specification in a perturbed stochastic system is then interpreted as follows: the solution process measured in the correspondingly worst but accurate probability law still satisfies the probabilistic specification. The work [155] demonstrated the robust Lyapunov-stability for discrete-time stochastic systems with perturbations. The authors in [137] considered the same type of systems and developed a robust algorithm to guarantees practical convergence to a Nash equilibrium in non-cooperative games. Continuous-time stochastic differential inclusions are also well studied [93, 94, 118].

Considering this complexity issue as well as the possible appearance of extra uncertainties, this chapter formulates stochastic Lyapunov and barrier functions to deal with sufficient conditions for robust probabilistic invariance and reachability related specifications.

Conventions for Notation:

For any stochastic processes $\{X_t\}_{t\geq 0}$ we use the shorthand notation $X := \{X_t\}_{t\geq 0}$. For any stopped process $\{X_{t\wedge\tau}\}_{t\geq 0}$, where τ is a stopping time, we use the shorthand notation X^τ. We denote the Borel σ-algebra of a set by $\mathscr{B}(\cdot)$ and the space of all probability measures on $\mathscr{B}(\cdot)$ by $\mathfrak{P}(\cdot)$. Given a probability space $(\mathbf{\Omega}, \mathcal{F}, \mathbf{P})$, we denote by $\|\cdot\|_1 := \mathbf{E}|\cdot|$ the $\mathcal{L}^1$-norm for $\mathbb{R}^n$-valued random variables, and let $\mathcal{B} := \{X : \mathbb{R}^n\text{-valued random variable with } \|X\|_1 < 1\}$.

3.1 Preliminaries

Before proceeding, we briefly introduce perturbed SDEs and the solution concept. We focus on the introduction of control-free cases, and the systems with controls should be similar.

Control-free system dynamics

Consider the following form of perturbed SDEs:

$$dX_t = f(X_t)dt + b(X_t)dW_t + \vartheta\xi(t)dt, \tag{3.1}$$

where W is an m-dimensional standard Wiener process; the process $\xi : \mathbb{R}_{\geq 0} \to \overline{\mathcal{B}}$ is a measurable signal independent with W, whose marginals $\xi(t) \in \overline{\mathcal{B}}$ are independent random variables with unknown distributions; $f : \mathbb{R}^n \to \mathbb{R}^n$ is a nonlinear vector field; $b : \mathbb{R}^n \to \mathbb{R}^{n \times m}$ is a smooth mapping. For future references, we denote systems driven by SDE (3.1) by $\mathscr{S}_\vartheta$, of which the ϑ represents the intensity of perturbations. Similar to Chapter 2, we denote by $\mathscr{S}$ the control-free stochastic system without extra perturbations.

Remark 3.1.1. *We can treat $\xi : \mathbb{R}_{\geq 0} \to \overline{\mathbb{B}}_1$ as a special case, where the marginals are independent point-mass perturbations that possess Dirac measures δ_x for unknown $x \in \overline{\mathbb{B}}_1$.*

Assumption 3.1.2. *We make the standing assumptions on the regularity of the system $\mathcal{S}_\vartheta$ for the rest of this chapter:*

(1) The mappings f, b satisfy local Lipschitz continuity.

(2) The eigenvalues $\lambda_i[(bb^T)(x)]$ of the matrix $bb^T(x)$ for $i = 1, 2, \cdots, n$ satisfy

$$\sup_{x \in \mathbb{R}^n} \min_{i=1,2,\cdots,n} \lambda_i[(bb^T)(x)] > 0.$$

Definition 3.1.3 (Solution concepts). *The system $\mathscr{S}_\vartheta$ admits*

(1) a strong solution on a given filtered probability space $(\Omega^\dagger, \mathscr{F}^\dagger, \{\mathscr{F}_t^\dagger\}, \mathbb{P}^\dagger)$, where the given Wiener process W is defined, if there exists an adapted process X satisfying the SDE (3.1) for any $\xi : \mathbb{R}_{\geq 0} \to \overline{\mathcal{B}}$;

(2) a weak solution if there exists a filtered probability space $(\Omega^\dagger, \mathscr{F}^\dagger, \{\mathscr{F}_t^\dagger\}, \mathbb{P}^\dagger)$ with a pair (X, W) of adapted stochastic processes, such that W is a Wiener process and X solves the SDE (3.1) for any $\xi : \mathbb{R}_{\geq 0} \to \overline{\mathcal{B}}$.

Note that from a modelling point of view, we usually do not specify in *a priori* a Wiener process [131]. In addition, for the purpose of verifying dynamical behaviors in probability laws, i.e. the probabilistic properties in the state space, it is not necessary to restrict ourselves to a specified probability space. For a system $\mathscr{S}_\vartheta$, we consider the weak sense of solutions and denote by $\mathfrak{S}_\vartheta(x, W)$ the set of all weak solutions with $X_0 = x$ a.s. for a given $x \in \mathbb{R}^n$. Likewise, for a given set $K \subseteq \mathbb{R}^n$, let $\mathfrak{S}_\vartheta(K, W)$ denote the set of all weak solutions with any initial distribution on $(K, \mathscr{B}(K))$.

Remark 3.1.4. *For $\vartheta \equiv 0$, the solution set $\mathfrak{S}_\vartheta(x, W)$ (resp. $\mathfrak{S}_\vartheta(K, W)$) becomes a singleton.*

Canonical space

We have a Wiener process W defined on some probability space $(\Omega^\dagger, \mathscr{F}^\dagger, \mathbb{P}^\dagger)$ for each weak solution. We transfer information to the canonical space, which gives us the convenience to study the law of the solution processes as well as the probabilistic behavior in the state space.

Definition 3.1.5 (Canonical space)**.** *Define* $\mathbf{\Omega} := C([0,\infty);\mathbb{R}^n)$ *with coordinate process* $\mathfrak{X}_t(\varpi) := \varpi(t)$ *for all* $t \geq 0$ *and all* $\varpi \in \mathbf{\Omega}$*. Define* $\mathcal{F}_t := \sigma\{\mathfrak{X}_s,\ 0 \leq s \leq t\}$ *for each* $t \geq 0$*, then the smallest* σ*-algebra containing the sets in every* $\mathcal{F}_t$*, i.e.* $\mathcal{F} := \bigvee_{t\geq 0} \mathcal{F}_t$*, turns out to be same as* $\mathscr{B}(\mathbf{\Omega})$*. For each* $X \in \mathfrak{S}_\vartheta(\mathbb{R}^n, W)$*, the induced measure (probability law)* $\mathbf{P}^X \in \mathfrak{P}(\mathbf{\Omega})$ *on* $\mathcal{F}$ *is such that* $\mathbf{P}^X[A] = \mathbb{P}^\dagger \circ X^{-1}(A)$ *for every* $A \in \mathscr{B}(\mathbf{\Omega})$*. We also denote* $\mathbf{E}^X$ *by the associated expectation operator w.r.t.* $\mathbf{P}^X$*. The canonical space for* X *is given as* $(\mathbf{\Omega}, \mathcal{F}, \mathbf{P}^X)$*.*

For the special case when $\vartheta \equiv 0$, we specifically denote $\mathbf{P}^x$ by the probability law of the weak solution X given $X_0 = x$ a.s. for some $x \in \mathbb{R}^n$. We also denote $\mathbf{E}^x$ by the associated expectation operator w.r.t. $\mathbf{P}^x$. Likewise, for any arbitrary initial distribution $\mu \in \mathfrak{P}(\Omega)$, the unique probability law is given as

$$\mathbf{P}^\mu[\,\cdot\,] := \int_{\mathbb{R}^n} \mathbf{P}^x[\,\cdot\,]\mu(dx).$$

Since we are unclear about the base probability space $(\Omega^\dagger, \mathscr{F}^\dagger, \mathbb{P}^\dagger)$ where the Wiener process W is defined, we prefer to transfer information to the canonical space, which gives us the convenience to study the probability law of the weak solutions and the probabilistic behavior in the state space.

Definition 3.1.6 (Weak convergence of measures and processes))**.** *Given any separable metric space* (E,ρ)*, a sequence of* $\{\mathbf{P}^n\}$ *of* $\mathfrak{P}(E)$ *is said to weakly converge to* $\mathbf{P} \in \mathfrak{P}(E)$*, denoted by* $\mathbf{P}^n \rightharpoonup \mathbf{P}$*, if for all* $f \in C_b(E)$ *we have*

$$\lim_{n\to\infty} \int_E f\ d\mathbf{P}^n = \int_E f\ d\mathbf{P}.$$

A sequence $\{X^n\}$ *of continuous processes* X^n *with law* $\mathbf{P}^n$ *is said to weakly converge (on* $[0,T]$*) to a continuous process* X *with law* $\mathbf{P}^X$*, denoted by* $X^n \rightharpoonup X$*, if for all* $f \in C_b(C([0,T];\mathbb{R}^n))$ *we have* $\lim_{n\to\infty} \mathbf{E}^n[f(X^n)] = \mathbf{E}^X[f(X)]$*.*

Remark 3.1.7. *We work on this weak topology to study a family of processes. More details on the weak topology as well as metric spaces of probability measures can be found in Appendix E. Note that the first part of Definition 3.1.6 is also provided in Appendix E. The weak convergence of a sequence of processes is dual to the weak convergence of their probability laws. Both concepts can be used interchangeably.*

Generator and characteristic operator

It is fundamental for many applications that we can associate a second order partial differential operator to a stochastic process [131]. We introduce the following concepts for system $\mathscr{S}$.

Definition 3.1.8 (Infinitesimal generator). *The infinitesimal generator $\mathscr{A}$ of $\mathscr{S}$[1] is defined by*

$$\mathscr{A}h(x) = \lim_{t\downarrow 0} \frac{\mathbf{E}^x[h(X_t)] - h(x)}{t}, \quad x \in \mathbb{R}^n.$$

We further denote $\operatorname{dom}(\mathscr{A})$ by the set of functions for which the limit exists for all $x \in \mathbb{R}^n$.

Proposition 3.1.9. *[131] Let X be driven by $\mathscr{S}$. Then, if $h \in C_b^2(\mathbb{R}^n)$, we have $h \in \operatorname{dom}(\mathscr{A})$ and*

$$\mathscr{A}h(x) = \nabla h(x) \cdot f(x) + \frac{1}{2} \operatorname{tr}\left[(bb^T)(x) \cdot h_{xx}(x)\right], \tag{3.2}$$

where $h_{xx} = (h_{x_i x_j})_{n\times n}$ and $\operatorname{tr}[\,\cdot\,]$ denotes the trace.

Definition 3.1.10 (Characteristic operator). *The characteristic operator $\mathfrak{L}$ of $\mathscr{S}$ is defined by*

$$\mathfrak{L}h(x) = \lim_{N\downarrow 0} \frac{\mathbf{E}^x[h(X_{\tau_N})] - h(x)}{\mathbf{E}^x[\tau_N]}, \quad x \in \mathbb{R}^n,$$

where the N's are sequence of open sets N_k decreasing to the the point x, i.e., $N_{k+1} \subset N_k$ and $\cap_k N_k = \{x\}$, and $\tau_N = \inf\{t > 0,\ X_t \notin N\}$. We further denote $\operatorname{dom}(\mathfrak{L})$ by the set of functions for which the limit exists for all $x \in \mathbb{R}^n$.

Proposition 3.1.11. *[131] Let X be driven by $\mathcal{S}_\vartheta$ with $\vartheta = 0$. Then, if $h \in C^2(\mathbb{R}^n)$, we have $h \in \operatorname{dom}(\mathfrak{L})$ and*

$$\mathfrak{L}h(x) = \nabla h(x) \cdot f(x) + \frac{1}{2} \operatorname{tr}\left[(bb^T)(x) \cdot h_{xx}(x)\right], \tag{3.3}$$

where $h_{xx} = (h_{x_i x_j})_{n\times n}$ and $\operatorname{tr}[\,\cdot\,]$ denotes the trace.

Theorem 3.1.12 (Dynkin's formula). *Let $h \in C_b^2(\mathbb{R}^n)$. Suppose τ is a stopping time such that $\mathbf{E}^x[\tau] < \infty$. Then,*

$$\mathbf{E}^x[h(X_\tau)] = h(x) + \mathbf{E}^x\left[\int_0^\tau \mathscr{A}h(X_s)\right]. \tag{3.4}$$

[1] Formally, $\mathscr{A}$ is the infinitesimal generator of the transition semigroup (see Appendix B for details) for system $\mathscr{S}$.

Remark 3.1.13. *Note that there is a subtle but crucial difference between $\mathscr{A}$ and $\mathfrak{L}$ in terms of the valid test functions, albeit the two operators possess the exact same form. However, if τ is the first exit time of a bounded set, then the Dynkin's formula can be used as*

$$\mathbf{E}^x[h(X_\tau)] = h(x) + \mathbf{E}^x\left[\int_0^\tau \mathfrak{L}h(X_s)\right], \quad h \in C^2(\mathbb{R}^n). \tag{3.5}$$

We do not usually distinguish the concept of $\mathscr{A}$ and $\mathfrak{L}$ when our goal is to use Dynkin's formula on a bounded domain with its associated first exit time.

We extend the above notions to systems $\mathscr{S}_\vartheta$ with $\vartheta \neq 0$. For each $d \in \overline{\mathbb{B}}$, we denote by $\mathscr{A}_d$ the generator of $\mathscr{S}_\vartheta$ and by $\mathfrak{L}_d$ the characteristic operator of $\mathscr{S}_\vartheta$. Furthermore, we have

$$\mathscr{A}_d h(x) = \nabla h(x) \cdot (f(x) + \vartheta d) + \frac{1}{2}\operatorname{tr}\left[(bb^T)(x) \cdot h_{xx}(x)\right], \ h \in C_b^2(\mathbb{R}^n)$$

and

$$\mathfrak{L}_d h(x) = \nabla h(x) \cdot (f(x) + \vartheta d) + \frac{1}{2}\operatorname{tr}\left[(bb^T)(x) \cdot h_{xx}(x)\right], \ h \in C^2(\mathbb{R}^n).$$

Definition 3.1.14 (Clarification of Notations)**.** *Since $\mathscr{A}$ and $\mathfrak{L}$ (resp. $\mathscr{A}_d$ and $\mathfrak{L}_d$) have the same form apart from the domain of test functions, to keep notation succinct, we do not use $\mathscr{A}$ (resp. $\mathscr{A}_d$) to this end unless specially emphasized. Readers are able to justify the meaning of the operators based on the contexts.*

Regularities of solutions

Since for each measurable signal ξ, the corresponding martingale problem is well posed under Assumption 3.1.2, by Markovian selection theorems [54, Theorem 5.19, Chap 4], the unique solution $\mathbf{P}^X$ to the martingale problem also makes the associated weak solution (X, W) Markovian.

We also do not exclude explosive solutions[2] in general. As a matter of fact, under the assumptions on f and g, for any $X \in \mathfrak{S}_\vartheta(\mathbb{R}^n, W)$, there exists a stopping time τ_{ex} such that $\mathbf{P}^X[\tau_{\text{ex}} > 0] = 1$. Each weak solution X exists locally for all $t \in (0, \tau_{\text{ex}})$ such that $\tau_{\text{ex}} = \infty$ (exists globally) or $\lim_{t \nearrow \tau_{\text{ex}}} \|X_t\| = \infty$ (explodes within finite time).

[2]See [16, Section 5.5] for details.

Systems with controls

Let the set of control $\mathcal{U}$ be given. Suppose that $\mathcal{U}$ is complete and separable metric spaces. Consider a nonlinear system of the form

$$dX_t = f(X_t)dt + g(X_t)\mathfrak{u}_t dt + b(X_t)dW_t + \vartheta\xi(t)dt, \tag{3.6}$$

where the mapping $g : \mathbb{R}^n \to \mathbb{R}^{n\times p}$ is smooth; $\mathfrak{u} : \mathbb{R}_{\geq 0} \to \mathcal{U}$ is a locally bounded non-randomized control signal, whilst the other notation remains the same.

In practice, step controls (see Definition C.2.1) are frequently used. It can be verified that the probability law of the process $(X, \mathfrak{u})$ can be constructed using a sequential definition as in the discrete-time case (see details in Appendix C.1 and C.2) given step controls.

For a given control process $\mathfrak{u}$ with step controls at sufficiently dense sampling time, and an arbitrary initial distribution $\mu \in \mathfrak{P}(\mathcal{X})$, we denote by $X^{\mathfrak{u}}$ the controlled process and by $\mathbf{P}^{\mu,\mathfrak{u}}$ the law of $X^{\mathfrak{u}}$. Note that, for each $t > 0$, the random variable $X_t^{\mathfrak{u}}$ is determined by $\mathfrak{u}_s$ and $X_s^{\mathfrak{u}}$ for $s \in [0, t)$.

In terms of control synthesis, we focus on deterministic Markov policies κ (see Definition C.3) that generate step controls based on the sampling points. We further denote $\mathscr{S}_\vartheta^\kappa$ by the control system driven by (3.6) that is comprised by $\mathfrak{u} = \kappa(x)$. We denote by X^κ the controlled process given the synthesized control and by $\mathbf{P}^{\mu,\kappa}$ the law of X^κ for any initial distribution $\mu \in \mathfrak{P}(\mathcal{X})$.

The corresponding generator/characteristic operator is given as

$$\mathfrak{L}_d^u h(x) := \mathfrak{L}_d h(x) + L_g h(x)u, \quad h \in \operatorname{dom}(\mathfrak{L}_d)$$

for each $d \in \overline{\mathbb{B}}$ and $u \in \mathcal{U}$, where $L_g h(x) = \nabla h(x) \cdot g(x)$. Similarly, when $\vartheta = 0$, we have

$$\mathfrak{L}^u h(x) := \mathfrak{L} h(x) + L_g h(x)u, \quad h \in \operatorname{dom}(\mathfrak{L}).$$

3.2 Stochastic Barrier Functions for Probabilistic Invariance Specifications

We have seen in Section 2.2 that two special types of barrier functions are frequently used to guarantee set invariance in the deterministic context. In particular, the barrier conditions associated to ZBFs are less strict than those of RBFs. For this reason, we extend the notion of ZBFs to CBFs for control synthesis of safety specifications aiming at generating bounded control inputs when trajectories approach the safe boundaries.

However, for diffusion processes, such a zeroing-type relaxation of barrier conditions for stochastic barrier functions will have a significant negative impact on the satisfaction probabilities. On the other hand, the non-relaxed reciprocal type conditions provide more severe control constraints than their deterministic counterpart. For these reasons, we propose a middle ground to characterize safety properties for (3.6). To better convey the idea, we directly work on the stochastic control barrier functions and only consider the special case when $\vartheta = 0$ throughout this section.

3.2.1 Problem Definition

For simplicity, we consider $X_0 = x$ a.s. for some $x \in \mathbb{R}^n$ as the initial condition. We first introduce the stochastic analogue of set invariance, i.e. probabilistic set invariance, defined as follows.

Definition 3.2.1 (Probabilistic set invariance)**.** *Let X be a stochastic process. A set $\mathcal{C} \subseteq \mathbb{R}^n$ is said to be invariant w.r.t. a tuple (x, T, p) for X, where $x \in \mathcal{C}$, $T \geq 0$, and $p \in [0, 1]$, if the initial condition $X_0 = x$ a.s. implies*

$$\mathbf{P}^x[X_t \in \mathcal{C},\ 0 \leq t \leq T] \geq p. \tag{3.7}$$

Moreover, if $\mathcal{C} \subseteq \mathbb{R}^n$ is invariant w.r.t. $(x, T, 1)$ for all $x \in \mathcal{C}$ and $T \geq 0$, then $\mathcal{C}$ is strongly invariant for X.

Definition 3.2.2 (Controlled probabilistic invariance)**.** *Given system (3.6) and a control signal $\mathfrak{u}$, a set $\mathcal{C} \subseteq \mathbb{R}^n$ is said to be controlled invariant under $\mathfrak{u}$ w.r.t. a tuple (x, T, p) for system (3.6), if $\mathcal{C}$ is invariant w.r.t. (x, T, p) for the controlled process $X^{\mathfrak{u}}$.*

Similarly, $\mathcal{C} \subseteq \mathbb{R}^n$ is strongly controlled invariant under $\mathfrak{u}$ if $\mathcal{C} \subseteq \mathbb{R}^n$ is controlled invariant under $\mathfrak{u}$ w.r.t. $(x, T, 1)$ for all $x \in \mathcal{C}$ and $T \geq 0$.

For the rest of this section, we consider a safe set of the form

$$\mathcal{C} := \{x \in \mathbb{R}^n : h(x) \geq 0\}, \quad h \in C^2(\mathbb{R}^n). \tag{3.8}$$

We also define the boundary and interior of $\mathcal{C}$ explicitly as below

$$\partial\mathcal{C} := \{x \in \mathbb{R}^n : h(x) = 0\}, \tag{3.9}$$

$$\text{Int}(\mathcal{C}) := \{x \in \mathbb{R}^n : h(x) > 0\}. \tag{3.10}$$

Problem 3.2.3 (Probabilistic set invariance control)**.** *Given a compact set* $\mathcal{C} \subseteq \mathbb{R}^n$ *defined in* (3.8), *a point* $x \in \operatorname{Int}(\mathcal{C})$, *and a tuple* (x, T, p), *design a control strategy* κ *such that under* $\mathfrak{u} = \kappa(x)$, *the interior* $\operatorname{Int}(\mathcal{C})$ *is controlled invariant w.r.t.* (x, T, p) *for the resulting strong solutions to* (3.6).

3.2.2 Safe-Critical Control Design via Barrier Functions

In this subsection, we propose stochastic barrier certificates that can be used to design a control strategy κ for Problem 3.2.3. Before proceeding, it is necessary to review stochastic control barrier functions to interpret probabilistic set invariance. Note that we consider the safe set as constructed in (3.8), where the function h is given *a priori*.

Stochastic Reciprocal and Zeroing Barrier Functions

Similar to the terminology for deterministic cases [5], we introduce the construction of stochastic control barrier functions as follows.

Definition 3.2.4 (Reciprocal stochastic control barrier function)**.** *A function* $B : \operatorname{Int}(\mathcal{C}) \to \mathbb{R}$ *is called a R-SCBF for system* (3.6) *if* $B \in C^2(\mathbb{R}^n)$ *and satisfies the following properties:*

(1) there exist class-$\mathcal{K}$ *functions* α_1, α_2 *such that for all* $x \in \mathbb{R}^n$ *we have*

$$\frac{1}{\alpha_1(h(x))} \leq B(x) \leq \frac{1}{\alpha_2(h(x))}; \tag{3.11}$$

(2) there exists a class-$\mathcal{K}$ *function* α_3 *such that*

$$\inf_{\mathfrak{u} \in \mathcal{U}} [\mathfrak{L}^u B(x) - \alpha_3(h(x))] \leq 0. \tag{3.12}$$

We refer to the set of control policies generated by (3.12) *as*

$$\mathfrak{K}_R(x) := \{u \in \mathcal{U} : \mathfrak{L}^u B(x) - \alpha_3(h(x)) \leq 0\}. \tag{3.13}$$

Proposition 3.2.5 ([40])**.** *Suppose that there exists an R-SCBF for system* (3.6). *If* $\mathfrak{u}(t) = \kappa(X_t^{\mathfrak{u}}) \in \mathfrak{K}_R(X_t^{\mathfrak{u}})$ *for all* $t \geq 0$, *then for all* $t \geq 0$ *and* $X_0 = x \in \operatorname{Int}(\mathcal{C})$, *we have* $\mathbf{P}^{x,\mathfrak{u}}[X_t^{\mathfrak{u}} \in \operatorname{Int}(\mathcal{C})] = 1$ *for all* $t \geq 0$.

Remark 3.2.6. *The result admits a $\mathbf{P}^{x,\mathfrak{u}}$-a.s. controlled invariant set for the marginals of $X^{\mathfrak{u}}$, and is easily extended to a pathwise $\mathbf{P}^{x,\mathfrak{u}}$-a.s. controlled set invariance. Note that the solution is right continuous. Let $\{t_n,\ n = 1, 2, ...\}$ be the set of all rational numbers in $[0, \infty)$, and put*

$$\mathbf{\Omega}^* := \bigcap_{1\leq n<\infty} \{\omega : X^{\mathfrak{u}}_{t_n} \in \mathrm{Int}(\mathcal{C})\},$$

then $\mathbf{\Omega}^* \in \mathcal{F}$ (a σ-algebra is closed w.r.t. countable intersections). Since $\mathbb{Q}$ (the set of rational numbers) is dense in $\mathbb{R}$, $X^{\mathfrak{u}}$ is right continuous, and h is continuous, we have

$$\mathbf{\Omega}^* := \{\varpi : X^{\mathfrak{u}}_t \in \mathrm{Int}(\mathcal{C}),\ \forall t \in [0, \infty)\}.$$

Consequently, we have $\mathbf{P}^{x,\mathfrak{u}}[\mathbf{\Omega}^*] \equiv 1$ from the marginal result.

Definition 3.2.7. *(Zeroing stochastic control barrier function) A function $B : \mathcal{C} \to \mathbb{R}$ is called a Z-SCBF for system (3.6) if $B \in C^2(\mathbb{R}^n)$ and*

(1) $B(x) \geq 0$ for all $x \in \mathcal{C}$;

(2) $B(x) < 0$ for all $x \notin \mathcal{C}$;

(3) there exists an extended $\mathcal{K}_\infty$ function α such that

$$\sup_{u\in\mathcal{U}} [\mathfrak{L}^u B(x) + \alpha(B(x))] \geq 0. \tag{3.14}$$

We refer the set of control policies generated by (3.14) as

$$\mathfrak{K}_Z(x) := \{u \in \mathcal{U} : \mathfrak{L}^u B(x) + \alpha(B(x)) \geq 0\}. \tag{3.15}$$

Proposition 3.2.8 (Worst-case probabilistic quantification)**.** *Suppose the mapping h as in (3.8) is a Z-SCBF with $\alpha(x) = kx$ and $k > 0$. Let $c = \sup_{x\in\mathcal{C}} h(x)$ and $X_0 = x \in \mathrm{Int}(\mathcal{C})$. Then, if $\mathfrak{u}_t = \kappa(X^{\mathfrak{u}}_t) \in \mathfrak{K}_Z(X^{\mathfrak{u}}_t)$ for all $t \in [0, T]$, where $\mathfrak{K}_Z(y) = \{u \in \mathcal{U} : \mathcal{A}h(y) + kh(y) \geq 0\}$, we have the following worst-case probability estimation:*

$$\mathbf{P}^{x,\mathfrak{u}} [X^{\mathfrak{u}}_t \in \mathrm{Int}(\mathcal{C}),\ 0 \leq t \leq T] \geq \left(\frac{h(x)}{c}\right) e^{-cT}. \tag{3.16}$$

Proof. Let $s = c - h(x)$ and $V(x) = c - h(x)$, then $V(x) \in [0, c]$ for all $x \in \mathcal{C}$. It is clear that $\mathfrak{L}^u V(x) = -\mathfrak{L}^u h(x)$ for any $u \in \mathcal{U}$. For $\mathfrak{u}(t) \in \mathfrak{K}_Z(X^{\mathfrak{u}}_t)$ for all $t \in [0, T]$, we have

$$\mathfrak{L}^u V(X^{\mathfrak{u}}_t) \leq -kV(X^{\mathfrak{u}}_t) + kc.$$

By [100, Theorem 3.1],

$$\mathbf{P}^{x,\mathfrak{u}}\left[\sup_{t\in[0,T]} V(X_t^{\mathfrak{u}}) \geq c\right] \leq 1 - \left(1 - \frac{s}{c}\right) e^{-cT}. \tag{3.17}$$

The result follows directly after this. □

Stochastic control barrier functions and high-order stochastic control barrier functions

In proposing stochastic control barrier functions for high-order control systems, the above R-SCBF and Z-SCBF are building blocks. The authors in [148] constructed high-order R-SCBFs and have found the sufficient conditions to guarantee pathwise set invariance with probability one. While the results seem strong, they come with significant costs. At the safety boundary, the control inputs need to be unbounded (as shown in Example 3.2.9 and in the numerical experiments in [167]). On the other hand, the synthesis of controller for a high-order system via a Z-SCBF has mild constraints. The trade-off is that the probability estimation of set invariance is of low quality (note that the worst-case probability estimation using first-order barrier function is already discounted over a relatively long time period). For these reasons, we propose a different type of stochastic control barrier functions other than the above mentioned two types.

Example 3.2.9. *Consider a stochastic system*

$$dX_t = (X_t + \mathfrak{u}_t)dt + bdW_t, \quad b \in \mathbb{R}.$$

Suppose that $\mathcal{C} = \{x \in \mathbb{R} : \; x \leq 1\}$ *with* $h(x) = 1 - x$*. Accordingly, an R-SCBF for the system is* $B(x) = \frac{1}{h(x)}$ *and we have* $\frac{\partial^2 B}{\partial x^2} = \frac{2}{h^3}$*. Let* $\alpha = 1$*. Then the R-SCBF condition is*

$$\mathfrak{L}^u B(x) = \frac{1}{h^2(x)}(x+u) + \frac{b^2}{h^3(x)} \leq h(x)$$

and the control inputs should satisfy

$$u \leq h^3(x) - x - \frac{b^2}{h(x)}.$$

As x *approaches 1, the control approaches* $-\infty$*. This implies that in order to guarantee safety, we requires an unbounded control around the boundary of the safe set for stochastic systems, which can be difficult to satisfy for some practical applications.*

To obtain non-vanishing worst-case probability estimation (compared to Z-SCBFs) in infinite horizon, we propose a safety certificate via a stochastic Lyapunov-like control barrier function [100], which we still name it as a SCBF.

Definition 3.2.10 (Stochastic control barrier function)**.** *A function $B \in C^2(\mathbb{R}^n)$ is said to be a SCBF if $L_g B(x) \neq 0$ (recall notation in Definition 2.1.3) and the following conditions are satisfied:*

(1) $B(x) \geq 0$ *for all* $x \in \mathcal{C}$*;*

(2) $B(x) < 0$ *for all* $x \notin \mathcal{C}$*;*

(3) $\sup_{u \in \mathcal{U}} \mathfrak{L}^u B(x) \geq 0.$

We refer the set of control policies generated by (3) as

$$\mathfrak{K}(x) := \{u \in \mathcal{U} : \mathfrak{L}^u B(x) \geq 0\}. \tag{3.18}$$

Proposition 3.2.11. *Suppose the mapping h is an SCBF with the corresponding set of control policies $\mathfrak{K}(x)$. Let $c = \sup_{x \in \mathcal{C}} h(x)$ and $X_0 = x \in \operatorname{Int}(\mathcal{C})$. Then, if $\mathfrak{u}_t = \kappa(X_t^{\mathfrak{u}}) \in \mathfrak{K}(X_t^{\mathfrak{u}})$ for all $t \geq 0$, we have the following worst-case probability estimation:*

$$\mathbf{P}^{x,\mathfrak{u}}\left[X_t^{\mathfrak{u}} \in \operatorname{Int}(\mathcal{C}),\ 0 \leq t < \infty\right] \geq \frac{h(x)}{c}.$$

Proof. Let $V = c - h$, then $V(x) \geq 0$ for all $x \in \mathcal{C}$ and $\mathfrak{L}^u V(x) \leq 0$. The result is followed by [100, Lemma 2.1] using a supermartingale argument. Indeed, for every $t \geq 0$,

$$\begin{aligned} \mathbf{E}^{x,\mathfrak{u}}[V(X_{\tau\wedge t}^{\mathfrak{u}})] &\geq \mathbf{E}^{x,\mathfrak{u}}[\mathbb{1}_{\{\tau \leq t\}} V(X_{\tau\wedge t}^{\mathfrak{u}})] \\ &\geq \mathbf{P}^{x,\mathfrak{u}}[\tau \leq t] \cdot \mathbf{E}^{x,\mathfrak{u}}[V(X^{\mathfrak{u}}(\tau)] \\ &> c \cdot \mathbf{P}^{x,\mathfrak{u}}[\tau \leq t]. \end{aligned} \tag{3.19}$$

However, $\mathbf{E}^{x,\mathfrak{u}}[V(X_{\tau\wedge t}^{\mathfrak{u}})] \leq V(x)$ for all $t \geq 0$. Therefore,

$$\mathbf{P}^{x,\mathfrak{u}}[\tau \leq t] < \frac{V(x)}{c},\ \forall t \geq 0. \tag{3.20}$$

Sending $t \to \infty$ we get $\mathbf{P}^{x,\mathfrak{u}}[\tau < \infty] \leq \frac{V(x)}{c}$ for all $x \in \operatorname{Int}(\mathcal{C})$. The result follows immediately. □

Definition 3.2.12. *A function $B : C^{2r}(\mathbb{R}^n) \to \mathbb{R}$ is called a stochastic control barrier function with relative degree r for system (3.6) if*

(1) B satisfies (1) and (2) of Definition 3.2.10, and

(2) $L_g\mathfrak{L}^{r-1}B(x) \neq 0$ and $L_g\mathfrak{L}^{j-1}B(x) = 0$ for $j = 1, 2, \dots, r-1$ and $x \in \mathcal{C}$.

(3) $\sup_{u\in\mathcal{U}}(\mathfrak{L}^u)^r B(x) \geq 0$.

We refer the set of control policies generated by (3) as

$$\mathfrak{K}^r(x) := \{u \in \mathcal{U} : (\mathfrak{L}^u)^r B(x) \geq 0\}. \tag{3.21}$$

If the system (3.6) is an r^{th}-order stochastic control system, to steer the process $X^{\mathfrak{u}}$ to satisfy probabilistic set invariance w.r.t. $\mathrm{Int}(\mathcal{C})$, we recast the mapping h as an SCBF with relative degree r. For $h \in C^{2r}(\mathbb{R}^n)$, we define a series of functions $b_0, b_j : \mathbb{R}^n \to \mathbb{R}$ such that for each $j = 1, 2, \dots, r$, we have $b_0, b_j \in \mathrm{dom}(\mathfrak{L})$ and

$$\begin{aligned} &b_0(x) = h(x), \\ &b_j(x) = \mathfrak{L}^{\mathfrak{u}} \circ (\mathfrak{L}^{\mathfrak{u}})^{j-1} b_0(x). \end{aligned} \tag{3.22}$$

We further define the corresponding superlevel sets $\mathcal{C}_j$ for $j = 1, 2, \dots, r$ as

$$\mathcal{C}_j = \{x \in \mathbb{R}^n : b_j(x) \geq 0\}. \tag{3.23}$$

Theorem 3.2.13. *If the mapping h is an SCBF with relative degree r, and the control process $\mathfrak{u}$ is such that $\mathfrak{u}_t \in \mathfrak{K}^r(X_t^{\mathfrak{u}})$ for all $t \geq 0$. Let $c_j =: \sup_{x\in\mathcal{C}_j} b_j(x)$ for each $j = 0, 1, ..., r$ and $X_0 = x \in \bigcap_{j=0}^{r} \mathcal{C}_j^\circ$. Then we have the following worst-case probability estimation:*

$$\mathbf{P}^{x,\mathfrak{u}}[X_t^{\mathfrak{u}} \in \mathrm{Int}(\mathcal{C}),\ 0 \leq t < \infty] \geq \prod_{j=0}^{r-1} \frac{b_j(x)}{c_j}.$$

Remark 3.2.14. *We omit the proof due to its less relevance to the main topic. The details can be found in [167, Theorem III.10]. The above result estimates the lower bound of the safety probability given the constrained control signals $\mathfrak{u}$. Under some extreme conditions, the worst case may happen. Indeed, a conservative assumption is made in the proof such that within finite time $X^{\mathfrak{u}}$ will cross the boundary of each $\mathcal{C}_j$. Another implicit condition that may cause the worst-case lower bound is when the event $\bigcup_{j=0}^{r-1}\{\mathfrak{L}^u b_j = 0,\ 0 \leq t < \infty\}$ is a $\mathbf{P}^{x,\mathfrak{u}}$-null set. This, however, is practically possible since the controller indirectly influences the value of $\mathfrak{L}^u b_j$ for all $j < r$, the strong invariance of the level set $\{\mathfrak{L}^u b_j = 0\}$ is not guaranteed using QP scheme.*

On the other hand, a non-zero probability of $\{\mathfrak{L}^u b_j = 0\}$ for any $j < r$ makes a compensate for the lower bound estimation. On $\{\mathfrak{L}^u b_j = 0\}$, the process $\{b_j(X_{t\wedge\tau_j})\}_{t\geq 0}$ is a lower bounded *martingale and therefore converges with probability 1. We can estimate (using standard exit-time problem arguments) that* $\{b_j(X_{t\wedge\tau_j})\}_{t\geq 0}$ *reaching* 0 *within a finite time has a probability* $1-\frac{b_j(X_{\tau_{j+1}})}{c_j}$.

A nice selection of controller is to implicitly reduce the total time a sample path spends in $\{\mathfrak{L}^u b_j \leq 0\}$ *for each* j*. However, this is a challenging task by only steering the bottom-level flow, which in turn gives us a future research direction.*

Two examples are provided in [167] to validate our results. It can be shown numerically that the proposed SCBFs have smaller control efforts compared to R-SCBFs and higher safe probability compared to Z-SCBFs. In particular, given a bounded constraint on control inputs, the safe (conditional) probability using SCBF is much higher than using R-SCBF. Such an empirical difference is enlarged for control systems with high-relative degrees. On the other hand, the result demonstrates an overall better performance of SCBF compared to Z-SCBF given randomly selected initial points, which indicates a potentially larger winning set.

3.3 Stochastic Lyapunov-Barrier Functions for Robust Probabilistic Reach-Avoid-Stay Specifications

In this section, we aim to provide a stochastic version for the probabilistic reach-avoid-stay problems in consideration of robustness.

Motivated by the deterministic robust abstractions [112, 110] and the comparisons with robust Lyapunov-type characterizations of reach-avoid-stay specifications in Section 2.4.2, to better understand how these two perform in the stochastic context, this section formulates stochastic Lyapunov-barrier functions to deal with sufficient conditions for robust probabilistic reach-avoid-stay specifications. The studies on robust stochastic abstractions will be carried out in the following section.

To this end, we first establish a connection between stochastic stability with safety constraints and reach-avoid-stay specifications based on (3.1). We then prove that stochastic Lyapunov and Lyapunov-barrier functions provide sufficient conditions for the target objectives. We apply Lyapunov-barrier conditions in control synthesis for reach-avoid-stay specifications based on (3.6), and show its effectiveness in a case study.

3.3.1 Stability and Safety Concepts

We first consider the control-free dynamics driven by (3.1) and introduce the problem definitions. The systems with controls should be similar.

Apart from Assumption 3.1.2, we also impose the following requirements on f and g.

Assumption 3.3.1. *There exists a trivial solution x_e for system $\mathscr{S}$ (i.e. $\mathscr{S}_\vartheta$ when $\vartheta = 0$) such that $f(x_e) = g(x_e) = 0$.*

We provide definitions for probabilistic set stability given a closed set $A \subseteq \mathbb{R}^n$.

Definition 3.3.2 (Uniform stability in probability law)**.** *The set A is said to be Pr-US for $\mathscr{S}_\vartheta$ if for each $\epsilon \in (0,1)$ there exists $\varphi_\epsilon \in \mathcal{K}$ such that*

$$\inf_{X \in \mathfrak{S}_\vartheta(x,W)} \mathbf{P}^X[|X_t|_A \leq \varphi_\epsilon(|x|_A) \;\; \forall t \geq 0] \geq 1 - \epsilon, \tag{3.24}$$

where x is the initial condition.

Remark 3.3.3. *Equation (3.24) is equivalent to the following: for any $\epsilon \in (0,1)$ and $r > 0$, there exists an $\eta = \eta(\epsilon, r) \in (0, r)$ such that*

$$\inf_{X \in \mathfrak{S}_\vartheta(x,W)} \mathbf{P}^X[|X_t|_A \leq r \;\; \forall t \geq s(\varpi)] \geq 1 - \epsilon, \tag{3.25}$$

whenever $|X_{s(\varpi)}|_A \leq \eta$ for some random time $s(\varpi)$. We can simply pick $\eta = \varphi_\epsilon^{-1}$.

Definition 3.3.4 (Uniform attractivity in probability law)**.** *The set A is said to be Pr-UA for $\mathscr{S}_\vartheta$ if there exists some $\eta > 0$ such that, for each $\epsilon \in (0,1)$, $r > 0$, there exists some $T > 0$ such that whenever $|x|_A < \eta$,*

$$\inf_{X \in \mathfrak{S}_\vartheta(x,W)} \mathbf{P}^X[|X_t|_A < r, \; \forall t \geq T] \geq 1 - \epsilon. \tag{3.26}$$

Definition 3.3.5 (Uniformly asymptotic stability in probability law)**.** *The set A is said to be Pr-UAS for $\mathscr{S}_\vartheta$ if it is Pr-US and Pr-UA for $\mathscr{S}_\vartheta$.*

Next we introduce several definitions pertinent to probabilistic stability with safety guarantees. To this end, we consider a closed unsafe set $U \subseteq \mathbb{R}^n$.

Definition 3.3.6 (Work place)**.** *Since the solutions are not generally non-explosive without stability assumptions, a bounded workplace $\mathscr{N} := \mathbb{B}_{\tilde{R}}(x_e)$ with sufficiently large $\tilde{R} > 0$ is added as an extra constraint. We name $\mathcal{D} = \mathcal{D}(\mathscr{N}, U) := \mathscr{N} \cap U^c$.*

We define the following first-exit times for the regions of our interest.

Definition 3.3.7 (Explosion and safety). *For any solution $X \in \mathfrak{S}_\vartheta(\mathbb{R}^n, W)$, we define the explosion time $\tau^* = \tau^*(\mathscr{N}) := \inf\{t \geq 0 : X_t \in \mathscr{N}^c\}$ and safety time*

$$\tau_{\mathcal{D}} := \inf\{t \geq 0 : X_t \in \mathcal{D}^c\}.$$

It is clear that for each X, for $\mathbf{P}^X$-a.s. we have $\tau^* \leq \tau_{\mathcal{D}}$.

Remark 3.3.8. *Safety is usually the priority in practice. Given safety requirement w.r.t. $\mathcal{D}$ (resp. $\mathscr{N}$), to study conditional probabilistic properties before reaching the unsafe boundary of some process X, it is equivalent to just working with the law of $X^{\tau_{\mathcal{D}}}$ (resp. X^{τ^*}). Note that for systems with trivial Pr-US sets, the indicator $\mathbb{1}_{\{\tau^*=\infty\}} \to 1$ as $\tilde{R} \to \infty$ and does not render 'too much harm' to replace the law of X^{τ^*} by $\mathbf{P}^X$.*

Suppose we 'kill' the weak solutions whenever they become unsafe and consider the modified probability law with supports in the safe region, the following proposition verifies a notion of weak compactness (rather than the conventional compactness) of stopped weak solutions of $\mathscr{S}_\vartheta$.

Proposition 3.3.9. *Under the Assumption 3.1.2, given any compact set K, the set of all stopped processes X^{τ^*} is nonempty and sequentially weakly compact (w.r.t. the weak convergence) on every filtered probability space $(\mathbf{\Omega}, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]})$, where $X \in \bigcup_{x\in K} \mathfrak{S}_\vartheta(x, W)$ (resp. $X \in \mathfrak{S}_\vartheta(K, W)$). That is, given any sequence of weak solutions $\{X^n\}_{n=1}^\infty$ in the above sense, there is a subsequence $\{X^{n_k}\}$, a process $X \in \bigcup_{x\in K} \mathfrak{S}_\vartheta(x, W)$ (resp. $X \in \mathfrak{S}_\vartheta(K, W)$) such that $(X^{n_k})^{\tau^*} \rightharpoonup X^{\tau^*}$.*

Remark 3.3.10. *Note that the above results also hold for stopped processes $X^{\tau_{\mathcal{D}}}$, where $X \in \bigcup_{x\in K} \mathfrak{S}_\vartheta(x, W)$ (resp. $X \in \mathfrak{S}_\vartheta(K, W)$).*

The conclusions follow immediately by [93, Theorem 1] and [94, Corollary 1.1, Chap 3]. The proof falls in standard procedures. We can first show that the truncated laws $\{\mathbf{P}^{n,\tau^*}\}$ of the stopped processes $\{(X^n)^{\tau^*}\}$ form a tight family of measures[3] on $(\mathbf{\Omega}, \mathcal{F}, \{\mathcal{F}_t\}_{t\in[0,T]})$. Then the relatively weak compactness follows since $(X^{n_k})^{\tau^*} \rightharpoonup X^{\tau^*}$ if and only if $\mathbf{P}^{n_k,\tau^*} \rightharpoonup \mathbf{P}^{\tau^*}$. The weak closedness comes from compactness of the reachable sets of the stopped processes.

Now we introduce two closely-related specifications pertaining to stability and safety issues.

[3] See Appendix E for details.

Definition 3.3.11 (Probabilistic stability with safety guarantees)**.** *Given a closed set $U \subseteq \mathbb{R}^n$, let $\mathcal{D}$ and $\tau_{\mathcal{D}}$ be defined as in Definitions 3.3.6 and 3.3.7, respectively. Given $\mathcal{X}_0, A \subseteq \mathcal{D}$ and $p \in [0,1]$, $\mathscr{S}_\vartheta$ is said to satisfy a probabilistic stability under safety specification w.r.t. $(\mathcal{X}_0, A, U)$ with probability at least p, denoted by $(\mathcal{X}_0, A, U, p)$, if*

(1) A is closed and Pr-UAS for $\mathscr{S}_\vartheta$;

(2) For all $X \in \bigcup_{x\in\mathcal{X}_0} \mathfrak{S}_\vartheta(x, W)$,

$$\mathbf{P}^X \left[\tau_{\mathcal{D}} = \infty \text{ and } \lim_{t\to\infty} |X_t|_A = 0\right] \geq p.$$

The above definition is the probabilistic analogue of the stability with safety guarantee specification in the deterministic settings. The condition (1) of Definition 3.3.11 is to qualify the behavior of solutions near A, whereas condition (2) is to require that the solution can be attracted asymptotically to some neighborhood of A and meanwhile maintain safe with probability at least p.

The following events (sets of sample paths in view of canonical space) are defined in favor of defining the second specification.

Definition 3.3.12. *Given $\mathcal{X}_0, \Gamma \subseteq \mathcal{D}$. On $(\mathbf{\Omega}, \mathcal{F})$, for each $X \in \bigcup_{x\in\mathcal{X}_0} \mathfrak{S}_\vartheta(x, W)$, we define the following events.*

(1) The reach-and-stay event

$$RS(\mathcal{X}_0, \Gamma, \mathcal{D}) := \{\varpi \in \mathbf{\Omega} : \tau_\gamma < \infty \text{ and } X_{t\wedge\tau_{\mathcal{D}}} \in \Gamma,\ \forall t \geq \tau_\gamma\},$$

where $\tau_\gamma := \inf\{t \geq 0 : X_t \in \Gamma\}$ is the first hitting time of Γ for the given X;

(2) The reach-avoid-stay event $RAS(\mathcal{X}_0, \Gamma, \mathcal{D}) := RS(\mathcal{X}_0, \Gamma, \mathcal{D}) \cap \{\tau_{\mathcal{D}} = \infty\}$.

Definition 3.3.13. *(Probabilistic reach-avoid-stay specification): Given a closed set $U \subseteq \mathbb{R}^n$, let $\mathcal{D}$ and $\tau_{\mathcal{D}}$ be defined as in Definitions 3.3.6 and 3.3.7, respectively. Given $\mathcal{X}_0, \Gamma \subseteq \mathcal{D}$ and $p \in [0,1]$, $\mathscr{S}_\vartheta$ is said to satisfy a probabilistic reach-avoid-stay specification w.r.t. $(\mathcal{X}_0, \Gamma, U)$ with probability at least p, denoted by $(\mathcal{X}_0, \Gamma, U, p)$, if for every weak solution $X \in \bigcup_{x\in\mathcal{X}_0} \mathfrak{S}_\vartheta(x, W)$, we have $\mathbf{P}^X[RAS(\mathcal{X}_0, \Gamma, \mathcal{D})] \geq p$.*

The probabilistic reach-avoid-stay specification is the linear temporal property that we mainly focus on. In the stochastic context, we explore the connection between the target specification and the probabilistic specification of stability with safety guarantees in the next subsection.

3.3.2 A Connection to Probabilistic Stability with Safety Guarantees

Probabilistic stability with safety implies probabilistic reach-avoid-stay

We first show that if a closed set A is Pr-US for $\mathscr{S}_\vartheta$, then the attraction to A in probability law possesses the uniformity.

Proposition 3.3.14. *Suppose that a closed set $A \subset \mathcal{D}$ is Pr-US for $\mathscr{S}_\vartheta$. Let K be a compact set and $p \in (0,1)$. Then the following two statements are equivalent:*

(1) For any solution $X \in \bigcup_{x\in K} \mathfrak{S}_\vartheta(x, W)$,

$$\mathbf{P}^X \left[\lim_{t\to\infty} |X_{t\wedge\tau_\mathcal{D}}|_A = 0\right] \geq p.$$

(2) For every $r > 0$, there exists $T = T(r, \varepsilon)$ such that for any $X \in \bigcup_{x\in K} \mathfrak{S}_\vartheta(x, W)$,

$$\mathbf{P}^X \left[|X_{t\wedge\tau_\mathcal{D}}|_A < r, \ \forall t \geq T\right] \geq p.$$

Proof. Clearly (2) implies (1). We only show the converse. Suppose that (2) is not true. Then there exists some $r > 0$ such that for all $n > 0$ there exists $x_n \in K$, and $X^n \in \mathfrak{S}_\vartheta(x_n, W)$ with law $\mathbf{P}^n$ such that

$$\mathbf{P}^n[|X^n_{t\wedge\tau_\mathcal{D}}|_A \leq r, \ \forall t \geq n] < p. \tag{3.27}$$

Now let $\tau^n = \inf\{t \geq 0 : X^n_{t\wedge\tau_\mathcal{D}} \in \mathbb{B}_\eta(A)\}$, where η is to be chosen later. Rearranging (3.27) we have (for each n),

$$\begin{aligned} p >&\mathbf{P}^n[|X^n_{t\wedge\tau_\mathcal{D}}|_A \leq r, \ \forall t \geq n] \\ \geq&\mathbf{P}^n[\tau^n < n \text{ and } |X^n_{t\wedge\tau_\mathcal{D}}|_A \leq r, \ \forall t \geq n] \\ =&\mathbf{P}^n[\tau^n < n]\mathbf{P}[|X^n_{t\wedge\tau_\mathcal{D}}|_A \leq r, \ \forall t \geq n \mid \tau^n < n] \\ \geq&\mathbf{P}^n[\tau^n < n]\mathbf{P}[|X^n_{t\wedge\tau_\mathcal{D}}|_A \leq r, \ \forall t \geq \tau^n], \end{aligned} \tag{3.28}$$

By the definition of Pr-US in view of Remark 3.3.3, there exists an $\eta = \eta(r, \epsilon) < r$ such that $\mathbf{P}^n[|X^n_{t\wedge\tau_\mathcal{D}}|_A \leq r, \ \forall t \geq s(\varpi)] \geq \epsilon$ whenever $|X_{s(\varpi)\wedge\tau_\mathcal{D}}|_A \leq \eta$ for some random time $s(\varpi)$ and arbitrary ϵ. Clearly, τ^n satisfies the requirement of $s(\varpi)$. We choose ϵ sufficiently close to 1 so that

$$\mathbf{P}[|X^n_{t\wedge\tau_\mathcal{D}}|_A \leq r, \ \forall t \geq \tau^n] \geq \epsilon$$

and hence, by (3.28),

$$\mathbf{P}^n[\tau^n < n] = p - \hat{p} < p, \tag{3.29}$$

where $\hat{p} = \hat{p}(\epsilon) \ll 1$. Note that we have implicitly defined η and τ^n based on the choice of ϵ such that (3.29) holds.

However, by Remark 3.3.8 and Proposition 3.3.9, there exists a subsequence, still denoted by $X^n \in \mathfrak{S}_\vartheta(x_n, W)$, such that $x_n \to x$ and $(X^n)^{\tau_\mathcal{D}} \rightharpoonup X^{\tau_\mathcal{D}}$ with $X \in \mathfrak{S}_\vartheta(x, W)$ on any compact interval of $\mathbb{R}_{\geq 0}$. By Skorohod [43, Theorem 2.4], there exists a probability space $(\tilde{\Omega}^\dagger, \tilde{\mathscr{F}}^\dagger, \{\tilde{\mathscr{F}}_t^\dagger\}, \tilde{\mathbb{P}}^\dagger)$, a process $\tilde{X}^{\tau_\mathcal{D}}$ and a sequence of processes $\{(\tilde{X}^n)^{\tau_\mathcal{D}}\}$ with laws $\mathbf{P}$ and $\{\mathbf{P}^n\}$, respectively, such that

$$\lim_{n\to\infty} (\tilde{X}^n)^{\tau_\mathcal{D}} = \tilde{X}^{\tau_\mathcal{D}}, \quad \tilde{\mathbb{P}}^\dagger - \text{a.s.}. \tag{3.30}$$

Let $\tau = \inf\{t \geq 0 : |\tilde{X}_{t\wedge\tau_\mathcal{D}}|_A \leq \eta/2\}$, due to the asymptotic behavior from (1), we have

$$\tilde{\mathbb{P}}^\dagger[\tau < \infty] \geq p. \tag{3.31}$$

By (3.30) and (3.31), there exists some sufficiently large $N_1(\eta, q_1)$ and $N_2(\eta, q_2)$ such that for any arbitrary $q_1, q_2 \in (0, 1)$,

$$\tilde{\mathbb{P}}^\dagger \left[\sup_{t\in[0,N_1]} |\tilde{X}^n_{t\wedge\tau_\mathcal{D}} - \tilde{X}_{t\wedge\tau_\mathcal{D}}| \leq \eta/2 \right] \geq q_1, \tag{3.32}$$

and

$$\tilde{\mathbb{P}}^\dagger \left[\tau < N_2\right] \geq pq_2. \tag{3.33}$$

Note that the events in (3.32) and (3.33) are independent, combining these and choosing $n \geq \max(N_1, N_2)$, we have

$$\tilde{\mathbb{P}}^\dagger \left[\exists t < n \text{ s.t. } \tilde{X}^n_{t\wedge\tau_\mathcal{D}} \in \mathbb{B}_\eta(A)\right] \geq \tilde{\mathbb{P}}^\dagger\left[Q\right] \geq pq_1q_2,$$

where

$$Q := \left\{ \sup_{t\in[0,N_1]} |\tilde{X}^n_{t\wedge\tau_\mathcal{D}} - \tilde{X}_{t\wedge\tau_\mathcal{D}}| \leq \eta/2 \right\} \cap \{|\tilde{X}_{\tau\wedge\tau_\mathcal{D}}|_A \leq \eta/2 \text{ for } \tau < N_2\}.$$

We let $q_1q_2 > \frac{p-\hat{p}}{p}$, then there exists an n such that

$$\mathbf{P}^n[\tau_n < n] > p - \hat{p}. \tag{3.34}$$

Equation (3.34) contradicts (3.29), which completes the proof. □

Corollary 3.3.15. *If $\mathscr{S}_\vartheta$ satisfies a stability with safety guarantee specification $(\mathcal{X}_0, A, U, p)$ and $\mathcal{X}_0$ is compact, then for every $\epsilon > 0$, $\mathscr{S}_\vartheta$ satisfies the reach-avoid-stay specification $(\mathcal{X}_0, \overline{\mathbb{B}}_\epsilon(A), U, p)$.*

Proof. We add the condition $\{\tau_\mathcal{D} = \infty\}$, then (1) and (2) are still equivalent in Proposition 3.3.14. The conclusion follows directly by the definitions of the two specifications. □

The converse side

The converse side is intended to show probabilistic stability with safety is necessary to probabilistic reach-avoid-stay specifications. Unfortunately, due to the diffusion effects and the concept of weak solutions, probabilistic reach-avoid-stay specifications, other than reach-avoid-stay with probability one, may fail to be related to probabilistic stability with safety guarantees w.r.t. some subset of the target set. For this reason, we only convey the main idea in this subsection and complete the proofs in Section 3.3.5.

Throughout this subsection, we suppose that $\mathcal{X}_0, \Gamma \subseteq \mathcal{D}$ and $\mathscr{S}_\vartheta$ satisfies a reach-avoid-stay specification $(\mathcal{X}_0, \Gamma, U, p)$. We first make a quick judgement that there exists a probability-p invariant compact subset of Γ.

Lemma 3.3.16. *Suppose that Γ is compact and $\mathcal{X}_0$ is nonempty. If $\mathscr{S}_\vartheta$ satisfies a reach-avoid-stay specification $(\mathcal{X}_0, \Gamma, U, p)$ with $p \in (0,1]$, then the set*

$$A = \{x \in \Gamma : \forall X \in \mathfrak{S}_\vartheta(x, W), \mathbf{P}^X[X_t \in \Gamma,\ \forall t \geq 0] \geq p\}$$

is a nonempty and compact set with

$$\mathbf{P}^X[X_t \in A,\ \forall t \geq 0] \geq p \tag{3.35}$$

for all $X \in \mathfrak{S}_\vartheta(A, W)$.

The next lemma shows that given an arbitrary solution X of $\mathscr{S}_{\vartheta'}$, we can construct a weak solution Z for $\mathscr{S}_\vartheta$ that solves the martingale problem and is relatively close to X.

Lemma 3.3.17. *Let $\vartheta' \in (0, \vartheta)$ and τ be such that $\tau < \tau_\mathcal{D}$ a.s.. Then there exists some $r = r(\tau, \vartheta, \vartheta')$ such that for every $X \in \mathfrak{S}_{\vartheta'}(x, W)$ with $x \in \mathcal{D}$, and for all $z \in \overline{\mathbb{B}}_r(x)$, there exists a weak solution $Z \in \mathfrak{S}_\vartheta(z, W)$ such that $Z_{T\wedge\tau_\mathcal{D}} \in \overline{\mathbb{B}}_r(X_{T\wedge\tau_\mathcal{D}})$ a.s. for $T \in [\tau, \infty)$.*

It can be shown that under the construction of Lemma 3.3.16 and 3.3.17, the set A (generated by solutions of $\mathscr{S}_\vartheta$) is Pr-UA for any weak solution of $\mathscr{S}_{\vartheta'}$ with $\vartheta' \in (0, \vartheta)$.

Proposition 3.3.18. *Suppose that $\mathscr{S}_\vartheta$ satisfies reach-avoid-stay specification $(\mathcal{X}_0, \Gamma, U, p)$. Let A be the set given in Lemma 3.3.16. Then A is uniformly attractive for $\mathscr{S}_{\vartheta'}$ with probability at least p, i.e., for every $\epsilon > 0$, there exists $T = T(\epsilon, p)$ such that for any $X \in \bigcup_{x\in\mathcal{X}_0} \mathfrak{S}_{\vartheta'}(x, W)$,*

$$\mathbf{P}\left[|X_{t\wedge\tau_\mathcal{D}}|_A < \epsilon,\ \forall t \geq T\right] \geq p.$$

On the other hand, for non-strictly invariant sets ($p < 1$), we are not able to show the Pr-US property due to a geometric gap where we cannot arbitrarily set ϵ and r as in Definition 3.3.2.

However, if there exists an invariant subset of Γ with probability one, nice properties appear. This is not a surprise given Proposition 2.3.16. The possibility of such an existence occurs when the system admits a family of a.s. stable Dirac invariant measures for each signal ξ, which are strictly contained in Γ. We convert the statement into the stochastic context in the next proposition. The proofs for the following two statements are omitted.

Proposition 3.3.19. *For system $\mathscr{S}_\vartheta$, any nonempty compact set $A \subset \mathcal{D}$ with $\mathbf{P}^X[X_t \in A,\ \forall t \geq 0] = 1$ is Pr-UAS for $\mathscr{S}_{\vartheta'}$ whenever $\vartheta' \in [0, \vartheta)$.*

Corollary 3.3.20. *If $\mathscr{S}_\vartheta$ satisfies a reach-avoid-stay specification $(\mathcal{X}_0, \Gamma, U, 1)$ with compact $\mathcal{X}_0$, then there exists a nonempty compact set $A \subseteq \Gamma$ with $\mathbf{P}^X[X_t \in A,\ \forall t \geq 0] = 1$ such that for any $\vartheta' \in [0, \vartheta)$, $\mathscr{S}_{\vartheta'}$ satisfies a stability with safety specification $(\mathcal{X}_0, A, U, 1)$.*

3.3.3 Lyapunov-Barrier Conditions for Probabilistic Stability With Safety

We aim to show how Lyapunov-Barrier functions can sufficiently guarantee the probabilistic stability with safety in this subsection. We first introduce stochastic Lyapunov functions and then extend the deterministic analogue in Chapter 2 to the stochastic counterpart.

Recall region $\mathcal{D}$ and $\mathscr{N}$ in Definition 3.3.6.

Definition 3.3.21 (Stochastic Lyapunov functions)**.** *Let $A \subseteq \mathcal{D}$ be a closed set. A function $V \in (C^2(\mathbb{B}_R(A)); \mathbb{R}_{\geq 0})$ is said to be a SLF w.r.t. A if there exist $\alpha_1, \alpha_2, \alpha_3 \in \mathcal{K}$ such that, for all $x \in \mathbb{B}_R(A)$,*

$$\alpha_1(|x|_A) \leq V(x) \leq \alpha_2(|x|_A) \tag{3.36}$$

and

$$\sup_{d \in \vartheta\overline{\mathbb{B}}} \mathfrak{L}_d V(x) \leq -\alpha_3(|x|_A). \tag{3.37}$$

We first make a quick extension of the existing Lyapunov theorems to systems with extra $\mathcal{L}_1$-bounded perturbations.

Lemma 3.3.22 (Uniform recurrence)**.** *Given an SLF V, there exists some $\eta > 0$ such that, for every $\epsilon \in (0, 1)$ and $r \in (0, R/2)$, there exists some $T = T(\epsilon, \eta, r) > 0$ such that for any $x \in \mathbb{B}_\eta(A)$,*

$$\inf_{X \in \mathfrak{S}_\vartheta(x, W)} \mathbf{P}^X[\tau < T] \geq 1 - \epsilon,$$

where $\tau = \inf\{t \geq 0 : X_t \in \mathbb{B}_r(A)\}$ is the first hitting time of $\mathbb{B}_r(A)$ for each $X \in \mathfrak{S}_\vartheta(x, W)$.

Proof. We just show the sketch. The proof falls in a similar procedure as in the proof of [119, Theorem 2.7]. We define the first hitting times τ_1, τ_2 of $\mathbb{B}_{r_1}(A)$ and $\mathbb{B}^c_{r_2}(A)$, where $0 < r_1 < r_2 \leq R/2$. By Dynkin's formula, for each $X \in \mathfrak{S}_\vartheta(x, W)$ with $x \in \mathbb{B}_\eta(A)$ (η is to be selected), we have

$$\begin{aligned}0 &\leq V(x) + \mathbf{E}^X \int_0^{\tau_1\wedge\tau_2\wedge t} \mathfrak{L}_d V(X(s))ds \\ &\leq \alpha_2(\eta) - \alpha_3(r_1)\mathbf{E}^X[\tau_1 \wedge \tau_2 \wedge t]\end{aligned}$$

On the other hand,

$$\begin{aligned}&\mathbf{E}^X[\tau_1 \wedge \tau_2 \wedge t] \\ \geq &\int_\Omega \mathbb{1}_{\{\tau_1\wedge\tau_2\geq t\}} \cdot (\tau_1(\varpi) \wedge \tau_2(\varpi) \wedge t)\, d\mathbf{P}^X(\varpi) \\ =&t\mathbf{P}^X[\tau_1 \wedge \tau_2 \geq t].\end{aligned}$$

Combining the above, we have $\mathbf{P}^X[\tau_1 \wedge \tau_2 \geq t] \leq \alpha_2(\eta)/t\alpha_3(r_1)$ for each t, which holds for all $X \in \mathfrak{S}_\vartheta(x, W)$. By this relation, we construct $T := T(\epsilon, \eta, r_1) = 2\alpha_2(\eta)/\epsilon\alpha_3(r_1)$ and see

$$\inf_{X\in\mathfrak{S}_\vartheta(x,W)} \mathbf{P}^X[\tau_1 \wedge \tau_2 < T] \geq 1 - \frac{\epsilon}{2}.$$

Now, let $\eta = \eta(\epsilon, r_2)$ be selected according to Remark 3.3.3 based on the Pr-US property, such that $\mathbf{P}^X[\tau_2 = \infty] \geq 1 - \epsilon/2$ whenever $|x|_A \leq \eta$. Therefore, for $|x|_A \leq \eta$, we have for all $X \in \mathfrak{S}_\vartheta(x, W)$,

$$1 - \frac{\epsilon}{2} \leq \mathbf{P}^X[\tau_1 \wedge \tau_2 < T] \leq \mathbf{P}^X[\tau_1 < T] + \mathbf{P}^X[\tau_2 < T] \leq \mathbf{P}^X[\tau_1 < T] + \frac{\epsilon}{2},$$

which complete the proof by letting $r = r_1$.

□

Proposition 3.3.23. *Suppose $A \subset \mathcal{D}$ is compact. If there exists an SLF V w.r.t. A, then A is Pr-UAS for $\mathscr{S}_\vartheta$.*

Proof. By a standard supermartingale argument [100, Lemma 1, Chap II], we can show that the existence of SLF implies Pr-US. To show Pr-UA, let $r \in (0, R/2)$, then by Pr-US and Remark 3.3.3, there exists a $k \in (0, r)$ such that $\sup_{X\in\mathfrak{S}_\vartheta(x,W)} \mathbf{P}^X[|X_t|_A \leq r\ \forall t \geq s(\varpi)] \geq 1 - \frac{\epsilon}{2}$ whenever

$|X_s|_A \le k$. Now let $\tau = \inf\{t \ge 0 : X_t \in \mathbb{B}_k(A)\}$. By Lemma 3.3.22, there exists some $\eta > 0$ such that we can find a $T = T(\epsilon/2, \eta, k)$ to make

$$\inf_{X \in \mathfrak{S}_\vartheta(x,W)} \mathbf{P}^X[\tau < T] \ge 1 - \frac{\epsilon}{2}.$$

Therefore, for all $|x|_A < \eta$ and for all $X \in \mathfrak{S}_\vartheta(x, W)$,

$$\begin{aligned} &\mathbf{P}^X[|X_t|_A \le r,\ \forall t \ge T] \\ \ge &\mathbf{P}^X[\tau < T \text{ and } |X_t|_A \le r,\ \forall t \ge T] \\ \ge &\mathbf{P}^X[\tau < T]\mathbf{P}^X[|X_t|_A \le r,\ \forall t \ge \tau \mid \tau < T] \\ \ge &\mathbf{P}^X[\tau < T](1 - \epsilon/2) \ge (1 - \epsilon/2)^2 \ge 1 - \epsilon. \end{aligned} \tag{3.38}$$

□

The following result demonstrates that the existence of an SLF is sufficient to guarantee probabilistic stability with safety specifications with probabilities depending on initial conditions.

Theorem 3.3.24. *Suppose that $A \subset \mathcal{D}$ is compact and $\mathbb{B}_R(A) \subset \mathscr{N}$. If there exists an SLF $V \in (C^2(\mathbb{B}_R(A)); \mathbb{R}_{\ge 0})$ and some $G := \mathbb{B}_r(A)$ such that*

(1) $r \in (0, R]$ and $G \subset \mathcal{D}$,

(2) $\mathcal{X}_0 \subset G$,

then $\mathscr{S}_\vartheta$ satisfies the probabilistic stability with safety specification $\left(\mathcal{X}_0, A, U, 1 - \frac{\sup_{x \in \mathcal{X}_0} V(x)}{\alpha_1(r)}\right)$.

We need the following lemma to accomplish the proof.

Lemma 3.3.25. *For each $X \in \bigcup_{x \in \mathcal{X}_0} \mathfrak{S}_\vartheta(x, W)$, set $\tau := \inf\{t \ge 0 : X_t \in G^c\}$. Then for all $X \in \bigcup_{x \in \mathcal{X}_0} \mathfrak{S}_\vartheta(x, W)$,*

$$\mathbf{P}^X\left[\lim_{t \to \infty} |X_t|_A = 0 \mid \tau = \infty\right] = 1.$$

Proof. By a similar approach to Lemma 3.3.22, we set an arbitrary $r^* \in (0, r)$. By the Pr-US property, for all $X \in \mathfrak{S}_\vartheta(x, W)$ there should exist $\eta \in (0, r^*)$ such that for any $\epsilon \in (0, 1)$, $X_{\tau_\eta} \in \overline{\mathbb{B}}_\eta(A)$ implies $\mathbf{P}^X[X_t \in \mathbb{B}_{r^*}(A),\ \forall t \ge \tau_\eta] \ge 1 - \epsilon$, where

$$\tau_\eta = \inf\{t \ge 0 : X_t \in \mathbb{B}_\eta(A)\}.$$

By Dynkin's formula, for each weak solution we have

$$\begin{aligned} 0 &\le V(x) + \mathbf{E}^X \int_0^{\tau_\eta\wedge\tau\wedge t} \mathfrak{L}_d V(X(s))ds \\ &\le V(x) - \alpha_3(\eta)\mathbf{E}^X[\tau_\eta \wedge \tau \wedge t] \end{aligned} \tag{3.39}$$

Since that on $\{\tau_\eta \wedge \tau \ge t\}$ we have $\tau_\eta \wedge \tau \wedge t = t$, thus

$$\mathbf{E}^X[\tau_\eta \wedge \tau \wedge t] \ge \int_\Omega \mathbb{1}_{\{\tau_\eta\wedge\tau\ge t\}} \cdot t\, d\mathbf{P}^X(\omega) = t\mathbf{P}^X[\tau_\eta \wedge \tau \ge t],$$

combining with (3.39) we have

$$\mathbf{P}^X[\tau_\eta \wedge \tau \ge t] \le V(x)/t\alpha_3(\eta), \quad \text{for each } t, \tag{3.40}$$

which implies $\mathbf{P}^X[\tau_\eta \wedge \tau < \infty] = 1$ for all $X \in \bigcup_{x\in\mathcal{X}_0} \mathfrak{S}_\vartheta(x, W)$. On $\{\tau = \infty\}$, for all weak solution, we have $\mathbf{P}^X[\tau_\eta < \infty] = 1$ and

$$\begin{aligned} &\mathbf{P}^X[\limsup_{t\to\infty} |X_t|_A \le r^*] \\ \ge &\mathbf{P}^X[|X_t|_A \le r^*, \ \forall t \ge \tau_\eta \mid \tau_\eta < \infty] \ge 1 - \epsilon. \end{aligned}$$

Since ϵ and r^* are arbitrary, the conclusion follows. □

Remark 3.3.26. *Lemma 3.3.25 shows that SLF eliminate the possibility of safe sample paths up/down-crossing any neighborhood of A infinitely often. [100, Theorem 2, Chap II] demonstrates the same result by constructing the total time spent in $G \setminus \mathbb{B}_\epsilon(A)$ after time t and showing that it converges a.s. to 0 as $t \to \infty$.*

Proof of Theorem 3.3.24. The existence of SLF shows that A is Pr-UAS for $\mathscr{S}_\vartheta$. Now, for all $X \in \mathfrak{S}_\vartheta(x, W)$ with $x \in \mathcal{X}_0$, define $\tau := \inf\{t \ge 0 : X_t \in G^c\}$. Then, for all $t \ge 0$ and for all $X \in \mathfrak{S}_\vartheta(x, W)$,

$$\mathbf{E}^X[V(X_{\tau\wedge t})] = V(X_0) + \mathbf{E}^X\left[\int_0^{\tau\wedge t} \mathfrak{L}_d V(X_s)ds\right] \le V(x), \tag{3.41}$$

and, for all $t \ge 0$,

$$\mathbf{E}^X[V(X_{\tau\wedge t})] \ge \mathbf{E}^X[\mathbb{1}_{\{\tau\le t\}}V(X_\tau)] > \alpha_1(r)\mathbf{P}^X[\tau \le t], \tag{3.42}$$

which imply

$$\mathbf{P}^X[\tau \le t] < V(x)/\alpha_1(r), \quad \forall t \ge 0. \tag{3.43}$$

Sending $t \to \infty$ we get for all $X \in \bigcup_{x \in \mathcal{X}_0} \mathfrak{S}_\vartheta(x, W)$, $\mathbf{P}^X[\tau < \infty] < V(x)/\alpha_1(r)$, i.e.,

$$\inf_{X \in \mathfrak{S}_\vartheta(x,W)} \mathbf{P}^X[\tau = \infty] \geq 1 - \frac{\sup_{x \in \mathcal{X}_0} V(x)}{\alpha_1(r)}. \tag{3.44}$$

Since $\{\tau = \infty\} \subseteq \{\tau_\mathcal{D} = \infty\}$ and by Lemma 3.3.25, the conclusion follows. □

We have seen in the proof that conditions $\alpha_1(x) \leq V(x)$ and $\sup_{d \in \vartheta\overline{\mathbb{B}}} \mathfrak{L}_d V \leq 0$ play the role of guaranteeing the probabilistic set invariance. We refer these conditions as the stochastic barrier certificates. An application in control synthesis, termed as stochastic control barrier functions, has been shown in Section 3.2 with better safety probability compared to the zeroing-type barrier certificates [139], however, less effectiveness than the reciprocal-type barrier certificates. To provide stability with safety with probability one, one can combine SLF with the reciprocal-type barrier functions.

Theorem 3.3.27. *Let the same assumption in Theorem 3.3.24 be satisfied. Suppose there exists an SLF* $V \in (C^2(\mathbb{B}_R(A)); \mathbb{R}_{\geq 0})$*, some* $G := \mathbb{B}_r(A)$ *such that* $G \in \mathcal{D}$ *and* $\mathcal{X}_0 \subset G$*, as well as a function* $B \in (C^2(G); \mathbb{R}_{\geq 0})$ *satisfying*

(1) $\exists \alpha_1, \alpha_2 \in \mathcal{K}$ *s.t.*

$$\frac{1}{\alpha_1(|x|_A)} \leq B(x) \leq \frac{1}{\alpha_2(|x|_A)}, \quad \forall x \in G; \tag{3.45}$$

(2) $\exists \alpha_3 \in \mathcal{K}$ *s.t.*

$$\sup_{d \in \vartheta\overline{\mathbb{B}}} [\mathfrak{L}_d B(x) - \alpha_3(|x|_A)] \leq 0, \quad \forall x \in G. \tag{3.46}$$

Then $\mathscr{S}_\vartheta$ *satisfies the probabilistic stability with safety specification* $(\mathcal{X}_0, A, U, 1)$.

Proof. The proof is similar to Theorem 3.3.24. We rely on the SLF to provide the property shown in Lemma 3.3.25. Then the reciprocal type barrier function B guarantees that $\mathbf{P}[\tau = \infty] = 1$ [40, Theorem 1] for each weak solution. □

Remark 3.3.28. *Suppose* $U^c = \{x \in \mathbb{R}^n : h(x) \geq 0\}$ *where* h *is smooth, one can possibly enlarge* G *such that* $G \cap U \neq \emptyset$ *with* $\partial(G \cap U)$ *being piecewise smooth. To see the satisfaction of stability with safety specifications, along with the old conditions, one can introduce an extra reciprocal-type barrier function, denoted by* $\tilde{B}$*, and verify extra conditions that are similar to* (3.45) *and* (3.46) *by replacing* $|x|_A$ *with* $h(x)$.

3.3.4 Applications in Control Problems

In this section, based on the results from Section 3.3.2 and 3.3.3, we make a straightforward extension to a stochastic control Lyapunov-barrier characterization for $\mathscr{S}_\vartheta$ satisfying a probabilistic reach-avoid-stay specification $(\mathcal{X}_0, \Gamma, U, p)$ under controls. As a continuation of [122], we conduct a case study on enhancing the performance of jet engine compressors, under both noisy disturbances and bounded point mass perturbations, based on a reduced Moore-Greitzer nonlinear SDE model.

Probabilistic reach-avoid-stay control via stochastic control Lyapunov-barrier functions

We consider systems with controls as in Section 3.1. Recall (3.6) a nonlinear system of the form

$$dX_t = f(X_t)dt + g(X_t)\mathfrak{u}_t dt + b(X_t)dW_t + \vartheta\xi(t)dt.$$

Definition 3.3.29. *(Probabilistic reach-avoid-stay controllable): Given $\mathcal{X}_0, \Gamma \subseteq \mathcal{D}$ and $p \in [0, 1]$, $\mathscr{S}_\vartheta$ is said to be probabilistic reach-avoid-stay controllable w.r.t. $(\mathcal{X}_0, \Gamma, U, p)$, if there exists a Lipschitz continuous control strategy κ such that the system $\mathscr{S}_\vartheta^\kappa$ satisfies the specification $(\mathcal{X}_0, \Gamma, U, p)$.*

The following result is a straightforward extension of Theorem 3.3.24.

Proposition 3.3.30. *Given $\mathcal{X}_0, \Gamma \subseteq \mathcal{D}$, if there exists a smooth function $V \in (C^2(\mathbb{B}_R(A)); \mathbb{R}_{\geq 0})$ and some $G := \mathbb{B}_r(A)$, such that*

(1) $r \in (0, R]$, $G \subset \mathcal{D}$ and $\mathcal{X}_0 \subset G$;

(2) $\alpha_1(|x|_A) \leq V(x) \leq \alpha_2(|x|_A)$ and

$$\inf_{u\in\mathcal{U}} \sup_{x\in S} \sup_{d\in\vartheta\overline{\mathbb{B}}} [\mathfrak{L}_d^u V(x) + \alpha_3(|x|_A)] \leq 0,$$

for some $\alpha_1, \alpha_2, \alpha_3 \in \mathcal{K}$, where $\mathfrak{L}_d^u V(x) := \mathfrak{L}_d V(x) + \nabla V(x) \cdot g(x)u$ (see Page 80).

Then $\mathscr{S}_\vartheta$ is probabilistically reach-avoid-stay controllable w.r.t. $\left(\mathcal{X}_0, \Gamma, U, 1 - \frac{\sup_{x\in\mathcal{X}_0} V(x)}{\alpha_1(r)}\right)$.

Similarly, one can extend the above proposition to find sufficient conditions for a 'probability 1' reach-avoid-stay based on Theorem 3.3.27. Apart from the conditions in Proposition 3.3.30, one need to additionally verify if there exists a $B \in (C^2(G); \mathbb{R}_{\geq 0})$ satisfying

(1) $\frac{1}{\tilde{\alpha}_1(|x|_A)} \leq B(x) \leq \frac{1}{\tilde{\alpha}_2(|x|_A)}, \quad \forall x \in G$ for some class-$\mathcal{K}$ functions $\tilde{\alpha}_1, \tilde{\alpha}_2$;

(2) $\inf_{u\in\mathcal{U}} \sup_{x\in S} \sup_{d\in\vartheta\overline{\mathbb{B}}} [\mathfrak{L}_d^u B(x) - \tilde{\alpha}_3(|x|_A)] \leq 0$, for some class-$\mathcal{K}$ function $\tilde{\alpha}_3$.

Remark 3.3.31. *In view of Remark 3.3.28, the region G can be further relaxed if ∂U is smooth enough. Correspondingly, some extra conditions given by another reciprocal control barrier function are needed to guarantee the sufficiency of $(\mathcal{X}_0, \Gamma, U, 1)$ controllability.*

Case study

We use the reduced Moore-Greitzer SDE model with an additive control input $[v, 0]^T$ and a multiplicative noise to illustrate the effectiveness. The model is given as:

$$\begin{aligned} \frac{d}{dt}\begin{bmatrix}\Phi(t)\\ \Psi(t)\end{bmatrix} &= \begin{bmatrix}\frac{1}{l_c}(\psi_c - \Psi(t))\\ \frac{1}{16 l_c}\left(\Phi(t) - \gamma\sqrt{\Psi(t)}\right)\end{bmatrix} + \vartheta\begin{bmatrix}\xi^{(1)}(t)\\ \xi^{(2)}(t)\end{bmatrix} \\ &\quad + \varepsilon\begin{bmatrix}(\Phi(t) - \Phi_e(\gamma))\beta^{(1)}(t)\\ (\Psi(t) - \Psi_e(\gamma))\beta^{(2)}(t)\end{bmatrix} + \begin{bmatrix}v(t)\\ 0\end{bmatrix}, \end{aligned} \tag{3.47}$$

where $\beta^{(1)}, \beta^{(2)}$ are i.i.d. Brownian motions, $(\Phi_e(\gamma), \Psi_e(\gamma)) =: X_e(\gamma)$ are equilibrium points for $\xi^{(1)}, \xi^{(2)}, v \equiv 0$. The engine parameters (as in Section 1.2) are the same as the settings in Section 2.4.2. The intensity of noises are such that

$$\varepsilon = 0.08,\ \ \vartheta = 0.001.$$

For $\xi^{(1)}, \xi^{(2)}, v \equiv 0$, the system admits a family of equilibrium points $X_e(\gamma)$ depending on the tunable parameter γ. As γ drops in the neighborhood of the deterministic Hopf bifurcation point, the system undergoes a D-bifurcation (the stability of the invariant measure $\delta_{\{X_e\}}$ changes and a new invariant measure in $\mathbb{R}^n \setminus \{X_e\}$ is built up) and a P-bifurcation (the shape of density of the new measure changes). The full stochastic Hopf bifurcation diagram in [8, Fig 9.13] conveys the brief idea.

The pictures in Figure 3.1 and 3.2 depict sample paths given initial condition

$$(\Phi_0, \Psi_0) = (0.5343, 0.6553)$$

with $\xi^{(1)}, \xi^{(2)}, v \equiv 0$ under different values of γ.

Within the a.s. exponentially stable region, any bounded perturbation ξ causes a bounded long-term perturbation of $X_e(\gamma)$, and ultimately formulate a compact set containing $X_e(\gamma)$. For unstable $\delta_{\{X_e\}}$, especially for those after P-bifurcation, we are interested in stabilizing the robust system to a compact set.

Problem 3.3.32. *We aim to manipulate γ and v simultaneously such that the state (Φ, Ψ) are regulated to satisfy reach-avoid-stay specification $(\mathcal{X}_0, \Gamma, U, 1)$. We require that $\mu : \mathbb{R}_{\geq 0} \to [0.5, 1]$ is time-varied with $\gamma(0) \in [0.62, 0.66]$ and $|\gamma(t + \mathfrak{t}) - \gamma(t)| \leq 0.01\mathfrak{t}$ for any $\mathfrak{t} > 0$. We define $\mathcal{X}_0 = \{(\Phi_e(\gamma(0)), \Psi_e(\gamma(0)))\}$; Γ to be the ball that centered at $\mathfrak{r} = (0.4519, 0.6513)$ with radius $r = 0.013$, i.e. $\Gamma = \mathfrak{r} + r\overline{\mathbb{B}}$; the unsafe set $U = \{(x, y) : h_1 \leq 0\} \cap \{(x, y) : h_2 \leq 0\}$, where $h_1(x, y) = -|(x, y) - (0.49, 0.64)| + 0.055$, $h_2(x, y) = |(x, y) - (0.50, 0.65)| - 0.003$. We set $v \in \mathcal{U} = [-0.05, 0.05] \cap \mathbb{R}$.*

For each SDE, the signals $\xi^{(1)}, \xi^{(2)}$ of each sampling time is generated randomly from $\{-1, 1\}$. We choose SLF $V(x, y) = \frac{\mathfrak{l}_c}{2}(x - \mathfrak{r}_1)^2 + 8\mathfrak{l}_c(y - \mathfrak{r}_2)^2$ and $\alpha_3(x) = 0.1x$; set $B_i = -\log\left(\frac{h_i}{1+h_i}\right)$ for $i = 1, 2$. The settings for the quadratic programming keep the same as Section 2.4.2. We mix sample paths under different $\xi^{(1)}, \xi^{(2)}$ and show the simulation results in Figure 3.3.

Remark 3.3.33. *Note that we have adopted reciprocal type barrier functions, which potentially generates impulse-like control signals (to cancel the diffusion effects) and terminates the programming. However, once the synthesis succeeds, the feasible controlled sample paths satisfy the specification.*

3.3.5 Proofs of Results from Section 3.3.2

Proof of Lemma 3.3.16. The proof for non-emptiness and probability-p invariance property (3.35) of A is similar to Lemma 2.3.15, we can show that the reachable set within (ramdom) time interval $[\tau_1, \tau_2]$ is a valid choice by a strong Markov property argument, where $\tau_1 := \inf\{t \geq 0 : X_t \in \Gamma\}$ and $\tau_2 := \inf\{t > \tau_1 : X_t \in \Gamma^c\}$.

Indeed, one can easily show that the reachable set $\bigcup_{X \in \bigcup_{x\in\mathcal{X}_0} \mathfrak{S}_\vartheta(x,W)} \mathscr{R}_\vartheta^{\tau_1 \leq t \leq \tau_2}(X) \subset \Gamma$, and by the strong Markov property, for every $X \in \bigcup_{x\in\mathcal{X}_0} \mathfrak{S}_\vartheta(x, W)$, any restarted solution, denoted by $\tilde{X} \in \mathfrak{S}_\vartheta(X_{\tau_3}, W))$ where $\tau_3 \in [\tau_1, \tau_2)$, has the same law as X_{τ_3+s} for all $s \geq 0$. Continuing the above, for all $X \in \bigcup_{x\in\mathcal{X}_0} \mathfrak{S}_\vartheta(x, W)$,

$$\begin{aligned} &\mathbf{P}[\tilde{X}_{t\wedge\tau_{\mathcal{D}}} \in \Gamma, \ \forall t \geq 0] \\ &\geq \inf_{X \in \bigcup_{x\in\mathcal{X}_0} \mathfrak{S}_\vartheta(x,W)} \mathbf{P}[\tau_1 < \infty \text{ and } \tau_2 = \infty] \geq p. \end{aligned} \tag{3.48}$$

By (3.48), $\bigcup_{X\in\mathfrak{S}_\vartheta(\mathcal{X}_0,W)} \mathscr{R}_\delta^{\tau_1 \leq t \leq \tau_2}(X) \subset A$. The probability-$p$ invariance is again by a standard strong Markov property argument and the definition of A.

To show that A is closed, let x_n be a sequence in A such that $x_n \to x \in \Gamma$. We need to show that x is also in A. Suppose the opposite, then there exists some $X \in \mathfrak{S}_\vartheta(x, W)$ such that

$$\mathbf{P}[X_{t\wedge\tau_\mathcal{D}} \in \Gamma,\ \forall t \geq 0] < p \Leftrightarrow \mathbf{P}[\tau < \infty] \geq 1 - p,$$

where $\tau := \inf\{t \geq 0 : X_t \in \Gamma^c\}$. Due to the weak compactness of the solution, by Skorohod, there exists a probability space $(\tilde{\Omega}^\dagger, \tilde{\mathscr{F}}^\dagger, \{\tilde{\mathscr{F}}_t^\dagger\}, \tilde{\mathbb{P}}^\dagger)$, a process $\tilde{X}^{\tau_\mathcal{D}}$ and a sequence of processes $\{(\tilde{X}^n)^{\tau_\mathcal{D}}\}$ with laws $\mathbf{P}$ and $\{\mathbf{P}^n\}$, respectively, such that

$$\lim_{n\to\infty} (\tilde{X}^n)^{\tau_\mathcal{D}} = \tilde{X}^{\tau_\mathcal{D}}, \quad \tilde{\mathbb{P}}^\dagger - \text{a.s.}$$

on every $[0, T]$. On each $\{\tau \leq T\}$, $\tilde{X}_\tau^n \to \tilde{X}_\tau \notin \Gamma$, and since Γ^c is open, for n sufficiently large, we have $\tilde{X}_\tau^n \notin \Gamma$. The above shows that $\tau \leq T \implies \exists t \in [0, T]$ s.t.$X^n \notin \Gamma$ for all T. Therefore, for sufficiently large n, sending T to infinity, we have $\mathbf{P}^n[\exists t \geq 0 \text{ s.t. } X_t^n \notin \Gamma] \geq 1 - p$, which violates the probabilistic invariance of A. Hence, $x \in A$. The boundedness of A is from the compactness of Γ. □

Proof of Lemma 3.3.17. The proof is similar to [114, Lemma 15] and (hybrid case) except in the context of weak solutions. We construct $Z_{s\wedge\tau_\mathcal{D}} = X_{s\wedge\tau_\mathcal{D}} + \frac{s}{T}[Z_{T\wedge\tau_\mathcal{D}} - X_{T\wedge\tau_\mathcal{D}} + (X_0 - Z_0)] + Z_0 - X_0$ for all $s \in [0, T]$. Then

$$|Z_{s\wedge\sigma} - X_{s\wedge\sigma}| \leq |Z_{T\wedge\sigma} - X_{T\wedge\sigma}|\frac{s}{T} + |X_0 - X_0|(1 - \frac{s}{T}) \leq r \tag{3.49}$$

For any text function $\phi \in C^\infty(\mathbb{R}^n)$, we define processes

$$M^\phi(t) = \phi(Z_{t\wedge\tau_\mathcal{D}}) - \phi(Z_0) - \int_0^{t\wedge\tau_\mathcal{D}} \nabla\phi(Z_s) \cdot f(Z_s) + \frac{1}{2}\operatorname{tr}\left[(bb^T)(Z_s) \cdot \phi_{xx}(Z_s)\right] ds \tag{3.50}$$

$$N^\phi(t) = \phi(X_{t\wedge\tau_\mathcal{D}}) - \phi(X_0) - \int_0^{t\wedge\tau_\mathcal{D}} \nabla\phi(X_s) \cdot f(X_s) + \frac{1}{2}\operatorname{tr}\left[(bb^T)(X_s) \cdot \phi_{xx}(X_s)\right] ds \tag{3.51}$$

as well as a martingale

$$\hat{M}^\phi(t) = \phi(X_{t\wedge\tau_\mathcal{D}}) - \phi(X_0) - \int_0^{t\wedge\tau_\mathcal{D}} \mathfrak{L}_d\phi(X_s)ds \tag{3.52}$$

One can show that $|M^\phi(t) - \hat{M}^\phi(t)|$ has a bound $B(T, r, \phi, \vartheta')$ based on the properties of f, g, ϕ and $Z^{\tau_\mathcal{D}}$, i.e.,

$$\begin{aligned} |M^\phi(t) - \hat{M}^\phi(t)| &\leq |M^\phi(t) - N^\phi(t)| + |N^\phi(t) - \hat{M}^\phi(t)| \\ &\leq (2C_1 r + TC_2 r) + T\sup_{x\in\mathcal{D}} \nabla\phi(x) \cdot \vartheta', \end{aligned} \tag{3.53}$$

where C_1 is generated due to the choice of ϕ, C_2 is due to the properties of f, g and ϕ.

To make M^ϕ a family of martingales under the family of laws of the stopped process of $\mathscr{S}_\vartheta$, one also needs to guarantee that $|M^\phi(t) - \hat{M}^\phi(t)| \leq T \sup_{x\in\mathcal{D}} \nabla\phi(x)\cdot\vartheta$ for all $t \in [0, T]$, since any two martingales of the martingale problem (under the laws of the corresponding stopped processes of $\mathscr{S}_\vartheta$) should not be differed larger than the above bound. Feasible ranges of r can be obtained based on the requirement

$$(2C_1 r + TC_2 r) + T\sup_{x\in\mathcal{D}} \nabla\phi(x)\cdot\vartheta' \leq T\sup_{x\in\mathcal{D}}\nabla\phi(x)\cdot\vartheta.$$

The process Z then satisfies the requirement given the feasible r. □

Proof of Proposition 3.3.18 We just show the sketch. Without loss of generality, we consider $\mathbb{B}_\epsilon(A) \subseteq \Gamma$. Suppose the claim is not true, then there exists some $\epsilon > 0$ such that for all $n > 0$ there exists $x_n \in \mathcal{X}_0$, $X^n \in \mathfrak{S}_{\vartheta'}(x_n, W)$ such that

$$\mathbf{P}^n[|X_t^n|_A > \epsilon,\ \exists t \geq n] > 1 - p.$$

We now show this leads to a contradiction. By the assumption, $\mathscr{S}_{\vartheta'}$ also satisfies the reach-avoid-stay specification $(\mathcal{X}_0, \Gamma, U, p)$. By a similar argument of weak compactness as in the proof of Lemma 3.3.16, there exists a sufficiently large N such that for all $n \geq N$, defining the Γ-entering time $\mathfrak{h} := \inf\{t \geq 0 : X_t^n \in \Gamma\}$, we are able to show that

$$\mathbf{P}^n[\exists t \geq n :\ X_{t+\mathfrak{t}}^n \notin \overline{\mathbb{B}}_\epsilon(A)] \geq 1 - p \tag{3.54}$$

Let $\tau := N$ in Lemma 3.3.17. Note that, by the construction in Lemma 3.3.17, $\inf_{x\in A}|x|_\Gamma = 0$ and we are able to find a process Z with $Z_0 \in A$ and $Z_0 \in \overline{\mathbb{B}}_r(X^n_{\mathfrak{h}\wedge\tau_\mathcal{D}})$ a.s.. By Lemma 3.3.17, there exists a process $Z \in \mathfrak{S}_\vartheta(Z_0, W)$ such that $Z_{t\wedge\tau_\mathcal{D}}$ and $X^n_{(t+\mathfrak{h})\wedge\tau_\mathcal{D}}$ share the same law for all $t \geq n$. However, $Z_{t\wedge\tau_\mathcal{D}} \in A$ for all $t \geq 0$ by the definition of A. Therefore,

$$\mathbf{P}[\exists t \geq n : Z_{t\wedge\tau_\mathcal{D}} \notin A] \leq \mathbf{P}[\exists t \geq 0 : Z_{t\wedge\tau_\mathcal{D}} \notin A] < 1 - p,$$

which contradicts (3.54). □

3.4 A Discussion on Lyapunov-Barrier Approaches for Unknown Dynamics

We have seen the existence of control stochastic Lyapunov-barrier (resp. stochastic barrier) functions is sufficient to guarantee probabilistic reach-avoid-stay (resp. safety) specifications given full knowledge of (3.6). However, in practice, we do not usually have precise information about the system dynamics, whence the generator $\mathfrak{L}^u$ (see definition in Page 80) as well as the set of valid control strategies are not determined.

In [126], the authors propose a framework of estimating the value of $\mathfrak{L}h(x)$ for some test function h at one specific point x within the domain. In this section, we extend this idea and discuss in what sense a data-driven method can approximate the function $\mathfrak{L}h$ (or $\mathfrak{L}^u h$) for some h in the whole domain, such that the approximation can be embedded in the stochastic Lyapunov-barrier framework.

To better convey the methodology, we simply set $\vartheta = 0$ and consider h as a potential SCBF (aiming at probabilistic safety controlling) for system $\mathscr{S}$. As a consequence, under some reasonable assumptions, the scheme introduce an extra robustness to the Itô derivative of an SCBF condition as in Section 3.2.

3.4.1 Worst-Case Probabilistic Quantification

Note that we only assume that the diffusion term is unknown in (3.6) with out loss of generality. Since our goal in this section is to use the approximated SCBF to regulate the safety direction, we consider a compact safe set $\mathcal{C}$, a bounded open set $G \supseteq \mathcal{C}$, and a function $h \in C^2(G) \cap C_b^2(\mathbb{R}^n)$ playing the role in (3.8). Due to the partial knowledge of the system dynamics, we are unable to capture the correction term in the Itô derivative of the nominal barrier function h along sample paths. In other words, the second term in $\mathfrak{L}h(x)$ (or $\mathfrak{L}^u h(x)$) is unknown. We use a data-driven method to approximate the function $\mathfrak{L}h$, and impose similar barrier conditions on the approximated $\hat{\mathfrak{L}}h$ for safety-critical control.

Based on the partial observation of data, we show that a degree of robustness (in $\mathcal{L}^1$ sense) in the barrier condition is necessary to balance the inaccuracy of data. A similar approach can be applied to derive the robustness for the other types of stochastic control Lyapunov-barrier functions.

We suppose that data is sampled without control inputs. Then for each x, the law $\mathbf{P}^x$ of the process X is independent of u. However, the approximated $\hat{\mathfrak{L}}h$, and hence $\hat{\mathfrak{L}}^u h$, is used to generate barrier conditions for the control problem. We further define the stopping time

$$\tau := \inf\{t \geq 0 : X_t \in \partial\mathcal{C}\}$$

for each sampled process. Let $\mathfrak{C}$ denote a finite subset of $\mathcal{C}$.

Probability estimation based on partially observed data

We make the following assumptions for the rest of the derivation. We will show later that the assumptions are feasible for compact $\mathcal{C}$.

Assumption 3.4.1. *Let $\hat{\mathfrak{L}}h$ be the approximation of $\mathfrak{L}h$ based on the training set $\mathfrak{C}$. We assume that*

(1) For any $y \in \mathcal{C}$ and any $\epsilon > 0$, there exists an $x \in \mathfrak{C}$ such that[4]

$$\mathbf{E}^y \sup_{t\in[0,\tau]} |\hat{\mathfrak{L}}h(X_t) - \mathfrak{L}h(X_t)| \leq \mathbf{E}^x \sup_{t\in[0,\tau]} |\hat{\mathfrak{L}}h(X_t) - \mathfrak{L}h(X_t)| + \epsilon. \tag{3.55}$$

(2) For any $\varsigma \in (0, 1]$, there exists a probability measure $\mathbb{P}$ with marginals $\mathbf{P}^x$ for all $x \in \mathfrak{C}$ such that

$$\mathbb{E} \sup_{x\in\mathcal{C}} |\mathfrak{L}h(x) - \hat{\mathfrak{L}}h(x)| \leq \varsigma. \tag{3.56}$$

Furthermore, we assume that both $\hat{\mathfrak{L}}h$ and $\mathfrak{L}h$ are Lipschitz continuous on the compact set $\mathcal{C}$.

We apply the approximated function $\hat{\mathfrak{L}}h$ and show the worst-case safety probability of the controlled process under policy generated by the following robust scheme.

Proposition 3.4.2. *Suppose we are given arbitrary $\varsigma > 0, \varepsilon > 0$ and training set $\mathfrak{C}$. Let $\hat{\mathfrak{L}}h$ be generated as in Assumption 3.4.1. Suppose that $\sup_{u\in\mathcal{U}} \hat{\mathfrak{L}}^u h(x) \geq \varsigma + \epsilon$ for all $x \in \mathcal{C}$. Let $\mathfrak{K}(x) = \{u \in \mathcal{U} : \hat{\mathfrak{L}}^u h(x) \geq \varsigma + \epsilon\}$. Then for any $x \in \text{Int}(\mathcal{C})$ and a control signal $\mathfrak{u}$ such that $\mathfrak{u}_t = \kappa(X_t) \in \mathfrak{K}(X_t)$, we have*

$$\mathbf{P}^{x,\mathfrak{u}}[X_t^{\mathfrak{u}} \in \text{Int}(\mathcal{C}),\ 0 \leq t < \infty] \geq \frac{h(x)}{\sup_{y\in\mathcal{C}} h(y)}.$$

Proof. Let $c = \sup_{y\in\mathcal{C}} h(y)$ and set $V = c - h$. Then for all $x \in \mathcal{C}^\circ$, we have $V(x) > 0$ and $\hat{\mathfrak{L}}^{\mathfrak{u}} V(x) \leq -(\varsigma + \epsilon)$. Note that

$$\mathbf{E}^{x,\mathfrak{u}}[V(X_{\tau\wedge t}^{\mathfrak{u}})] = V(x) + \mathbf{E}^{x,\mathfrak{u}}\left[\int_0^{\tau\wedge t} \mathfrak{L}^{\mathfrak{u}} V(X_s^{\mathfrak{u}}) ds\right] \tag{3.57}$$

[4]Note that $\tau < \tau_{\text{ex}}$ with probability 1.

and by assumption,

$$\begin{aligned}
&\mathbf{E}^{x,\mathfrak{u}}\left[\int_0^{\tau\wedge t}\mathfrak{L}^{\mathfrak{u}}V(X_s^{\mathfrak{u}})ds\right]\\
=&\mathbf{E}^{x,\mathfrak{u}}\left[\int_0^{\tau\wedge t}\mathfrak{L}^{\mathfrak{u}}V(X_s^{\mathfrak{u}})-\hat{\mathfrak{L}}^{\mathfrak{u}}V(X_s^{\mathfrak{u}})\,ds\right]+\mathbf{E}^{x,\mathfrak{u}}\left[\int_0^{\tau\wedge t}\hat{\mathfrak{L}}^{\mathfrak{u}}V(X_s^{\mathfrak{u}})\,ds\right]\\
\leq&\int_0^{\tau\wedge t}\mathbf{E}^{x,\mathfrak{u}}|\mathfrak{L}^{\mathfrak{u}}V(X_s^{\mathfrak{u}})-\hat{\mathfrak{L}}^{\mathfrak{u}}V(X_s^{\mathfrak{u}})|\,ds-(\varsigma+\epsilon)\cdot(\tau\wedge t)\\
\leq&\int_0^{\tau\wedge t}\mathbb{E}\sup_{s\in[0,\tau]}|\mathfrak{L}^{\mathfrak{u}}V(X_s^{\mathfrak{u}})-\hat{\mathfrak{L}}^{\mathfrak{u}}V(X_s^{\mathfrak{u}})|\,ds-\varsigma\cdot(\tau\wedge t)\\
\leq&\int_0^{\tau\wedge t}\mathbb{E}\sup_{x\in\mathcal{C}}|\mathfrak{L}^{\mathfrak{u}}V(x)-\hat{\mathfrak{L}}^{\mathfrak{u}}V(x)|\,ds-\varsigma\cdot(\tau\wedge t)\leq 0,
\end{aligned} \tag{3.58}$$

where the fourth line of the above is to transfer information from arbitrary $x \in \mathcal{C}$ to the data used in $\mathfrak{C}$. The ϵ is cancelled based on (1) of Assumption 3.4.1, where the notation $\mathfrak{L}$ is also replaced by $\mathfrak{L}^{\mathfrak{u}}$ in that $b(x)$ in (3.6) is known and does not render any error of measurement. Hence, by (3.57), we have

$$\mathbf{E}^{x,u}[V(X_{\tau\wedge t}^{u})]\leq V(x),\quad \forall t\geq 0. \tag{3.59}$$

On the other hand, for all $t \geq 0$,

$$\begin{aligned}
\mathbf{E}^{x,\mathfrak{u}}[V(X_{\tau\wedge t}^{\mathfrak{u}})] &\geq \mathbf{E}^{x,\mathfrak{u}}[\mathbb{1}_{\{\tau\leq t\}}V(X_{\tau\wedge t}^{\mathfrak{u}})]\\
&\geq \mathbf{P}^{x,\mathfrak{u}}[\tau\leq t]\cdot\mathbf{E}^{x,\mathfrak{u}}[V(X^{\mathfrak{u}}(\tau)]\\
&> c\cdot\mathbf{P}^{x,\mathfrak{u}}[\tau\leq t].
\end{aligned} \tag{3.60}$$

Therefore, by (3.59) and (3.60), we have

$$\mathbf{P}^{x,\mathfrak{u}}[\tau\leq t]<\frac{V(x)}{c},\ \forall t\geq 0. \tag{3.61}$$

Sending $t \to \infty$ we get $\mathbf{P}^{x,\mathfrak{u}}[\tau < \infty] \leq \frac{V(x)}{c}$ for all $x \in \mathcal{C}^{\circ}$. Rearranging this we can obtain the conclusion. □

Remark 3.4.3. *Note that (2) in Assumption 3.4.1 indicates that the error of estimation should converge in $\mathcal{L}^1$, and cannot be replaced by 'in probability' in the sense that, for every ς, there exists a $\delta = \delta(\varsigma)$ such that* $\mathbf{P}\left[\sup_{x\in\mathcal{C}}|\mathfrak{L}^{\mathfrak{u}}h(x)-\hat{\mathfrak{L}}^{\mathfrak{u}}h(x)|>\delta\right]<\varsigma$. *The latter is not sufficient to show the last line of (3.57) in general.*

3.4.2 Feasibility of Assumptions

Note that, for the compact set $\mathcal{C}$ and for sufficiently dense training data, the conditions in Assumption 3.4.1 can be satisfied theoretically. We will show that both (1) and (2) of Assumption 3.4.1 require the selection of the training data but separately.

1) Justification of Assumption 3.4.1(1): We observe that for each x in a compact set $\mathcal{C}$, for any fixed $T > 0$, the quantity $\sup_{t\in[0,\tau\wedge T]} |\mathfrak{L}h(\cdot) - \hat{\mathfrak{L}}h(\cdot)|$ is a bounded function on the canonical space generated by $\mathcal{C}$ with measure $\mathbf{P}^x$. In view of Definition 3.1.6 and Proposition 3.3.9, the quantity

$$\left\{\mathbf{E}^x \sup_{t\in[0,\tau\wedge T]} |\mathfrak{L}h(X_t) - \hat{\mathfrak{L}}h(X_t)|\right\}_{x\in\mathcal{C}}$$

forms a compact set (in the conventional sense). By the boundedness assumption on $\mathcal{C}$, we have $\tau < \infty$ $\mathbf{P}^x$-a.s. for every $x \in \mathcal{C}$. Therefore, sending T to infinity, we still have the compactness for

$$\left\{\mathbf{E}^x \sup_{t\in[0,\tau]} |\mathfrak{L}h(X_t) - \hat{\mathfrak{L}}h(X_t)|\right\}_{x\in\mathcal{C}}.$$

By choosing $\mathfrak{C}$ sufficiently dense in $\mathcal{C}$, for each given $\epsilon > 0$, we are able to build the ϵ-net with centers in $\mathfrak{C}$ such that for any arbitrary $y \in \mathcal{C}$, there exists an $x \in \mathfrak{C}$ such that $\mathfrak{L}h - \hat{\mathfrak{L}}h$ are weakly ϵ-close to each other in the sense of (3.55).

We then verify the feasibility of (2) of Assumption 3.4.1.

2) Approximating $\mathfrak{L}h$ over a finite set: Note that, following a similar procedure as in [126], we are able to approximate $\mathfrak{L}h$ by some $\tilde{\mathfrak{L}}h$ at one single point $x \in \mathbb{R}^n$ at a time, whose precision is measured in the corresponding probability[5] $\mathbb{P}^x := \otimes_{i=1}^{\infty}\mathbf{P}^x$. However, to fit the assumption, we need the precision to be measured in $\mathcal{L}^1$ sense.

By [126, Theorem 6], for each $x \in \mathbb{R}^n$, we can utilize Dynkin's formula, the Lipschitz continuity of f, g, b, and the relation

$$\tilde{\mathfrak{L}}_1 h(x) = \frac{\mathbf{E}^x[h(X_{\tau_s})] - h(x)}{\tau_s}$$

at some deterministic sampling time τ_s to obtain the first-step approximation

$$|\tilde{\mathfrak{L}}_1 h(x) - \mathfrak{L}h(x)| \le \delta, \tag{3.62}$$

[5] In [126], the authors used $\mathbb{P}$, but in our context it is recast to be $\mathbb{P}^x$. The uniqueness of $\mathbb{P}^x$ is by Kolmogrov's extension theorem.

where $\delta = C_1\tau_s + C_2\sqrt{\tau_s}$, and $C_1, C_2 > 0$ are constants generated by Lipschitz continuity. The precision δ can be arbitrarily small.

The authors in [126] then applied the LLN to approximate $\mathbf{E}^x[h(X_{\tau_s})]$ (in the definition of $\tilde{\mathfrak{L}}_1$) by $\frac{1}{n}\sum_{i=1}^n h(X_{\tau_s}^{(i)})$ with i.i.d. $h(X_{\tau_s}^{(i)})$ draw from $\mathbf{P}^x$ at the marginal time τ_s. The approximation

$$\tilde{\mathfrak{L}}h = \frac{\frac{1}{n}\sum_{i=1}^n h(X_{\tau_s}^{(i)}) - h(x)}{\tau_s}$$

creates errors in probability w.r.t. $\mathbb{P}^x$ as in [126, Theorem 12], i.e., for each $\beta \in (0, 1]$, there exists a $\tilde{\delta}$ such that $\mathbb{P}^x[|\mathfrak{L}h(x) - \tilde{\mathfrak{L}}h(x)| \leq \tilde{\delta}] > 1 - \beta$.

Note that, the only place that we introduce $\mathbb{P}^x$ is when we use LLN. We need to leverage the convergence in the $\mathcal{L}^1$ sense, i.e., as $n \to \infty$,

$$\mathbb{E}^x\left|\frac{1}{n}\sum_{i=1}^n h(X_{\tau_s}^{(i)}) - \mathbf{E}^x[h(X_{\tau_s})]\right| \to 0. \tag{3.63}$$

This is indeed the case as an existing result (see Appendix B for details), even though it is seldom mentioned.

Combining (3.63) and (3.62), we can easily obtain that for each $x \in \mathbb{R}^n$, for any $\delta > 0$ (we abuse the notation that is different from the one used in (3.62)), there exists a sufficiently large n such that

$$\mathbb{E}^x\left|\tilde{\mathfrak{L}}h(x) - \mathfrak{L}h(x))\right| \leq \delta. \tag{3.64}$$

Repeating the same process for x over a finite set $\mathfrak{C}$ gives

$$\sup_{x\in\mathfrak{C}} \mathbb{E}\left[|\mathfrak{L}h(x) - \tilde{\mathfrak{L}}h(x)|\right] \leq \delta, \tag{3.65}$$

where $\mathbb{E}$ is the associated expectation w.r.t. $\mathbb{P} := \otimes_{x\in\mathfrak{C}}\mathbb{P}^x$.

Remark 3.4.4. *The above $\tilde{\mathfrak{L}}_1$ converges to the infinitesimal generator as $\tau_s \to 0$. To use Dynkin's formula together with $\tilde{\mathfrak{L}}_1$ to get the δ bound in the first-step approximation, the above scheme only fits for test functions $h \in C_b^2(\mathbb{R}^n)$. We noticed that the authors in [126] misused the condition and provided a case study based on $h(x) = x^2$. This partially explains the unintuitive phenomenon in [126, Fig. 2] that the error increases as the sampling time is refined. We can fix this problem by smoothing the h to be $C_b^2(\mathbb{R}^n \setminus G)$ such that the approximation of $\mathfrak{L}h(x)$ has guarantees inside $G \supseteq \mathcal{C}$, which is all we need. Alternatively, since our purpose is not to compare the influence of the sampling time τ_s, we can set $G \supseteq \mathcal{C}$ sufficiently large and make τ_s arbitrarily small in practice, such that the 'out-of-domain' samples generated by C^2 functions do not account for much of the mass.*

3) Optimization error: For any $\eta > 0$, we assume there exists an optimizer that can learn an approximation $\hat{\mathfrak{L}}h$ based on data $\left\{\tilde{\mathfrak{L}}h(x) : \ x \in \mathfrak{C}\right\}$ such that

$$\sup_{x\in\mathfrak{C}} |\hat{\mathfrak{L}}h(x) - \tilde{\mathfrak{L}}h(x)| < \eta. \tag{3.66}$$

4) Generalization error: By continuity of $\hat{\mathfrak{L}}h(x)$ and $\mathfrak{L}h(x)$, there exists some $x^* \in \mathcal{C}$ such that

$$\sup_{x\in\mathcal{C}} |\hat{\mathfrak{L}}h(x) - \mathfrak{L}h(x)| = |\hat{\mathfrak{L}}h(x^*) - \mathfrak{L}h(x^*)|.$$

For any $\theta > 0$, by choosing $\mathfrak{C}$ to be sufficiently dense in $\mathcal{C}$ and the Lipschitz continuity of $\hat{\mathfrak{L}}h(x)$ and $\mathfrak{L}h(x)$ on $\mathcal{C}$, there exists some $y \in \mathfrak{C}$ such that

$$|\hat{\mathfrak{L}}h(x^*) - \hat{\mathfrak{L}}h(y)| \le \theta, \quad |\mathfrak{L}h(x^*) - \mathfrak{L}h(y)| \le \theta.$$

It follows that

$$\begin{aligned}
&\mathbb{E}\sup_{x\in\mathcal{C}} |\hat{\mathfrak{L}}h(x) - \mathfrak{L}h(x)| = \mathbb{E}|\hat{\mathfrak{L}}h(x^*) - \mathfrak{L}h(x^*)| \\
=&\mathbb{E}|\hat{\mathfrak{L}}h(y) - \mathfrak{L}h(y) + \hat{\mathfrak{L}}h(x^*) - \hat{\mathfrak{L}}h(y) + \mathfrak{L}h(y) - \mathfrak{L}h(x^*)| \\
\le&\mathbb{E}|\hat{\mathfrak{L}}h(y) - \mathfrak{L}h(y)| + 2\theta \\
=&\mathbb{E}|\hat{\mathfrak{L}}h(y) - \tilde{\mathfrak{L}}h(y) + \tilde{\mathfrak{L}}h(y) - \mathfrak{L}h(y)| + 2\theta \\
\le&\mathbb{E}\sup_{y\in\mathfrak{C}} |\hat{\mathfrak{L}}h(y) - \tilde{\mathfrak{L}}h(y)| + \mathbb{E}|\tilde{\mathfrak{L}}h(y) - \mathfrak{L}h(y)| + 2\theta \\
\le&\eta + \sup_{y\in\mathfrak{C}} \mathbb{E}|\tilde{\mathfrak{L}}h(y) - \mathfrak{L}h(y)| + 2\theta \\
\le&\eta + \delta + 2\theta \le \varsigma,
\end{aligned}$$

where ς is from Assumption 3.4.1(2), provided that we choose η, δ, and θ sufficiently small.

Remark 3.4.5. *The final $\mathfrak{C}$ should be chosen based on all of the above criteria such that (1) and (2) of Assumption 3.4.1 can both be satisfied. An algorithm for neural network approximation of $\hat{\mathfrak{L}}h$ is developed in [168]. Case studies are also provided to validate the result in Proposition 3.4.2. The controlled sample paths using the training result $\hat{\mathfrak{L}}h$ embedded in the SCBF framework demonstrate effectiveness of the proposed method in fulfilling the probabilistic safety specification.*

3.5 Summary

In this chapter, we first compare the pros and cons of the existing frequently used R-SCBFs and Z-SCBFs. We propose SCBFs for probabilistic safety control and extend the worst-case safety probability estimation to high-order SCBFs. We show that the proposed SCBFs provide compromised trade-offs between the imposed control constraints and the conservatism in the estimation of safety probability, which are demonstrated both theoretically and empirically. To accurately obtain the probabilistic winning sets, it is necessary to capture how the probability measure is distorted by the input processes. However, this may be computationally challenging for stochastic control systems.

We then formulated stochastic Lyapunov-barrier functions to develop sufficient conditions on probabilistic reach-avoid-stay specifications. Given uncertainties of the model, robustness was taken into account such that a worst-case scenario is guaranteed. We characterized a general topological structure of the initial sets, target sets and unsafe sets under the stochastic settings, and discussed relaxations given the smoothness of the unsafe boundary. We investigated the effectiveness in a case study of jet engine compressor control problem. Despite of the potentially unbounded control inputs, the control version of SLF along with reciprocal-type barrier functions guarantee a (conditionally) probability-one satisfaction.

However, just like deterministic Lyapunov-like functions only providing a stability characterization of the solutions, the stochastic Lyapunov-type argument can only estimate a lower bound of 'satisfaction in probability/law' without solving the evolving states and distributions. It renders more difficulties of selecting Lyapunov/barrier functions under the restrictive geometric requirements of the initial conditions and unsafe sets. Another drawback of the current results is the lack of converse stochastic Lyapunov-barrier function theorems. We cannot construct the converse connection between the two specifications defined in Section 3.3.1 as well due to the diffusion effects.

We finally investigated how a data-driven approach can be embedded in the Lyapunov-barrier scheme dealing with safety-critical control of unknown stochastic systems. We demonstrated the possibility of combining approximations of the Itô derivative of potential SCBFs and the barrier scheme conditions, and showed that an $\mathcal{L}^1$-robustness should be considered to guarantee the probabilistic safety. A similar procedure can be applied to derive the robustness for the other types of stochastic control Lyapunov-barrier functions. One drawback of the current analysis is the lack of convergence rate w.r.t. the size of the training data and the sampling time.

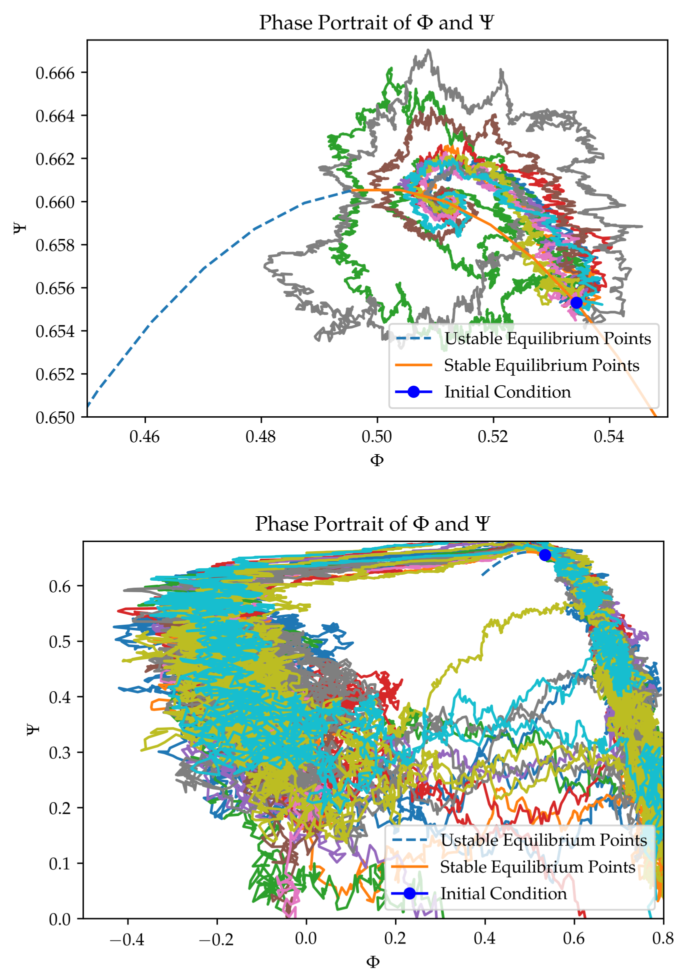


Figure 3.1: Sample paths of (3.47) given $\gamma = 0.63$ and $\gamma = 0.59$ respectively.

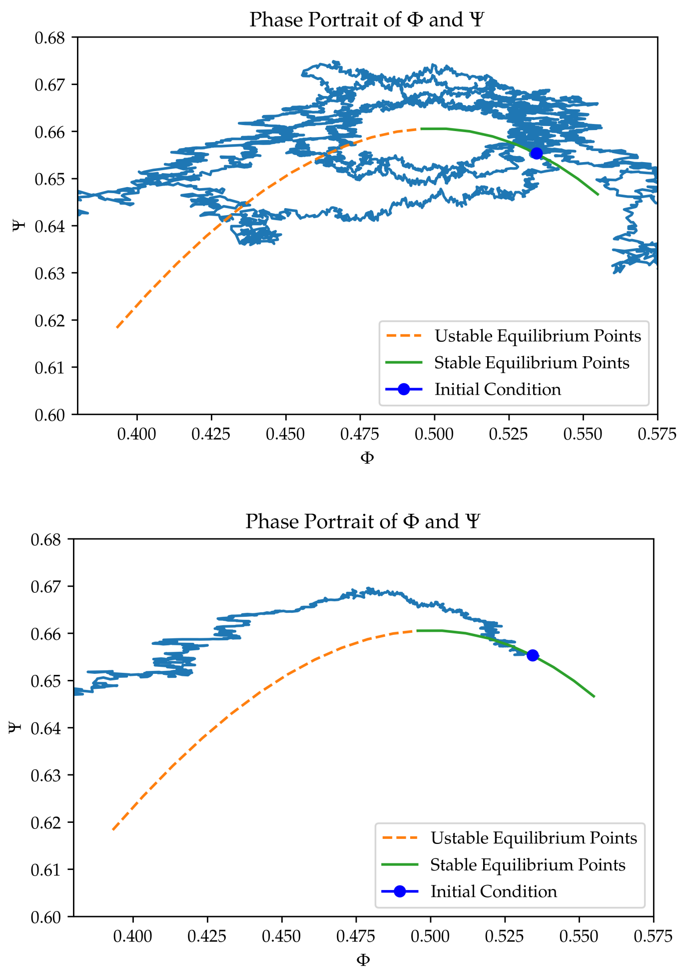


Figure 3.2: Convergent and divergent sample paths of (3.47) given $\gamma = 0.609$.

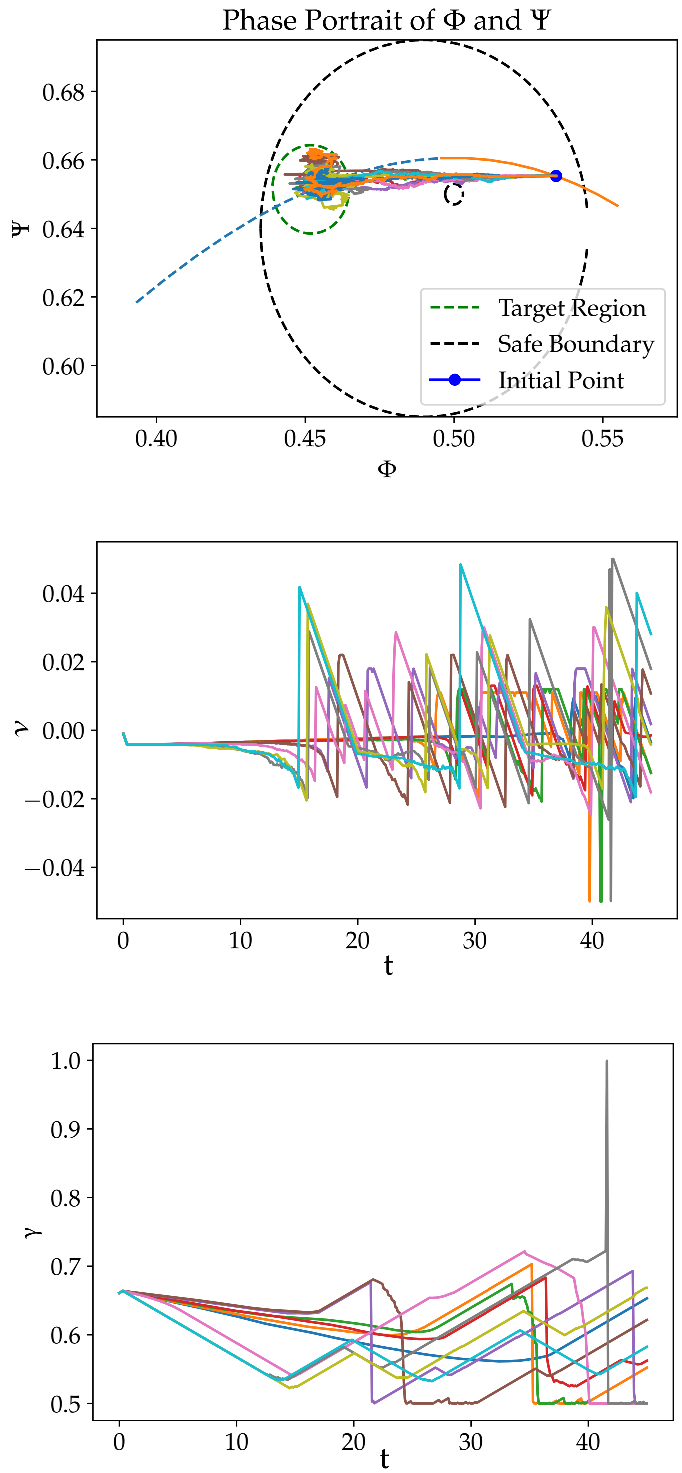


Figure 3.3: Controlled sample paths, control input v, γ of Problem 3.3.32.

Chapter 4

Robustly Complete Finite-State Abstractions for Verification and Control Synthesis of Stochastic Systems

We have seen in Chapter 2 the merits and drawbacks of abstraction-based formal methods and Lyapunov-barrier approaches, respectively, for verification and control synthesis of certain fundamental specifications in perturbed deterministic nonlinear systems. We also extended Lyapunov-barrier approaches for the stochastic counterpart and justified its sufficiency. Several concerns are raised due to the Itô diffusion: (1) it remains difficult to construct converse stochastic Lyapunov-barrier functions; (2) it becomes less accurate than its deterministic counterpart to estimate the set of initial conditions from which the solution processes or controlled processes satisfy the specifications of our interests. It is necessary to carry out abstraction-based formal methods to address those problems.

As motivated in Page 3, Section 1.1, the mathematical regularities of stochastic abstractions are not yet well understood. We investigate soundness and propose the concept of robust completeness for stochastic abstractions based on the topology of metrizable space of (uncertain) probability measures. We show that the technique proves more powerful than purely discussing the value of probabilities. We also would like to clarify that the main purpose of this chapter is not on providing more efficient algorithms for computing abstractions. We aim to provide a theoretical foundation of stochastic abstractions for continuous-state stochastic systems with additional uncertainties and hope to shed some light on designing more powerful algorithms.

4.1 Background

In Example 2.0.1, the philosophy of constructing finite-state abstraction[1] for deterministic systems was illustrated. The idea is to discretize the continuous state-space into finite grids as the state of the abstracted transition system (e.g. the r.h.s. picture in Figure 2.1) via some relation. The relation should guarantee that the transition between labels in the original systems should be included in the abstraction. The abstraction, which contains finite states and most-likely non-deterministic transitions, is then used to product with an automation generated by an LTL specification for graph search utilizing the existing model checking algorithms (recall Remark A.0.8). The soundness is guaranteed in the sense that, if a specification φ is realizable for the abstraction, so is for the original system.

On the other hand, the efficiency of the algorithm depends on the size of the product graph of the abstraction and the automation of an LTL specification. Apart from the complexity of the specified LTL formula, the size of the abstraction is a dominating factor. Whilst bisimilar or equivalent symbolic abstraction models exist for linear deterministic systems [96, 158], only approximately complete (that is incompleteness with a bright side) finite abstractions can be achieved via stability assumptions [136, 69] or robustness (in terms of Dirac perturbations) [112, 110, 113] for deterministic nonlinear systems. That is, over-approximation for nonlinear systems is inevitable when constructing the abstraction systems, i.e., unnecessary transitions between labels are included in the abstraction. However, [112] provided a perspective that such an inclusion can be arbitrarily asymptotically precise if we keep refining the grids. The analysis, however, depends on the introduction of robustness. To be more precise, given two transition systems S_1 and S_2 driven by point-mass perturbed discrete-time nonlinear difference equations

$$x(t+1) = f(x(t), u(t)) + \vartheta_i \xi(t),\ i = 1, 2,$$

where $\xi : \mathbb{R}_{\geq 0} \to \overline{\mathbb{B}}$ is a measurable signal with intensity ϑ_i for S_i, respectively. Briefly, [112, Theorem 2] indicates that, as long as $0 \leq \vartheta_1 < \vartheta_2$ and f is locally Lipschitz continuous in both arguments, we can find a relation and hence a finite-state transition system T such that

$$S_1 \preceq T \preceq S_2,$$

where $\preceq$ reads as 'abstracted by'. As $(\vartheta_2 - \vartheta_1) \to 0$, the T should intuitively contain more states and more precise transitions. A direct consequence of this result is that, if a given LT specification realizable, we can algorithmically realize the same specification for a (potentially

[1] In deterministic settings, we simply use 'finite abstraction' in terms of the finite number of states/nodes/grids in the abstracted transition systems. However, in the stochastic context, there are two levels of finiteness, i.e., the state/grid level and measure level. Detailed explanation is provided throughout this chapter.

less) robust system. Since the proof of above theorem is constructive, we can algorithmically verify or synthesize a control strategy for S_1 by computing T first and then solving a verification or control synthesis problem for T with the specification [112].

To this end, we aim to deliver the complete analysis for stochastic system in determining the mathematical size of the stochastic abstractions. We focus on the control-free systems, then we extend the methodology to control systems.

Conventions for Notation: We denote by $\prod$ the product of ordinary sets, spaces, or function values. Denote by $\otimes$ the product of collections of sets, or sigma algebras, or measures. The n-times repeated product of any kind is denoted by $(\cdot)^n$ for simplification. Denote by $\pi_j : \prod_{i=0}^{\infty}(\cdot)_i \to (\cdot)_j$ the projection to the j^{th} component.

Let $|\cdot|$ denote the inifinity norm in $\mathbb{R}^n$. Given a matrix M, we denote by M_i its i^{th} row and by M_{ij} its entry at i^{th} row and j^{th} column. The other commonly used notations follow the previous sections.

4.2 Robustly Complete Finite-State Abstractions for Verification of Stochastic Systems

4.2.1 Preliminaries

We consider $\mathbb{N} = \{0, 1, \cdots\}$ as the discrete time index set, and a general Polish (complete and separable metric) space $\mathcal{X}$ as the state space. For any discrete-time $\mathcal{X}^\infty$-valued stochastic process X, we introduce some standard concepts as follows.

Canonical spaces for discrete-time control-free stochastic systems

We have seen the canonical spaces for continuous-time control-free stochastic systems in Chapter 3. This notion can be defined in an even easier way for the discrete-time settings. Given a stochastic process X defined on some (most likely unknown) probability space $(\Omega^\dagger, \mathscr{F}^\dagger, \mathbb{P}^\dagger)$. For $\varpi \in \mathcal{X}^\infty =: \mathbf{\Omega}$ and $t \in \mathbb{N}$, we define $\varpi_t := \pi_t(\varpi)$ and the coordinate process $\mathfrak{X}_t : \mathcal{X}^\infty \to \mathcal{X}$ as $\mathfrak{X}_t(\varpi) := \varpi_t$ associated with $\mathcal{F} := \sigma\{\mathfrak{X}_0, \mathfrak{X}_1, \cdots\}$. Then $\Omega^\dagger \longrightarrow \mathcal{X}^\infty$ $(\omega^\dagger \longmapsto \prod_{t=0}^{\infty} X_t(\omega^\dagger))$ is a measurable map from $(\Omega^\dagger, \mathscr{F}^\dagger)$ to $(\mathbf{\Omega}, \mathcal{F})$. In particular, $\mathcal{F} = \sigma\{\mathfrak{X}_t \in \Gamma, \;\; \Gamma \in \mathscr{B}(\mathcal{X}), \;\; t \in \mathbb{N}\} = \mathscr{B}(\mathcal{X}^\infty) = \mathscr{B}^\infty(\mathcal{X}) = \sigma\{\mathcal{C}\}$, where $\mathcal{C}$ is the collection of all finite-dimensional cylinder set of the following form:

$$\prod_{i=1}^{n}\Gamma_i = \{\varpi : \mathfrak{X}_{t_1}(\varpi) \in \Gamma_1, \cdots, \mathfrak{X}_{t_n}(\varpi) \in \Gamma_n,\ \Gamma_i \in \mathscr{B}(\mathcal{X}), t_i \in \mathbb{N}, i = 1, \cdots, n\}.$$

The measure $\mathbf{P} := \mathbb{P}^{\dagger} \circ X^{-1}$ of the defined coordinate process $\mathfrak{X}$ is then uniquely determined and admits the probability law of the process X on the product state space, i.e.,

$$\mathbf{P}[\mathfrak{X}_{t_1} \in \Gamma_1, \cdots, \mathfrak{X}_{t_n} \in \Gamma_n] = \mathbf{P}\left(\prod_{i=1}^{n}\Gamma_i\right) = \mathbb{P}^{\dagger}[X_{t_1} \in \Gamma_1, \cdots, X_{t_n} \in \Gamma_n]. \tag{4.1}$$

for any finite-dimensional cylinder set $\prod_{i=1}^{n}\Gamma_i \in \mathcal{F}$. We call $(\mathbf{\Omega}, \mathcal{F}, \mathbf{P})$ the canonical space of X and denote by $\mathbf{E}$ the associated expectation operator.

Definition 4.2.1 (Clarification of Notations)**.** *In the specific context of discrete state space $\mathcal{X}$, we use the notation $(\Omega, \mathscr{F}, \mathcal{P})$ for the discrete canonical spaces of some discrete-state process. We would like to still use the notation $(\mathbf{\Omega}, \mathcal{F}, \mathbf{P})$ if the topology of $\mathcal{X}$ is not clear or not emphasized.*

Remark 4.2.2. *We usually denote by ν_i the marginal distribution of $\mathbf{P}$ at some $i \in \mathbb{N}$. We can informally write the n-dimensional distribution (on n-dimensional cylinder set) as $\mathbf{P}(\cdot) = \otimes_{i=1}^{n}\nu_i(\cdot)$ regardless of the dependence.*

Markov transition systems

Markov processes are defined in Appendix B. We recall the notation that, for discrete-time Markov processes, the one-step transition function at every $t \in \mathbb{T}$ is defined as

$$\Theta_t(x, \Gamma) := \mathbf{P}[X_{t+1} \in \Gamma \mid X_t = x],\ \ \Gamma \in \mathscr{B}(\mathcal{X}). \tag{4.2}$$

We denote correspondingly $\Theta_t := \{\Theta_t(x, \Gamma) :\ x \in \mathcal{X},\ \Gamma \in \mathscr{B}(\mathcal{X})\}$ as the family of one-step transition probabilities at time t. Homogeneous (autonomous) Markov processes are such that $\Theta_t = \Theta_s$ for all $t \neq s$, and the n-step transition can be recursively defined by $\Theta^{n+1}(x, \cdot) = \int_{\mathcal{X}} \Theta(y, \cdot)\Theta^n(x, dy)\nu_0(dx)$.

We are interested in Markov processes with discrete observations of states, which is done by assigning abstract labels over a finite set of atomic propositions. We define an abstract family of labelled Markov processes as follows.

Definition 4.2.3 (Markov system)**.** *A Markov system is a tuple $\mathbb{X} = (\mathcal{X}, [\![\Theta]\!], \mathrm{AP}, L)$, where*

- ✧ $\mathcal{X} = \mathcal{W} \cup \Delta$, *where* $\mathcal{W}$ *is a bounded working space,* $\Delta := \mathcal{W}^c$ *represents all the out-of-domain states;*
- ✧ $[\![\Theta]\!]$ *is a collection of transition probabilities from which* Θ_t *is chosen for every* t*;*
- ✧ AP *is the finite set of atomic propositions;*
- ✧ $L : \mathcal{X} \to 2^{\mathrm{AP}}$ *is the (Borel-measurable) labelling function.*

For $X \in \mathbb{X}$ *with* $X_0 = x_0$ *a.s., we denote by* $\mathbf{P}_X^{x_0}$ *the law, and* $\{\mathbf{P}_X^{x_0}\}_{X\in\mathbb{X}}$ *by its collection. Similarly, for any initial distribution* $\nu_0 \in \mathfrak{P}(\mathcal{X})$*, we define the law by* $\mathbf{P}_X^{\nu_0}(\cdot) = \int_{\mathcal{X}} \mathbf{P}_X^{x}(\cdot)\nu_0(dx)$*, and denote* $\{\mathbf{P}_X^{\nu_0}\}_{X\in\mathbb{X}}$ *by its collection. We denote by* $\{\mathbf{P}_n^{q_0}\}_{n=0}^{\infty}$ *(resp.* $\{\mathbf{P}_n^{\nu_0}\}_{n=0}^{\infty}$*) a sequence of* $\{\mathbf{P}_X^{x_0}\}_{X\in\mathbb{X}}$ *(resp.* $\{\mathbf{P}_X^{\nu_0}\}_{X\in\mathbb{X}}$*). We simply use* $\mathbf{P}_X$ *(resp.* $\{\mathbf{P}_X\}_{X\in\mathbb{X}}$*) if we do not emphasize the initial condition.*

For a path $\varpi := \varpi_0\varpi_1\varpi_2\cdots \in \mathbf{\Omega}$, define by $L_\varpi := L(\varpi_0)L(\varpi_1)L(\varpi_2)\cdots$ its trace (also see Appendix A). The space of infinite words is denoted by

$$(2^{\Pi})^{\omega} = \{A_0A_1A_2\cdots : A_i \in 2^{\Pi},\ i = 0, 1, 2\cdots\}.$$

A LT property is a subset of $(2^{\Pi})^{\omega}$. We are only interested in LT properties Ψ such that $\Psi \in \mathscr{B}((2^{\Pi})^{\omega})$, i.e., those are Borel-measurable.

Remark 4.2.4. *Note that, by [161] and [162, Proposition 2.3], any* ω*-regular language of labelled Markov processes is measurable. The proof relies on the properties of the canonical space with the fact that* $\mathcal{F} = \sigma\{\mathcal{C}\}$*, as well as the connection with Büchi automation. It follows that, for any Markov process* X *of the given* $\mathbb{X}$*, the traces* L_ϖ *generated by measurable labelling functions are also measurable. For each* $\Psi \in \mathscr{B}((2^{\Pi})^{\omega})$*, we have the event* $L_\varpi^{-1}(\Psi) \in \mathcal{F}$*.*

A particular subclass of LT properties can be specified by LTL[2] (see Appendix A for details). To connect with LTL specifications, we introduce the semantics of path satisfaction as well as probabilistic satisfaction as follows.

Definition 4.2.5. *Suppose* Ψ *is an LTL formulae. For a given labelled Markov process* X *from* $\mathbb{X}$ *with initial distribution* ν_0*, we formulate the canonical space* $(\mathbf{\Omega}, \mathcal{F}, \mathbf{P}_X^{\nu_0})$*. For a path* $\varpi \in \mathbf{\Omega}$*, we define the path satisfaction as*

$$\varpi \vDash \Psi \iff L_\varpi \vDash \Psi.$$

We denote by $\{X \vDash \Psi\} := \{\varpi :\ \varpi \vDash \Psi\} \in \mathcal{F}$ *the events of path satisfaction. Given a specified probability* $\rho \in [0, 1]$*, we define the probabilistic satisfaction of* Ψ *as*

$$X \vDash \mathcal{P}_{\bowtie\rho}^{\nu_0}[\Psi] \iff \mathcal{P}_X^{\nu_0}\{X \vDash \Psi\} \bowtie \rho,$$

where $\bowtie \in \{\leq, <, \geq, >\}$*.*

[2] While we consider LTL due to our interest, it can be easily seen that all results of this chapter in fact hold for any measurable LT property, including ω-regular specifications.

Discrete-time continuous-state stochastic systems

We define Markov processes determined by the difference equation

$$X_{t+1} = f(X_t) + b(X_t)\mathbf{w}_t + \vartheta\xi_t \tag{4.3}$$

where the state $X_t(\varpi) \in \mathcal{X} \subseteq \mathbb{R}^n$ for all $t \in \mathbb{N}$, the stochastic inputs $\{\mathbf{w}_t\}_{t\in\mathbb{N}}$ are i.i.d. Gaussian random variables with covariance $\mathrm{id}_{k\times k}$ without loss of generality. Mappings $f : \mathbb{R}^n \to \mathbb{R}^n$ and $b : \mathbb{R}^n \to \mathbb{R}^{n\times k}$ are locally Lipschitz continuous. The memoryless perturbation $\xi_t \in \overline{\mathcal{B}}$ are independent random variables with intensity $\vartheta \geq 0$ and unknown distributions.

For $\vartheta \neq 0$, (4.3) defines a family $\mathbb{X}$ of Markov processes X. A special case of (4.3) is such that ξ has Dirac (point-mass) distributions $\{\delta_x : x \in \mathbb{B}\}$ centered at some uncertain points within a unit ball.

Remark 4.2.6. *The discrete-time stochastic dynamic is usually obtained from numerical schemes of stochastic differential equations driven by Brownian motions to simulate the probability laws at the observation times. Gaussian random variables are naturally selected to simulate Brownian motions at discrete times. Note that in [51], random variables are used with known unimodal symmetric density with an interval as support. Their choice is in favor of the mixed-monotone models to provide a more accurate approximation of transition probabilities. Other than the precision issue, such a choice does not bring us more of the other $\mathcal{L}^1$ properties. Since we focus on formal analysis based on $\mathcal{L}^1$ properties rather than providing accurate approximation, using Gaussian randomnesses as a realization does not lose any generality.*

We only care about the behaviors in the bounded working space $\mathcal{W}$. By defining stopping time $\tau := \inf\{t \in \mathbb{N} : X \notin \mathcal{W}\}$ for each X, we are able to study the probability law of the corresponding stopped (killed) process X^τ for any initial condition x_0 (resp. ν_0), which coincides with $\mathbf{P}_X^{x_0}$ (resp. $\mathbf{P}_X^{\nu_0}$) on $\mathcal{W}$. To avoid any complexity, we use the same notation X and $\mathbf{P}_X^{x_0}$ (resp. $\mathbf{P}_X^{\nu_0}$) to denote the stopped processes and the associated laws. Such processes driven by (4.3) can be written as a Markov system

$$\mathbb{X} = (\mathcal{X}, [\![\mathcal{T}]\!], \mathrm{AP}, L_{\mathbb{X}}), \tag{4.4}$$

where for all $x \in \mathcal{X}\setminus\mathcal{W}$, the transition probability should satisfy $\mathcal{T}(x,\Gamma) = 0$ for all $\Gamma\cap\mathcal{W} \neq \emptyset$; $[\![\mathcal{T}]\!]$ is the collection of transition probabilities. For ξ having Dirac distributions, the transition $\mathcal{T}$ is of the following form:

$$\mathcal{T}(x,\cdot) \in \begin{cases} \{\mu \sim \mathcal{N}(f(x) + \vartheta\xi,\ \ b(x)b^T(x)),\ \xi \in \mathbb{B}\}, & \forall x \in \mathcal{W}, \\ \{\mu :\ \mu(\Gamma) = 0,\ \ \forall\Gamma \cap \mathcal{W} \neq \emptyset\}, & \forall x \in \mathcal{X} \setminus \mathcal{W}. \end{cases} \tag{4.5}$$

Assumption 4.2.7. *We assume that **in** $\in L(x)$ for any $x \notin \Delta$ and **in** $\notin L(\Delta)$. We can also include 'always (**in**)' in the specifications to observe sample paths for 'inside-domain' behaviors, which is equivalent to verifying $\{\tau = \infty\}$.*

Robust abstractions

We define a notion of abstraction between continuous-state and finite-state Markov systems via state-level relations and measure-level relations.

Definition 4.2.8. *A (binary) relation Σ from A to B is a subset of $A \times B$ satisfying*

(1) for each $a \in A$, $\Sigma(a) := \{b \in B : (a, b) \in \Sigma\}$;

(2) for each $b \in B$, $\Sigma^{-1}(b) := \{a \in A : (a, b) \in \Sigma\}$;

(3) for $A' \subseteq A$, $\Sigma(A') = \cup_{a \in A'} \Sigma(a)$;

(4) and for $B' \subseteq B$, $\Sigma^{-1}(B') = \cup_{b \in B'} \Sigma^{-1}(b)$.

Definition 4.2.9. *Given a continuous-state Markov system*

$$\mathbb{X} = (\mathcal{X}, [\![\mathcal{T}]\!], \mathrm{AP}, L_{\mathbb{X}})$$

and a finite-state Markov system

$$\mathbb{I} = (\mathcal{Q}, [\![\Theta]\!], \mathrm{AP}, L_{\mathbb{I}}),$$

where $\mathcal{Q} = (q_1, \cdots, q_n)^T$ and $[\![\Theta]\!]$ stands for a collection of $n \times n$ stochastic matrices.

We say that $\mathbb{I}$ abstracts $\mathbb{X}$, and write $\mathbb{X} \preceq_{\Sigma_\alpha} \mathbb{I}$, if there exist

(1) a state-level relation $\alpha \subseteq \mathcal{X} \times \mathcal{Q}$ from $\mathbb{X}$ to $\mathbb{I}$ such that, for all $x \in \mathcal{X}$, there exists $q \in \mathcal{Q}$ such that $(x, q) \in \alpha$ ($\alpha(x) \neq \emptyset$) and $L_{\mathbb{I}}(q) = L_{\mathbb{X}}(x)$;

(2) a measure-level relation $\Sigma_\alpha \subseteq \mathfrak{P}(\mathcal{X}) \times \mathfrak{P}(Q)$ from $\mathbb{X}$ to $\mathbb{I}$ such that, for all $i \in \{1, 2, \cdots, n\}$, all $\mathcal{T} \in [\![\mathcal{T}]\!]$ and all $x \in \alpha^{-1}(q_i)$, there exists $\Theta \in [\![\Theta]\!]$ such that $(\mathcal{T}(x, \cdot), \Theta_i) \in \Sigma_\alpha$ and that $\mathcal{T}(x, \alpha^{-1}(q_j)) = \Theta_{ij}$ for all $j \in \{1, 2, \cdots, n\}$.

Similarly, we say that $\mathbb{X}$ abstracts $\mathbb{I}$, and write $\mathbb{I} \preceq_{\Sigma_\alpha} \mathbb{X}$, if there exist

(1) a state-level relation $\alpha \subseteq \mathcal{Q} \times \mathcal{X}$ from $\mathbb{I}$ to $\mathbb{X}$ such that, for all $q \in \mathcal{Q}$, there exists an $x \in \mathcal{X}$ such that $(q, x) \in \alpha$ ($\alpha(q) \neq \emptyset$) and $L_{\mathbb{I}}(q) = L_{\mathbb{X}}(x)$;

(2) a measure-level relation $\Sigma_\alpha \subseteq \mathfrak{P}(Q) \times \mathfrak{P}(\mathcal{X})$ from $\mathbb{I}$ to $\mathbb{X}$ such that, for all $i \in \{1, 2, \cdots, n\}$, all $\Theta \in [\![\Theta]\!]$ and all $x \in \alpha^{-1}(q_i)$, there exists $\mathcal{T} \in [\![\mathcal{T}]\!]$ such that $(\Theta_i, \mathcal{T}(x, \cdot)) \in \Sigma_\alpha$ and that $\mathcal{T}(x, \alpha^{-1}(q_j)) = \Theta_{ij}$ for all $j \in \{1, 2, \cdots, n\}$.

Assumption 4.2.10. *Without loss of generality, we assume that the labelling function is amenable to a rectangular partition*[3]*. In other words, a state-level abstraction can be obtained from a rectangular partition.*

Remark 4.2.11. *Heuristically, we stand from the side of the original system and require an abstraction to*

- *contain states with the same labels as states of the original system;*
- *include transitional measures with the same measuring results on all the discrete states given any starting point of the original system that can be mapped to an abstract state.*

Given a rectangular partition, one immediate consequence of the existence of an abstraction is that the transition matrices are able to recover all possible transition probabilities (of the original system) from a grid to another.

4.2.2 Soundness of Robust IMC Abstractions

Definition 4.2.12. *An IMC is a tuple $\mathcal{I} = (\mathcal{Q}, \check{\Theta}, \hat{\Theta}, \mathrm{AP}, L_{\mathcal{I}})$, where*

- *$\mathcal{Q}$ is an N-dimensional state-space;*
- *AP and L are the same as in Definition 4.2.3;*
- *$\check{\Theta}$ is an $N \times N$ matrix such that $\check{\Theta}_{ij}$ is the lower bound of transition probability from the state number i to j for each $i, j \in \{1, 2, \cdots, N\}$;*
- *$\hat{\Theta}$ is an $N \times N$ matrix such that $\hat{\Theta}_{ij}$ is the upper bound of transition probability from the state number i to j for each $i, j \in \{1, 2, \cdots, N\}$.*

IMCs are 'quasi' Markov systems on a discrete state space with upper/under approximations ($\hat{\Theta}/\check{\Theta}$) of the real transition matrices. To abstract the transition probabilities of continuous-state Markov systems (4.4), $\hat{\Theta}$ and $\check{\Theta}$ are obtained from over/under approximations of $\mathcal{T}$ based on the state space partition. In this case, the state-level abstraction α is such that $\nu(\alpha^{-1}(q_j)) = \mathrm{vol}(q_j)$ for $\nu \in \mathfrak{P}(\mathcal{X})$ and each j. Throughout this section, we assume that $\hat{\Theta}$ and $\check{\Theta}$ have been correspondingly constructed.

[3] See e.g. [51, Definition 1].

Remark 4.2.13. *Suppose $\mathcal{Q} = (q_1, \cdots, q_N)^T$, then the over/under approximations are such that $\check{\Theta}_{ij} \le \int_{\alpha^{-1}(q_j)} \mathcal{T}(x, dy) \le \hat{\Theta}_{ij}$ for all $x \in q_i$ and $i, j \in \{1, \cdots, N\}$.*

Given an IMC $\mathcal{I}$, we recast it to a finite-state Markov system (recall Definition 4.2.3)

$$\mathbb{I} = (\mathcal{Q}, [\![\Theta]\!], \text{AP}, L_{\mathbb{I}}), \tag{4.6}$$

where

- ✧ $\mathcal{Q}$ is the finite state-space partition with dimension $N + 1$ containing $\{\Delta\}$, i.e., $\mathcal{Q} = (q_1, q_2, \cdots, q_N, \Delta)^T$;
- ✧ $[\![\Theta]\!]$[4] is a set of stochastic matrices satisfying

$$[\![\Theta]\!] = \{\Theta : \text{stochastic matrices with } \check{\Theta} \le \Theta \le \hat{\Theta} \text{ componentwisely}\}; \tag{4.7}$$

- ✧ AP is as before, and $L_{\mathbb{I}} = L_{\mathcal{I}}$.

To make $\mathbb{I}$ an abstraction for $\mathbb{X}$ as in (4.4) in the sense of Definition 4.2.9, we need the approximation to be such that $\check{\Theta}_{ij} \le \int_{\alpha^{-1}(q_j)} \mathcal{T}(x, dy) \le \hat{\Theta}_{ij}$ for all $x \in q_i$ and $i, j = 1, \cdots, N$, as well as $\Theta_{N+1} = (0, 0, \cdots, 1)$. We further require that the partition should respect the boundaries induced by the labeling function, i.e., for any $q \in \mathcal{Q}$,

$$L_{\mathbb{I}}(q) = L_{\mathbb{X}}(x), \; \forall x \in q.$$

Clearly, the above connections on the state and transition probabilities satisfy Definition 4.2.9.

Example 4.2.14. *Consider a continuous state space $\mathcal{X} \in \mathbb{R}^2$ and a discretization into 3-grids state space $\mathcal{Q} = (q_1, q_2, q_3)^T$, where q_i's are the symbolic nodes representing the whole area of $x \in \mathcal{X}$ in the grids. The space $\mathfrak{P}(\mathcal{Q})$, which is the hyperplane $q_1 + q_2 + q_3 = 1$, is illustrated in Figure 4.1.*

Let $\mathbb{X} = (\mathcal{X}, [\![\mathcal{T}]\!], \text{AP}, L_{\mathbb{X}})$ be a Markov system and $\mathbb{I} = (\mathcal{Q}, [\![\Theta]\!], \text{AP}, L_{\mathbb{I}})$ be the IMC abstraction, where $[\![\Theta]\!]$ is to be determined, and, for any $i \in \{1, 2, 3\}$ and for all $x \in \alpha^{-1}(q_i)$, $L_{\mathbb{X}}(x) = L_{\mathbb{I}}(q_i)$. Now we graphically illustrate how each row of $[\![\Theta]\!]$ is determined. We take the third row Θ_3 as an example.

Let t_1 and t_2 be any two consecutive time, then the transition probability from any $x \in \alpha^{-1}(q_3)$ to any $\alpha^{-1}(q_i)$ for $i \in \{1, 2, 3\}$ is given as $\mathcal{T}(x, \; \alpha^{-1}(q_i))$. Figure 4.2 shows an example of one-step transition probability $T(x, \alpha^{-1}(q_1))$ for any $x \in \alpha^{-1}(q_3)$. The under/over approximation

[4] This is a necessary step to guarantee proper probability measures in (4.8). Algorithms can be found in [83] or [102, Section V-A].

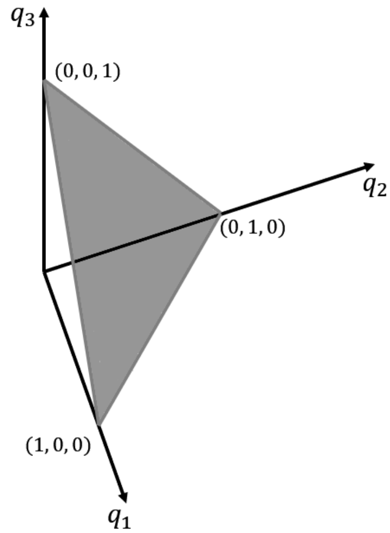


Figure 4.1: The space of probability measure $\mathfrak{P}(\mathcal{Q})$: the shaded area.

$\check{\Theta}_3$/$\hat{\Theta}_3$ can be generated accordingly using $\check{\Theta}_{3j} \leq \int_{\alpha^{-1}(q_j)} \mathcal{T}(x, dy) \leq \hat{\Theta}_{3j}$ for $j = 1, 2, 3$ and $x \in \alpha^{-1}(q_3)$. The third row of $[\![\Theta]\!]$ is such that $[\![\Theta_3]\!] = [\check{\Theta}_3, \hat{\Theta}_3] \cap \mathfrak{P}(\mathcal{Q})$, which is shown in Figure 4.3. The other rows of $[\![\Theta]\!]$ are obtained in a similar way. The tighter estimation of the upper/lower bounds of the transition probabilities, the smaller shaded area we can obtain for each row of $[\![\Theta]\!]$. $\square$

The Markov system $\mathbb{I}$ is understood as a family of 'perturbed' Markov chain generated by the uncertain choice of Θ for each t. The n-step transition matrices are derived based on $[\![\Theta]\!]$ as

$$\begin{aligned} [\![\Theta^{(2)}]\!] &= \{\Theta_0\Theta_1 : \quad \Theta_0, \Theta_1 \in [\![\Theta]\!]\}, \\ &\vdots \\ [\![\Theta^{(n)}]\!] &= \{\Theta_0\Theta_1 \cdots \Theta_n : \quad \Theta_i \in [\![\Theta]\!],\ i = 0, 1, \cdots, n\}. \end{aligned}$$

We aim to show the soundness of robust IMC as abstractions in this subsection. The proofs of results in this section are completed in Section .

Weak compactness of the set of marginal probabilities

We first investigate the topological property of the set of marginal measures of some $\mathbb{I}$. Suppose $\mathbb{I}$ is given as an abstraction of $\mathbb{X}$ as in (4.4).

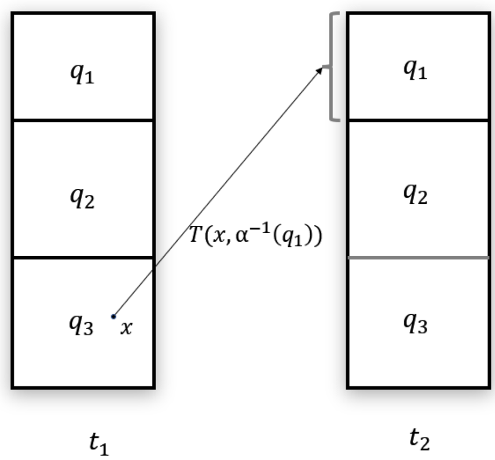


Figure 4.2: One-step transition probability $\mathcal{T}(x, \alpha^{-1}(q_1))$ of $\mathbb{X}$ from any $x \in \alpha^{-1}(q_3)$ to the area represented by q_1.

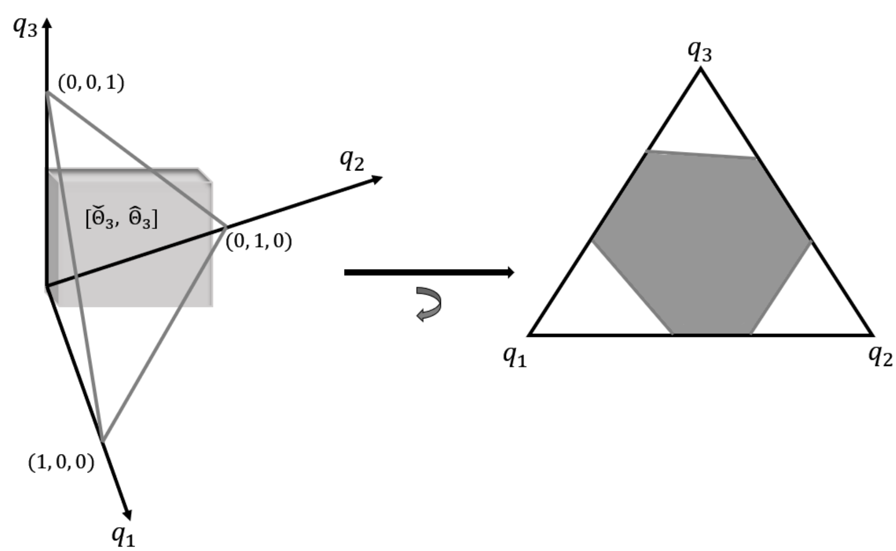


Figure 4.3: The intersection (shaded) of the box $[\check{\Theta}_3, \hat{\Theta}_3]$ and the space of measure $\mathfrak{P}(\mathcal{Q})$.

Definition 4.2.15. *Given an initial distribution $\mu_0 \in \mathfrak{P}(\mathcal{Q})$, the marginal probability measure at each t forms a set*

$$\mathfrak{P}(\mathcal{Q}) \supseteq \mathscr{M}_t^{\mu_0} := \{\mu_t = (\Theta^{(t)})^T \mu_0 : \quad \Theta^{(t)} \in [\![\Theta^{(t)}]\!]\}. \tag{4.8}$$

If we do not emphasize the initial distribution μ_0, we also use $\mathscr{M}_t$ to denote the marginals for short.

The following lemma is rephrased from [163, Theorem 2] and shows the structure of the $\mathscr{M}_t^{\mu_0}$ for each $t \in \mathbb{N}$ and any initial distribution μ_0.

Lemma 4.2.16. *Let $\mathbb{I}$ be a Markov system of the form (3.6) that is derived from an IMC. Then the set $\mathscr{M}_t$ of all possible probability measures at each time $t \in \mathbb{N}$ is a convex polytope, and immediately is compact. The vertices*[5] *of $\mathscr{M}_t$ are of the form*

$$(V_{i_t})^T \cdots (V_{i_2})^T (V_{i_1})^T \mu_0 \tag{4.9}$$

for some vertices V_{i_j} of $[\![\Theta]\!]$ and $j \in \{1, \cdots, t\}$.

Example 4.2.17. *Let $Q = (q_1, q_2, q_3)^T$ and $\mu_0 = (1, 0, 0)^T$. The under/over estimations of transition matrices are given as*

$$\check{\Theta} = \begin{bmatrix} \frac{1}{4} & 0 & \frac{1}{4} \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \hat{\Theta} = \begin{bmatrix} \frac{3}{4} & 0 & \frac{3}{4} \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

Then $[\![\Theta]\!]$ forms a convex set of stochastic matrices with vertices

$$V_1 = \begin{bmatrix} \frac{1}{4} & 0 & \frac{3}{4} \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad V_2 = \begin{bmatrix} \frac{3}{4} & 0 & \frac{1}{4} \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}.$$

Therefore, the vertices of $\mathscr{M}_1^{\mu_0}$ are

$$\nu_1^{(1)} = (V_1)^T \mu_0 = (\frac{1}{4}, 0, \frac{3}{4})^T, \quad \nu_1^{(2)} = (V_2)^T \mu_0 = (\frac{3}{4}, 0, \frac{1}{4})^T.$$

Hence, $\mathscr{M}_1^{\mu_0} = \{\mu : \mu = a\nu_1^{(1)} + (1-a)\nu_1^{(2)},\ a \in [0,1]\}$. Similarly, the vertices of $\mathscr{M}_2^{\mu_0}$ are

$$\nu_2^{(1)} = (V_1)^T (V_1)^T \mu_0 = (\frac{1}{16}, \frac{12}{16}, \frac{3}{16})^T, \quad \nu_2^{(2)} = (V_2)^T (V_1)^T \mu_0 = (\frac{3}{16}, \frac{12}{16}, \frac{1}{16})^T,$$
$$\nu_2^{(3)} = (V_1)^T (V_2)^T \mu_0 = (\frac{3}{16}, \frac{4}{16}, \frac{9}{16})^T, \quad \nu_2^{(4)} = (V_2)^T (V_2)^T \mu_0 = (\frac{9}{16}, \frac{4}{16}, \frac{3}{16})^T,$$

[5] This is a rather standard concept in geometry. More analytical details can be found in [83, Chapter 2].

and

$$\mathscr{M}_2^{\mu_0} = \{\mu : \mu = ab\nu_2^{(1)} + a(1-b)\nu_2^{(2)} + b(1-a)\nu_2^{(3)} + (1-a)(1-b)\nu_2^{(4)},\ a, b \in [0,1]\}.$$

The calculation of the rest of $\mathscr{M}_t^{\mu_0}$ follows the same procedure. □

Now we introduce the total variation distance $\|\cdot\|_{\mathrm{TV}}$ and see how $(\mathscr{M}_t, \|\cdot\|_{\mathrm{TV}})$ (at each t) implies the weak topology. We kindly refer readers to Appendix E for more details. In words, the weak convergence concept in the stochastic settings is an analogue of the pointwise convergence in the deterministic cases (see Example E.0.2 and Remark E.0.4). To describe the convergence (in probability law) of $\{X_n\}$ in $\mathcal{X}$, it is equivalent to investigate the weak convergence of their associated measures $\{\mu_n\}$. We emphasize the definition of total variation distance as follows.

Definition 4.2.18 (Total variation distance)**.** *Given two probability measures μ and ν on $\mathcal{X}$, the total variation distance is defined as*

$$\|\mu - \nu\|_{\mathrm{TV}} = 2 \sup_{\Gamma \in \mathscr{B}(\mathcal{X})} |\mu(\Gamma) - \nu(\Gamma)|. \tag{4.10}$$

In particular, if $\mathcal{X}$ is a discrete space,

$$\|\mu - \nu\|_{\mathrm{TV}}^d = \|\mu - \nu\|_1 = \sum_{q \in \mathcal{X}} |\mu(q) - \nu(q)|. \tag{4.11}$$

Corollary 4.2.19. *Let $\mathbb{I}$ be a Markov system of the form (3.6) that is derived from an IMC. Then at each time $t \in \mathbb{N}$, for for each $\{\mu_n\} \subseteq \mathscr{M}_t$, there exists a $\mu \in \mathscr{M}_t$ and a subsequence $\{\mu_{n_k}\}$ such that $\mu_{n_k} \rightharpoonup \mu$. In addition, for each $h \in C_b(\mathcal{X})$ and $t \in \mathbb{N}$, the set $H = \{\sum_{\mathcal{X}} h(x)\mu(x),\ \mu \in \mathscr{M}_t\}$ forms a convex and compact subset in $\mathbb{R}$.*

Remark 4.2.20. *The above shows that $\|\cdot\|_{\mathrm{TV}}$ implies the weak topology of measures on $\mathcal{Q}$. Note that since $\mathcal{Q}$ is bounded and finite, any metrizable family of measures on $\mathcal{Q}$ is compact. For example, let $\mathcal{Q} = \{q_1, q_2\}$, and $\{(0,1)^T, (1,0)^T\}$ be a set of singular measures on $\mathcal{Q}$. Then every sequence of the above set has a weakly convergent subsequence. However, these measures do not have a convex structure as $\mathscr{M}_t$. Hence, the corresponding H that is generated by $\{(0,1)^T, (1,0)^T\}$ only provides vertices in $\mathbb{Z}$.*

Weak compactness of probability laws of $\mathbb{I}$ on infinite horizon

To verify the probability of ω-regular properties, we need to consider the weak compactness of probability laws instead of only marginal properties. Taking the advantages of the compactness and convexity of $[\![\Theta]\!]$, it turns out the weak topology of the probability laws also possesses the same property as the marginal ones. The results are demonstrated as follows, and the proofs are completed in Section 4.2.6.

Without loss of generality, we focus on the case where $I_0 = q_0$ a.s. for any $q_0 \in \mathcal{Q}\backslash\{\Delta\}$. The cases for arbitrary initial distribution should be similar. We formally denote $\mathscr{M}^{q_0} := \{\mathcal{P}_I^{q_0}\}_{I\in\mathbb{I}}$ by the set of probability laws of every discrete-state Markov processes $I \in \mathbb{I}$ with initial state $q_0 \in \mathcal{Q}$.

Proposition 4.2.21. *For any $q_0 \in \mathcal{Q}$, every sequence $\{\mathcal{P}_n^{q_0}\}_{n=0}^{\infty}$ of $\mathscr{M}^{q_0}$ has a weakly convergent subsequence.*

Theorem 4.2.22. *Let $\mathbb{I}$ be a Markov system of the form (3.6) that is derived from an IMC. Then for any LTL formula Ψ, the set $S^{q_0} = \{\mathcal{P}_I^{q_0}(I \vDash \Psi)\}_{I\in\mathbb{I}}$ is a convex and compact subset in $\mathbb{R}$, i.e., a compact interval.*

Soundness of IMC abstractions

We are now ready to show the soundness of IMC abstractions.

Proposition 4.2.23. *Let $\mathbb{X}$ be a Markov system driven by (4.4). Then every sequence $\{\mathbf{P}_n^{x_0}\}_{n=0}^{\infty}$ of $\{\mathbf{P}_X^{x_0}\}_{X\in\mathbb{X}}$ has a weakly convergent subsequence. Consequently, for any LTL formula Ψ, the set $\{\mathbf{P}_X^{x_0}(X \vDash \Psi)\}_{X\in\mathbb{X}}$ is a compact subset in $\mathbb{R}$.*

Lemma 4.2.24. *Let $X \in \mathbb{X}$ be any Markov process driven by (4.4) and $\mathbb{I}$ be the finite-state IMC abstraction of $\mathbb{X}$. Suppose the initial distribution ν_0 of X is such that $\nu_0(q_0) = 1$. Then, there exists a unique law $\mathcal{P}_I^{q_0}$ of some $I \in \mathbb{I}$ such that, for any LTL formula Ψ,*

$$\mathbf{P}_X^{\nu_0}(X \vDash \Psi) = \mathcal{P}_I^{q_0}(I \vDash \Psi).$$

Combining Lemma 4.2.24 and Theorem 4.2.22, we obtain the main result of this subsection.

Theorem 4.2.25. *Assume the settings in Lemma 4.2.24. For any LTL formula Ψ, we have*

$$\mathbf{P}_X^{\nu_0}(X \vDash \Psi) \in \{\mathcal{P}_I^{q_0}(I \vDash \Psi)\}_{I\in\mathbb{I}}.$$

Corollary 4.2.26. *Let $\mathbb{X}$, its IMC abstraction $\mathbb{I}$, an LTL formula Ψ, and a constant $\rho \in [0, 1]$ be given. Suppose $I \vDash \mathcal{P}_{\bowtie\rho}^{q_0}[\Psi]$ for all $I \in \mathbb{I}$, we have $X \vDash \mathbf{P}_{\bowtie\rho}^{\nu_0}[\Psi]$ for all $X \in \mathbb{X}$ with $\nu_0(q_0) = 1$.*

Remark 4.2.27. *Note that we do not have $\mathbf{P}_X^{\nu_0} \in \{\mathcal{P}_I^{q_0}\}_{I\in\mathbb{I}}$ since each $\mathcal{P}_I^{q_0}$ is a discrete measure whereas $\mathbf{P}_X^{\nu_0}$ is not. They only coincide when measuring Borel subset of $\mathscr{F}$. It would be more accurate to state that $\mathbf{P}_X^{\nu_0}(X \vDash \Psi)$ is a member of $\{\mathcal{P}_I^{q_0}(I \vDash \Psi)\}_{I\in\mathbb{I}}$ rather than say "the true distribution (the law as what we usually call) of the original system is a member of the distribution set represented by the abstraction model" [102].*

Remark 4.2.28. *We have seen that, in view of Lemma 4.2.24, the 'post-transitional' measures are automatically related only based on the relations between transition probabilities. We will see in the next subsection that such relations can be constructed to guarantee an approximate completeness of $\mathbb{I}$.*

Proposition 4.2.29. *Let $\epsilon := \max_i \|\hat{\Theta}_i - \check{\Theta}_i\|_{TV}$. Then for each LTL formula Ψ, as $\epsilon \to 0$, the length $\lambda(S^{q_0}) \to 0$.*

Remark 4.2.30. *By Lemma 4.2.24, for each $X \in \mathbb{X}$, there exists exactly one $\mathcal{P}_I$ of some $I \in \mathbb{I}$ by which satisfaction probability equals to that of X. The precision of $\hat{\Theta}$ and $\check{\Theta}$ determines the size of S^{q_0}. Once we are able to calculate the exact law of X, the S^{q_0} becomes a singleton by Proposition 4.2.29. For example, let each $\mathbf{w}_t$ become δ_0, we have each $\mathscr{M}_t$ reduced to a singleton $\{\delta_{f(x_t)}\}$ automatically. The verification problem becomes checking whether $L(f(x_t)) \vDash \Psi$ given the partition $\mathcal{Q}$. The probability of satisfaction is either 0 or 1. Another example would be $X_{t+1} = AX_t + B\mathbf{w}_t$, where A, B are linear matrices. We are certain about the exact law of this system, and there is no need to introduce IMC for approximations at the beginning. IMC abstractions prove more useful when coping with systems whose distributions are uncertain or not readily computable.*

4.2.3 Robust Completeness of IMC Abstractions

In this section, we are given a Markov system $\mathbb{X}_1$ driven by

$$X_{t+1} = f(X_t) + b(X_t)\mathbf{w}_t + \vartheta_1 \xi_t^{(1)}, \;\; \xi_t^{(1)} \in \overline{\mathbb{B}}, \tag{4.12}$$

with point-mass perturbations of strength $\vartheta_1 \geq 0$. Based on $\mathbb{X}_1$, we first construct an IMC abstraction $\mathbb{I}$. We then show that $\mathbb{I}$ can be abstracted by a system $\mathbb{X}_2$ with more general $\mathcal{L}^1$-bounded noise of any arbitrary strength $\vartheta_2 > \vartheta_1$, i.e.,

$$X_{t+1} = f(X_t) + b(X_t)\mathbf{w}_t + \vartheta_2 \xi_t^{(2)}, \;\; \xi_t^{(2)} \in \overline{\mathcal{B}}, \tag{4.13}$$

Recalling the soundness analysis of IMC abstractions in Section 4.2.2, the relation of satisfaction probability is induced by a relation between the continuous and discrete transitions. To capture the probabilistic properties of stochastic processes, reachable set of probability measures is the analogue of the reachable set in deterministic cases. We rely on a similar technique in this section to discuss how transition probabilities of different uncertain Markov systems are related. To metricize sets of Gaussian measures and to connect them with discrete measures, we prefer to use Wasserstein metric $\|\cdot\|_{\mathrm{W}}$ (see Definition E.0.10). The details of probability metrics can be found in Appendix E.

Before proceeding, we define the set of transition probabilities of $\mathbb{X}_i$ from any box[6] $[x] \subseteq \mathbb{R}^n$ as

$$\mathbb{T}_i([x]) = \{\mathcal{T}(x,\cdot):\ \mathcal{T} \in [\![\mathcal{T}]\!]_i,\ x \in [x]\},\ i = 1, 2,$$

and use the following lemma to approximate $\mathbb{T}_1([x])$.

Lemma 4.2.31. *Fix any $\vartheta_1 > 0$, any box $[x] \subseteq \mathbb{R}^n$. For all $k > 0$, there exists a finitely terminated algorithm to compute an over-approximation of the set of (Gaussian) transition probabilities from $[x]$, such that*

$$\mathbb{T}_1([x]) \subseteq \widehat{\mathbb{T}_1([x])} \subseteq \mathbb{T}_1([x]) + k\overline{\mathcal{B}}_W,$$

where $\widehat{\mathbb{T}_1([x])}$ is the computed over-approximation set of Gaussian measures, and $\overline{\mathcal{B}}_W$ defined in Definition E.0.10.

Remark 4.2.32. *The proof is completed in Section 4.2.6. The lemma renders the inclusions with larger Wasserstein distance to ensure no missing information about the covariances.*

We introduce the following concept only for analysis.

Definition 4.2.33. *For $i = 1, 2$, we introduce the modified transition probabilities for $\mathbb{X}_i = (\mathcal{X}, [\![\mathcal{T}]\!]_i, x_0, \Pi, L)$ based on (4.5). For all $\mathcal{T}_i \in [\![\mathcal{T}]\!]_i$, let*

$$\widetilde{\mathcal{T}_i}(x,\Gamma) = \begin{cases} \mathcal{T}_i(x,\Gamma),\ \forall \Gamma \subseteq \mathcal{W},\ \forall x \in \mathcal{W}, \\ \mathcal{T}_i(x,\mathcal{W}^c),\ \Gamma = \partial\mathcal{W},\ \forall x \in \mathcal{W}, \\ 1,\ \Gamma = \partial\mathcal{W},\ x \in \partial\mathcal{W}. \end{cases} \tag{4.14}$$

Correspondingly, let $\widetilde{[\![\mathcal{T}]\!]}$ denote the collection. Likewise, we also use $\widetilde{(\,\cdot\,)}$ to denote the induced quantities of any other types w.r.t. such a modification.

Remark 4.2.34. *The above modification does not affect the law of the stopped processes since we do not care about the 'out-of-domain' transitions. We use a weighted point mass to represent the measures at the boundary, and the mean should remain the same. It can be easily shown that the Wasserstein distance between any two measures in $\widetilde{[\![\mathcal{T}]\!]}(x,\cdot)$ is upper bounded by that of the non-modified ones.*

[6] This is also called an interval or a hyperrectangle.

Theorem 4.2.35. *For any $0 \leq \vartheta_1 < \vartheta_2$, we set $\mathbb{X}_i = (\mathcal{X}, \widetilde{[\![\mathcal{T}]\!]}_i, \mathrm{AP}, L_{\mathbb{X}})$, $i = 1, 2$, where $\mathbb{X}_1$ and $\mathbb{X}_2$ is are driven by* (4.12) *and* (4.13)*, respectively. Then, under Assumption 4.2.10, there exists a rectangular partition $\mathcal{Q}$ (state-level relation $\alpha \subseteq \mathcal{X} \times \mathcal{Q}$),a measure-level relation Σ_α and a collection of transition matrices $[\![\Theta]\!]$, such that the system $\mathbb{I} = (\mathcal{Q}, [\![\Theta]\!], \mathrm{AP}, L_{\mathbb{I}})$ abstracts $\mathbb{X}_1$ and is abstracted by $\mathbb{X}_2$ by the following relation:*

$$\mathbb{X}_1 \preceq_{\Sigma_\alpha} \mathbb{I}, \ \ \mathbb{I} \preceq_{\Sigma_\alpha^{-1}} \mathbb{X}_2. \tag{4.15}$$

Proof. We construct a finite-state IMC with partition $\mathcal{Q}$ and an inclusion of transition matrices $[\![\Theta]\!]$ as follows. By Assumption 4.2.10, we use uniform rectangular partition on $\mathcal{W}$ and set $\alpha = \{(x, q) : q = \eta\lfloor \frac{x}{\eta} \rfloor\} \cup \{(\Delta, \Delta)\}$, where $\lfloor \cdot \rfloor$ is the floor function and η is to be chosen later. Denote the number of discrete nodes by $N + 1$.

Note that any family of (modified) Gaussian measures $\widetilde{[\![\mathcal{T}]\!]}_1$ is induced from $[\![\mathcal{T}]\!]_1$ and should contain its information. For any $\widetilde{\mathcal{T}} \in \widetilde{[\![\mathcal{T}]\!]}_1$ and $q \in \mathcal{Q}$,

1) for all $\tilde{\nu} \sim \widetilde{\mathcal{N}}(m, s^2) \in \widetilde{\mathbb{T}}_1(\alpha^{-1}(q), \cdot)$, store $\{(m_l, s_l) = (\eta\lfloor \frac{m}{\eta} \rfloor, \eta^2\lfloor \frac{s^2}{\eta^2} \rfloor)\}_l$;
2) for each l, define $\tilde{\nu}_l^{\mathrm{ref}} \sim \widetilde{\mathcal{N}}(m_l, s_l)$ (implicitly, we need to compute $\nu_l^{\mathrm{ref}}(\alpha^{-1}(\Delta))$); compute $\tilde{\nu}_l^{\mathrm{ref}}(\alpha^{-1}(q_j))$ for each $q_j \in \mathcal{Q} \setminus \Delta$;
3) for each l, define $\mu_l^{\mathrm{ref}} = [\tilde{\nu}_l^{\mathrm{ref}}(\alpha^{-1}(q_1)), \cdots, \tilde{\nu}_l^{\mathrm{ref}}(\alpha^{-1}(q_N)), \tilde{\nu}_l^{\mathrm{ref}}(\alpha^{-1}(\Delta)]$;
4) compute $\mathbf{ws} := (\sqrt{2N} + 2)\eta$ and $\mathbf{tv} := N\eta \cdot \mathbf{ws}$;
5) construct $[\![\mu]\!] = \bigcup_l \{\mu : \left\|\mu - \mu_l^{\mathrm{ref}}\right\|_{\mathrm{TV}} \leq \mathbf{tv}(\eta), \ \mu(\Delta) + \sum_j^N \mu(q_j) = 1\}$;
6) let $\Sigma_\alpha := \{(\tilde{\nu}, \mu), \ \mu \in [\![\mu]\!]\}$ be a relation between $\tilde{\nu} \in \widetilde{\mathbb{T}}(\alpha^{-1}(q))$ and the generated $[\![\mu]\!]$.

Repeat the above step for all q, the relation Σ_α is obtained. The rest of the proof falls in the following steps. For $i \leq N$, we simply denote $\mathfrak{G}_i := \widetilde{\mathbb{T}}_1(\alpha^{-1}(q_i), \cdot)$ and $\widehat{\mathfrak{G}}_i := \widetilde{\mathbb{T}}_1(\widehat{\alpha^{-1}(q_i)}, \cdot)$.

Claim 1: For $i \leq N$, let $[\![\Theta_i]\!] = \Sigma_\alpha(\widehat{\mathfrak{G}}_i)$. Then the finite-state IMC $\mathbb{I}$ with the transition collection $[\![\Theta]\!]$ abstracts $\mathbb{X}_1$.

Indeed, for each $i = 1, \cdots, N$ and each $\widetilde{\mathcal{T}}$, we have $\Sigma_\alpha(\mathfrak{G}_i) \subseteq \Sigma_\alpha(\widehat{\mathfrak{G}}_i)$. We pick any modified Gaussian $\tilde{\nu} \in \widehat{\mathfrak{G}}_i$, there exists a $\tilde{\nu}^{\mathrm{ref}}$ such that (by Proposition E.0.12) $\left\|\tilde{\nu} - \tilde{\nu}^{\mathrm{ref}}\right\|_{\mathrm{W}} \leq \left\|\nu - \nu^{\mathrm{ref}}\right\|_{\mathrm{W}} \leq \sqrt{2N}\eta$. We aim to find all discrete measures μ induced from $\tilde{\nu}$ (such that their probabilities match on discrete nodes as requirement by Definition 4.2.9). All such μ should satisfy[7],

$$\begin{aligned}\left\|\mu-\mu^{\text{ref}}\right\|_{\mathrm{W}}^{d} &= \left\|\mu-\mu^{\text{ref}}\right\|_{\mathrm{W}} \\ &\leq \|\mu-\tilde{\nu}\|_{\mathrm{W}}+\left\|\tilde{\nu}-\tilde{\nu}^{\text{ref}}\right\|_{\mathrm{W}}+\left\|\tilde{\nu}^{\text{ref}}-\mu^{\text{ref}}\right\|_{\mathrm{W}} \\ &\leq (2+\sqrt{2N})\eta,\end{aligned} \tag{4.16}$$

where the first term of line 2 is bounded by,

$$\begin{aligned}\|\mu-\tilde{\nu}\|_{\mathrm{W}} &= \sup_{h\in C(\mathcal{X}),\mathrm{Lip}(h)\leq 1}\left|\int_{\mathcal{X}} h(x)d\mu(x)-\int_{\mathcal{X}} h(x)d\tilde{\nu}(x)\right| \\ &\leq \sup_{h\in C(\mathcal{X}),\mathrm{Lip}(h)\leq 1}\sum_{j=1}^{n}\int_{\alpha^{-1}(q_j)}|h(x)-h(q_j)|d\tilde{\nu}(x) \\ &\leq \eta\sum_{j=1}^{n}\int_{\alpha^{-1}(q_j)} d\tilde{\nu}(x)\leq\eta,\end{aligned} \tag{4.17}$$

and the third term of line 2 is bounded in a similar way. By step 5), 6) and Proposition E.0.13, all possible discrete measures μ induced from $\tilde{\nu}$ should be included in $\Sigma_\alpha(\widehat{\mathfrak{G}}_i)$. Combining the above, for any $\tilde{\nu}\in\mathfrak{G}_i$ and hence in $\widehat{\mathfrak{G}}_i$, there exists a discrete measures in $\Theta_i\in\Sigma_\alpha(\widehat{\mathfrak{G}}_i)$ such that for all q_j we have $\tilde{\nu}(\alpha^{-1}(q_j))=\Theta_{ij}$. This satisfies the requirements in Definition 4.2.9.

Claim 2: $\Sigma_\alpha^{-1}(\Sigma_\alpha(\mathfrak{G}_i))\subseteq\mathfrak{G}_i+(2\eta+N\eta\cdot\mathbf{tv}(\eta))\cdot\overline{\mathcal{B}}_W$. This is to recover all possible (modified) measures $\tilde{\nu}$ from the constructed $\Sigma_\alpha(\mathfrak{G}_i)$, such that their probabilities on discrete nodes coincide. Note that, the 'ref' information is recorded when computing $\Sigma_\alpha(\mathfrak{G}_i)$ in the inner parentheses. Therefore, for any $\mu\in\Sigma_\alpha(\mathfrak{G}_i)$ there exists a μ^{ref} within a total variation radius $\mathbf{tv}(\eta)$. We aim to find corresponding measure $\tilde{\nu}$ that matches μ by their probabilities on discrete nodes. All such $\tilde{\nu}$ should satisfy,

$$\begin{aligned}\left\|\tilde{\nu}-\tilde{\nu}^{\text{ref}}\right\|_{\mathrm{W}} &\leq \|\tilde{\nu}-\mu\|_{\mathrm{W}}+\left\|\mu-\mu^{\text{ref}}\right\|_{\mathrm{W}}^{d}+\left\|\mu^{\text{ref}}-\tilde{\nu}^{\text{ref}}\right\|_{\mathrm{W}} \\ &\leq 2\eta+N\eta\cdot\mathbf{tv}(\eta),\end{aligned} \tag{4.18}$$

where the bounds for the first and third terms are obtained in the same way as (4.17). The second term is again by a rough comparison in Proposition E.0.13. Note that $\tilde{\nu}^{\text{ref}}$ is already recorded

[7] Note that we also have $\left\|\mu-\mu^{\text{ref}}\right\|_{\mathrm{W}}^{d}\leq\|\mu-\tilde{\nu}\|_{\mathrm{W}}^{d}+\left\|\tilde{\nu}-\tilde{\nu}^{\text{ref}}\right\|_{\mathrm{W}}^{d}+\left\|\tilde{\nu}^{\text{ref}}-\mu^{\text{ref}}\right\|_{\mathrm{W}}^{d}=\left\|\tilde{\nu}-\tilde{\nu}^{\text{ref}}\right\|_{\mathrm{W}}^{d}$, but it is hard to connect directly from $\left\|\tilde{\nu}-\tilde{\nu}^{\text{ref}}\right\|_{\mathrm{W}}^{d}$ to $\left\|\tilde{\nu}-\tilde{\nu}^{\text{ref}}\right\|_{\mathrm{W}}$ for general measures. This connection can be done if we only compare Dirac or discrete measures.

in $\mathfrak{G}_i$. The inequality in (4.18) provides an upper bound of Wasserstein deviation between any possible satisfactory measure and some $\tilde{\nu}^{\text{ref}} \in \mathfrak{G}_i$.

Claim 3: If we can choose η and k sufficiently small such that

$$2\eta + N\eta \cdot \mathbf{tv}(\eta) + k \leq \vartheta_2 - \vartheta_1, \tag{4.19}$$

then $\mathbb{I} \preceq_{\Sigma_\alpha^{-1}} \mathbb{X}_2$.

Indeed, the $[\![\Theta]\!]$ is obtained by $\Sigma_\alpha(\widehat{\mathfrak{G}}_i)$ for each i. By Claim 2 and Lemma 4.2.31, we have

$$\Sigma_\alpha^{-1}(\Sigma_\alpha(\widehat{\mathfrak{G}}_i)) \subseteq \widehat{\mathfrak{G}}_i + (2\eta + N\eta \cdot \mathbf{tv}(\eta)) \cdot \overline{\mathcal{B}}_W \subseteq \mathfrak{G}_i + (2\eta + N\eta \cdot \mathbf{tv}(\eta) + k) \cdot \overline{\mathcal{B}}_W$$

for each i. By the construction, we can verify that $\widetilde{\mathbb{T}}_2(\alpha^{-1}(q_i)) = \mathfrak{G}_i + (\vartheta_2 - \vartheta_1) \cdot \overline{\mathcal{B}}_W$. The selection of η makes $\Sigma_\alpha^{-1}(\Sigma_\alpha(\widehat{\mathfrak{G}}_i)) \subseteq \widetilde{\mathbb{T}}_2(\alpha^{-1}(q_i))$, which completes the proof. □

Remark 4.2.36. *The relation Σ_α (resp. Σ_α^{-1}) provides a procedure to include all proper (continuous, discrete) measures that connect with the discrete probabilities. The key point is to record $\tilde{\nu}^{\text{ref}}$, μ^{ref}, and the corresponding radius. These are nothing but finite coverings of the space of measures. This also explains the reason why we use 'finite-state' rather than 'finite' abstraction. The latter has a meaning of using finite numbers of representative measures to be the abstraction.*

To guarantee a sufficient inclusion, conservative estimations, e.g. the bound $\sqrt{2N}\eta$ in Claim 1 and the bound in Proposition E.0.13, are made. This estimation can be done more accurately given more assumptions. For example, the deterministic systems (where $\mathbf{w}$ becomes δ) provide Dirac transition measures, the $\|\mu - \mu^{\text{ref}}\|_{\text{W}}^d = 0$ and hence the second term in (4.18) is 0.

Remark 4.2.37. *Note that, to guarantee the second abstraction based on Σ_α^{-1}, we search all possible measures that has the same discrete probabilities as $\mu \in \Sigma_\alpha(\widehat{\mathfrak{G}}_i)$, not only those Gaussians with the same covariances as $\mathfrak{G}_i$ (or $\widehat{\mathfrak{G}}_i$). Such a set of measures provide a convex ball w.r.t. Wasserstein distance. This actually makes sense because in the forward step of creating $\mathbb{I}$, we have used both Wasserstein and total variation distance to find a convex inclusion of all Gaussian or Gaussian related measures (see Figure 4.4). There ought to be some measures that are 'non-recoverable' to Gaussians, unless we extract some 'Gaussian recoverable' discrete measures in $[\![\Theta_i]\!]$, but this loses the point of over-approximation. In this view, IMC abstractions provide unnecessarily larger inclusions than needed.*

For the deterministic case, the above mentioned 'extraction' is possible, since the transition measures do not have diffusion, the convex inclusion becomes a collection of vertices themselves (also see Remark 4.2.20). Based on these vertices, we are able to use Σ_α to find the δ measures within a convex ball w.r.t. Wasserstein distance.

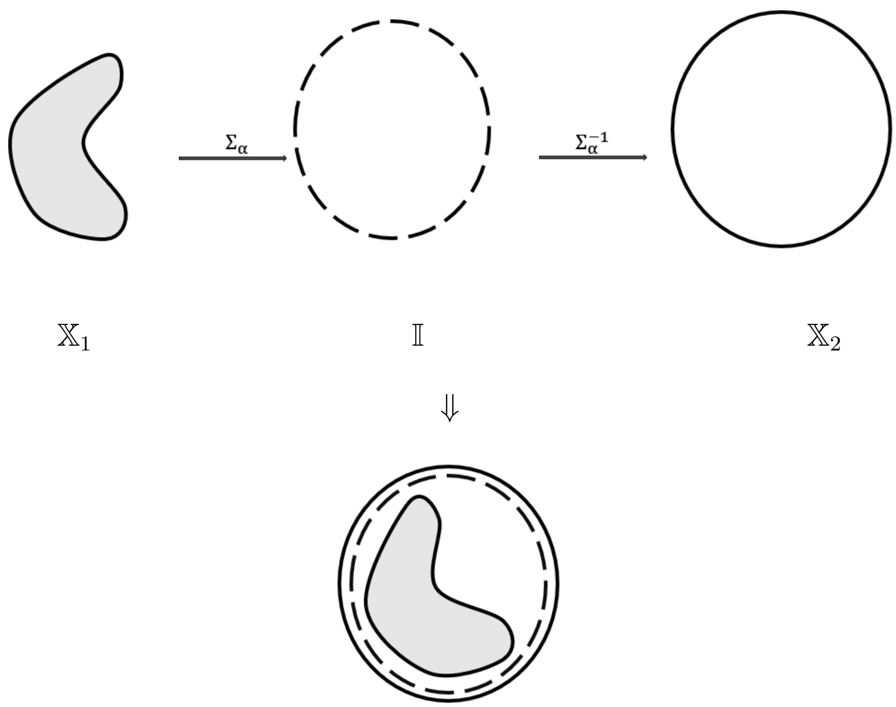


Figure 4.4: The size comparison of systems $\mathbb{X}_1$, $\mathbb{I}$ and $\mathbb{X}_2$ in probability metrics. The dotted line for $\mathbb{I}$ indicates discreteness. The three systems connect via their measurable labelling functions. The system $\mathbb{X}_1$ reduces to a singleton if no extra uncertain perturbations.

In contrast to the above special case [112], where the uncertainties are bounded w.r.t. the infinity norm, we can only guarantee the approximated completeness via a robust $\mathcal{L}^1$-bounded perturbation with strictly larger intensity than the original point-mass perturbation. However, this indeed describes a general type of uncertainties for the stochastic systems to guarantee $\mathcal{L}^1$-related properties, including probabilistic properties. Unless higher-moment specifications are of interests, uncertain $\mathcal{L}^1$-random variables are what we need to be the analogue of perturbations in [112].

Corollary 4.2.38. *Given an LTL formula Ψ, let $S_i^{\nu_0} = \{\mathbf{P}_X^{\nu_0}(X \vDash \Psi)\}_{X\in\mathbb{X}_i}$ ($i = 1, 2$) and $S_{\mathbb{I}}^{q_0} = \{\mathcal{P}_I^{q_0}(I \vDash \Psi)\}_{I\in\mathbb{I}}$, where the initial conditions are such that $\nu_0(\alpha^{-1}(q_0)) = 1$. Then all the above sets are compact and $S_1^{\nu_0} \subseteq S_{\mathbb{I}}^{q_0} \subseteq S_2^{\nu_0}$.*

The proof in shown in Section 4.2.6.

4.2.4 A Comparison with Numerical Approximations

The purpose of abstraction-based formal methods is in general different from constructing numerical solutions for SDEs. Numerical SDE schemes use the time discretization and simulate sample paths at discrete time steps, which is also known as time discrete approximations. The numerical analysis for SDEs is to determine how good the approximation is and in what sense it is close to the exact solution [95].

Aside from the analysis based on the time discretization, the stochastic driving forces in discrete-time numerical simulations are given with discrete distributions *in a priori*. For example, a spatial step size should be provided to generate a pseudo random number from a Gaussian distribution. Consequently, there is a unique solution in the discrete canonical space driven by this discrete noise. In view of (4.17), the discretized measure of any random variable already provides a deviation from the real measure to begin with. The numerical simulation provides a much smaller set of measurable sample paths, i.e. a natural filtration $\mathscr{F}^{\mathbf{w},d}$ w.r.t. the discrete version of noise $\mathbf{w}$ rather than $\mathcal{F}$ (recall Definition 4.2.1). The missing transitions or measurable sample paths from $\mathcal{F}$ cannot be recovered given a fixed discretized noise at a time.

In addition, we cannot see any information of the (discrete) probability law of sample paths unless we enumerate all the realizations (a full observation), which is impossible.

On the other hand, from the dual problem point of view, a finite difference approximation for the associated Fokker-Plank equation (parabolic equation)

$$\frac{\partial \rho_t}{\partial t} = \mathfrak{L}^* \rho_t,$$

where $\mathfrak{L}^*$ is the adjoint operator of the generator $\mathfrak{L}$ as in (3.3), provides approximated discrete marginal densities of the probability laws of the solution processes. However, in view of finite-dimensional distribution as in (B.8), this is not sufficient for the evaluation of the probability of sample paths satisfying some LT properties over the time horizon. We need to consider the approximation of the associated transition semigroups $e^{t\mathcal{L}^*}$ to fulfill such a type of evaluation.

In comparison with the numerical solutions and the dual approximation of the probability distributions, the stochastic abstractions in this chapter start with discrete-time systems but do not use the spatially discretized noise as the driving force. Instead, we directly work on generating a relation based on the state-space discretization such that the transition kernel of the original system is 'included' in the discrete family of transition matrices in the sense of Corollary 4.2.38 or Corollary 4.3.14 which we will see in the next section. The formal guarantee of these inclusions avoids the approximation error given by the finite difference numerical scheme and is robust to extra $\mathcal{L}^1$-bounded uncertainties.

Even though a refinement of grid size can lead to a convergence for both numerical simulations and stochastic abstractions, they converge from different 'directions'. In other words, the family of the discrete probability laws from an abstraction reduces to a singleton whilst the missing transitions in a numerical simulation become empty as the size of the grids converges to 0.

Furthermore, unlike specification guided algorithm design, the stochastic abstraction considered in this chapter can be utilized to verification and control synthesis of any probabilistic ω-regular specifications regardless of the computational efforts.

4.2.5 Discussion

In this section, we constructed an IMC abstraction for continuous-state stochastic systems with possibly bounded point-mass (Dirac) perturbations. We showed that such abstractions are not only sound, in the sense that the set of satisfaction probability of LT properties contains that of the original system, but also approximately complete in the sense that the constructed IMC can be abstracted by another system with stronger but more general $\mathcal{L}^1$-bounded perturbations. Consequently, the winning set of the probabilistic specifications for a more perturbed continuous-state stochastic system contains that of the less Dirac perturbed system. Similar to most of the existing converse theorems, e.g. converse Lyapunov functions, the purpose is not to provide an efficient approach for finding them, but rather to characterize the theoretical possibilities of having such existence.

It is interesting to compare with robust deterministic systems, where no random variables are involved. In [112], both perturbed systems are w.r.t. bounded point masses. More heavily perturbed systems abstract less perturbed ones and hence preserve robust satisfaction of linear-time properties. However, when we try to obtain the approximated completeness via uncertainties in stochastic system, the uncertainties should be modelled by more general $\mathcal{L}^1$ random variables. Note that the probabilistic properties of random variables is dual to the weak topology of measures, we study the measures and hence probability laws of processes instead of the state space *per se*. The state-space topology is not sufficient to quantify the regularity of IMC abstractions. In contrast, the $\mathcal{L}^1$ uncertain random variables is a perfect analogue of the uncertain point masses (in $|\cdot|$) for deterministic systems. If we insist on using point masses as the only type of uncertainties for stochastic systems, the IMC type abstractions would possibly fail to guarantee the completeness. For example, suppose the point-mass perturbations represents the lack of precision caused by deterministic control inputs [117, Definition 2.3], the winning set decided by the ϑ_2-precision deterministic policies is not enough to cover that of the IMC abstraction, which fails to ensure an approximated bi-similarity of IMCs compared to [112].

4.2.6 Proofs of Results

Proof of Corollary 4.2.19.

Proof. It is clear that $\mathcal{Q}$ under discrete metric is complete and separable. In addition, for each t, the space $(\mathscr{M}_t, \|\cdot\|_{\text{TV}})$ is complete and separable. By Lemma 4.2.16, each $(\mathscr{M}_t, \|\cdot\|_{\text{TV}})$ is also compact. For any sequence $\{\mu_n\} \subseteq \mathscr{M}_t$, a quick application of Theorem E.0.6 leads to the existence of a weakly convergent subsequence $\{\mu_{n_k}\}$ and a weak limit point μ in $\mathscr{M}_t$. By the definition of weak convergence and the discrete structure of $\mathcal{Q}$, it is clear that for each $h \in C_b(\mathcal{X})$ and $t \in \mathbb{N}$, we have

$$\sum_{\mathcal{X}} h(x)\mu_{n_k}(x) \to \sum_{\mathcal{X}} h(x)\mu$$

in a strong sense, which concludes the compactness of H. Now we choose $\mu_1, \mu_2 \in \mathscr{M}_t$, then $a\mu_1 + (1-a)\mu_2 \in \mathscr{M}_t$ for all $a \in [0,1]$. Therefore,

$$a\sum_{\mathcal{X}} h(x)\mu_1(x) + (1-a)\sum_{\mathcal{X}} h(x)\mu_2(x) = \sum_{\mathcal{X}} h(x)[a\mu_1 + (1-a)\mu_2](x) \in H$$

for all $a \in [0,1]$. This shows the convex structure of H. □

Proof of Proposition 4.2.21.

Proof. We make a bit abuse of notation and define $\pi_T : \mathcal{Q}^\infty \to \prod_0^T \mathcal{Q}$ as the projection onto the finite product space of $\mathcal{Q}$ up to time T. Since we do no emphasize the initial conditions, we also use $\mathcal{P}$, $\mathscr{M}$ and $\mathscr{M}_t$ for short. By Tychonoff theorem, any product of Q is also compact w.r.t. the product topology. Therefore, any family of measures on Q^T is tight and hence compact. By Remark 4.2.2, for every $\mathcal{P} \in \mathscr{M}$, we have $\mathcal{P} \circ \pi_T^{-1} = \otimes_{t=0}^T \mu_t$ (recall Remark 4.2.2) for some $\mu_t \in \mathscr{M}_t$, and $\{\mathcal{P} \circ \pi_T^{-1}\}_{I\in\mathbb{I}}$ forms a compact set. Hence, every sequence $\{\mathcal{P}_n \circ \pi_T^{-1}\}_n \subseteq \{\mathcal{P} \circ \pi_T^{-1}\}_{I\in\mathbb{I}}$ with any finite T contains a weakly convergent subsequence. We construct the convergent subsequence of $\{\mathcal{P}_n\}_n$ in the following way.

We initialize the procedure by setting $T = 0$. Then $\mathscr{M}_0$ is compact, and there exists a weakly convergent subsequence $\{\mathcal{P}_{0,n} \circ \pi_0^{-1}\}$. Based on $\{\mathcal{P}_{0,n}\}$, we are able to see that it contains a weakly convergent subsequence, denoted by $\{\mathcal{P}_{1,n}\}$, such that $\{\mathcal{P}_{1,n} \circ \pi_1^{-1}\}$ weakly converges. By induction, we have $\{\mathcal{P}_{k+1,n}\} \subseteq \{\mathcal{P}_{k,n}\}$ for each $k \in \mathbb{N}$. Repeating this argument and picking the diagonal subsequence $\{\mathcal{P}_{n,n}\}$, then $\{\mathcal{P}_{n,n}\}$ has the property that $\{\mathcal{P}_{n,n} \circ \pi_T^{-1}\}$ is weakly convergent for each T . We denote the weak limit point of each $\{\mathcal{P}_{n,n} \circ \pi_T^{-1}\}$ by $\otimes_{t=0}^T \mu_t$. By the way of construction, we have

$$\otimes_{t=0}^T \mu_t(\cdot) = \otimes_{t=0}^{T+1} \mu_t(\cdot \times Q), \quad \forall T \in \mathbb{N}.$$

By Kolmogorov's extension theorem, there exists a unique $\mathcal{P}$ on Q^∞ such that $\otimes_{t=0}^T \mu_t(\cdot) = \mathcal{P} \circ \pi_T^{-1}(\cdot)$ for each T.

We have seen that for each $\{\mathcal{P}_n\}$, the constructed subsequence satisfies $\mathcal{P}_{n,n} \rightharpoonup \mathcal{P}$, which concludes the claim. □

Proof of Theorem 4.2.22

Proof. Since we do not emphasize the initial conditions, we simply drop the superscripts q_0 for short. Given $I \in \mathbb{I}$ with any initial condition, the corresponding canonical space is $(\Omega, \mathscr{F}, \mathcal{P}_I)$. By Proposition 4.2.21, every sequence $\{\mathcal{P}_n\} \subseteq \mathscr{M}$ has a weakly convergent subsequence, denoted by $\{\mathcal{P}_{n_k}\}$, to a $\mathcal{P} \in \mathscr{M}$ of some I. Note that for any I, the measurable set $\{I \vDash \Psi\} = \{\varpi : \varpi \vDash \Psi\} \in \mathscr{F}$ is the same due to the identical labelling function. It is important to notice that due to the discrete topology of Ω, every Borel measurable set $A \in \mathscr{F}$ is such that $\partial A = \emptyset$. By Definition E.0.3 we have $\mathcal{P}_{n_k}(I_{n_k} \vDash \Psi) \to \mathcal{P}(I \vDash \Psi)$ for all Ψ. The compactness of S^{q_0} follows immediately. To show the convexity of S^{q_0}, we notice that, for any $q_0, \cdots q_{n_t} \in \mathcal{Q}$ and $I \in \mathbb{I}$,

$$\mathcal{P}_I\left(I_0 = q_0, \cdots, I_t = q_{n_t}, I_{t+1} = q_{n_{t+1}}\right) \in \{\Theta_{n_{t+1},n_t}\Theta_{n_t,n_{t-1}} \cdots \Theta_{n_1,0}\delta_{q_0} : \Theta \in [\![\Theta]\!]\}$$

and hence forms a convex set. Immediately, the convexity holds for $\{\mathcal{P}_I(\prod_{i=1}^n \Gamma_i)\}_{I\in\mathbb{I}}$ for any cylinder set $\prod_{i=1}^n \Gamma_i$. By a standard monotone class argument, $\{\mathcal{P}_I(A)\}_{I\in\mathbb{I}}$

is also convex for any Borel measurable set $A \in \mathscr{F}$, which implies the convexity of S^{q_0} in the statement. $\square$

Proof of Proposition 4.2.23

Proof. Note that the laws are associated with X with $X_0 = x_0$, which actually means the stopped process X^τ (Recall notations in Section 4.2.1). Now that $X_{t\wedge\tau} \in \overline{\mathcal{W}}$ for each t, the state space of X is compact, so is the countably infinite product. By a similar argument as Proposition 4.2.21, we can conclude the first part of the statement. Note that by assumption, the partition respects the boundary of the labelling function. Hence, for all formula Ψ, the boundary of $\{X \vDash \Psi\} \in \mathscr{F}$ has measure 0. The second part can be concluded directly by Definition E.0.3. $\square$

Proof of Lemma 4.2.24

Proof. Note that X is on $(\mathbf{\Omega}, \mathcal{F}, \mathbf{P}_X^{\nu_0})$ and I is on $(\Omega, \mathscr{F}, \mathcal{P}_I^{q_0})$. We first show the case when $\nu_0 = \delta_{x_0}$ for any $x_0 \in \alpha^{-1}(q_0)$. That is, if for $X_0 = x_0$ a.s. with any $x_0 \in \alpha^{-1}(q_0)$, there exists a unique law of some $I \in \mathbb{I}$ such that $\mathbf{P}_X^{x_0}(X \vDash \Psi) = \mathcal{P}_I^{q_0}(I \vDash \Psi)$ for any Ψ.

Let ν_t denote the marginal distribution of $\mathbf{P}_X^{x_0}$ at each t. Let $\mathscr{M}_t = \{\mu_t\}_{I\in\mathbb{I}}$ denote the set of marginal distributions of $\{\mathcal{P}_I^{q_0}\}_{I\in\mathbb{I}}$. Now, at $t = 1$, $\nu_1(\alpha^{-1}(q_j)) = \mathcal{T}(x_0, \alpha^{-1}(q_j))\delta_{x_0}$ for all $j \in \{1, 2, \cdots, N+1\}$. Suppose q_0 is the i^{th} element of $\mathcal{Q}$, by the construction of IMC, we have

$$\check{\Theta}_{ij} \le \nu_1(\alpha^{-1}(q_j)) = \int_{\alpha^{-1}(q_j)} \delta_{x_0}\mathcal{T}(x_0, dy) \le \hat{\Theta}_{ij}, \;\; \forall x_0 \in \alpha^{-1}(q_0) \text{ and } \forall j \in \{1, 2, \cdots, N+1\}.$$

Since $\sum_{q\in Q} \nu_1(\alpha^{-1}(q)) = 1$, by letting $\mu_1 = (\nu_1(\alpha^{-1}(q_1)), \nu_1(\alpha^{-1}(q_2)), \cdots, \nu_1(\alpha^{-1}(q_{N+1}))^T$, we have automatically $\mu_1 \in \mathscr{M}_1$ by definition. Note that μ_1 is unique w.r.t. $\|\cdot\|_{\text{TV}}$, and has the property that $\mu(q) = \nu(\alpha^{-1}(q))$ for each $q \in Q$.

Similarly, at $t = 2$, we have

$$\check{\Theta}_{ij}\mu_1(q_i) \le \int_{\alpha^{-1}(q_j)} \int_{\alpha^{-1}(q_i)} \nu_1(dx)\mathcal{T}(x, dy) \le \hat{\Theta}_{ij}\mu_1(q_i), \;\forall i, j \in \{1, 2, \cdots, N+1\},$$

where $\mathcal{T}$ may not be the same as that of $t = 1$. Therefore, for any $j \in \{1, 2, \cdots, N+1\}$,

$$\nu_2(\alpha^{-1}(q_j)) = \sum_{i=1}^{N+1} \int_{\alpha^{-1}(q_j)} \int_{\alpha^{-1}(q_i)} \nu_1(dx)\mathcal{T}(x, dy)$$

and there exists a μ_2 such that $\sum_i \check{\Theta}_{ij}\mu_1(q_i) \leq \mu_2(q_j) = \nu_2(\alpha^{-1}(q_j)) \leq \sum_i \hat{\Theta}_{ij}\mu_1(q_i)$, which means (by (4.8)) $\mu_2 \in \mathscr{M}_2$. In addition, there also exists a $\mathcal{P}^{q_0}$ such that its one-dimensional marginals up to $t = 2$ admit μ_1 and μ_2, and satisfies

$$\mathcal{P}^{q_0}[I_0 = q_0, I_1 = q_i, I_2 = q_j] = \mathbf{P}_X^{x_0}[X_0 = x_0, X_1 \in \alpha^{-1}(q_i), X_2 \in \alpha^{-1}(q_j)].$$

Repeating this procedure, there exists a unique $\mu_t \in \mathscr{M}_t$ w.r.t. $\|\cdot\|_{\text{TV}}$ for each t, such that $\mu_t(q) = \nu_t(\alpha^{-1}(q))$ for each $q \in \mathcal{Q}$. It is also clear that for each given $x_0 \in q_0$ and each t, the selected $\mathcal{P}^{q_0}$ satisfies

$$\mathbf{P}_X^{x_0}\left(\prod_0^{t-1} A_i\right) = \mathcal{P}^{q_0}\left(\prod_0^{t-1} A_i\right) = \mathcal{P}^{q_0}\left(\prod_0^{t-1} A_i \times \mathcal{Q}\right), \quad A_i \in \mathscr{B}(\mathcal{Q}).$$

By Kolmogrov extension theorem, there exists a unique law $\mathcal{P}_I^{q_0}$ of some $I \in \mathbb{I}$ such that each T-dimensional distribution coincides with $\otimes_0^T \mu_i$, and, for each given $x_0 \in \alpha^{-1}(q_0)$, $\mathbf{P}_X^{x_0}(\Gamma) = \mathcal{P}_I^{q_0}(\Gamma)$ for all $\Gamma \in \mathscr{B}(\mathcal{Q}^\infty) = \mathscr{F}$. Due to the assumption that $L_{\mathbb{X}}(x) = L_{\mathbb{I}}(q)$ for all $x \in \alpha^{-1}(q)$ and $q \in \mathcal{Q}$, we have

$$\{L_{\mathbb{X}}^{-1}(\Psi)\} = \{L_{\mathbb{I}}^{-1}(\Psi)\} \in \mathscr{F}$$

for all LTL formula Ψ, which implies $\{X \vDash \Psi\} = \{I \vDash \Psi\}$ by definition. Thus, given $x_0 \in \alpha^{-1}(q_0)$, the above $\mathcal{P}_I^{q_0}$ should satisfy $\mathbf{P}_X^{x_0}(X \vDash \Psi) = \mathcal{P}_I^{q_0}(I \vDash \Psi)$.

Based on the above conclusion, as well as the definition of $\mathbf{P}_X^{\nu_0}$ and the convexity of S^{q_0} (recall Theorem 4.2.22), the result for more general initial distribution ν_0 with $\nu_0(\alpha^{-1}(q_0)) = 1$ can be obtained. □

Proof of Proposition 4.2.29.

Proof. Let $\mu, \nu \in \mathscr{M}_t$ for each t, and $V, K \in [\![\Theta]\!]$ be any stochastic matrices generated by $\mathbb{I}$. Then, for each t, we have

$$\begin{aligned}\left\|V^T\mu - K^T\nu\right\|_{\text{TV}} &\leq \left\|V^T\mu - V^T\nu\right\|_{\text{TV}} + \max_i \|V_i - K_i\|_{\text{TV}} \|\nu\|_{\text{TV}} \\ &\leq \frac{1}{2}\max_{i,j} \|V_i - K_j\|_{\text{TV}} \|\mu - \nu\|_{\text{TV}} + \epsilon \|\nu\|_{\text{TV}} \\ &\leq \|\mu - \nu\|_{\text{TV}} + \epsilon.\end{aligned} \tag{4.20}$$

This implies that the total deviation of any $\tilde{\mu}, \tilde{\nu} \in \mathscr{M}_{t+1}$ is bounded by

$$\max_{\mathscr{M}_t} \|\mu - \nu\|_{\text{TV}} + \epsilon.$$

Note that at $t = 0$, $\max_{\mathscr{M}_0} \|\mu - \nu\|_{\mathrm{TV}} = 0$. Hence, at each $t > 0$, as $\epsilon \to 0$,

$$\max_{\mathscr{M}_t} \|\mu - \nu\|_{\mathrm{TV}} \to 0.$$

By the product topology and Kolmogrov extension theorem, the set $\{\mathcal{P}_I\}_{I \in \mathbb{I}}$ is reduced to a singleton. The conclusion follows after this. □

Proof of Lemma 4.2.31.

Proof. It can be proved, for example, using inclusion functions. Let $\mathbb{IR}^n$ denote the set of all boxes in $\mathbb{R}^n$. Let $[f] : \mathbb{IR}^n \to \mathbb{IR}^n$ be a convergent inclusion function of f satisfying (i) $f([x]) \subseteq [f]([x])$ for all $[x] \in \mathbb{IR}^n$; (ii) $\lim_{\lambda([y]) \to 0} \lambda([f]([x]) = 0$, where λ denote the width. Similarly, let $[B] : \mathbb{IR}^n \to \mathbb{IR}^{n \times k}$ be a convergent inclusion matrix of $b(x)$ and satisfy (i) $b([x]) \subseteq [B]([x])$ for all $[x] \in \mathbb{IR}^n$; (ii) $\lim_{\lambda([B]) \to 0} \lambda([B]([x]) = 0$, where $\lambda([B]([x]) := \max_{i,j} \lambda([B_{ij}])$.

Without loss of generality, we assume that $k < 1$. Due to the Lipschitz continuity of f and b, we can find inclusions such that $\lambda([f]([y]) \le L^f \lambda([y])$ for any subintervals of $[x]$, and similarly $\lambda([B]([y]) \le L^b \lambda([y])$, where L^f and L^b are the Lipschitz constants for f and b, respectively. For each such interval $[y]$, we can obtain the interval $[m] = [f]([y])$ and $[s^2] = [B]([y])[B]^*([y])$. Let T denote the collection of Gaussian measures with mean and covariance of all such intervals ($[m]$ and $[s^2]$), and $\widehat{\mathbb{T}_1([x])}$ be its union. Then $\widehat{\mathbb{T}_1([x])}$ satisfies the requirement. Indeed, we have $\mathbb{T}_1([x]) \subseteq \widehat{\mathbb{T}_1([x])}$. For the second part of inclusion, we have for any $\mu \in \widehat{\mathbb{T}_1([x])}$ and $\nu \in \mathbb{T}_1([x])$,

$$\|\mu - \nu\|_{\mathrm{W}}^2 \le [m]^2 + [s^2] \le (L^f \lambda([y]))^2 + (L^b \lambda([y]))^2, \tag{4.21}$$

where we are able to choose $[y]$ arbitrarily small such that $(L^f \lambda([y]))^2 + (L^b \lambda([y]))^2 < k^2$. The second part of inclusion can be completed by such a choice of $[y]$. □

Proof of Corollary 4.2.38

Proof. The first part of the proof is provided in Section 4.2.2. The second inclusion is done in a similar way as Lemma 4.2.24 and Theorem 4.2.25. Indeed, by the definition of abstraction, for any $\mu \in \mathscr{M}_t$, there exists a marginal measure ν of some $X \in \mathbb{X}_2$ such that their probabilities match on discrete nodes. By the same induction as Lemma 4.2.24, we have that for any law $\mathcal{P}_I^{q_0}$ of some $I \in \mathbb{I}$, there exists a $\mathbf{P}_X^{\nu_0}$ of some $X \in \mathbb{X}_2$ such that the probabilities of any $\Gamma \in \mathscr{B}(Q^\infty)$ match. The second inclusion follows after this. The compactness also follows a similar way as Proposition 4.2.23. Note that, S_1 may not be convex, but $S_{\mathbb{I}}$ and S_2 are (also see details in Remark 4.2.37). □

4.3 Robustly Complete Finite-State Abstractions for Control Synthesis of Stochastic Systems with Full Observation

We have seen in control-free discrete-time stochastic systems that IMCs can be constructed as abstractions to the original system with soundness and robust completeness. The philosophy is to discretize the state space $\mathcal{X}$ into $\mathcal{Q}$ obeying the labelling boundaries, such that any ω-regular property measured in the probability laws generated by the family of discrete transition functions/matrices contain the satisfaction probability of the original system. In this section, we extend the methodology to the construction of robustly complete abstractions for stochastic control systems. The finite-state abstraction model can be used for solving the control synthesis problem. It will be shown that there is a decision procedure to answer whether the original system is controllable w.r.t. some LTL specification.

4.3.1 Preliminaries

Let $\mathcal{U} \subseteq \mathbb{R}^p$ be a compact space of control inputs. The canonical setup for discrete-time controlled processes is provided in Appendix C.1. Given any measurable process $\mathfrak{u}$, the probability law of the joint process $(X, \mathfrak{u}) := \{(X_t, \mathfrak{u}_t)\}_{t\geq 0}$ can be determined. We also denote $X^{\mathfrak{u}}$ by the controlled process if we emphasize on the state-space marginal of $(X, \mathfrak{u})$. Furthermore, if $\mathfrak{u}$ is generated based on a control policy κ, we replace the notation $X^{\mathfrak{u}}$ by X^{κ}.

We usually consider $(X, \mathfrak{u})$ to be obtained from Markov models, whose transition probabilities, unlike control-free systems, have an extra dependence of the current control input, i.e.,

$$\Theta_t^u(x, \Gamma) = \mathbf{P}[X_{t+1} \in \Gamma \mid X_t = x, \mathfrak{u}_t = u].$$

Assumption 4.3.1. *We assume that $\mathfrak{u}$ is deterministic in a priori or generated by deterministic control policies (see Appendix C.3 for details), i.e., for each t, $\mathfrak{u}_t \in \mathcal{U}$ rather than $\mathfrak{u}_t \in \mathfrak{P}(\mathcal{U})$.*

Now we consider an abstract family of labelled controlled Markov processes as follows.

Definition 4.3.2 (Controlled Markov system)**.** *A Markov decision system is a tuple $\mathbb{XU} = (\mathcal{X}, \mathcal{U}, \{\Theta\}, \mathrm{AP}, L)$, where*

- ✧ $\mathcal{X}$, AP*, and* L *are as previously mentioned.*
- ✧ $\mathcal{U}$ *is the set of actions;*

✧ $\{\Theta\} := \{[\![\Theta^u]\!]\}_{u\in\mathcal{U}}$ *contains all collections of control-dependent transition probabilities; for every* t*, given an action* $u \in \mathcal{U}$*, the transition* Θ^u_t *is chosen from* $[\![\Theta^u]\!]$ *accordingly.*

Note that for every given $\mathfrak{u}$ and initial condition $X_0 = x_0$ (resp. initial distribution ν_0), we can generate a process $X^{\mathfrak{u}} \in \mathbb{XU}^{\mathfrak{u}}$, whose probability law is denoted by $\mathbf{P}_X^{x_0,\mathfrak{u}}$ (resp. $\mathbf{P}_X^{\nu_0,\mathfrak{u}}$), and $\mathbb{XU}^{\mathfrak{u}}$ denotes all the processes that are generated given $\mathfrak{u}$. The collection of all the probability laws of such controlled processes is denoted by $\{\mathbf{P}_X^{x_0,\mathfrak{u}}\}_{X^{\mathfrak{u}}\in\mathbb{XU}^{\mathfrak{u}}}$ (resp. $\{\mathbf{P}_X^{\nu_0,\mathfrak{u}}\}_{X^{\mathfrak{u}}\in\mathbb{XU}^{\mathfrak{u}}}$). If $\mathfrak{u}$ is known to be generated according to some deterministic control policy κ, the previously mentioned notations are changed correspondingly by replacing the superscripts $\mathfrak{u}$ by κ.

We consider controlled Markov processes determined by the following fully-observed Markov system

$$X_{t+1} = f(X_t, \mathfrak{u}_t) + b(X_t)\mathbf{w}_t + \vartheta\xi_t, \tag{4.22}$$

where $f : \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^n$ is locally Lipschitz continuous in both arguments. The rest of notations are the same as previously mentioned. We can translate (4.22) into a controlled Markov system

$$\mathbb{XU} = (\mathcal{X}, \mathcal{U}, \{\mathcal{T}\}, \mathrm{AP}, L_{\mathbb{XU}}), \tag{4.23}$$

where $\{\mathcal{T}\} := \{[\![\mathcal{T}^u]\!]\}_{u\in\mathcal{U}}$. Given each control process $\mathfrak{u}$, we deal with the out-of-domain behaviors of $X^{\mathfrak{u}}$ in the same way as in the control-free cases.

Definition 4.3.3. *Given a continuous-state controlled Markov system*

$$\mathbb{XU} = (\mathcal{X}, \mathcal{U}, \{\mathcal{T}\}, \mathrm{AP}, L_{\mathbb{XU}})$$

with a compact $\mathcal{U} \in \mathbb{R}^p$*, and a finite-state Markov system*

$$\mathbb{IA} = (\mathcal{Q}, \mathit{Act}, \{\Theta\}, \mathrm{AP}, L_{\mathbb{IA}}),$$

where $\mathcal{Q} = (q_1, \cdots, q_N)^T$*,* $\mathit{Act} = \{a_1, \cdots, a_M\}$*, and* $\{\Theta\} := \{[\![\Theta^a]\!]\}_{a\in\mathit{Act}}$ *contains all collections of* $n \times n$ *stochastic matrices that are also dependent on* a*.*

We say that $\mathbb{IA}$ *abstracts* $\mathbb{XU}$*, and write* $\mathbb{XU} \preceq_{\Sigma_\alpha} \mathbb{IA}$*, if there exist*

(1) a state-level relation $\alpha \subseteq \mathcal{X} \times \mathcal{Q}$ *from* $\mathbb{XU}$ *to* $\mathbb{IA}$ *such that, for all* $x \in \mathcal{X}$*, there exists* $q \in \mathcal{Q}$ *such that* $(x, q) \in \alpha$ *(*$\alpha(x) \neq \emptyset$*) and* $L_{\mathbb{IA}}(q) = L_{\mathbb{XU}}(x)$*;*

(2) a measure-level relation $\Sigma_\alpha \subseteq \mathfrak{P}(\mathcal{X}) \times \mathfrak{P}(Q)$ *from* $\mathbb{XU}$ *to* $\mathbb{IA}$ *such that, for all* $i \in \{1, 2, \cdots, N\}$ *and* $a \in \mathit{Act}$*, there exists* $u \in \mathcal{U}$ *such that for any* $\mathcal{T}^u \in [\![\mathcal{T}^u]\!]$ *and all* $x \in \alpha^{-1}(q_i)$*, there exists* $\Theta^a \in [\![\Theta^a]\!]$ *satisfying* $(\mathcal{T}^u(x,\cdot), \Theta^a_i) \in \Sigma_\alpha$ *and* $\mathcal{T}^u(x, \alpha^{-1}(q_j)) = \Theta^a_{ij}$ *for all* $j \in \{1, 2, \cdots, n\}$*.*

The converse abstraction is defined in a similar way.

4.3.2 Soundness of Robust BMDP Abstractions

Definition 4.3.4. *A BMDP is a tuple $\mathcal{IA} = (\mathcal{Q}, Act, \{\check{\Theta}\}, \{\hat{\Theta}\}, \mathrm{AP}, L_{\mathcal{IA}})$, where*

- ✧ *$\mathcal{Q}$ is an N-dimensional state-space;*
- ✧ *Act is a finite-dimensional actions;*
- ✧ *AP and $L_{\mathcal{BA}}$ are the same as in Definition 4.2.3;*
- ✧ *$\{\check{\Theta}\} := \{\check{\Theta}^a\}_{a \in Act}$ is a family of $N \times N$ matrix such that $\check{\Theta}^a_{ij}$ is the lower bound of transition probability from the state number i to j for each $i, j \in \{1, 2, \cdots, N\}$ and action $a \in Act$;*
- ✧ *$\{\hat{\Theta}\} := \{\hat{\Theta}^a\}_{a \in Act}$ is a family of $N \times N$ matrix such that $\hat{\Theta}^a_{ij}$ is the upper bound of transition probability from the state number i to j for each $i, j \in \{1, 2, \cdots, N\}$ and action $a \in Act$.*

Similar to $\mathcal{I}$ defined in Definition 4.2.12, we are able to transfer a $\mathcal{IA}$ into a controlled Markov system $\mathbb{IA}$ as in Definition 4.3.3, whose $[\![\Theta^a]\!]$'s are well defined sets of stochastic matrices for each $a \in Act$.

Remark 4.3.5. *To make $\mathbb{IA}$ an abstraction for (4.23), we can discretize both $\mathcal{X}$ and $\mathcal{U}$, such that each node $a \in Act$ represents a grid of $u \in \mathcal{U}$. We then need the approximation to be such that $\check{\Theta}^a_{ij} \le \int_{\alpha^{-1}(q_j)} \mathcal{T}^u(x, dy) \le \hat{\Theta}^a_{ij}$ for a $u \in a$, for all $x \in \alpha^{-1}(q_i)$ and $i, j = 1, \cdots, N$, as well as $\Theta_{N+1} = (0, 0, \cdots, 1)$. We further require that the partition should respect the boundaries induced by the labeling function as previously mentioned.*

Definition 4.3.6. *Given a state-level abstraction α and a measure-level abstraction Σ_α from $\mathbb{XU}$ to $\mathbb{IA}$. Let ϕ and κ be some control policies of $\mathbb{XU}$ and $\mathbb{IA}$, respectively. Recall notations in (C.1). We call ϕ a Σ_α-implementation of κ if, for each $t \in \mathbb{N}$,*

$$\mathfrak{u}_t = \phi_t(X_{[0,t]}, \mathfrak{u}_{[0,t-1]}), \quad X \in \mathbb{XU}^\pi$$

is chosen according to

$$\mathfrak{a}_t = \kappa_t(I_{[0,t]}, \mathfrak{a}_{[0,t-1]}), \quad I \in \mathbb{IA}^\kappa$$

in a way that, for any realization u and a of $\mathfrak{u}_t$ and $\mathfrak{a}_t$, for any $\mathcal{T}^u \in [\![\mathcal{T}^u]\!]$ and all $x \in \alpha^{-1}(q_i)$, there exists $\Theta^a \in [\![\Theta^a]\!]$ satisfying $(\mathcal{T}^u(x, \cdot), \Theta^a_i) \in \Sigma_\alpha$ and $\mathcal{T}^u(x, \alpha^{-1}(q_j)) = \Theta^a_{ij}$ for all $j \in \{1, 2, \cdots, n\}$.

We can define the converse implementation from $\mathbb{IA}$ to $\mathbb{XU}$ based on a converse measure-level relation (from $\mathbb{IA}$ to $\mathbb{XU}$) in a similar way.

Remark 4.3.7. *Heuristically, a control policy κ is generated in the finite-state finite-action abstraction model within $\mathbb{IA}$ to ensure a probabilistic satisfaction of some LTL specification. The selection of the control policy π is subjected to κ and hence $\mathbb{IA}^\kappa$ according to the abstraction relation, such that (2) of Definition 4.3.3 is guaranteed.*

Theorem 4.3.8. *Let $\mathbb{IA}$ be a controlled Markov system that is derived from a BMDP with any initial distribution μ_0. Then for any LTL formula Ψ, given any admissible deterministic control policy κ[8], the set $S^{\mu_0,\kappa} = \{\mathcal{P}_I^{\mu_0,\kappa}(I^\kappa \vDash \Psi)\}_{I^\kappa \in \mathbb{IA}^\kappa}$ is a compact interval.*

Proof. The proof is similar to Theorem 4.2.22. We only show the sketch. Let $\mathfrak{a}$ be the control input process generated by κ such that $\mathfrak{a}_t = \kappa_t(I_{[0,t]}, \mathfrak{a}_{[0,t-1]})$ for each t. Note that $\mathfrak{a} \in \mathscr{B}(Act^\infty)$ and $\mathfrak{a}_t \in \mathscr{B}(Act)$, where the set of actions Act admits a discrete topology. The weak compactness of the probability law $\{\mathcal{P}_I^{\mu_0,\kappa}\}_{I^\kappa \in \mathbb{IA}^\kappa}$ follows exactly the same reasoning as in Proposition 4.2.21. The convexity of every finite-dimensional distribution of $I^\mathfrak{a}$ can be obtained in similar way as in Theorem 4.2.22 based on the transition procedure (C.6), i.e., for any $q_0, q_{n_1}, \cdots, q_{n_t} \in \mathcal{Q}$,

$$\begin{aligned}&\mathcal{P}_I^{q_0,\kappa}\left[I_0 = q_0, \cdots, I_t = q_{n_t}, I_{t+1} = q_{n_{t+1}}\right]\\ \in&\{\Theta^{a_t}_{n_{t+1},n_t}\Theta^{a_{t-1}}_{n_t,n_{t-1}}\cdots\Theta^{a_0}_{n_1,0}\delta_{q_0} : \Theta^{a_i} \in [\![\Theta^{a_i}]\!], i \in \{0,\cdots,t\}, \text{ and } a_t = \kappa(I_{[0,t]} = q_{[0,t]}, a_{[0,t-1]})\}.\end{aligned}$$

By a standard monotone class argument, the convexity for any Borel measurable set $A \in \mathscr{F}$ measured in the set of laws $\mathcal{P}_I^{q_0,\kappa}$ are guaranteed, which implies the convexity of $S^{q_0,\kappa}$, and hence that of $S^{\mu_0,\kappa}$. □

Theorem 4.3.9. *Let $\mathbb{XU}$ as in (4.23) be a controlled Markov system driven by (4.22). Suppose that there exist a state-level abstraction α, a measure-level abstraction Σ_α, and a BMDP abstraction $\mathbb{IA}$ such that $\mathbb{XU} \preceq_{\Sigma_\alpha} \mathbb{IA}$. Let Ψ be an LTL formula. Suppose the initial distribution ν_0 of $\mathbb{XU}$ is such that $\nu_0(\alpha^{-1}(q_0)) = 1$. Then, given an admissible deterministic control policy κ, there exists a Σ_α-implementation policy ϕ of κ such that*

$$\mathbf{P}_X^{\nu_0,\phi}(X^\phi \vDash \Psi) \in \{\mathcal{P}_I^{q_0,\kappa}(I^\kappa \vDash \Psi)\}_{I^\kappa \in \mathbb{IA}^\kappa}, \quad X^\phi \in \mathbb{XU}^\phi.$$

Proof. The proof should be similar to Lemma 4.2.24 and Theorem 4.2.25. We only show the sketch. We consider $\nu_0 = \delta_{x_0}$ a.s. for simplicity. Note that, at $t = 1$, by the definition of BMDP and Remark 4.3.5, there exists a $u \in a$ such that,

$$\check{\Theta}^a_{ij} \le \nu_1^u(\alpha^{-1}(q_j)) = \int_{\alpha^{-1}(q_j)} \delta_{x_0}\mathcal{T}^u(x_0, dy) \le \hat{\Theta}^a_{ij}, \quad \forall x_0 \in q_0 \text{ and } \forall j \in \{1, 2, \cdots, N+1\},$$

[8]See Appendix C.3 for details

where $a = \kappa(q_0)$, and u is selected accordingly such that the above relation is satisfied. We can easily check that $\mu_1^a = (\nu_1^u(\alpha^{-1}(q_1)), \cdots, \nu_1^u(\alpha^{-1}(q_{N+1}))^T$ is a proper marginal distribution of $\mathbb{IA}$. We then propagate the process inductively according to (C.6) by

1) selecting u at each time according to the realization $a = \kappa_t$;

2) selecting $\mathcal{T}^u$ and $\Theta^a \in [\![\Theta^a]\!]$ at each time via the connection as the above.

We can verify that, by the above selection procedure, there exists a $\mathcal{P}^{q_0,\kappa}$ such that

$$\begin{aligned}&\mathcal{P}^{q_0,\kappa}[I_0 = q_0, \mathfrak{a}_0 = a_0, I_1 = q_i, \mathfrak{a}_1 = a_1, \cdots]\\ =&\mathbf{P}_X^{x_0,\mathfrak{u}}[X_0 = x_0, \mathfrak{u}_0 = u_0, X_1 \in \alpha^{-1}(q_i), \mathfrak{u}_2 = u_2, \cdots]\end{aligned}$$

holds for any finite-dimensional distribution, where $u_t \in a_t$ for all t. By Kolmogrov extension theorem, there exists a unique law $\mathcal{P}_I^{q_0,\kappa}$ for $(I, \mathfrak{a})$ or $I^\kappa \in \mathbb{IA}^\kappa$ such that it has the same measuring results on any $\mathscr{F}$-measurable sets as the law $\mathbf{P}_X^{x_0,\mathfrak{u}}$ of the generated process $(X, \mathfrak{u})$ or $X^\mathfrak{u}$. The Σ_α-implementation π exists and is given as $\phi_t(\cdot \mid X_{[0,t]}, \mathfrak{u}_{[0,t-1]}) = \mathbf{P}_X^{x_0,\mathfrak{u}}[\mathfrak{u}_t = (\cdot) \mid X_{[0,t]}, \mathfrak{u}_{[0,t-1]}]$ in view of (C.8). □

Corollary 4.3.10. *Let $\mathbb{XU}$, its BMDP abstraction $\mathbb{IA}$, an LTL formula Ψ, and a constant $\rho \in [0, 1]$ be given. Suppose there exists a control policy κ such that $I^\kappa \vDash \mathcal{P}_{\bowtie\rho}^{q_0}[\Psi]$ for all $I^\kappa \in \mathbb{IA}^\kappa$, then there exists a policy ϕ such that $X^\phi \vDash \mathbf{P}_{\bowtie\rho}^{\nu_0,\phi}[\Psi]$ for all $X^\phi \in \mathbb{XU}^\phi$ with $\nu_0(\alpha^{-1}(q_0)) = 1$.*

4.3.3 Construction of Robustly Complete BMDP Abstractions

In this subsection, we consider two continuous-state systems with $0 \leq \vartheta_1 \leq \vartheta_2$: $\mathbb{XU}_1$, which is driven by

$$X_{t+1} = f(X_t, \mathfrak{u}_t) + b(X_t)\mathbf{w}_t + \vartheta_1\xi_t^{(1)}, \quad \xi_t^{(1)} \in \overline{\mathbb{B}}, \tag{4.24}$$

and $\mathbb{XU}_2$, which is driven by

$$X_{t+1} = f(X_t, \mathfrak{u}_t) + b(X_t)\mathbf{w}_t + \vartheta_2\xi_t^{(2)}, \quad \xi_t^{(2)} \in \overline{\mathcal{B}}, \tag{4.25}$$

We construct a sound and robustly complete BMDP abstraction $\mathbb{IA}$ for $\mathbb{XU}_1$ in a similar way as in Section 4.2.3, i.e., we build a state-level relation α and a measure-level Σ_α, such that

$$\mathbb{XU}_1 \preceq_{\Sigma_\alpha} \mathbb{IA}, \;\; \mathbb{IA} \preceq_{\Sigma_\alpha^{-1}} \mathbb{XU}_2.$$

We introduce a similar notion for each fixed $u \in \mathcal{U}$ as in Section 4.2.3 and define the set of transition probabilities of $\mathbb{XU}_i$ from any box $[x] \subseteq \mathbb{R}^n$ as

$$\mathbb{T}_i^u([x]) = \{\mathcal{T}^u(x, \cdot) : \; \mathcal{T}^u \in [\![\mathcal{T}^u]\!]_i, \; x \in [x]\}, \; i = 1, 2.$$

The following lemma is straightforward based on Lemma 4.2.31.

Lemma 4.3.11. *Fix any $\vartheta_1 > 0$, any box $[x] \subseteq \mathbb{R}^n$, and $u \in \mathcal{U}$. For all $k > 0$, there exists a finitely terminated algorithm to compute an over-approximation of the set of (Gaussian) transition probabilities from $[x]$, such that*

$$\mathbb{T}_1^u([x]) \subseteq \widehat{\mathbb{T}_1^u([x])} \subseteq \mathbb{T}_1^u([x]) + k\overline{\mathcal{B}}_W,$$

where $\widehat{\mathbb{T}_1^u([x])}$ is the computed over-approximation set of Gaussian measures.

Remark 4.3.12. *We consider modification of the transitions in the same way as Definition 4.2.33 for each fixed $u \in \mathcal{U}$. Similarly, as an extension of notations, we use $\widetilde{(\,\cdot\,)}^u$ to denote the induced quantities of any other types w.r.t. such a modification.*

Theorem 4.3.13. *For any $0 \leq \vartheta_1 < \vartheta_2$, we consider $\mathbb{XU}_i = (\mathcal{X}, \mathcal{U}, \{\widetilde{\mathcal{T}}\}_i, \mathrm{AP}, L_{\mathbb{XU}})$, $i = 1, 2$, that are driven by (4.24) and (4.25), respectively. Then, under Assumption 4.2.10, there exists a rectangular partition $\mathcal{Q}$ (state-level relation $\alpha \subseteq \mathcal{X} \times \mathcal{Q}$), a measure-level relation Σ_α and a finite-state abstraction system $\mathbb{I} = (\mathcal{Q}, Act, \{\Theta\}, \mathrm{AP}, L_{\mathbb{IA}})$ such that*

$$\mathbb{XU}_1 \preceq_{\Sigma_\alpha} \mathbb{IA}, \quad \mathbb{IA} \preceq_{\Sigma_\alpha^{-1}} \mathbb{XU}_2. \tag{4.26}$$

Proof. We construct a finite-state BMDP in a similar way as in Theorem 4.2.35. By Assumption 4.2.10, we use uniform rectangular partition $\mathcal{Q}$ on $\mathcal{W}$. We then let the state-level relation be $\alpha = \{(x, q) : q = \eta\lfloor\frac{x}{\eta}\rfloor\} \cup \{(\Delta, \Delta)\}$, and $Act = \{a : \varrho\lfloor\frac{u}{\varrho}\rfloor\}$, where $\lfloor\cdot\rfloor$ is the floor function. The parameters η, ϱ are to be chosen later. Denote the number of discrete nodes by $N + 1$.

We construct the measure-level abstraction by the same procedure as in Theorem 4.2.35. We repeat the procedure with updated notations for the control systems. For any fixed $u = a \in Act$, for any $\widetilde{\mathcal{T}}^u \in \widetilde{[\![\mathcal{T}^u]\!]}_1$ and $q \in \mathcal{Q}$,

1) for all $\tilde{\nu}^u \sim \widetilde{\mathcal{N}}(m, s^2) \in \widetilde{\mathbb{T}}_1^u(\alpha^{-1}(q), \cdot)$, store $\{(m_l, s_l) = (\eta\lfloor\frac{m}{\eta}\rfloor, \eta^2\lfloor\frac{s^2}{\eta^2}\rfloor)\}_l$;

2) for each l, define $\tilde{\nu}_l^{u,\mathrm{ref}} \sim \widetilde{\mathcal{N}}(m_l, s_l)$ (implicitly, we need to compute $\nu_l^{u,\mathrm{ref}}(\alpha^{-1}(\Delta))$); compute $\tilde{\nu}_l^{u,\mathrm{ref}}(\alpha^{-1}(q_j))$ for each $q_j \in \mathcal{Q} \setminus \Delta$;

3) for each l, define $\mu_l^{u,\mathrm{ref}} = [\tilde{\nu}_l^{u,\mathrm{ref}}(\alpha^{-1}(q_1)), \cdots, \tilde{\nu}_l^{u,\mathrm{ref}}(\alpha^{-1}(q_N)), \tilde{\nu}_l^{u,\mathrm{ref}}(\alpha^{-1}(\Delta)]$;

4) compute $\mathbf{ws} := (\sqrt{2N} + 2)\eta$ and $\mathbf{tv} := N\eta \cdot \mathbf{ws}$;

5) construct $[\![\mu^u]\!] = \bigcup_l \{\mu : \left\|\mu - \mu_l^{u,\mathrm{ref}}\right\|_{\mathrm{TV}} \leq \mathbf{tv}(\eta),\ \mu(\Delta) + \sum_j^N \mu(q_j) = 1\}$;

6) let $\Sigma_\alpha := \{(\tilde{\nu}^u, \mu^u),\ \ \mu^u \in [\![\mu^u]\!]\}$ be a relation between $\tilde{\nu}^u \in \widetilde{\mathbb{T}}^u(\alpha^{-1}(q))$ and the generated $[\![\mu^u]\!]$.

Repeat the above step for all q and then for all $u = a \in \mathit{Act}$, the relation Σ_α is obtained. We denote $\mathfrak{G}_i^u := \widetilde{\mathbb{T}}_1^u(\alpha^{-1}(q_i), \cdot)$ and $\widehat{\mathfrak{G}}_i^u := \widetilde{\mathbb{T}}_1^u(\widehat{\alpha^{-1}(q_i)}, \cdot)$.

For each $u = a \in \mathit{Act}$, for $i \leq N$, let $[\![\Theta_i^u]\!] = \Sigma_\alpha(\widehat{\mathfrak{G}}_i^u)$ and the transition collection be $[\![\Theta^u]\!]$. It is clear that the finite-state BMDP $\mathbb{IA}$ abstracts $\mathbb{XU}_1$ based on Definition 4.3.3: for each a, there exists $u \in \mathcal{U}$ (where we set it to be a), such that for any $\tilde{\nu^u} \in \mathfrak{G}_i^u$ and hence in $\widehat{\mathfrak{G}}_i^u$, there exists a discrete measures in $\Theta_i^u \in \Sigma_\alpha(\widehat{\mathfrak{G}}_i^u)$ such that for all q_j we have $\tilde{\nu}^u(\alpha^{-1}(q_j)) = \Theta_{ij}^u$. The proof is done by the exact same way as the proof of Claim 1, Theorem 4.2.35 for each fixed control input.

Now we consider the size of η and ϱ such that the constructed BMDP can be abstracted by $\mathbb{XU}_2$ via the converse relation Σ_α^{-1}. Note that $a \in u + \varrho\overline{\mathbb{B}}$ for any $u \in \mathcal{U}$. We need to choose η, ϱ and k sufficiently small such that

$$2\eta + N\eta \cdot \mathbf{tv}(\eta) + L\rho + k \leq \vartheta_2 - \vartheta_1, \tag{4.27}$$

where L is the Lipschitz constant of f, then we have

$$\begin{aligned}\Sigma_\alpha^{-1}(\Sigma_\alpha(\widehat{\mathfrak{G}}_i^u)) &\subseteq \widehat{\mathfrak{G}}_i^u + (2\eta + N\eta \cdot \mathbf{tv}(\eta)) \cdot \overline{\mathcal{B}}_W + L\varrho\overline{\mathbb{B}} \\ &\subseteq \mathfrak{G}_i^u + (2\eta + N\eta \cdot \mathbf{tv}(\eta) + L\varrho + k) \cdot \overline{\mathcal{B}}_W\end{aligned} \tag{4.28}$$

for each i. The only difference from Claim 2 and 3 of Theorem 4.2.35 is in terms of the control. By the construction, we can verify that for each $u \in \mathcal{U}$, there exists an $a \in \mathit{Act}$ (which is guaranteed by the finite covering relation $a \in u + \varrho\overline{\mathbb{B}}$) such that the choice in (4.28) makes $\Sigma_\alpha^{-1}(\Sigma_\alpha(\widehat{\mathfrak{G}}_i^u)) \subseteq \widetilde{\mathbb{T}}_2^u(\alpha^{-1}(q_i))$, which completes the proof. □

Since the proof of the robust completeness of the BMDP $\mathbb{IA}$ is constructive, we can algorithmically synthesize a control strategy for $\mathbb{XU}_1$ by generating $\mathbb{IA}$ and then solving a discrete synthesis problem for $\mathbb{IA}$ with some probabilistic LTL specification. In view of Corollary 4.3.10, if a control strategy κ exists to fulfill the probabilistic specification for $\mathbb{IA}^\kappa$, then there exists a policy ϕ to guarantee the satisfaction of $\mathbb{XU}_1^\phi$. On the other hand, if there is no policies to realize a specification for $\mathbb{IA}$, then the system $\mathbb{XU}_2$ is also not controllable w.r.t. the same specification. The latter is implied by the following corollary.

Corollary 4.3.14. *Given an LTL formula Ψ, let $S_2^{\nu_0,\phi} = \{\mathbf{P}_X^{\nu_0,\phi}(X \vDash \Psi)\}_{X^\phi \in \mathbb{XU}_2}$ be the set of the satisfaction probability of Ψ under a control policy ϕ for the system $\mathbb{XU}_2$. Then, for each* *control policy ϕ of $\mathbb{XU}_2$, there exists a policy κ for $\mathbb{IA}$ such that $S_{\mathbb{IA}}^{q_0,\kappa} \subseteq S_2^{\nu_0,\phi}$ for any initial conditions satisfying $\nu_0(\alpha^{-1}(q_0)) = 1$, where $S_{\mathbb{IA}}^{q_0,\kappa} = \{\mathcal{P}_I^{q_0,\kappa}(I \vDash \Psi)\}_{I^\kappa \in \mathbb{IA}}$. Both $S_{\mathbb{IA}}^{q_0,\kappa}$ and $S_2^{\nu_0,\phi}$ are compact.*

Proof. The inclusion is done in a similar way as Theorem 4.3.9 by the inductive construction of probability laws. The compactness also follows a similar way as Proposition 4.2.23 for each policy. □

4.4 A Discussion on Stochastic Control Systems with Noisy Observation

In this section, we discuss the case when the observations of the sample paths are corrupted by noise. Since there is no direct access to the exact sample path information, we aim to obtain optimal estimates of the sample path signal based on noisy observations, which is known as the optimal filter. Apart from the nonlinear filtering, the philosophy of constructing sound and robustly complete abstractions for such systems maintain the same. We hence do not reiterate the procedure in this section but rather deliver a discussion on the mathematical complexity of the potential abstractions. Before we proceed, we briefly introduce the theory of nonlinear filtering.

4.4.1 Nonlinear Filtering for Discrete-Time Systems

Consider the discrete-time signal and observation of the following form

$$X_{t+1} = f(X_t, \mathfrak{u}_t) + b(X_t)\mathbf{w}_t, \tag{4.29a}$$

$$Y_t = h(X_t) + \beta_t, \tag{4.29b}$$

where Y is a $\mathcal{Y}$-valued observation via a continuous Borel measurable function h and i.i.d. Gaussian process β with proper dimensions. We also set $\mathbf{w}$ and β to be mutually independent.

Similar to (C.1), for any fixed $T > 0$, we define the short hand notation for the history of observation

$$Y_{[0,T]} := \{Y_t\}_{t\in[0,T]} \tag{4.30}$$

Unlike the system without corrupted observations, it is natural to suppose that the selection of a control at time T is based on $Y_{[0,T]}$ and $\mathfrak{u}_{[0,T-1]}$. An admissible control policy $\kappa = \{\kappa_t\}$ in this case is such that, for each fixed $t > 0$, we have

$$\kappa_t(\mathfrak{C} \mid Y_{[0,t]}; \mathfrak{u}_{[0,t-1]}) = \mathbf{P}[\mathfrak{u}_t \in \mathfrak{C} \mid Y_{[0,t]}; \mathfrak{u}_{[0,t-1]}], \quad \mathfrak{C} \in \mathscr{B}(\mathcal{U}). \tag{4.31}$$

A deterministic admissible policy κ is such that $\mathfrak{u}_t = \kappa_t(Y_{[0,t]}; \mathfrak{u}_{[0,t-1]})$.

Let $H(Y_t \in A \mid X_t = x_t)$, $A \in \mathscr{B}(\mathcal{Y})$, be the observation channel, which is the transition kernel generated by (4.29b). Given any initial distribution μ_0 of X, the probability law $\mathbf{P}^{\mu_0,\kappa}$ of $(X, Y, \mathfrak{u}) := \{X_t, Y_t, \mathfrak{u}_t\}_{t\in\mathbb{N}}$ can be uniquely determined, whose finite-dimensional distributions are given in the form of

$$\begin{aligned}
&\mathbf{P}^{\mu_0,\kappa}[X_0 \in \Gamma_0, Y_0 \in A_0, \mathfrak{u}_0 \in \mathfrak{C}_0, \cdots, X_t \in \Gamma_t, Y_t \in A_t, \mathfrak{u}_t \in \mathfrak{C}_t] \\
&= \int_{\Gamma_0} \mu_0(dx_0) \int_{A_0} H(dy_0 \mid X_0 = x_0) \int_{\mathfrak{C}_0} \kappa_0(du_0 \mid X_0 = x_0) \cdots \\
&\quad \times \int_{\Gamma_t} \Theta_{t-1}^{u_{t-1}}(dx_t \, | X_{t-1} = x_{t-1}; \mathfrak{u}_{t-1} = u_{t-1}) \times \int_{A_t} H(dy_t \mid X_t = x_t) \\
&\quad \times \int_{\mathfrak{C}_t} \kappa_T(d\mathfrak{u}_t \, | Y_0 = y_0, \cdots Y_t = y_t; \mathfrak{u}_0 = u_0, \cdots, \mathfrak{u}_{t-1} = u_{t-1}).
\end{aligned} \tag{4.32}$$

Given a policy κ (we set it to be deterministic without loss of generality), the estimation of X_t given $Y_{[0,t]}$ that minimizes the mean square error loss is given as

$$\Pi_t(\Gamma) := \mathbf{P}^{\mu_0,\kappa}[X_t \in \Gamma \mid Y_{[0,t]}, \mathfrak{u}_{[0,t-1]}], \quad \Gamma \in \mathscr{B}(\mathcal{X}).$$

We call this random measure $\Pi_t \in \mathfrak{P}(\mathcal{X})$ for each t the optimal filter. Using Bayes rule, we have

$$\begin{aligned}
\Pi_t(\Gamma) &= \mathbf{P}^{\mu_0,\kappa}[X_t \in \Gamma \mid Y_{[0,t]}, \mathfrak{u}_{[0,t-1]}] \\
&= \frac{\int_{\mathcal{X}} H(Y_t | X_t = x_t) \Theta_{t-1}^{\mathfrak{u}_{t-1}}(x_{t-1}, \Gamma) \cdot \Pi_{t-1}(dx_{t-1})}{\int_{\mathcal{X}} \int_{\mathcal{X}} H(Y_t | X_t = x_t) \Theta_{t-1}^{\mathfrak{u}_{t-1}}(x_{t-1}, dx_t) \cdot \Pi_{t-1}(dx_{t-1})} \\
&=: F(\Pi_{t-1}, Y_{t-1}, \mathfrak{u}_{t-1})(\Gamma).
\end{aligned} \tag{4.33}$$

where $\mathfrak{u}_{t-1}$ is determined by $\kappa_{t-1}(Y_{[0,t-1]}, \mathfrak{u}_{[0,t-2]})$. We can simply regard the denominator of $F(\Pi_{t-1}, Y_{t-1}, \mathfrak{u}_{t-1})$ as a normalizer

$$\mathfrak{n}(Y_t) := \int_{\mathcal{X}} \int_{\mathcal{X}} H(Y_t | X_t = x_t) \Theta_{t-1}^{\mathfrak{u}_{t-1}}(x_{t-1}, dx_t) \cdot \Pi_{t-1}(dx_{t-1})$$

since the dependence on $\mathcal{X}$ is averaged out. It can also be shown that the process $(\Pi, \mathfrak{u}) := \{\Pi_t, \mathfrak{u}_t\}_{t\in\mathbb{N}}$ is a controlled Markov process [146] with transition probability

$$\mathbf{P}[\Pi_{t+1} \in D \mid \Pi_t = \pi_t, \mathfrak{u}_t = u_t] = \int_{\mathcal{Y}} \mathbb{1}_{\{F(\pi_t, y_t, u_t) \in D\}} \cdot \mathfrak{n}(dy_t), \quad D \in \mathscr{B}(\mathfrak{P}(\mathcal{X})).$$

We also use $\Pi^{\mathfrak{u}}$ to emphasize the marginal behavior of the process $(\Pi, \mathfrak{u})$. Given the observations and the adaptively generated control signal, the optimal estimation of the conditional probability of satisfying any LTL formula Ψ is given by

$$\mathbf{P}_{\Pi}^{\mu_0,\mathfrak{u}}[X \vDash \Psi] := \mathbf{P}^{\mu_0,\mathfrak{u}}[X \vDash \Psi \mid Y] = \int_{\mathcal{X}^\infty} \mathbb{1}_{\{X \vDash \Psi\}} \, \Pi^{\mathfrak{u}}(dx). \tag{4.34}$$

Note that it is difficult to obtain the full knowledge of Y, our goal is to generate control policies such that the optimal estimation $\mathbf{P}^{\mu_0,\mathfrak{u}}[X \vDash \Psi \mid Y] \in \mathbb{R}$ possesses certain confidence of satisfying the probabilistic requirement given any realization of observation. The above derivation converts the problem into a fully observed controlled Markov process $(\Pi, \mathfrak{u})$ via an enlargement of the state space, where control policies and even optimal control policies can be synthesized accordingly for the (hypothetically) fully observed Π [146]. The policy fulfilling the goal mentioned above is thereby decidable.

The construction of the optimal filter process (or the function F in (4.33)) can be decomposed into a two-step recursion based on the transition relation in (4.33).

Prediction (Prior): At time t, $\Pi_{t-1}(dx)$ is feed into the r.h.s. of the prior knowledge of the dynamics for X, i.e., (4.29a). The prediction of X_t based on $Y_{[0,t-1]}$ as well as the u determined at t is such that

$$\hat{\Pi}_t(dx) = \int_{\mathcal{X}} \Theta_t^u(\tilde{x}, dx)\Pi_{t-1}(d\tilde{x}). \tag{4.35}$$

Filtering (Posterior): This step is to assimilate the observation at the instant t, which is given as

$$\Pi_t(dx) = \mathfrak{n}(Y_t)H(Y_t|X_t = x)\hat{\Pi}_t(dx), \tag{4.36}$$

where $\mathfrak{n}(Y_t) = \int_{\mathcal{X}} H(Y_t|X_t = x)\hat{\Pi}_t(dx)$ is the normalizer.

For numerical approximation, we simulate and propagate the optimal filter process using matrix approximations of each step's transition kernel, whereas for formal abstractions, we need to find the 'inclusion' of the transitions for each step as usual.

4.4.2 A Brief Discussion on Stochastic Abstractions for Control Systems with Noisy Observations

Motivated by generating optimal control policies using the knowledge filter process $(\Pi, \mathfrak{u})$, the stochastic abstractions for partially observed processes can be reduced to obtain a sound and robustly complete abstraction for the process $(\Pi, \mathfrak{u})$. To convey the idea, we simply consider the following two systems with noisy observations

$$X_{t+1} = f(X_t, \mathfrak{u}_t) + b(X_t)\mathbf{w}_t + \vartheta_1 \xi_t^{(1)}, \tag{4.37a}$$

$$Y_t = X_t + \beta_t + \varsigma_1 \zeta_t^{(1)}, \tag{4.37b}$$

and

$$X_{t+1} = f(X_t, \mathfrak{u}_t) + b(X_t)\mathbf{w}_t + \vartheta_2 \xi_t^{(2)}, \tag{4.38a}$$

$$Y_t = X_t + \beta_t + \varsigma_2 \zeta_t^{(2)}, \tag{4.38b}$$

where $\zeta_t^{(i)} \in \overline{\mathcal{B}}$ are i.i.d. for each t and each $i \in \{1, 2\}$, the intensities satisfy $0 \le \varsigma_1 < \varsigma_2$. The rest of the notations are as previously mentioned.

We convert the filter processes that are generated by (4.37) and (4.38) into the expression of controlled Markov systems

$$\mathbb{FU}_i = (\mathcal{X}, \mathcal{Y}, \mathcal{U}, \{T\}_i, [\![H]\!]_i, \mathrm{AP}, L_{\mathbb{FU}}), \quad i = 1, 2, \tag{4.39}$$

where the additional $\mathcal{Y}$ and its collection of observation channel $[\![H]\!]_i$ are needed in the filtering step for generating the controlled filter process $\Pi^{\mathfrak{u}}$. The other notions are the same as previously mentioned.

To find an abstraction for $\mathbb{FU}_1$, we need a state-level relation or discretization α as usual. Then, we need both $\{T\}_1$ and $[\![H]\!]_1$ to be abstracted via some measure-level relation, so that the transition probability of Π is abstracted by a set of discrete transition probabilities given the same set of discrete observations in the sense of (2) of Definition 4.3.3.

Now we denote the BMDP abstraction for $\mathbb{FU}_1$ as

$$\mathbb{IA} = (\mathcal{Q}, \mathcal{Y}_{\mathcal{Q}}, \mathit{Act}, \{\Theta\}, [\![H_{\mathcal{Q}}]\!], \mathrm{AP}, L_{\mathbb{IA}}), \tag{4.40}$$

where $\mathcal{Y}_{\mathcal{Q}}$ is the discretized observation states that are obtained by the state-level relation α, and $[\![H_{\mathcal{Q}}]\!]$ is the collection of the discrete observation channels that are obtained based on some measure-level relation Σ_α. The intuition of $\mathbb{IA}$ is that we need 'more' transitions in the abstraction for the prior knowledge of the dynamics that are related via the measurablility of labelled nodes, as well as 'more' transitions for the filtering step to obtain enough observations for decision making.

Remark 4.4.1. *Note that the collection $\{\Theta\}$ for each u can be obtained in the same way as the case without noisy observation. To obtain $[\![H_{\mathcal{Q}}]\!]$, we notice that*

$$H(dy \mid x) = \frac{1}{\sqrt{2\pi}} \exp\left[\frac{(y-x)^2}{2}\right] dy.$$

The over/under-approximation for any $x \in \alpha^{-1}(q_i)$ to the observation $\alpha^{-1}(q_j)$ can be evaluated accordingly.

The soundness of $\mathbb{IA}$ for the controlled filter process system $\mathbb{FU}_1$ is in the following sense: given any initial distribution, for each κ (based on the discrete observation $Y_{\mathcal{Q}}$) for $\mathbb{IA}$, there exists a control policy ϕ such that for each $\Pi^\phi \in \mathbb{FU}_1^\phi$,

✧ there exists a $\Pi^{d,\kappa} \in \mathbb{IA}^\kappa$ whose observation process $Y_{\mathcal{Q}}$ has the same probability with Y of Π^ϕ on each discrete node $q \in \mathcal{Q}$, and $\Pi^{d,\kappa}$ has the same evaluation on all the discrete measurable sets $A \in \mathscr{F}$ with Π^ϕ;

✧ the discrete probability law $\mathbb{P}^{d,\kappa}$ for $\Pi^{d,\kappa} \in \mathbb{IA}^\kappa$ forms a convex and weakly compact set;

✧ the optimal estimation satisfies, for a given $p \in [0,1]$,

$$\begin{aligned}&\int_{\Pi^\phi \in \mathscr{B}(\mathfrak{P}(\mathcal{X}))} \mathbb{1}_{\left\{p \bowtie \int_{\mathcal{X}^\infty} \mathbb{1}_{\{X^\phi \models \Psi\}} \Pi^\phi(dx)\right\}} \mathbb{P}^\phi(d\Pi^\phi) \\ &\in \left\{ \int_{\Pi^{d,\kappa} \in \mathscr{B}(\mathfrak{P}(\mathcal{Q}))} \mathbb{1}_{\left\{p \bowtie \int_{\mathcal{Q}^\infty} \mathbb{1}_{\{I^\kappa \models \Psi\}} \Pi^{d,\kappa}(dx)\right\}} \mathbb{P}^{d,\kappa}(d\Pi^{d,\kappa}) \right\}_{\Pi^{d,\kappa} \in \mathbb{IA}^\kappa},\end{aligned}$$

which is equivalent as

$$\begin{aligned}&\int_{\Pi^\phi \in \mathscr{B}(\mathfrak{P}(\mathcal{X}))} \mathbb{1}_{\left\{\mathbf{P}_\Pi^{\nu_0,\phi}[X^\phi \models \Psi] \bowtie p\right\}} \mathbb{P}^\phi(d\Pi^\phi) \\ &\in \left\{ \int_{\Pi^{d,\kappa} \in \mathscr{B}(\mathfrak{P}(\mathcal{Q}))} \mathbb{1}_{\left\{\mathcal{P}_\Pi^{\mu_0,\kappa}[X^\phi \models \Psi] \bowtie p\right\}} \mathbb{P}^{d,\kappa}(d\Pi^{d,\kappa}) \right\}_{\Pi^{d,\kappa} \in \mathbb{IA}^\kappa}\end{aligned} \tag{4.41}$$

A proper task is to find an control policy such that the optimal estimation of the probabilistic specification of $X \models \Psi$ has a confidence at least $q \in [0,1]$, i.e., $\mathbb{P}^\phi\left(\mathbf{P}^{\mu_0,\phi}[X \models \Psi \mid Y] \bowtie p\right) \geq q$. Then we can search control policies κ in $\mathbb{IA}$ for all the filter process Π^d, such that strategy can make the lower bound of

$$\left\{ \int_{\Pi^{d,\kappa} \in \mathscr{B}(\mathfrak{P}(\mathcal{Q}))} \mathbb{1}_{\left\{\mathcal{P}_\Pi^{\mu_0,\kappa}[X^\phi \models \Psi] \bowtie p\right\}} \mathbb{P}^{d,\kappa}(d\Pi^{d,\kappa}) \right\}_{\Pi^{d,\kappa} \in \mathbb{IA}^\kappa}$$

greater than or equal to q.

The robust completeness can be verified in a similar way as Section 4.3.3, except now we need to decompose the procedure to guarantee the robust completeness for both prediction and filtering steps. The discretization need to rely on the value of $\vartheta_2 - \vartheta_1$ and $\varsigma_2 - \varsigma_1$.

Remark 4.4.2. *Recall Section 4.2.4, where we have compared the abstraction with the numerical simulation of the probability measure using finite-difference schemes for Fokker-Planck equations. The counterpart of Fokker-Planck equations for evaluating the probability law of the optimal filter in systems with noisy observations is the famous Zakai's equation*[9]*, which is an SPDE. The approximation of such a solution already suffers from the curse of dimensionality. Using formal abstractions to enlarge the partially observed processes to the filter processes with full observations, based on which control policies can be determined and be utilized onto the partially observed cases, seems tedious and impractical. Besides the theoretical formal guarantee of a confidence of a satisfaction probability (i.e., a probabilistic requirement of the probabilistic specification), the abstraction essentially solves the continuous probability law of a continuous conditional expectation (or a random measure) upon some process with discrete labels using discrete inclusions. We hence do not recommend readers to complicate the problem.*

4.5 Summary

In this chapter, we investigated the mathematical properties of formal abstractions for discrete-time control-free and controlled stochastic systems. We discussed the motivation of constructing formal stochastic abstractions in Section 4.1 and the philosophy in comparison to numerical approximations in Section 4.2.4. The difference from the formal abstractions and the associated mathematical properties for deterministic systems are discussed in Section 4.2.5. A brief discussion on the extension of stochastic abstractions for controlled stochastic systems with noisy observation was provided in Section 4.4. The construction of such abstractions can be analogous to solving a discretized version of Zakai's SPDE equation via formal inclusions, which suffers from a curse of excessive dimensionality.

In words, through the proofs, we implicitly showed the pros and cons of IMC or BMDP abstractions. Using IMC or BMDP as abstractions can indeed guarantee the soundness, provided that conservative estimates for the probability measures are taken. However, they slightly over-approximate the original measure space by subtly changing its non-convexity. This makes the completeness achievable only via more perturbed $\mathcal{L}^1$ uncertainties. Hence the IMC abstractions in [52, 50] are sound but not necessarily complete in this sense. We view the most important contribution of our work to be providing an appropriate mathematical language to discuss soundness and completeness of abstractions of stochastic systems. Our work showed that abstractions for stochastic systems with extra uncertainties are generally not straightforward extensions from their non-stochastic counterparts.

[9] We omit the content here and kindly refer readers to [34] for details.

Chapter 5

Hopf Bifurcations of Moore-Greitzer PDE Model with Additive Noise

We have seen the Moore-Greitzer model in Section 1.2 describing flow and pressure changes in axial-flow jet engine compressors, whose state $u(t)$ takes values in a product Hilbert space $\mathscr{H} = \mathcal{H} \times \mathbb{R}^2$. The abstract form of evolution equation is given in (1.6), which we iterate as below:

$$\partial_t u = \mathcal{A}(\gamma)u + f(\gamma, u).$$

Three types of Hopf bifurcations may occur as the parameter γ varies due to the complicated spectrum properties of $\mathcal{A}(\gamma)$ subjected to the fluid's viscosity and the geometric configuration of the compressors (see Remark 1.2.2). However, in contrast to the PDEs, ODEs, and 2-dimension SDEs, the Hopf bifurcation in SPDEs is not well understood. The goal of this chapter is to extend the work of [29] and rigorously develop low-dimensional approximations using a multiscale analysis approach near the deterministic stall bifurcation[1] point γ_c in the presence of additive noise acting on the fast modes. We also show that the reduced-dimension approximation model (SDEs) contains multiplicative noise.

To make the analysis less cumbersome, we work on the localized model with topological equivalence

$$\partial_t \hat{u} = \mathcal{A}(\gamma)\hat{u} + B(\hat{u}, \hat{u}) + F(\hat{u}, \hat{u}, \hat{u}), \tag{5.1}$$

where $\hat{u} = u - u_e(\gamma) =: [\hat{\mathrm{g}}(t), \hat{\Phi}_\delta(t), \hat{\Psi}_\delta(t)]^T$ is the perturbation around $u_e(\gamma)$, the operators $B(\cdot, \cdot)$ and $F(\cdot, \cdot, \cdot)$ represent respectively bilinear and trilinear mappings. As for the system

[1] Recall that the stall bifurcation occurs in the subspace $\mathcal{H}$. We choose the configuration parameter of the compressor and the fluid's viscosity such that the indicator $\Delta > 0$ as in (1.10).

(5.1), the new equilibrium point satisfies $\hat{u}_e(\gamma) = 0$ (i.e. the R.H.S. of (5.1) is always 0) for all $\gamma = \gamma_c + \varepsilon^2\mathfrak{q}$ with $\varepsilon \in (0,1)$ and $\mathfrak{q} \in \mathbb{R}$.

Note that the existence of the center manifold for the deterministic Moore-Greitzer model is well understood [176]. The evolution of states on the center manifold is studied by naturally separating the dynamics into critical slowly-varying modes and fast modes. If we denote the orthogonal projection by $P_c : \mathscr{H} \to \mathscr{H}_1^c$ as well as $P_s := I - P_c$, the solution can be represented as $\mathscr{H} \ni \hat{u} = \hat{x} + \hat{y}$ with $\hat{x} \in P_c\mathscr{H}$ and $\hat{y} \in P_s\mathscr{H}$. Therefore, (5.1) can be converted into an equivalent form:

$$\begin{aligned} \partial_t \hat{x} &= \mathcal{A}(\gamma)\hat{x} + P_c[B(\hat{u},\hat{u}) + F(\hat{u},\hat{u},\hat{u})], \\ \partial_t \hat{y} &= \mathcal{A}(\gamma)\hat{y} + P_s[B(\hat{u},\hat{u}) + F(\hat{u},\hat{u},\hat{u})]. \end{aligned} \tag{5.2}$$

In the neighborhood of γ_c, the state $\hat{y}$ evolves much faster than $\hat{x}$. The analytical center manifold can be locally represented as a graph of a smooth function h, which is asymptotically attractive in the sense that $\lim_{t\to\infty} \hat{y}(t) = h(\hat{x}(t))$ [78]. Therefore, the dominating dynamics in the center manifold depends only on $\hat{x}$:

$$\partial_t \hat{x} = \mathcal{A}(\gamma)\hat{x} + P_c[B(\hat{x}+h(\hat{x}), \hat{x}+h(\hat{x})) + F(\hat{x}+h(\hat{x}), \hat{x}+h(\hat{x}), \hat{x}+h(\hat{x}))]. \tag{5.3}$$

The real part and imaginary part of the quantity $\hat{z} = c\langle e^{-i\theta}, \hat{x}\rangle \in \mathbb{C}$ solve a 2-dimensional amplitude equation (ODE) that is equivalent to (5.3), where c denotes a normalizer.

Since $h(\hat{x})$ is generally not easy to obtain with precise information, the driving dynamics in (5.3) are usually approximated in proper ways. Multiscale method [133] has shown its maturity in application to local bifurcation analysis and deriving the approximated dynamics of (5.3). To do this, we approximate $\mathcal{A}(\gamma)$ around γ_c by $\mathcal{A}^{\mathfrak{c}} + \varepsilon^2\mathcal{A}^{\mathfrak{q}}$, where $\mathcal{A}^{\mathfrak{c}} := \mathcal{A}(\gamma_c)$ and $\mathcal{A}^{\mathfrak{q}} := \mathfrak{q}\mathcal{A}'(\gamma_c)$. We make the impact of $\mathcal{A}^{\mathfrak{q}}$ significant by 'zooming' into a small neighborhood of $\hat{u} = 0$ and using the 'fast play mode'. That is, we consider $x(t) = \varepsilon^{-1}\hat{x}(\varepsilon^{-2}t)$ and $y(t) = \varepsilon^{-1}\hat{y}(\varepsilon^{-2}t)$, and look at

$$\begin{aligned} \partial_t x &= \mathcal{A}^{\mathfrak{q}}x + P_c[\varepsilon^{-1}B(x+y,x+y) + F(x+y,x+y,x+y)], \\ \partial_t y &= \varepsilon^{-2}\mathcal{A}^{\mathfrak{c}}y + \mathcal{A}^{\mathfrak{q}}y + P_s[\varepsilon^{-1}B(x+y,x+y) + F(x+y,x+y,x+y)], \end{aligned} \tag{5.4}$$

where $\mathcal{A}^{\mathfrak{c}}x = 0$ when it is restricted to the critical modes. The fast varying modes y are homogenized (or averaged) into the slowly-varying modes x. The homogenization (or averaging) result restricted to $P_c\mathscr{H}$, which is also only dependent on x, can be obtained based on [133, Chapter 13]. This approximate result should be asymptotically convergent to (5.3) as $\varepsilon \to 0$ with the same topological stability.

Now we introduce small randomness on some probability space $(\mathbf{\Omega}, \mathcal{F}, \mathbf{P})$. The evolution of the axial flow in an engine compressor with unsteady turbulence is modelled by the abstract Moore-Greitzer stochastic PDE, written locally as

$$d\hat{u} = \mathcal{A}(\gamma)\hat{u}dt + B(\hat{u}, \hat{u})dt + F(\hat{u}, \hat{u}, \hat{u})dt + \varepsilon d\widetilde{W}_t, \quad \hat{u}(0) = \hat{u}_0, \tag{5.5}$$

where $\widetilde{W}$ represents the effect of turbulence [92, 73], modeled by an additive Gaussian noise (white in time, either white or colored in space; see details in Definition D.2.1 and D.2.2) with a small strength ε (same as the ε in $\gamma = \gamma_c + \varepsilon^2\mathfrak{q}$). The random perturbations are small, but over a long time their effect can be significant on the slow dynamics of the amplitudes of the critical modes (center manifold).

In this chapter, a complex-valued SDE or a two-dimensional real-valued SDE, regarded as the stochastic amplitude equations of the dominant dynamics, is derived for the stall bifurcation. We achieve this by investigating $v(t) := \varepsilon^{-1}\hat{u}(\varepsilon^{-2}t)$ that solves

$$dv = \varepsilon^{-2}\mathcal{A}^{\mathfrak{c}}vdt + \mathcal{A}^{\mathfrak{q}}vdt + \varepsilon^{-1}B(v,v)dt + F(v,v,v)dt + \varepsilon^{-1}dW_t, \quad v(0) = v_0, \tag{5.6}$$

where $W_t := \varepsilon\widetilde{W}_{\varepsilon^{-2}t}$ is a new Wiener process. To this end, we Denote the solution of (5.6) by $v(t) = [\mathfrak{g}(t), \Phi_\delta(t), \Psi_\delta(t)]^T$, where $\mathfrak{g}(t) = \varepsilon^{-1}\,\mathrm{g}(\varepsilon^{-2}t)$, $\Phi_\delta(t) = \varepsilon^{-1}(\Phi(\varepsilon^{-2}t) - \Phi_e(\varepsilon^{-2}t))$, and $\Psi_\delta(t) = \varepsilon^{-1}(\Psi(\varepsilon^{-2}t) - \Psi_e(\varepsilon^{-2}t))$.

Due to the natural separation of the temporal scales close to the deterministic bifurcation points γ_c, we restrict our attention to the small region of the parameter $\gamma = \gamma_c + \varepsilon^2\mathfrak{q}$ in the vicinity of stall bifurcation point γ_c, and show how the heavily damped stable modes enter the critical modes through the nonlinearities. The work, as an extension of [29], is based on a multiscale analysis of the coupling between the slow and fast modes.

Before proceeding, we provide the locally critical and stable dynamics for the stall case. A similar procedure can be used to study the surge as well as the stall-surge cases.

5.1 Preliminaries

5.1.1 Projection and Simplifications

We first provide the explicit form of $\mathcal{A}^{\mathfrak{c}}$, $\mathcal{A}^{\mathfrak{q}}$, B and F in (5.4) or (5.6) for $\gamma = \gamma_c + \varepsilon^2\mathfrak{q}$ and $\mathfrak{q} \in \mathbb{R}$. Recall that $\mathcal{A}(\gamma) = A + Df_{u_e}(\gamma)$, where $Df_{u_e}(\gamma)$ is the Fréchet derivative about u_e. We then have $\mathcal{A}^{\mathfrak{c}} := \mathcal{A}(\gamma_c) = A + Df_{u_e}(\gamma_c)$, $\mathcal{A}^{\mathfrak{q}} := \mathfrak{q}\mathcal{A}'(\gamma_c) = A + \mathfrak{q}Df'_{u_e}(\gamma_c)$; the associated eigenvalues (see 5 of Section 1.2.2) of $\mathcal{A}^{\mathfrak{c}}$ and $\mathcal{A}^{\mathfrak{q}}$ are respectively denoted by $(\cdot)^{\mathfrak{c}}$ and $(\cdot)^{\mathfrak{q}}$.

Thus,

$$Df_{u_e}(\gamma_c) = \begin{bmatrix} \mathfrak{a}\psi'_{c,\gamma_c}\,\mathrm{K}^{-1} & 0 & 0 \\ 0 & \frac{1}{\mathfrak{l}_c}\psi'_{c,\gamma_c} & -\frac{1}{\mathfrak{l}_c} \\ 0 & \frac{1}{4\mathfrak{B}^2\mathfrak{l}_c} & \frac{1}{4\mathfrak{B}^2\mathfrak{l}_c}\mathcal{S}'_{\mu_c} \end{bmatrix}, \tag{5.7}$$

$$\mathfrak{q}Df'_{u_e}(\gamma_c) = \begin{bmatrix} \mathfrak{a}(\psi''_{c,\gamma_c}\Phi'_{e,c}\mathfrak{q})\,\mathrm{K}^{-1} & 0 & 0 \\ 0 & \frac{1}{\mathfrak{l}_c}(\psi''_{c,\gamma_c}\Phi'_{e,c}\mathfrak{q}) & 0 \\ 0 & 0 & \frac{1}{4\mathfrak{B}^2\mathfrak{l}_c}(\mathcal{S}''_{\gamma_c}\Psi'_{e,c}\mathfrak{q}) \end{bmatrix}, \tag{5.8}$$

where $\Phi'_{e,c} := \Phi'_e(\gamma_c)$, $\Psi'_{e,c} := \Psi'_e(\gamma_c)$; $\psi'_{c,\gamma} := \psi'_c(\Phi_e(\gamma)) = \frac{3\iota}{2\mathfrak{M}}\left[1 - \left(\frac{\Phi_e(\gamma)}{\mathfrak{M}} - 1\right)^2\right]$, $\mathcal{S}'_\gamma = -\frac{\gamma}{2\sqrt{\Psi_e(\gamma)}}$; $\psi''_{c,\gamma} := \psi''_c(\Phi_e(\gamma)) = -\frac{3\iota}{\mathfrak{M}^2}(\frac{\Phi_e(\gamma)}{\mathfrak{M}} - 1)$, $\mathcal{S}''_\gamma = \frac{\gamma}{4\sqrt{\Psi_e(\gamma)}^3}$.

The bilinear operator is given as

$$B(\zeta,\eta) = \frac{1}{2}\begin{bmatrix} \mathfrak{a}(\psi''_{c,\gamma_c})\left[\mathrm{K}^{-1}(\zeta_1\eta_1 - \Pi^{(2)}\zeta_1\eta_1 + \zeta_1\eta_2) + \zeta_2\,\mathrm{K}^{-1}\eta_1\right] \\ \frac{1}{\mathfrak{l}_c}(\psi''_{c,\gamma_c})(\zeta_2\eta_2 + \Pi^{(2)}\zeta_1\eta_1) \\ \frac{1}{4\mathfrak{B}^2\mathfrak{l}_c}(\mathcal{S}''_{\gamma_c})\zeta_3\eta_3 \end{bmatrix}, \tag{5.9}$$

where $\zeta, \eta \in \mathscr{H}$ and are written as $\zeta = [\zeta_1, \zeta_2, \zeta_3]$ and $\eta = [\eta_1, \eta_2, \eta_3]$, the operator $\Pi^{(2)}$ is from (1.7), the other notations are same as the above Fréchet derivative case.

The trilinear operator is given as

$$F(v,v,v) = \frac{1}{6}\begin{bmatrix} \mathfrak{a}(\psi'''_c)[\mathrm{K}^{-1}(v_1^3 - \Pi^{(3)}v_3) + 3\,\mathrm{K}^{-1}(v_1^2v_2 - \Pi^{(2)}v_1^2v_2 + v_1v_2^2)] \\ \frac{1}{\mathfrak{l}_c}(\psi'''_c)(v_2^3 + \Pi^{(3)}v_1^3 + 3\Pi^{(2)}v_1^2v_2) \\ \frac{1}{4\mathfrak{B}^2\mathfrak{l}_c}(\mathcal{S}'''_{\gamma_c})v_3^3 \end{bmatrix}, \tag{5.10}$$

where $v := [v_1, v_2, v_3] \in \mathscr{H}$, $\psi'''_c := \psi'''_c(\Phi_e(\gamma)) = -\frac{3\iota}{\mathfrak{M}^3}$, $\mathcal{S}'''_\gamma = -\frac{3\gamma}{8\sqrt{\Psi_e(\gamma)}^5}$. The operator $\Pi^{(3)}$ is from (1.7).

Remark 5.1.1. *For short, we also denote $B(v,v)$ and $F(v,v,v)$ by $B(v)$ and $F(v)$, respectively.*

Now that the explicit form of eigenfunctions of $\mathcal{A}^{\mathfrak{c}}$ are known, we revisit the projection in (5.2) and derive the simplified decomposed dynamics.

Definition 5.1.2 (Projections). *Let $\mathfrak{h} := e^{i\theta}$ and $\bar{\mathfrak{h}} := e^{-i\theta}$ denote the critical eigenvectors. Then the corresponding adjoint eigenfunctions are $\mathfrak{h}^* := [\frac{\mathrm{K}^{-1}}{2\pi}e^{-i\theta}, 0, 0]^T$ and $\overline{\mathfrak{h}}^* := [\frac{\mathrm{K}^{-1}}{2\pi}e^{i\theta}, 0, 0]^T$, and the corresponding eigenvalues are $\rho^{\mathfrak{c}}_{\pm 1} + \varepsilon^2\rho^{\mathfrak{q}}_{\pm 1}$. Note that by the definition of inner product in Remark 1.2.1, we have $\langle \mathfrak{h}, \mathfrak{h}^*\rangle_{\mathscr{H}} = 1$, $\langle \mathfrak{h}, \overline{\mathfrak{h}}^*\rangle_{\mathscr{H}} = 0$; $\langle \bar{\mathfrak{h}}, \mathfrak{h}^*\rangle_{\mathscr{H}} = 0$, $\langle \bar{\mathfrak{h}}, \overline{\mathfrak{h}}^*\rangle_{\mathscr{H}} = 1$. The critical projection operator is explicitly defined by $P_c := \langle \mathfrak{h}^*, \cdot\rangle_{\mathscr{H}}\mathfrak{h} + \langle \overline{\mathfrak{h}}^*, \cdot\rangle_{\mathscr{H}}\bar{\mathfrak{h}}$, and the stable projection $P_s = I - P_c$.*

Definition 5.1.3 (Projection related notations)**.** *For simplicity,*

(1) We introduce shorthand notation $B_c := P_c B$. We define F_c, F_s, $\mathcal{A}_c^{\mathfrak{c}}$, $\mathcal{A}_c^{\mathfrak{q}}$, $\mathcal{A}_s^{\mathfrak{c}}$, $\mathcal{A}_s^{\mathfrak{q}}$, $\mathscr{H}_c$ and $\mathscr{H}_s$ in a similar way.

(2) The associated eigenvalues of $\mathcal{A}_c^{\mathfrak{c}}$ and $\mathcal{A}_c^{\mathfrak{q}}$ are respectively denoted by

$$\rho_c^{\mathfrak{c}} = ib_c^{\mathfrak{c}} := ib_1(\gamma_c), \quad \bar{\rho}_c^{\mathfrak{c}} = -ib_c^{\mathfrak{c}} := -ib_1(\gamma_c),$$
$$\rho_c^{\mathfrak{q}} = a_c^{\mathfrak{q}} + ib_c^{\mathfrak{q}} := \mathfrak{q}(a_1'(\gamma_c) + ib_1'(\gamma_c)),$$

and

$$\bar{\rho}_c^{\mathfrak{q}} = a_c^{\mathfrak{q}} - ib_c^{\mathfrak{q}} := \mathfrak{q}(a_1'(\gamma_c) - ib_1'(\gamma_c)).$$

(3) We use simple notations for the amplitudes of the critical projection, $B_{c,1} := \langle \mathfrak{h}^*, B \rangle_{\mathscr{H}}$ as well as $F_{c,1} := \langle \mathfrak{h}^*, F \rangle_{\mathscr{H}}$.

(4) We also introduce the operator $\mathcal{A}_s := \mathcal{A}_s^{\mathfrak{c}} + \varepsilon^2 \mathcal{A}_s^{\mathfrak{q}}$, and the index set $\mathbb{Z}_s := \mathbb{Z}_0 \setminus \{\pm 1\}$.

(5) The associated eigenvalues of $\mathcal{A}_s^{c}$ and $\mathcal{A}_s^{\mathfrak{q}}$ (restricted to $\mathcal{H}_s$) are denoted by $\rho_k^{\mathfrak{c}} := a_k^{\mathfrak{c}} + ib_k^{\mathfrak{c}}$ and $\rho_k^{\mathfrak{q}} := a_k^{\mathfrak{q}} + ib_k^{\mathfrak{q}}$ for $k \in \mathbb{Z}_s$.

We represent the solution $\hat{u} \in \mathscr{H}$ as $\hat{u} = \hat{x} + \hat{y}$ for $\hat{x} \in P_c\mathscr{H}$ and $\hat{y} \in P_s\mathscr{H}$. Let $\hat{z} = \langle \mathfrak{h}^*, \hat{x} \rangle_{\mathscr{H}}$ and $\bar{\hat{z}} = \langle \mathfrak{h}, \hat{x} \rangle_{\mathscr{H}}$ be the complex-valued amplitudes. By the above separation of spectrum, we obtain the local critical and stable dynamics as:

$$d\hat{z} = [(\rho_1^{\mathfrak{c}} + \varepsilon\rho_1^{\mathfrak{q}})\hat{z} + B_{c,1}(\hat{x} + \hat{y}, \hat{x} + \hat{y}) + F_{c,1}(\hat{x} + \hat{y})]\, dt; \tag{5.11a}$$
$$d\hat{y} = [\mathcal{A}_s\hat{y} + B_s(\hat{x} + \hat{y}, \hat{x} + \hat{y}) + F_s(\hat{x} + \hat{y})]\, dt. \tag{5.11b}$$

Since $\hat{z}$ and $\bar{\hat{z}}$ are complex conjugates, showing the dynamics of either one of them is sufficient to represent the critical amplitude dynamics.

Remark 5.1.4. *Note that P_c can be interpreted as a two-fold projection: (a) projection from $\mathscr{H}$ onto $\mathcal{H}$; (b) projection from $\mathcal{H}$ onto $\mathscr{H}_1^{c}$.*

Furthermore, the bilinear operator B possesses the following properties.

1. $B_{c,1}(\hat{x}, \hat{x}) = \langle \mathfrak{h}^*, B(\hat{x}, \hat{x}) \rangle_{\mathscr{H}} = \langle \mathfrak{h}^*, B(\hat{z}\mathfrak{h}, \hat{z}\mathfrak{h}) + 2B(\hat{z}\mathfrak{h}, \overline{\hat{z}\mathfrak{h}}) + B(\overline{\hat{z}\mathfrak{h}}, \overline{\hat{z}\mathfrak{h}}) \rangle_{\mathscr{H}}$, but we can justify that $\langle \mathfrak{h}^*, B(\hat{z}\mathfrak{h}, \hat{z}\mathfrak{h}) \rangle_{\mathscr{H}} = \langle \mathfrak{h}^*, B(\overline{\hat{z}\mathfrak{h}}, \overline{\hat{z}\mathfrak{h}}) \rangle_{\mathscr{H}} = \langle \mathfrak{h}^*, B(\hat{z}\mathfrak{h}, \overline{\hat{z}\mathfrak{h}}) \rangle_{\mathscr{H}} = 0$.

2. $B_{c,1}(\hat{y}, \hat{y}) = \langle \mathfrak{h}^*, B(\hat{y}, \hat{y}) \rangle_{\mathscr{H}} = \frac{\mathfrak{a}(\psi''_{c,\gamma_c})}{1+\mathfrak{a}\mathfrak{m}} \sum_{k \in \{-2,-3,\ldots\}}^{k+l=1} \hat{g}_k \hat{g}_l$.

3. $B_{c,1}(\hat{x} + \hat{y}, \hat{x} + \hat{y}) = 2B_{c,1}(\hat{x}, \hat{y}) + B_{c,1}(\hat{y}, \hat{y})$,
   $B_c(\hat{x} + \hat{y}, \hat{x} + \hat{y}) = 2B_c(\hat{x}, \hat{y}) + B_c(\hat{y}, \hat{y})$.

4. $B_s(\hat{x} + \hat{y}, \hat{x} + \hat{y}) = B_s(\hat{x}, \hat{x}) + 2B_s(\hat{x}, \hat{y}) + B_s(\hat{y}, \hat{y})$.

5.1.2 Notations and Assumptions for the Stochastic Model

The main purpose of this chapter is to investigate the dominating dynamics in the critical subspace of stall in the neighbourhood of γ_c and $v = \mathbf{0}$ with the presence of additive noise. To better understand the long-term effect of the noise to the process within the small region around the deterministic equilibrium point u_e, we examine the behavior of $v(t) := \varepsilon^{-1}\hat{u}(\varepsilon^{-2}t)$ and consider the following truncated Cauchy problem

$$dv = \varepsilon^{-2}\mathcal{A}^{\mathfrak{c}}vdt + \mathcal{A}^{\mathfrak{q}}vdt + \varepsilon^{-1}B(v,v)dt + \mathrm{F}(v,v,v)dt + \varepsilon^{-1}dW_t, \quad v(0) = v_0. \tag{5.12}$$

The associated semigroup is given as $S(t) = e^{\varepsilon^{-2}t\mathcal{A}^{\mathfrak{c}}}$ (recall notations from Section 5.1.1).

We have seen that the deterministic solutions belong to $\mathbf{H}^2_{\mathrm{per}} \times \mathbb{R} \times \mathbb{R}$ (see Remark 1.2.1), which coincides with $\mathrm{dom}(\mathcal{A}^{\mathfrak{c}})$. Now we define the fractional spaces w.r.t. $\mathrm{dom}(\mathcal{A}^{\mathfrak{c}})$ and $\mathbf{H}^2_{\mathrm{per}}$ for the stochastic settings in order to have a more flexible scale of regularity.

Definition 5.1.5 (Fractional Power Space)**.** *For $\alpha \in \mathbb{R}$, define the interpolation fractional power (Hilbert) space [134] $\mathscr{H}_\alpha := \mathrm{dom}((\mathcal{A}^{\mathfrak{c}})^\alpha)$ endowed with inner product $\langle u, v\rangle_\alpha = \langle (\mathcal{A}^{\mathfrak{c}})^\alpha u, (\mathcal{A}^{\mathfrak{c}})^\alpha v\rangle_{\mathscr{H}}$ and the induced norm $\|\cdot\|_\alpha := \|\mathcal{A}^{\mathfrak{c}}\cdot\|$. Similarly, denote $\mathcal{H}_\alpha := \mathrm{dom}(\mathcal{A}^{\mathfrak{c}}|^\alpha_{\mathcal{H}})$. We also denote the dual space of $\mathscr{H}_\alpha$ (resp. $\mathcal{H}_\alpha$) by $\mathscr{H}_{-\alpha}$ (resp. $\mathcal{H}_{-\alpha}$) w.r.t. the inner product in $\mathscr{H}$ (resp. $\mathcal{H}$).*

More properties about $\mathcal{A}^{\mathfrak{c}}$ and $S(t)$ can be found in Appendix D.3. We particularly mention Remark D.3.2 as well as Proposition D.3.3. We also define the fractional Sobolev spaces for $\mathbf{H}^2_{\mathrm{per}}[0, 2\pi]$, which can be found in Definition D.3.5. The equivalence between $\mathrm{dom}((\mathcal{A}^{\mathfrak{c}})^\alpha)$ and the fractional Sobolev Hilbert space $\mathbf{H}^{2\alpha}_{\mathrm{per}}[0, 2\pi]$ (see Definition D.3.5) is given in Lemma D.3.6.

Definition 5.1.6 (Model of disturbances)**.** *For the Moore-Greitzer model, we restrict attention to $\mathcal{H}$ and construct Hilbert-space valued Wiener processes (see details in Appendix D.1 and D.2),*

$$W|_{\mathcal{H}}(t) = \sum_{k\in\mathbb{Z}^+\setminus\{1\}} \sqrt{q_k}(\beta_k(t) + i\beta_{-k}(t))e^{ik\theta} + \sum_{k\in\mathbb{Z}^-\setminus\{-1\}} \sqrt{q_k}(\beta_{-k}(t) - i\beta_k(t))e^{ik\theta}, \tag{5.13}$$

where $q_k = |k|^{-(4\zeta+1)-\upsilon}$ for any fixed $\upsilon > 0$, β_k are i.i.d. Brownian motions. Then the process $W|_{\mathcal{H}}(t)$ belongs to $\mathcal{H}_\zeta$ a.s..

The following examples are special cases of the engine disturbances:

(1) (White in time, colored in space) when $\zeta \geq 0$, q_k decays as k increases, then Q is a trace class operator (i.e. $\mathrm{tr}(Q) = \sum_{k\in\mathbb{Z}_s} q_k < \infty$ [43]) in $\mathcal{H}_\zeta \subset \mathcal{H}$, and $W|_{\mathcal{H}}(t)$ is automatically an $\mathcal{H}$-valued Q-Wiener process (Definition D.2.1);

(2) (Space-time white noise) when $\zeta = -1/4 - \upsilon/4$, $Q = \mathrm{id}$, $\mathrm{tr}(Q) = \infty$ and (5.13) does not converge in $\mathcal{H}$. However, when $\mathcal{H}$ is extended to $\mathcal{H}_\zeta \supset \mathcal{H}$ by a Hilbert-Schmidt inclusion operator, the $W|_{\mathcal{H}}(t)$ is well defined as an $\mathcal{H}_\zeta$-valued Gaussian process, also known as a generalized $\mathcal{H}$-valued cylindrical Wiener process (Definition D.2.2).

Assumption 5.1.7. *For $\alpha \in (0,1]$, let $W = [W|_{\mathcal{H}}, \beta_\Phi, \beta_\Psi]$ where $W|_{\mathcal{H}}$ is a generalized Q-Wiener process to be constructed by (5.13), β_Φ and β_Ψ are i.i.d. Brownian motions in $\mathbb{R}$. We assume that for a fixed sufficiently small $\upsilon > 0$, the choice of Q (in terms of the ζ in (5.13)) is such that*

$$\|Q^{1/2}\mathcal{A}^{\mathfrak{c}}|_{\mathcal{H}}^{\alpha-\frac{1}{2}+\upsilon} v\| < \infty, \ \forall v \in \mathcal{H}. \tag{5.14}$$

Remark 5.1.8. (5.14) *is to require that the choice of the Q-Wiener process should at worst be in $\mathcal{H}_{-\frac{1}{4}-\upsilon}$, which justifies a cylindrical Wiener process in $\mathcal{H}$.*

The construction (5.13) implies $\langle Q\mathfrak{h}, \mathfrak{h}^* \rangle = q_k = 0$ for $k \in \{1, -1\}$, which means that the additive noise does not act on $P_c\mathcal{H}$. This is in that the additive stochastic components in the stable, heavily damped modes also contribute to the critical modes. These contributions enter the critical modes as multiplicative noise. If additional additive noise is acting directly on the critical modes, it will be more intense than the multiplicative effects generated by the interaction between critical and stable modes. However, the stochastic stability is only affected by the presence of multiplicative noise in the critical modes. The proposed model of disturbances eliminate this strong additive effect to better understand and quantify the bifurcation behavior.

Lemma 5.1.9. *Given $\alpha \in (0,1]$, we assume $v_0 \in \mathcal{H}_\alpha$ and there exists $\beta \in (\alpha-1, \alpha]$ such that the bounded operators $B : \mathscr{H}_\alpha \times \mathscr{H}_\alpha \to \mathscr{H}_\beta$ and $F : \mathscr{H}_\alpha \times \mathscr{H}_\alpha \times \mathscr{H}_\alpha \to \mathscr{H}_\beta$.*

Proof. We only need to look at the regularities restricted to $\mathcal{H}$. By the product rule in Sobolev spaces, if $\alpha > 1/4$, then the continuous multiplication operators B, F map from $\mathcal{H}_\alpha$ to $\mathcal{H}_\alpha$. For $\alpha \leq 1/4$, we show the case for B and the proof for F is similar. For $u, v \in \mathcal{H}_\alpha$, there exists a $\beta \leq \alpha$ together with $\beta < 2\alpha - \frac{1}{4}$, such that $uv \in \mathcal{H}_\beta$. In addition, for the same u, v, there exists a $\kappa \in (0, \alpha]$ such that $uv \in \mathcal{W}^{2\kappa,1}_{\text{per}}$ [19]. Since $2\kappa - 1 \geq 2\alpha - 1 - \frac{1}{2}$, by the Sobolev embedding theorem, we have $\mathcal{W}^{2\kappa,1}_{\text{per}}$ embedded in $\mathcal{W}^{2\alpha-2,2}_{\text{per}} = \mathcal{H}_{\alpha-1}$, which verifies the worst case. Combining all of the above cases, we have $uv \in \mathcal{H}_\beta$ for some $\beta \in (\alpha-1, \alpha]$. By the definition of B, the components are of the form uv; the regularity of B follows. ☐

Proposition 5.1.10. *Suppose that Assumption 5.1.7 holds, then for each $\mathfrak{q}$ and $v(0) \in \mathscr{H}_\alpha$,* (5.12) *has a unique local[2] mild solution*

$$v(t) = S(t)v_0 + \int_0^t S(t-s)[\mathcal{A}^{\mathfrak{q}} + \varepsilon^{-1}B + F](v(s))ds + \varepsilon^{-1}\int_0^t S(t-s)dW_s \tag{5.15}$$

[2] The notion of local solution is the same as the finite-dimensional case.

such that $v(t) \in \mathcal{H}_\alpha$ *a.s.*.

Proof. We show a sketch of the proof. By isometry and Assumption 5.1.7, we have the following bound

$$\begin{aligned}\mathbf{E}\left[\left\|\int_0^t S(t-s)dW_s\right\|_\alpha^2\right] &\le \int_0^t \|Q^{1/2}\mathcal{A}^{\mathfrak{c}}|_{\mathcal{H}}^{\alpha} S(t-s)\|^2 ds + \int_0^t \|\mathcal{A}^{\mathfrak{c}}|_{\mathbb{R}^2}^{\alpha} S(t-s)\|^2 ds \\ &\le \|Q^{1/2}\mathcal{A}^{\mathfrak{c}}|_{\mathcal{H}}^{\alpha-\frac{1}{2}+\upsilon}\|^2 \int_0^t \|\mathcal{A}^{\mathfrak{c}}|_{\mathcal{H}}^{1/2-\upsilon} S(t-s)\|^2 ds \\ &+ \int_0^t \|\mathcal{A}^{\mathfrak{c}}|_{\mathbb{R}^2}^{\alpha} S(t-s)\|^2 ds < \infty.\end{aligned} \tag{5.16}$$

By Lemma 5.1.9 and Proposition D.3.3, it can be easily shown that $\|S(t-s)B\|_\alpha$ and $\|S(t-s)F\|_\alpha$ exist. However, B and F do not have the global Lipschitz or Dissipative properties. We use a cut-off argument as in [27] to define the existence time τ_{ex}. The mild solution exists uniquely and only up to τ_{ex}. □

We also need to specify a stopping time, such that the approximation processes will stop before the solution v blows up.

Definition 5.1.11 (Stopping time). *Given the terminal time* T *for* (D.10) *and a fixed arbitrarily small* $\kappa \in (0,1)$, *consider the stopping time*[3]

$$\tau^* := T \wedge \inf\{t > 0 : \ \|v(t)\|_\alpha \ge \varepsilon^{-\kappa}\}.$$

Definition 5.1.12. *We also define the order of error. A process* $R = \mathcal{O}(\varepsilon^{k-})$ *for some* $k > 0$ *if for some arbitrarily small* $\upsilon > 0$ *and any* $p \ge 1$, *there exists a* $C > 0$ *and such that* $\mathbf{E}\sup_{0\le\sigma\le\tau^*} |R(\sigma)|^p \le C\varepsilon^{pk-\upsilon}$.

Definition 5.1.13 (Other notations). *Recall notations from Definition 5.1.2 and 5.1.3. We introduce the following notations for future references.*

(1) For $n \in \mathbb{Z}_s$, *let* $\tilde{y} = [\sum_n \tilde{\mathfrak{g}}_{k\in\mathbb{Z}_s} e^{ik\theta}, \ \tilde{\Phi}_\delta, \tilde{\Psi}_\delta]^T$ *denotes the solution to*

$$d\tilde{y}(t) = \varepsilon^{-2}\mathcal{A}_s^{\mathfrak{c}}\tilde{y}dt + \varepsilon^{-1}P_s dW_t, \quad \tilde{y}(0) = y(0),$$

and $y^\star = \left[\sum_{k\in\mathbb{Z}_s} \mathfrak{g}^\star e^{ik\theta}, \ \Phi_\delta^\star, \Psi_\delta^\star\right]^T$ *denotes the associated stationary solution.*

[3] It is clear that $\tau^* < \tau_{\text{ex}}$ a.s. for all $\kappa > 0$.

(2) For convenience, we introduce

$$\mathcal{K}_n = \frac{\mathfrak{a}\psi''_{c,\gamma_c}|n|}{|n| + \mathfrak{am}} \quad \textit{for } \; n \in \mathbb{Z}_0 \tag{5.17}$$

and

$$\mathcal{G}_n = \frac{\mathfrak{a}\psi''_{c,\gamma_c}|n|}{2(|n| + \mathfrak{am})} \left(2(\Phi_\delta)\mathfrak{g}_n + \sum_{h \in \mathbb{Z}_s}^{j=n-h} \mathfrak{g}_h \mathfrak{g}_j \right) \quad \textit{for } \; n \in \mathbb{Z}_s. \tag{5.18}$$

(3) When $\mathcal{A}|_{\mathbb{R}^2}$ provides a stable spectrum, we symbolically represent the inverse operator as

$$\mathcal{A}_s^{\mathfrak{c}}|_{\mathbb{R}^2}^{-1} := \begin{bmatrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{bmatrix}. \tag{5.19}$$

Moreover, we denote the eigenvalue $\mathfrak{r}_{\pm 1}$ (recall 5 of Section 1.2.2) at γ_c as $\mathfrak{r}_{\pm 1} = a_{\mathfrak{r}} \pm i b_{\mathfrak{r}}$.

5.2 Dimension Reduction of Stochastic Moore-Greitzer PDE Model

We set $x(t) = \varepsilon^{-1}(\hat{x}(\varepsilon^{-2}t)$, $y(t) = \varepsilon^{-1}(\hat{y}(\varepsilon^{-2}t)$, as well as the complex valued rescaled amplitude $z(t) = \varepsilon^{-1}(\hat{z}(\varepsilon^{-2}t))$ and $\bar{z}(t) = \varepsilon^{-1}(\bar{\hat{z}}(\varepsilon^{-2}t))$ as in (5.4) to investigate the decomposed dynamics. Then, it is clear that $v(t) = x(t) + y(t) = z(t)\mathfrak{h} + \bar{z}(t)\bar{\mathfrak{h}} + y(t)$. When (5.12) is close to the critical point, the local critical and fast-varying stable dynamics are as follows:

$$dz = \left[\rho_c^{\mathfrak{q}} z + 2\varepsilon^{-1} B_{c,1}(x, y) + \varepsilon^{-1} B_{c,1}(y, y) + F_{c,1}(x + y)\right] dt, \quad z(0) = z_0, \tag{5.20a}$$

$$dy = \left[\varepsilon^{-2} \mathcal{A}_s y + \varepsilon^{-1} B_s(x + y) + F_s(x + y)\right] dt + \varepsilon^{-1} dW_t, \; y(0) = y_0, \tag{5.20b}$$

where $W = [W|_{\mathcal{H}}, \beta_\Phi, \beta_\Psi]$ is a Q-Wiener process such that $P_s W = W$ as introduced in Definition 5.1.6 and Assumption 5.1.7.

5.2.1 Coupling of Stable Modes Though Bilinear Terms

To obtain a finite-dimensional approximation for v based on (5.20a), we first investigate how the stochastically perturbed fast-varying y enters the terms of intermediate order ε^{-1}. The approach follows the idea provided in [29, Proposition 3.9].

Lemma 5.2.1. *Let $L := (\mathcal{A}_s^{\mathfrak{c}})^{-1}$. Then, for every stopping time $\sigma \le \tau^*$, we have*

$$\begin{aligned} \int_0^\sigma B_{c,1}(x,y)dt = -\varepsilon \int_0^\sigma B_{c,1}(x, LB_s(x+y))dt - 2\varepsilon \int_0^\sigma B_{c,1}(B_c(x,y),\ Ly)dt \\ - \varepsilon \int_0^\sigma B_{c,1}(B_c(y,y),\ Ly)dt - \varepsilon \int_0^\sigma B_{c,1}(x, LdW_t) + R_1(\sigma), \end{aligned} \tag{5.21}$$

where the remainder term R_1 is of order $\mathcal{O}(\varepsilon^{2-})$.

Proof. Expand the Q-Wiener process as

$$W = \left[2 \sum_{k \in \mathbb{Z}^+ \setminus \{1\}} \sqrt{q_k}(\beta_k \cos(k\theta) - \beta_{-k} \sin(k\theta)), \beta_\Phi, \beta_\Psi \right]^T.$$

Note that L is a bounded linear operator by the hypothesis on the spectrum. Now apply the infinite-dimensional Itô's formula,

$$dB(x,Ly) = B(dx, Ly) + B(x, Ldy) + \frac{1}{2} \sum_{i,j} \frac{\partial^2 B(x,Ly)}{\partial u_i \partial u_j} d\,\langle\langle U_i, U_j \rangle\rangle_t,$$

where $i,j \in \{1,2\}$, $U_1 = x$, $U_2 = y$, and $d\,\langle\langle \beta_k, \beta_l \rangle\rangle_t = \delta_{kl}dt$, $\langle\langle \beta_k, t \rangle\rangle_t = \langle\langle t, \beta_k \rangle\rangle_t = 0$ for all k,l in the index set $\mathbb{Z}^+ \setminus \{1\} \cup \{\Phi, \Psi\}$ and for all t. However, since $\frac{\partial^2 B(x,Ly)}{\partial x^2} = \frac{\partial^2 B(x,Ly)}{\partial y^2} = 0$, we have

$$\frac{1}{2} \sum_{i,j} \frac{\partial^2 B}{\partial u_i \partial u_j} d\,\langle\langle U_i, U_j \rangle\rangle_t = \frac{1}{2} \left(B(dx, dLy) + B(dx, dLy) \right) = B(dx, dLy).$$

By plugging in dx, dLy and eliminating all the $\langle\langle \beta_i, t \rangle\rangle$, $\langle\langle t, \beta_i \rangle\rangle$, $\langle\langle t, t \rangle\rangle$ terms,

$$\frac{1}{2} \sum_{i,j} \frac{\partial^2 B(x,Ly)}{\partial u_i \partial u_j} d\,\langle\langle U_i, U_j \rangle\rangle_t = B(dP_c dW_t, dP_s dW_t) = 0$$

Hence,

$$\begin{aligned} \varepsilon^2 dB(x,Ly) =& \varepsilon^2 B(dx, Ly) + \varepsilon^2 B(x, Ldy) \\ =& \varepsilon^2 B(\rho_c^{\mathfrak{c}} z \mathfrak{h} + \bar{\rho}_c^{\mathfrak{c}} \bar{z} \bar{\mathfrak{h}},\ Ly)dt + 2\varepsilon B(B_c(x,y),\ Ly)dt + \varepsilon B(B_c(y,y),\ Ly)dt \\ &+ \varepsilon^2 B(F_c(x+y),\ Ly))dt + B(x,y)dt + \varepsilon B(x, LB_s(x+y))dt \\ &+ \varepsilon^2 B(x, P_s LF(x+y))dt + \varepsilon^2 B(x, L\mathcal{A}_s^{\mathfrak{q}} y)dt + \varepsilon B(x, LdW_t), \end{aligned} \tag{5.22}$$

Notice that the above terms take values in $\mathscr{H}_\beta$. By the property of B given in Lemma 5.1.9 and the definition of stopping time τ^*, it is straightforward to verify that the term

$$\begin{aligned}\tilde{R}(\sigma) :=&\varepsilon^2 \left[\int_0^\sigma dB(x, Ly) - \int_0^\sigma [B(x, Ly) - B(F_c(x+y), Ly)]dt\right] \\ &- \varepsilon^2 \left[\int_0^\sigma [B(x, P_s LF(x+y)) - B(x, L\mathcal{A}_s^{\mathfrak{q}} y)]dt\right],\end{aligned} \tag{5.23}$$

which is the sum of all the term with scaling ε^2, satisfies $\|\tilde{R}(\sigma)\|_\beta = \mathcal{O}(\varepsilon^{2-})$. Rearranging (5.22) and taking projection using $\langle \mathfrak{h}^*, \cdot \rangle_{\mathscr{H}}$, the result can be concluded. □

Similarly, we have the following result. The proof is completed in Section 5.5.

Lemma 5.2.2. *For every stopping time $\sigma \leq \tau^*$, we have*

$$\begin{aligned}&\int_0^\sigma B_{c,1}(y, y)dt \\ &= -\varepsilon \int_0^\sigma \mathcal{K}_1 \sum_{k \in \{-2,-3\dots\}}^{k+l=1} \left(\frac{z(\mathcal{K}_k \mathfrak{g}_k \mathfrak{g}_{-k} + \mathcal{K}_l \mathfrak{g}_l \mathfrak{g}_{-l})}{\rho_k^{\mathfrak{c}} + \rho_l^{\mathfrak{c}}} dt + \frac{\mathfrak{g}_k \mathcal{G}_l + \mathfrak{g}_l \mathcal{G}_k}{\rho_k^{\mathfrak{c}} + \rho_l^{\mathfrak{c}}} dt \right) \\ &= -\varepsilon \int_0^\sigma \frac{\mathcal{K}_1 \mathcal{K}_2 \mathfrak{g}_3 \bar{z}^2}{2(\rho_{-2}^{\mathfrak{c}} + \rho_3^{\mathfrak{c}})} dt - \varepsilon \int_0^\sigma \mathcal{K}_1 \sum_{k \in \{-2,-3\dots\}}^{k+l=1} \frac{\bar{z}(\mathcal{K}_{l+1} \mathfrak{g}_k \mathfrak{g}_{l+1} + \mathcal{K}_{k+1} \mathfrak{g}_l \mathfrak{g}_{k+1})}{\rho_k^{\mathfrak{c}} + \rho_l^{\mathfrak{c}}} dt \\ &\quad - \varepsilon \int_0^\sigma \mathcal{K}_1 \sum_{k \in \{-2,-3\dots\}}^{k+l=1} \frac{\mathfrak{g}_k \sqrt{q_l}(d\beta_l(t) + id\beta_{-l}(t)) + \mathfrak{g}_l \sqrt{q_k}(d\beta_{-k}(t) - id\beta_k(t))}{\rho_k^{\mathfrak{c}} + \rho_l^{\mathfrak{c}}} \\ &\quad + R_2(\sigma),\end{aligned} \tag{5.24}$$

where $\mathcal{K}_j$,$\mathcal{G}_j$ and $\rho_j^{\mathfrak{c}}$ are defined in Definition 5.1.13, and the remainder term R_2 is such that $R_2(\sigma) = \mathcal{O}(\varepsilon^{2-})$.

Combining Lemma 5.2.1 and 5.2.2, we observe that the $\varepsilon^{-1} B_{c,1}$ terms in (5.20a) have the form of $\mathcal{O}(1)$ terms adding up with an error R of order $\mathcal{O}(\varepsilon^{1-})$. Hence, the amplitude equation (5.20a) is scaled such that the nonlinearities and the linear term are of the same order, which makes the analysis more amenable.

Remark 5.2.3. *In the case of a surge bifurcation, we would have $B_{c,1}(x, x) \neq 0$ with the same rescaling scheme. Since there is no contribution of homogenization from the stable modes, this term would dominate the rescaled critical mode with strength ε^{-1}. Hence, to yield a similar form* *as (5.20a), we should rescale the variables differently. One possibility would be to set* $z(t) := \varepsilon^{-2}\hat{z}(\varepsilon^{-2}t)$ *and* $y(t) := \varepsilon^{-2}\hat{y}(\varepsilon^{-2}t)$. *As for the stall-surge case, multiple rescaling schemes are needed to capture the bifurcation of* $\mathfrak{g}$ *and* $(\Phi_\delta, \Psi_\delta)$.

To keep the content succinct, we only demonstrate the methodology via the stochastic analysis for the stall instability. The cases for surge and stall-surge can be treated using similar methods.

5.2.2 Approximation of the Stable Modes

The purpose of this subsection is to find an approximation of the stable dynamics.

Lemma 5.2.4. *Let* $\tilde{y}(t)$ *solve the Ornstein-Uhlenbeck equation*

$$d\tilde{y}(t) = \varepsilon^{-2}\mathcal{A}^{\mathfrak{c}}\tilde{y}dt + \varepsilon^{-1}dW_t, \quad \tilde{y}(0) = y(0). \tag{5.25}$$

Then $\|y(t) - \tilde{y}(t)\|_\alpha = \mathcal{O}(\varepsilon^{1-})$.

Proof. Let $\tilde{S}(t) := e^{\varepsilon^{-2}\mathcal{A}_s^{\mathfrak{c}}t}$. By the definition of mild solutions, $y(t)-\tilde{y}(t) = \int_0^t \tilde{S}(t-s)[\varepsilon^{-1}B_s + F_s](x+y)ds + \int_0^t \tilde{S}(t-s)\mathcal{A}_s^{\mathfrak{q}}yds$. We show the bound for the error term $J(t) := \int_0^t \tilde{S}(t-s)\varepsilon^{-1}B_s(x+y)ds$, and the rest should be similar.

$$\begin{aligned}
\mathbf{E}\sup_{0\le\sigma\le\tau^*}\|J(\sigma)\|_\alpha^p &\le \varepsilon^{-p}\mathbf{E}\sup_{0\le\sigma\le\tau^*}\left[\int_0^\sigma \|\tilde{S}(\sigma-s)B_s(x+y)\|_\alpha\right]^p \\
&= \varepsilon^{-p}\mathbf{E}\sup_{0\le\sigma\le\tau^*}\left[\int_0^\sigma \|(\varepsilon^{-2}\mathcal{A}_s^{\mathfrak{c}})^\alpha\tilde{S}(\sigma-s)B_s(x+y)\|\right]^p \\
&\le C\varepsilon^{2\alpha-2\beta-p}\mathbf{E}\sup_{0\le\sigma\le\tau^*}\left[\int_0^\sigma \|e^{-\varepsilon^2\omega(\sigma-s)}(\sigma-s)^{-\alpha+\beta}B_s(x+y)\|_\beta ds\right]^p \\
&\le C\varepsilon^{2\alpha-2\beta-p-2\kappa p}\mathbf{E}\sup_{0\le\sigma\le\tau^*}\left[\int_0^\sigma e^{-\varepsilon^2\omega(\sigma-s)}(\sigma-s)^{-\alpha+\beta}ds\right]^p \\
&\le C\varepsilon^{2p-p-2\kappa p}\mathbf{E}\sup_{0\le\sigma\le\tau^*}\left[\int_0^{-\varepsilon^{-2}\omega\sigma} e^{-r}(r)^{-\alpha+\beta}dr\right]^p.
\end{aligned}$$

Note that there exists a constant $C' > 0$ such that $e^{-t}t^{-\alpha+\beta} \le C't^{-\alpha+\beta}$ for $t \le 1$, and $e^{-t}t^{-\alpha+\beta} \le C'e^{-t}$ for $t \ge 1$. We split the last integral by considering $\{\tau^* \le 1\}$ and $\{\tau^* > 1\}$, then it can be verified that the term $\mathbf{E}\sup_{0\le\sigma\le\tau^*}\left[\int_0^{-\varepsilon^{-2}\omega\sigma} e^{-r}(r)^{-\alpha+\beta}dr\right]^p$ is bounded by some C'. Therefore, $J = \mathcal{O}(\varepsilon^{1-})$ as expected.

□

Corollary 5.2.5. *For all $t \in (0, \tau^*]$, we have*

$$\|B(y(t), y(t)) - B(\tilde{y}(t), \tilde{y}(t))\|_\alpha = \mathcal{O}(\varepsilon^{2-})$$

and

$$\|F(x(t) + y(t)) - F(x(t) + \tilde{y}(t))\|_\alpha = \mathcal{O}(\varepsilon^{2-}).$$

Proof.

$$\begin{aligned}\|B(y, y) - B(\tilde{y}, \tilde{y})\|_\alpha &= \|B(y, y) - B(y, \tilde{y}) + B(y, \tilde{y}) - B(\tilde{y}, \tilde{y})\|_\alpha \\ &\leq \|B(y, y - \tilde{y})\|_\alpha + \|B(y - \tilde{y}, \tilde{y})\|_\alpha.\end{aligned}$$

By Lemma 5.2.4 and the continuity of B, the result follows. The proof for F is similar. □

Note that, by [29, Lemma 3.5] and [43], there is a version of $\tilde{y}$ with $\mathcal{H}_\alpha$-valued continuous sample paths. Further more, for every $\kappa_0 > 0$, $p > 0$, and $T > 0$, there exists a constant C such that $\mathbf{E} \sup_{t\in[0,T]} \|\tilde{y}(t)\|_\alpha^p \leq C e^{-\kappa_0}$. Consequently, the original process y can be verified to be of order $\mathcal{O}(\varepsilon^{-})$. We further average out the fast modes y over an invariant measure by considering the stationary behavior of $\tilde{y}$, i.e., we use the stationary solution $\tilde{y}^\star$ to submit into (5.21) and (5.24). The following result shows the homogenization error using $\tilde{y}^\star$.

Corollary 5.2.6. *By replacing y with $\tilde{y}^\star$ in (5.20), for each $t \in [0, \tau^*]$, we have*

$$\tilde{z}(t) = \tilde{z}(0) + \int_0^t \left[\rho_c^{\mathfrak{q}} z + 2\varepsilon^{-1} B_{c,1}(\tilde{x}, \tilde{y}^\star) + \varepsilon^{-1} B_{c,1}(\tilde{y}^\star, \tilde{y}^\star) + F_{c,1}(\tilde{x})\right] ds + \mathcal{O}(\varepsilon^{1-}). \tag{5.26}$$

Proof. Let R denote the error process, which contains $I(t) := \varepsilon^{-1} \int_0^t [B_{c,1}(\tilde{y}^\star, \tilde{y}^\star) - B_{c,1}(y, y)] ds$, $J(t) := \varepsilon^{-1} \int_0^t 2[B_{c,1}(\tilde{x}, \tilde{y}^\star) - B_{c,1}(\tilde{x}, y)] ds$, and $K(t) := \int_0^t [F_{c,1}(\tilde{x} + \tilde{y}^\star) - F_{c,1}(\tilde{x})] ds$. Recall that $\varepsilon^{-1} B_{c,1}$ terms are already of order $\mathcal{O}(1)$. The bound for I follows Corollary 5.2.5. For the error process J, we have

$$\begin{aligned}\mathbf{E} \sup_{t\in[0,\tau^*]} \|J(t)\|_\alpha^p &\leq C \mathbf{E} \sup_{t\in[0,\tau^*]} \|J(t)\|_\beta^p \\ &\leq C \mathbf{E} \sup_{t\in[0,\tau^*]} \left[\int_0^t \|x(t)\|_\alpha \|y(t) - \tilde{y}(t) + \tilde{y}(t) - \tilde{y}^\star(t)\|_\alpha ds\right]^p \\ &\leq C \varepsilon^{-\kappa p} \varepsilon^{p-} \leq C \varepsilon^{p-}.\end{aligned}$$

Note that the term $\|\tilde{y} - \tilde{y}^\star\|_\alpha$ is of order $\mathcal{O}(\varepsilon^1)$ in integrated form [29, Proposition 4.6]. We can expand K and evaluate the bound for each term of the expansion by a similar approach as the above. The conclusion follows by combining the above error bounds. □

5.3 Approximation Results

In this section, an explicit expression of $\tilde{y}^\star$ will be determined. Then, the dynamical behavior that is dominated by the critical mode is studied.

5.3.1 Calculation of the Stationary Stable Solutions

Equation (5.25) can be decomposed into

$$d\tilde{\mathfrak{g}}_k(t) = \varepsilon^{-2}\rho_k^{\mathfrak{c}}\tilde{\mathfrak{g}}_k dt + \varepsilon^{-1}\sqrt{q_k}(d\beta_k(t) + id\beta_{-k}(t)),\ \forall k \in \{2, 3, ...\}, \tag{5.27a}$$

$$d\tilde{\mathfrak{g}}_k(t) = \varepsilon^{-2}\rho_k^{\mathfrak{c}}\tilde{\mathfrak{g}}_k dt + \varepsilon^{-1}\sqrt{q_k}(d\beta_{-k}(t) - id\beta_k(t)),\ \forall k \in \{-2, -3, ...\}, \tag{5.27b}$$

$$d\begin{bmatrix}\tilde{\Phi}_\delta(t)\\ \tilde{\Psi}_\delta(t)\end{bmatrix} = \varepsilon^{-2}\mathcal{A}_s^{\mathfrak{c}}|_{\mathbb{R}^2}^{-1}\begin{bmatrix}\tilde{\Phi}_\delta(t)\\ \tilde{\Psi}_\delta(t)\end{bmatrix}dt + \varepsilon^{-1}\begin{bmatrix}d\beta_\Phi\\ d\beta_\Psi\end{bmatrix}. \tag{5.27c}$$

Note that the modes are pairwisely independent. We recall the notation in Definition 5.1.3-(5) that $\rho_k^{\mathfrak{c}} = a_k^{\mathfrak{c}} + ib_k^{\mathfrak{c}}$. If we express $\tilde{\mathfrak{g}}_k(t) = \tilde{\mathfrak{g}}_k^{\mathrm{R}}(t) + i\tilde{\mathfrak{g}}_k^{\mathrm{I}}(t),\ \forall k \in \mathbb{Z}_s$, then we can find the solution for each pair of $\tilde{\mathfrak{g}}_k^{\mathrm{R}}$ and $\tilde{\mathfrak{g}}_k^{\mathrm{I}}$ explicitly.

1. For every $k \in \{2, 3, ...\}$, the pair $[\tilde{\mathfrak{g}}_k^{\mathrm{R}}(t), \tilde{\mathfrak{g}}_k^{\mathrm{I}}(t)]^T$ are solved by

$$\begin{aligned}\begin{bmatrix}\tilde{\mathfrak{g}}_k^{\mathrm{R}}\\ \tilde{\mathfrak{g}}_k^{\mathrm{I}}\end{bmatrix}(t) &= e^{\frac{a_k^{\mathfrak{c}}(t-t_0)}{\varepsilon^2}}\begin{bmatrix}\cos(\frac{b_k^{\mathfrak{c}}(t-t_0)}{\varepsilon^2}) & -\sin(\frac{b_k^{\mathfrak{c}}(t-t_0)}{\varepsilon^2})\\ \sin(\frac{b_k^{\mathfrak{c}}(t-t_0)}{\varepsilon^2}) & \cos(\frac{b_k^{\mathfrak{c}}(t-t_0)}{\varepsilon^2})\end{bmatrix}\begin{bmatrix}\tilde{\mathfrak{g}}_k^{\mathrm{R}}(0)\\ \tilde{\mathfrak{g}}_k^{\mathrm{I}}(0)\end{bmatrix}\\ &+ \frac{\sqrt{q_k}e^{\frac{a_k^{\mathfrak{c}}t}{\varepsilon^2}}}{\varepsilon}\begin{bmatrix}\int_{t_0}^t e^{-\frac{a_k^{\mathfrak{c}}s}{\varepsilon^2}}cos(\frac{b_k^{\mathfrak{c}}s}{\varepsilon^2})d\beta_k(s) - \int_{t_0}^t e^{-\frac{a_k^{\mathfrak{c}}s}{\varepsilon^2}}sin(\frac{b_k^{\mathfrak{c}}s}{\varepsilon^2})d\beta_{-k}(s)\\ \int_{t_0}^t e^{-\frac{a_k^{\mathfrak{c}}s}{\varepsilon^2}}sin(\frac{b_k^{\mathfrak{c}}s}{\varepsilon^2})d\beta_k(s) + \int_{t_0}^t e^{-\frac{a_k^{\mathfrak{c}}s}{\varepsilon^2}}cos(\frac{b_k^{\mathfrak{c}}s}{\varepsilon^2})d\beta_{-k}(s)\end{bmatrix}.\end{aligned} \tag{5.28}$$

   The stationary solution (as $t_0 \to -\infty$) to (5.27a) and (5.27b) is given as $\tilde{\mathfrak{g}}_k^\star = (\tilde{\mathfrak{g}}_k^{\mathrm{R}})^\star + i(\tilde{\mathfrak{g}}_k^{\mathrm{I}})^\star$, where $(\tilde{\mathfrak{g}}_k^{\mathrm{R}})^\star$ and $(\tilde{\mathfrak{g}}_k^{\mathrm{I}})^\star$ are independent Gaussian processes with

$$\mathbf{E}[(\tilde{\mathfrak{g}}_k^{\mathrm{R}})^\star(t)] = \mathbf{E}[(\tilde{\mathfrak{g}}_k^{\mathrm{I}})^\star(t)] = 0$$

   and covariance matrix

$$\mathrm{Cov}(t, s) = \begin{bmatrix}\mathbf{E}[(\tilde{\mathfrak{g}}_k^{\mathrm{R}})^\star(t)(\tilde{\mathfrak{g}}_k^{\mathrm{R}})^\star(s)] & \mathbf{E}[(\tilde{\mathfrak{g}}_k^{\mathrm{R}})^\star(t)(\tilde{\mathfrak{g}}_k^{\mathrm{I}})^\star(s)]\\ \mathbf{E}[(\tilde{\mathfrak{g}}_k^{\mathrm{I}})^\star(t)(\tilde{\mathfrak{g}}_k^{\mathrm{R}})^\star(s)] & \mathbf{E}[(\tilde{\mathfrak{g}}_k^{\mathrm{I}})^\star(t)(\tilde{\mathfrak{g}}_k^{\mathrm{I}})^\star(s)]\end{bmatrix} = -\frac{q_k}{2a_k^{\mathfrak{c}}}e^{\frac{a_k^{\mathfrak{c}}|t-s|}{\varepsilon^2}}\mathrm{id}_{2\times 2}. \tag{5.29}$$

2. The solution to (5.27c) is given explicitly as,

$$\begin{bmatrix}\tilde{\Phi}_\delta(t)\\ \tilde{\Psi}_\delta(t)\end{bmatrix}(t) = e^{\frac{a_{\mathfrak{r}}(t-t_0)}{\varepsilon^2}} P \begin{bmatrix}\cos(\frac{b_{\mathfrak{r}}(t-t_0)}{\varepsilon^2}) & -\sin(\frac{b_{\mathfrak{r}}(t-t_0)}{\varepsilon^2})\\ \sin(\frac{b_{\mathfrak{r}}(t-t_0)}{\varepsilon^2}) & \cos(\frac{b_{\mathfrak{r}}(t-t_0)}{\varepsilon^2})\end{bmatrix} P^{-1} \begin{bmatrix}\tilde{\Phi}_\delta(0)\\ \tilde{\Psi}_\delta(0)\end{bmatrix} + \varepsilon^{-1}\int_{t_0}^{t} e^{\frac{a_{\mathfrak{r}}(t-s)}{\varepsilon^2}} P R_{t,s} P^{-1} \begin{bmatrix} d\beta_\Phi(s)\\ d\beta_\Psi(s)\end{bmatrix}, \tag{5.30}$$

where

$$P = \begin{bmatrix} 0 & 1\\ \mathrm{Im}(\nu_{\psi_1}) & \mathrm{Re}(\nu_{\psi_1})\end{bmatrix}, \quad R_{t,s} = \begin{bmatrix}\cos(\frac{b_{\mathfrak{r}}(t-s)}{\varepsilon^2}) & -\sin(\frac{b_{\mathfrak{r}}(t-s)}{\varepsilon^2})\\ \sin(\frac{b_{\mathfrak{r}}(t-s)}{\varepsilon^2}) & \cos(\frac{b_{\mathfrak{r}}(t-s)}{\varepsilon^2})\end{bmatrix}$$

and ν_{ψ_1} is defined in Sect. Secion 1.2.2-5. Therefore, the stationary solution (as $t_0 \to -\infty$) to (5.30) is given as

$$\mathbf{E}[\Phi^\star_\delta(t)] = \mathbf{E}[\Psi^\star_\delta(t)] = 0$$

and the covariance matrix

$$\mathrm{Cov}(t,s) = \varepsilon^{-2}\int_0^{t\wedge s} e^{\frac{a_{\mathfrak{r}}(t-r)}{\varepsilon^2}} (PR_{t,r}P^{-1})(PR_{t,r}P^{-1})^T dr. \tag{5.31}$$

Remark 5.3.1. *Note that the integral in (5.31) can be explicitly calculated. However, we use the implicit expression for the rest of the derivation.*

5.3.2 Evaluation of the Approximated Critical Amplitudes

Since every operator in (5.26), including B, K^{-1}, $\mathcal{A}_s^{-1}$ and $\langle \mathfrak{h}^*, \cdot\rangle_{\mathscr{H}}$, is given explicitly, after some cumbersome calculation applying the coupling results from Lemma 5.2.1 and 5.2.2, we obtain

$$-B_{c,1}(\tilde{x}, \mathcal{A}_s^{-1}B_s(\tilde{x},\tilde{x})) = -\frac{\mathcal{K}_1\mathcal{K}_2}{4\rho_2^{\mathfrak{c}}}\tilde{z}^2\overline{\tilde{z}} = -\frac{\mathcal{K}_1\mathcal{K}_2\rho^{\mathfrak{c}}_{-2}\tilde{z}^2\overline{\tilde{z}}}{4((a_2^{\mathfrak{c}})^2+(b_2^{\mathfrak{c}})^2)} =: h\rho^{\mathfrak{c}}_{-2}\tilde{z}^2\overline{\tilde{z}}, \tag{5.32}$$

where we have used notations defined in Def. 5.1.13-(2). Similarly,

$$-B_{c,1}(\tilde{x}, \mathcal{A}_s^{-1}B_s(y^*,y^*)) = N_1(\varpi)\overline{\tilde{z}} + N_2(\varpi)\tilde{z} - \frac{\mathcal{K}_1}{4\mathfrak{l}_c}\tilde{z}^2\overline{\tilde{z}}, \tag{5.33}$$

$$-B_{c,1}(\tilde{x}, \mathcal{A}_s^{-1}B_s(\tilde{x},\tilde{y}^\star)) = N_3(\varpi)\overline{z}^2, \tag{5.34}$$

$$-B_{c,1}(B_c(\tilde{x},\tilde{y}^\star),\ \mathcal{A}_s^{-1}\tilde{y}^\star) = N_4(\varpi)\tilde{z} + N_5(\varpi)\overline{\tilde{z}}, \tag{5.35}$$

$$-B_{c,1}(B_c(\tilde{y}^\star, \tilde{y}^\star),\ \mathcal{A}_s^{-1}\tilde{y}^\star) = N_6(\varpi). \tag{5.36}$$

From Lemma 5.2.2,

$$\varepsilon^{-1}B_{c,1}(\tilde{y}^\star, \tilde{y}^\star) =: N_7(\varpi)\tilde{z} + N_8(\varpi)\bar{\tilde{z}} + N_9(\varpi) + N_{10}(\varpi)\bar{\tilde{z}}^2. \tag{5.37}$$

We also have

$$F_{c,1}(\tilde{x}) = N_{11}(\varpi)\bar{\tilde{z}}^2 + N_{12}(\varpi)\tilde{z} + N_{13}(\varpi)\bar{\tilde{z}} + N_{14}(\varpi) - \frac{\mathcal{K}_1\psi_c'''}{2\psi_{c,\gamma_c}''}\tilde{z}^2\bar{\tilde{z}} \tag{5.38}$$

For the stochastic term,

$$-B_{c,1}(\tilde{x}, \mathcal{A}_s^{-1}dW_t) = -\frac{\mathcal{K}_1}{2}\left[\tilde{z}(l_{11}d\beta_\Phi + l_{12}d\beta_\Psi) + \frac{\bar{\tilde{z}}\sqrt{q_2}(d\beta_2 + id\beta_{-2})}{\rho_2^{\mathfrak{c}}}\right]. \tag{5.39}$$

The detailed information of the above shorthand notations $N_i(\omega)$ for $i \in \{1, 2, ..., 14\}$ are given in Section 5.5, where ϖ represents the randomness generated from the stable modes that are excited by stochastic terms. Making use of the results above (from Equation (5.32) to (5.39)), the approximated solution of $\tilde{z}$ can be determined by

$$\begin{aligned}\tilde{z}(t) = \tilde{z}(0) &+ \int_0^t (\rho_c^{\mathfrak{c}} + 2N_2(\varpi) + 4N_4(\varpi) + N_7(\varpi) + N_{12}(\varpi))\tilde{z}dt + \int_0^t (2h\rho_{-2}^{\mathfrak{c}} - j)\tilde{z}^2\bar{\tilde{z}}dt \\ &+ \int_0^t (2N_1(\varpi) + 4N_5(\varpi) + N_8(\varpi) + N_{13}(\varpi))\bar{\tilde{z}}dt \\ &+ \int_0^t (4N_3(\varpi) + N_{10}(\varpi) + N_{11}(\varpi))\bar{\tilde{z}}^2dt \\ &+ \int_0^t (2N_6(\varpi)\tilde{z} + N_9(\varpi) + N_{14}(\varpi))dt - 2\int_0^t B_{c,1}(\tilde{z}, \mathcal{A}_s^{-1}dW_t) + \mathcal{O}(\varepsilon^{1-}),\end{aligned} \tag{5.40}$$

where $j := \frac{\mathcal{K}_1\psi_c'''}{2\psi_{c,\gamma_c}''} + \frac{\mathcal{K}_1}{2\mathfrak{l}_c}$.

5.3.3 Final Approximation of the Critical Amplitudes

It is still not easy to evaluate (5.40). However, we observe that

$$\mathbf{E}[N_i(t)] = \mathbf{E}[N_i(0)] = 0,\ i \in \{1, 3, 5, 6, 8, 9, 10, 11, 13, 14\}, \tag{5.41}$$

$$\mathbb{R} \ni \mathbf{E}[N_2(t)] \neq 0 \text{ and } \mathbb{R} \ni \mathbf{E}[N_{12}(t)] \neq 0, \tag{5.42}$$

$$\mathbb{C} \ni \mathbf{E}[N_4(t)] \neq 0 \text{ and } \mathbb{C} \ni \mathbf{E}[N_7(t)] \neq 0. \tag{5.43}$$

Intuitively, we would like to replace N_i with $\overline{N}_i := \mathbf{E}[N_i(0)]$ for $i \in \{1, ..., 14\}$. The solution (5.40) can still be approximated in some sense with small error (the estimation relies on [29, Corollary 4.5]). We rephrase the statement of [29, Corollary 4.5] and provide it in the following theorem.

Theorem 5.3.2. *Let f be an $\tilde{\alpha}$-Hölder continuous function on $[0, \tau^*]$. Assume that for every $\varepsilon > 0$ and fixed $\kappa > 0$, there exist a constant C_1 such that*

$$\mathbb{E}\left[\left\|\int_s^t (N(r) - \overline{N}(r)dr\right\|_\alpha^p\right] \leq C_1(t-s)^{p/2}\varepsilon^p.$$

Then, for every $\mathfrak{y} < 2\tilde{\alpha}/(1+2\tilde{\alpha})$, there exists a constant C depending only on p and $\mathfrak{y}$ such that

$$\mathbf{E}\left[\sup_{t\in[0,\tau^*]}\left|\int_0^t f(s)(N(s) - \overline{N}(s))ds\right|^p\right] \leq C\varepsilon^{\mathfrak{y}p}\left(\mathbf{E}\left[\|f\|_{C^{\tilde{\alpha}}}\right]^{2p}\right)^{1/2},$$

where $\|\cdot\|_{C^{\tilde{\alpha}}}$ denotes the $\tilde{\alpha}$-Hölder norm.

Remark 5.3.3. *The above theorem can be used to approximate $\tilde{z}(t)$ by replacing N_i with $\overline{N}_i$ for each $i \in \{1, 2, ..., 14\}$, and the error is within $\mathcal{O}(\varepsilon^{\mathfrak{y}})$. In (5.40), $f_1 = f_5 = f_8 = f_{13} = \bar{\tilde{z}}$, $f_2 = f_4 = f_7 = f_{12} = \tilde{z}$, $f_3 = f_{10} = f_{11} = \bar{\tilde{z}}^2$, $f_6 = f_9 = f_{14} = 1$. Note that for $\tilde{\alpha} < 1/2$, we have f_i's satisfy the condition in Theorem 5.3.2. Consequently, we can choose $\mathfrak{y} < 1/2$. To use Theorem 5.3.2, it suffices to show the condition $\mathbb{E}\left[\left\|\int_s^t (N_i - \overline{N}_i)dr\right\|_\alpha^p\right] \leq C_1(t-s)^{p/2}\varepsilon^p$ holds [29]. We only show the cases when $k \in \mathbb{Z}_0$ (the case for $[\tilde{\Phi}_\delta^\star, \tilde{\Psi}_\delta^\star]^T$ is similar).*

Lemma 5.3.4. *For every $k \in \mathbb{Z} \setminus \{0\}$, we have*

$$\mathbf{E}\left[\left(\int_s^t \tilde{\mathfrak{g}}_k^\star(r)dr\right)^{2p}\right] \leq \frac{q_k^p\varepsilon^{2p}}{(a_k^{\mathfrak{c}})^{2p}}\varepsilon^{2p}(t-s)^p.$$

Proof. We also make a little abuse of notation and let $\tilde{\mathfrak{g}}_k^\star$ represent either $(\tilde{\mathfrak{g}}_k^{\mathrm{R}})^\star$ or $(\tilde{\mathfrak{g}}_k^{\mathrm{I}})^\star$ (from

(5.28)). Now let $p = 1$, then

$$\begin{aligned}
\mathbf{E}\left[\left(\int_s^t \tilde{\mathfrak{g}}_k^\star(r)dr\right)^2\right] &= \mathbf{E}\left[\left(\int_s^t \tilde{\mathfrak{g}}_k^\star(r)dr\right)\left(\int_s^t \tilde{\mathfrak{g}}_k^\star(u)du\right)\right] \\
&= \int_s^t \int_s^t \mathbf{E}[\tilde{\mathfrak{g}}_k^\star(r)\tilde{\mathfrak{g}}_k^\star(u)]drdu \\
&= -2\int_s^t \int_u^t \frac{q_k}{2a_k^{\mathfrak{c}}} e^{\frac{a_k^{\mathfrak{c}}(r-u)}{\varepsilon^2}} drdu \\
&= \frac{q_k\varepsilon^2}{(a_k^{\mathfrak{c}})^2}\left(t - s - \frac{\varepsilon^2}{(a_k^{\mathfrak{c}})^2}(1 - e^{\frac{a_k^{\mathfrak{c}}(t-s)}{\varepsilon^2}})\right) \le \frac{q_k\varepsilon^2}{(a_k^{\mathfrak{c}})^2}(t-s),
\end{aligned}$$

where the 2nd equality is by Fubini. Let $I_k := \int_s^t \tilde{\mathfrak{g}}_k^\star(r)dr$, then I_k is Gaussian with $\mathbf{E}[I_k] = 0$ and $\mathbf{E}[I_k^2] \le -\frac{q_k\varepsilon^2}{(a_k^{\mathfrak{c}})^2}(t-s)$. Therefore,

$$\mathbf{E}[|I_k|^{2p}] = \mathbf{E}[I_k^2]^p \le \left(\frac{q_k\varepsilon^2}{(a_k^{\mathfrak{c}})^2}(t-s)\right)^p$$

for every $p > 0$. □

Lemma 5.3.5. *For every $k \in \mathbb{Z}_s$, $k \neq l$ and $k + l \neq 0$, there exists a constant $C > 0$ such that*

$$\mathbf{E}\left[\left(\int_s^t \tilde{\mathfrak{g}}_k^\star(r)\tilde{\mathfrak{g}}_l^\star(r)dr\right)^{2p}\right] \le C\left(\frac{q_k q_l}{a_k^{\mathfrak{c}} a_l^{\mathfrak{c}}}\right)^p (t-s)^p \varepsilon^{2p}.$$

Lemma 5.3.6. *For every $k \in \mathbb{Z}_s$, $k = l$ or $k + l = 0$, there exists a constant $C > 0$ such that*

$$\mathbf{E}\left[\left(\int_s^t \tilde{\mathfrak{g}}_k^\star(s)\tilde{\mathfrak{g}}_l^\star(s) - \mathbf{E}[\tilde{\mathfrak{g}}_k^\star(s)\tilde{\mathfrak{g}}_l^\star(s)]ds\right)^{2p}\right] \le C\left(\frac{q_k q_l}{a_k^{\mathfrak{c}} a_l^{\mathfrak{c}}}\right)^p (t-s)^p \varepsilon^{2p}.$$

Lemma 5.3.7. *For every $k \in \mathbb{Z}_s$, there exists a constant $C > 0$ such that*

$$\mathbf{E}\left[\left(\int_s^t \tilde{\mathfrak{g}}_k^\star(s)\tilde{\mathfrak{g}}_l^\star(s)\tilde{\mathfrak{g}}_j^\star(s)ds\right)^{2p}\right] \le C\left(\frac{q_k q_l q_j}{a_k^{\mathfrak{c}} a_l^{\mathfrak{c}} a_j^{\mathfrak{c}}}\right)^p (t-s)^p \varepsilon^{2p}.$$

The proof for Lemma 5.3.5 to 5.3.7 is based on expanding the product of integrals that have Gaussian properties. The idea follows the proof of [29, Lemma 4.1]. We do not provide the proof in this section as we can simply treat the complex-valued $\tilde{\mathfrak{g}}_k^\star$ as we did in Lemma 5.3.4, and the rest follows exactly as [29, Lemma 4.1].

Corollary 5.3.8. *For every $i \in \{1, 2, ..., 14\}$, there exists a constant $C > 0$ such that*

$$\mathbf{E}\left[\left\|\int_s^t (N_i - \overline{N}_i)dr\right\|_\alpha^p\right] \leq C(t-s)^{p/2}\varepsilon^p.$$

Proof. By Definition 5.1.5 and Assumption 5.1.7, combining the definition of N_i and $\overline{N}_i$, it can be shown that the bounds generated from Lemma 5.3.5 to 5.3.7 converge. □

Renaming some constant quantities, we put (5.40) in a concise form. To this end, let

$$c_1 + ic_2 := \mathbf{E}[2N_2 + 4N_4 + N_7 + N_{12}].$$

and

$$\sigma_1 := -\frac{\mathcal{K}_1}{2}l_{11}, \quad \sigma_2 := -\frac{\mathcal{K}_1}{2}l_{12},$$
$$\sigma_3 := -\frac{\mathcal{K}_1 a_2^{\mathfrak{c}}\sqrt{q_2}}{2((a_2^{\mathfrak{c}})^2 + (b_2^{\mathfrak{c}})^2)}, \quad \sigma_4 := -\frac{\mathcal{K}_1 b_2^{\mathfrak{c}}\sqrt{q_2}}{2((a_2^{\mathfrak{c}})^2 + (b_2^{\mathfrak{c}})^2)},$$

as well as

$$\mathbf{M}(v^a) = \begin{bmatrix} \sigma_1 v_1^a & \sigma_2 v_1^a & \sigma_3 v_1^a - \sigma_4 v_2^a & \sigma_4 v_1^a + \sigma_3 v_2^a \\ \sigma_1 v_2^a & \sigma_2 v_2^a & -\sigma_3 v_2^a - \sigma_4 v_1^a & -\sigma_4 v_2^a + \sigma_3 v_1^a \end{bmatrix}_{2\times 4},$$

where $v^a := [\tilde{z}_1, \tilde{z}_2]^T$ represent the converted amplitudes. Moreover, we set

$$\mathbf{A}(\mathfrak{q}) := \begin{bmatrix} a_c^{\mathfrak{q}} + c_1 & -b_c^{\mathfrak{q}} - c_2 \\ b_c^{\mathfrak{q}} + c_2 & a_c^{\mathfrak{q}} + c_1 \end{bmatrix}_{2\times 2}, \tag{5.44}$$

$$\mathbf{B} := \begin{bmatrix} 2ha_2^{\mathfrak{q}} - j & 2hb_2^{\mathfrak{q}} \\ -2hb_2^{\mathfrak{q}} & 2ha_2^{\mathfrak{q}} - j \end{bmatrix}_{2\times 2}, \tag{5.45}$$

$$W_t = [\beta_\Phi(t), \beta_\Psi(t), \beta_2(t), \beta_{-2}(t)]^{\mathrm{T}}. \tag{5.46}$$

Recall that the h and j in (5.45) was aforementioned in (5.40). Then (5.40) is equivalent to

$$\begin{aligned} &v^a(t) = v^a(0) + \int_0^t \mathbf{A}(\mathfrak{q})v^a dt + \int_0^t |v^a|^2\mathbf{B}v^a dt + \int_0^t \mathbf{M}(v^a)dW_s + R^\varepsilon(t), \\ &v^a(0) = [\mathrm{Re}(\tilde{z}(0)), \mathrm{Im}(\tilde{z}(0))]^T, \ R^\varepsilon = \mathcal{O}(\varepsilon^{1/2-}). \end{aligned} \tag{5.47}$$

The $1/2$ reduction of the accuracy of the error term is due to the choice of $\mathfrak{y}$ in view of Remark 5.3.3. Note that the effects of additive noise that acts on the stable modes finally appear in the terms c_1, c_2, $ha_2^{\mathfrak{q}}$, and the multiplicative matrix $\mathbf{M}$. Not only the stability of the trivial solution $\tilde{z} = 0$ may be changed for any fixed $\mathfrak{q}$, but the dissipativity of the cubic nonlinearity may be different as well. The supercritical bifurcation structure could be destroyed by the noise.

5.3.4 Asymptotically Weak Convergence of the Approximation

We investigate how the solution to (5.47) (or the associated probability law) converges as $\varepsilon \to 0$. Given the canonical probability space $(\mathbf{\Omega}, \mathcal{F}, \mathbf{P}^\varepsilon)$[4] of the stopped process $\{v(t \wedge \tau^*)\}_{t\geq 0}$ driven by noises with intensity ε, the process $\{v^a(t\wedge\tau^*)\}_{t\geq 0} \in \mathbb{R}^2$ of (5.47) lies in the induced canonical space with probability law $\nu_c^\varepsilon = P_c\mathbf{P}^\varepsilon$ (recall Definition 5.1.11). Here we show that the unique limit ν_c of ν_c^ε solves the Martingale problem related to the 2-dimensional SDE:

$$\tilde{v}^a(t) = \tilde{v}^a(0) + \int_0^t \mathbf{A}(\mathfrak{q})\tilde{v}^a dt + \int_0^t |\tilde{v}^a|^2 \mathbf{B}\tilde{v}^a dt + \int_0^t \Sigma(\tilde{v}^a) d\beta_t, \tag{5.48}$$

where $\tilde{v}^a = [\tilde{v}_1^a, \tilde{v}_2^a]^T$, β stands for a two-dimensional Wiener process, and

$$\Sigma(\tilde{v}^a) := \begin{bmatrix} \left(\sum_{i=1}^4 \sigma_i\right)\tilde{v}_1^a + (\sigma_3 - \sigma_4)\,\tilde{v}_2^a \\ \left(\sum_{i=1}^2 \sigma_i - \sum_{i=3}^4 \sigma_i\right)\tilde{v}_2^a + (\sigma_3 - \sigma_4)\,\tilde{v}_1^a. \end{bmatrix} \tag{5.49}$$

Theorem 5.3.9. *Suppose $2h\alpha_2^s - f < 0$ in (5.45). For each fixed $T > 0$, the sequence of measures ν_c^ε converges weakly to ν_c, which is the law of the solution $\tilde{v}^a \in C([0,T];\mathbb{R}^2)$ to (5.48).*

To prove the above theorem, we need to demonstrate that: (1) the family of probability measure $\{\mathbf{P}^\varepsilon\}$ or $\{\nu_c^\varepsilon\}$ is tight, such that there exists a weakly convergent subsequence within that family, and (2) every accumulation point of ν_c^ε is the unique solution to the Martingale problem associated with (5.48).

Tightness of $\{\nu_c^\varepsilon\}$

The proof falls in standard procedures, we only provide the sketch. Let $f(\cdot) = \|\cdot\|^p$, and $h = v^a - R^\varepsilon$. Then, by [27, Lemma 4.9], we have

$$\operatorname{tr}[f''(h(\sigma))\mathbf{M}(h(\sigma) + R^\varepsilon(\sigma))\mathbf{M}(h(\sigma) + R^\varepsilon(\sigma))^*] \leq Cp(p-1)\|h(\sigma)\|^{p-2}\|h(\sigma) + R^\varepsilon(\sigma)\|^2. \tag{5.50}$$

[4]We emphasize the dependence of ε for the associated measure of solutions exited by noises with different intensities.

Applying Itô formula to $\|h\|^p$ for $p \geq 2$ and use the above inequality, for all $t \in [0, T]$, we have

$$\begin{aligned}\|h(t\wedge\tau^*)\|^p - \|h(0)\|^p \leq & p\int_0^{t\wedge\tau^*} \|h(s)\|^{p-2}\langle \mathbf{A}(\mathfrak{q})(h(s)+R(s)), h(s)+R(s)\rangle ds \\ & + \int_0^{t\wedge\tau^*} \|h(s)\|^{p-2}\langle |h(s)+R(s)|^2\mathbf{B}|h(s)+R(s)|, h(s)+R(s)\rangle ds \\ & + Cp(p-1)\int_0^{t\wedge\tau^*} \|h(s)\|^{p-2}\|h(s)+R(s)\|^2 ds \\ & + \int_0^{t\wedge\tau^*} \|h(s)\|^{p-2}\langle h(s), \mathbf{M}(h(s)+R(s))dW_s\rangle. \end{aligned} \tag{5.51}$$

By the assumption, we have $\langle x, |x|^2\mathfrak{B}x\rangle \geq b|x|^4$ for all $x \in \mathbb{R}^2$, and hence the first two terms can be bounded by $\tilde{C}\cdot(t\wedge\tau^*)$ for some $\tilde{C} > 0$. Applying Burkholder–Davis–Gundy inequality for the last term and then Young's inequality for the last two terms, combining the above, we can obtain

$$\mathbf{E}^{\nu_c^\varepsilon} \sup_{0\leq t\leq\tau^*} \|h(t)\|^p \leq C_1 \int_0^{T\wedge\tau^*} \mathbf{E}^{\nu_c^\varepsilon} \sup_{0\leq t\leq\tau^*} \|h(s)\|^p ds + C_2.$$

By Gronwall's inequality, we can verify that $\mathbf{E}^{\nu_c^\varepsilon}\sup_{0\leq t\leq\tau^*}\|h(t)\|^p \leq C$ and hence the quantity $\mathbf{E}^{\nu_c^\varepsilon}\sup_{0\leq t\leq\tau^*}\|v^a(t)\|^p$ is uniformly bounded, which implies the uniform tightness of $\{\nu_c^\varepsilon\}$.

Remark 5.3.10. *Note that by introducing the compact operator $G_\alpha : L^p([0,T];U)) \to C([0,T];U)$ for $0 < 1/p < \alpha \leq 1$ and $t \in [0,T]$:*

$$G_\alpha f(t) = \int_0^t (t-s)^{\alpha-1} S(t-s) f(s) ds, \quad f \in L^p([0,T], H),$$

as well as $Y_\alpha^\varepsilon(t) = \varepsilon\int_0^t (t-r)^{-\alpha} S(t-r) dW(r)$, the mild solution can be expressed as

$$v(t) = S(t)v_0 + G_1(\varepsilon^{-1}B + F)(t) + \frac{\sin\alpha\pi}{\pi} G_\alpha(Y_\alpha^\varepsilon)(t). \tag{5.52}$$

The compactness of G_α has been shown in [43, Proposition 8.4]. However, by a similar argument as in the Proof of Proposition 5.1.10, for a fixed $p \geq 2$ we can only find $C_1(\varepsilon), C_2(\varepsilon) > 0$ for each $\varepsilon > 0$ such that

$$\mathbf{E}^\varepsilon\left[\int_0^{t\wedge\tau^*} |Y_\alpha^\varepsilon(s)|^p ds\right] \leq C_1(\varepsilon), \quad \forall t \in [0,T], \tag{5.53}$$

and

$$\mathbf{E}^\varepsilon\left[\int_0^{t\wedge\tau^*} |\varepsilon^{-1}B + F(s)|^p ds\right] \leq C_2(\varepsilon) \;\; \forall t \in [0,T]. \tag{5.54}$$

The nonuniform bounds fail to guarantee the uniform tightness of $\{\mathbf{P}^\varepsilon\}$.

Martingale problem

Given a test function $\varphi \in C_0^\infty(P_c\mathscr{H})$, the generator $\mathfrak{L}$ of (5.48) is given by

$$\mathfrak{L}\varphi(x) = \langle \mathbf{A}(\mathfrak{q})x + |x|^2\mathbf{B}x,\ \nabla\varphi\rangle + \frac{1}{2}\sum_{i,j}^{2}\left(\Sigma\Sigma^T\right)_{ij}\frac{\partial^2\varphi}{\partial x_i \partial x_j}. \tag{5.55}$$

Then, by defining

$$M_t^\varepsilon := \varphi(v^a - R^\varepsilon)(t\wedge\tau^*) - \varphi(v^a)(0) - \int_0^{t\wedge\tau^*} \mathfrak{L}\varphi(v^a - R^\varepsilon)(s)ds,\ t\in[0,T], \tag{5.56}$$

it is clear that $\{M^\varepsilon\}$ is a family of (stopped) martingales. Due to the boundedness of $R^\varepsilon(t)$ in moments and the smoothness of the test function φ, there exists a process $\tilde{R}(t)$ such that

$$M_t^\varepsilon = \varphi(v^a)(t\wedge\tau^*) - \varphi(v^a)(0) - \int_0^{t\wedge\tau^*} \mathfrak{L}\varphi(v^a)(s)ds + \tilde{R}(t\wedge\tau^*),\ t\in[0,T], \tag{5.57}$$

and $\lim_{\varepsilon\to 0}\mathbf{E}^{\nu_c^\varepsilon}[\sup_{t\in[0,\tau^*]}\tilde{R}(t)] = 0$, where $\mathbf{E}^{\nu_c^\varepsilon}$ is the expectation operator w.r.t. the measure ν_c^ε. Therefore, for any $0 \le r_1 < r_2 < \ldots < r_n \le s < t \le T$ and $\{\varphi_j;\ j = 1,2,\ldots,n\} \subset C(P_c\mathscr{H})$, we alternatively have

$$\mathbf{E}^{\nu_c^\varepsilon}\left[\{M_t^\varepsilon - M_s^\varepsilon\}\prod_{j=1}^n \varphi_j(v_{r_j}^a)\right] = 0 \tag{5.58}$$

We also define the Martingale process w.r.t. (5.48) as

$$M_t = \varphi(v_t^a) - \varphi(v_0^a) - \int_0^t \mathfrak{L}\varphi(v_s^a)ds,\ t\in[0,T]. \tag{5.59}$$

Since the smooth test function has a compact support, we can also justify that $\{M_{t\wedge\tau^*}\}_{t\in[0,T]}$ is uniform integrable. It has been shown in [29, Corollary 3.7] that 'the event $\tau^* < T$ is caused by x (or z) getting too large'. More precisely, $\mathbf{P}[\tau^* < T] \le \mathbf{P}[K\|x_{\tau^*}\| \ge \varepsilon^{-\kappa}] + C\varepsilon^p$ for every $p > 0$ and $K > 1$. By the uniform boundedness of $\mathbf{E}^{\nu_c^\varepsilon}\sup_{0\le t\le\tau^*}\|v^a(t)\|^p$ for each p, as $\varepsilon\to 0$, we have $\mathbf{P}[\tau^* < T] \to 0$. Consequently, the indicator function $\mathbb{1}_{\{T<\tau^*\}} \to 1$ as $\varepsilon \to 0$.

Since $\{\nu_c^\varepsilon\}$ is a tight family of measures on $P_c\mathscr{H}$, we can find a convergent subsequence $\nu_c^{\varepsilon_n} \rightharpoonup \nu_c$ as $n\to\infty$ (where $\varepsilon_n \to 0$). Therefore,

$$\begin{aligned}\mathbf{E}^{\nu_c}\left[\{M_t - M_s\}\prod_{j=1}^n \varphi_j(v_{r_j}^a)\right] &= \lim_{n\to\infty}\mathbf{E}^{\nu_c^{\varepsilon_n}}\left[\{M_{t\wedge\tau^*} - M_{s\wedge\tau^*}\}\prod_{j=1}^n \varphi_j(v_{r_j}^a)\right] \\ &= \lim_{n\to\infty}\mathbf{E}^{\nu_c^{\varepsilon_n}}\left[\{M_t^{\varepsilon_n} - M_s^{\varepsilon_n}\}\prod_{j=1}^n \varphi_j(v_{r_j}^a)\right] = 0\end{aligned} \tag{5.60}$$

which means every limit of $\nu_c^{\varepsilon_n}$ solves the martingale problem w.r.t. (5.59). Note that given the local Lipschitz continuity of the vector fields, by Yamada-Watanabe, the solution to the Martingale problem is unique, which means every limit point ν_c is unique, and therefore the claim in Theorem 5.3.9 holds. Theorem 5.3.9 also implies that $\{v^a(t \wedge \tau^*)\}_{t\geq 0}$ converges to $\{\tilde{v}^a(t)\}_{t\in[0,T]}$ in probability law for any fixed $T > 0$.

5.4 Summary

Based on recent advances in stochastic PDEs given in [29], this section further develops the bifurcation analysis of the stochastic version of the Moore and Greitzer PDE model (5.5) for an axial flow compressor, in the presence of a Hopf bifurcation. Close to bifurcation, the null-space being finite-dimensional simplifies the analysis of such PDEs. We provided approximations for the state $\mathrm{g} \in \mathcal{H}$ for the stall case in the neighborhood of the deterministic bifurcation point. The evolution equation for slow-varying coordinates $\tilde{v}^a$, which can be treated as the normal form, is derived by a careful analysis of the coupling of slow-fast modes arising from the spectral gap.

As explained previously, in addition to the possible direct influence that the additive noise has on the critical modes (which we assumed to be identically zero in this study), the additive stochastic components in the stable, heavily damped modes also contribute to the critical modes. These contributions enter the critical modes as multiplicative noise through the terms $N_i' = \mathbb{E}[N_i(t)]$ for $i \in \{1, ..., 14\}$ in (5.40) and are eventually incorporated into the 2-dimensional SDE (5.47). Hence, the stochastic bifurcation points for stall are shifted due to the evolution (stochastic) of heavily damped modes. The dissipativity of the cubic symmetric nonlinearties may appear differently as well based on the choice of parameter. As the intensity $\varepsilon \to 0$, we justified a weak convergence of the probability measure of the slow-varying processes. The approximated slow processes also converge in probability law to the solution to (5.48).

5.5 Supplementary Results

Homogenization results

We provide the explicit form of random variables that appear in the homogenization results (Equation 5.40) in this section.

$$N_1(\varpi) = -\frac{\mathcal{K}_1\mathcal{K}_2}{4\rho_2^{\mathfrak{c}}}\left(2\tilde{\Phi}_\delta^\star\tilde{\mathfrak{g}}_2^\star + \sum_{k\in\mathbb{Z}_s}^{k+l=2}\tilde{\mathfrak{g}}_k^\star\tilde{\mathfrak{g}}_l^\star\right) \tag{5.61}$$

$$N_2(\varpi) = -\frac{\mathcal{K}_1}{4\mathfrak{l}_c}\left[(l_{11}\psi''_{c,\gamma_c})\left((\tilde{\Phi}_\delta^\star)^2 + 2\sum_{k\in\{-2,-3,\ldots\}}\tilde{\mathfrak{g}}_k^\star\tilde{\mathfrak{g}}_{-k}^\star\right) - (l_{12}\mathcal{S}'''_{\gamma_c})(\tilde{\Psi}_\delta^\star)^2\right] \tag{5.62}$$

$$N_3(\varpi) = -\frac{\mathcal{K}_1\mathcal{K}_2}{4\rho_2^{\mathfrak{c}}}\tilde{\mathfrak{g}}_3^\star \tag{5.63}$$

$$N_4(\varpi) = -\frac{\mathcal{K}_1^2}{4}\tilde{\mathfrak{g}}_{-2}^\star\left(\frac{\tilde{\mathfrak{g}}_2^\star}{\rho_2^{\mathfrak{c}}}\right) - \frac{\mathcal{K}_1^2}{4}(\tilde{\Phi}_\delta^\star)(l_{11}\tilde{\Phi}_\delta^\star + l_{12}\tilde{\Psi}_\delta^\star) \tag{5.64}$$

$$N_5(\varpi) = -\frac{\mathcal{K}_1^2}{4}(\tilde{\Phi}_\delta^\star)\left(\frac{\tilde{\mathfrak{g}}_2^\star}{\rho_2^{\mathfrak{c}}}\right) - \frac{\mathcal{K}_1^2}{4}\tilde{\mathfrak{g}}_2^\star(l_{11}\tilde{\Phi}_\delta^\star + l_{12}\tilde{\Psi}_\delta^\star) \tag{5.65}$$

$$N_6(\varpi) = -\frac{\mathcal{K}_1^2}{4}\left(\sum_{k\in\mathbb{Z}_s}^{k+l=1}\tilde{\mathfrak{g}}_k^\star\tilde{\mathfrak{g}}_l^\star\right)(l_{11}\tilde{\Phi}_\delta^\star + l_{12}\tilde{\Psi}_\delta^\star) - \frac{\mathcal{K}_1^2}{4}\left(\sum_{k\in\mathbb{Z}_s}^{k+l=-1}\tilde{\mathfrak{g}}_k^\star\tilde{\mathfrak{g}}_l^\star\right)\left(\frac{\tilde{\mathfrak{g}}_2^\star}{\rho_2^{\mathfrak{c}}}\right) \tag{5.66}$$

$$N_7(\varpi) = -\mathcal{K}_1\sum_{k\in\{-2,-3\ldots\}}^{k+l=1}\left(\frac{\mathcal{K}_k\tilde{\mathfrak{g}}_k^\star\tilde{\mathfrak{g}}_{-k}^\star + \mathcal{K}_l\tilde{\mathfrak{g}}_l^\star\tilde{\mathfrak{g}}_{-l}^\star}{\rho_k^{\mathfrak{c}} + \rho_l^{\mathfrak{c}}}\right) \tag{5.67}$$

$$N_8(\varpi) = -\mathcal{K}_1\sum_{k\in\{-2,-3\ldots\}}^{k+l=1}\frac{\mathcal{K}_{l+1}\tilde{\mathfrak{g}}_k^\star\tilde{\mathfrak{g}}_{l+1}^\star + \mathcal{K}_{k+1}\tilde{\mathfrak{g}}_l^\star\tilde{\mathfrak{g}}_{k+1}^\star}{\rho_k^{\mathfrak{c}} + \rho_l^{\mathfrak{c}}} \tag{5.68}$$

$$N_9(\varpi) = -\mathcal{K}_1\sum_{k\in\{-2,-3\ldots\}}^{k+l=1}\left(\frac{\tilde{\mathfrak{g}}_k^\star\mathcal{G}_l + \tilde{\mathfrak{g}}_l^\star\mathcal{G}_k}{\rho_k^{\mathfrak{c}} + \rho_l^{\mathfrak{c}}}\right) \tag{5.69}$$

$$N_{10}(\varpi) = -\frac{\mathcal{K}_1\mathcal{K}_2\tilde{\mathfrak{g}}_3^\star}{2(\rho_{-2}^{\mathfrak{c}} + \rho_3^{\mathfrak{c}})} \tag{5.70}$$

$$N_{11}(\varpi) = -\frac{\mathcal{K}_1\psi'''_c}{2\psi''_{c,\gamma_c}}\tilde{\mathfrak{g}}_3^\star \tag{5.71}$$

$$N_{12}(\varpi) = -\frac{\mathcal{K}_1\psi_c'''}{2\psi_{c,\gamma_c}''}\left[(\tilde{\Phi}_\delta^\star)^2 + 2\sum_{k\in\{-2,-3...\}}\tilde{\mathfrak{g}}_k^\star\tilde{\mathfrak{g}}_{-k}^\star\right] \tag{5.72}$$

$$N_{13}(\varpi) = -\frac{\mathcal{K}_1\psi_c'''}{2\psi_{c,\gamma_c}''}\left[2\tilde{\Phi}_\delta^\star\tilde{\mathfrak{g}}_2^\star + \sum_{k\in\mathbb{Z}_s}^{k+l=2}\tilde{\mathfrak{g}}_k^\star\tilde{\mathfrak{g}}_l^\star\right] \tag{5.73}$$

$$N_{14}(\varpi) = -\frac{\mathcal{K}_1\psi_c'''}{6\psi_{c,\gamma_c}''}\left[3\tilde{\Phi}_\delta^\star\left(\sum_{k\in\mathbb{Z}\setminus\{0,\pm1\}}^{k+l=1}\tilde{\mathfrak{g}}_k^\star\tilde{\mathfrak{g}}_l^\star\right) + \sum_{k,l\in\mathbb{Z}\setminus\{0,\pm1\}}^{k+l+m=1}\tilde{\mathfrak{g}}_k^\star\tilde{\mathfrak{g}}_l^\star\tilde{\mathfrak{g}}_m^\star\right] \tag{5.74}$$

Proof of Lemma 5.2.2

Proof. $B_{c,1}(y,y) = \frac{\mathfrak{a}(\psi_c'')}{1+\mathfrak{am}}\sum_{k\in\{-2,-3,...\}}^{k+l=1}\mathfrak{g}_k\mathfrak{g}_l$. From (7.2b) we keep the terms up to $\mathcal{O}(\varepsilon^{-1})$ and regard the rest as higher order terms (h.o.t.), then

$$\begin{aligned} d\mathfrak{g}_l &= \varepsilon^{-2}\rho_l^{\mathfrak{c}}\mathfrak{g}_l dt + \varepsilon^{-1}\langle e_l^*, P_s B(x+y,x+y)\rangle_{\mathscr{H}}dt \\ &\quad + \varepsilon^{-1}\sqrt{q_l}(d\beta_l(t) + id\beta_{-l}(t)) + \text{h.o.t.},\ \forall l\in\{3,4...\}, \end{aligned}$$

and

$$\begin{aligned} d\mathfrak{g}_k &= \varepsilon^{-2}\rho_k^{\mathfrak{c}}\mathfrak{g}_k dt + \varepsilon^{-1}\langle e_k^*, P_s B(x+y,x+y)\rangle_{\mathscr{H}}dt \\ &\quad + \varepsilon^{-1}\sqrt{q_k}(d\beta_{-k}(t) - id\beta_k(t)) + \text{h.o.t.}, \forall k\in\{-2,-3...\}. \end{aligned}$$

For $l\in\{3,4,...\}$ we have,

$$\begin{aligned} \langle e_l^*, P_s B(x+y,x+y)\rangle_{\mathscr{H}} &= \langle e_l^*, B(x,x)\rangle_{\mathscr{H}} + 2\langle e_l^*, P_s B(x,y)\rangle_{\mathscr{H}} + \langle e^*, P_s B(y,y)\rangle_{\mathscr{H}} \\ &= 0 + (z\mathcal{K}_{l-1}\mathfrak{g}_{l-1} + \bar{z}\mathcal{K}_{l+1}\mathfrak{g}_{l+1}) + \mathcal{G}_l, \end{aligned}$$

for $k=-2$,

$$\begin{aligned} \langle e_k^*, P_s B(x+y,x+y)\rangle_{\mathscr{H}} &= \langle e_k^*, B(x,x)\rangle_{\mathscr{H}} + 2\langle e_k^*, P_s B(x,y)\rangle_{\mathscr{H}} + \langle \nu_k^*, P_s B(y,y)\rangle_{\mathscr{H}} \\ &= \frac{\mathfrak{a}\psi_c''\bar{z}^2}{2+2\mathfrak{am}} + (z\mathcal{K}_{k-1}\mathfrak{g}_{k-1} + \bar{z}\mathcal{K}_{k+1}\mathfrak{g}_{k+1}) + \mathcal{G}_k \end{aligned}$$

and for $k\in\{-3,-4,...\}$, $\langle e_k^*, P_s B(x+y,x+y)\rangle_{\mathscr{H}} = (z\mathcal{K}_{k-1}\mathfrak{g}_{k-1} + \bar{z}\mathcal{K}_{k+1}\mathfrak{g}_{k+1}) + \mathcal{G}_k$.

Applying Itô's formula on $d(\mathfrak{g}_k\mathfrak{g}_l)$ for $k \in \{-3, -4...\}$ and $k + l = 1$ we have:

$$\begin{aligned} d(\mathfrak{g}_k\mathfrak{g}_l) &= \mathfrak{g}_k d\mathfrak{g}_l + \mathfrak{g}_l d\mathfrak{g}_k \\ &= \varepsilon^{-2}\rho_l^{\mathfrak{c}}\mathfrak{g}_k\mathfrak{g}_l dt + \varepsilon^{-1}\left(z\mathcal{K}_k\mathfrak{g}_k\mathfrak{g}_{-k} + \bar{z}\mathcal{K}_{l+1}\mathfrak{g}_k\mathfrak{g}_{l+1}\right)dt + \varepsilon^{-1}\mathfrak{g}_k\mathcal{G}_l dt \\ &+ \varepsilon^{-2}\rho_k^{\mathfrak{c}}\mathfrak{g}_k\mathfrak{g}_l dt + \varepsilon^{-1}\left(z\mathcal{K}_l\mathfrak{g}_l\mathfrak{g}_{-l} + \bar{z}\mathcal{K}_{k+1}\mathfrak{g}_l\mathfrak{g}_{k+1}\right)dt + \varepsilon^{-1}\mathfrak{g}_l\mathcal{G}_k dt \\ &+ \varepsilon^{-1}\mathfrak{g}_k\sqrt{q_l}(d\beta_l(t) + id\beta_{-l}(t)) + \varepsilon^{-1}\mathfrak{g}_l\sqrt{q_k}(d\beta_{-k}(t) - id\beta_k(t)); \end{aligned}$$

for $k = -2$ and $l = 3$ we have:

$$\begin{aligned} d(\mathfrak{g}_k\mathfrak{g}_l) &= \mathfrak{g}_k d\mathfrak{g}_l + \mathfrak{g}_l d\mathfrak{g}_k \\ &= \varepsilon^{-2}\rho_l^{\mathfrak{c}}\mathfrak{g}_k\mathfrak{g}_l dt + \varepsilon^{-1}\left(z\mathcal{K}_k\mathfrak{g}_k\mathfrak{g}_{-k} + \bar{z}\mathcal{K}_{l+1}\mathfrak{g}_k\mathfrak{g}_{l+1}\right)dt + \varepsilon^{-1}\mathfrak{g}_k\mathcal{G}_l dt + \varepsilon^{-1}\mathfrak{g}_l\mathcal{G}_k dt \\ &+ \varepsilon^{-2}\rho_k^{\mathfrak{c}}\mathfrak{g}_k\mathfrak{g}_l dt + \varepsilon^{-1}\left(z\mathcal{K}_l\mathfrak{g}_l\mathfrak{g}_{-l} + \bar{z}\mathcal{K}_{k+1}\mathfrak{g}_l\mathfrak{g}_{k+1}\right)dt + \varepsilon^{-1}\left(\frac{\mathfrak{a}\psi_c''\bar{z}^2\mathfrak{g}_3}{2 + 2\mathfrak{am}}\right)dt \\ &+ \varepsilon^{-1}\mathfrak{g}_k\sqrt{q_l}(d\beta_l(t) + id\beta_{-l}(t)) + \varepsilon^{-1}\mathfrak{g}_l\sqrt{q_k}(d\beta_{-k}(t) - id\beta_k(t)). \end{aligned}$$

Therefore, for $k \in \{-3, -4...\}$ and $k + l = 1$,

$$\begin{aligned} \mathfrak{g}_k\mathfrak{g}_l dt &= -\varepsilon\left(\frac{1}{\rho_k^{\mathfrak{c}} + \rho_l^{\mathfrak{c}}}\right)\left(z\mathcal{K}_k\mathfrak{g}_k\mathfrak{g}_{-k} + \bar{z}\mathcal{K}_{l+1}\mathfrak{g}_k\mathfrak{g}_{l+1} + \mathfrak{g}_k\mathcal{G}_l\right)dt \\ &- \varepsilon\left(\frac{1}{\rho_k^{\mathfrak{c}} + \rho_l^{\mathfrak{c}}}\right)\left(z\mathcal{K}_l\mathfrak{g}_l\mathfrak{g}_{-l} + \bar{z}\mathcal{K}_{k+1}\mathfrak{g}_l\mathfrak{g}_{k+1} + \mathfrak{g}_l\mathcal{G}_k\right)dt \\ &- \varepsilon\left[\frac{\mathfrak{g}_l\sqrt{q_k}(d\beta_{-k}(t) - id\beta_k(t)) + \mathfrak{g}_k\sqrt{q_l}(d\beta_l(t) + id\beta_{-l}(t))}{\rho_k^{\mathfrak{c}} + \rho_l^{\mathfrak{c}}}\right]; \end{aligned}$$

for $k = -2$ and $l = 3$,

$$\begin{aligned} \mathfrak{g}_k\mathfrak{g}_l dt &= -\varepsilon\left(\frac{1}{\rho_k^{\mathfrak{c}} + \rho_l^{\mathfrak{c}}}\right)\left(z\mathcal{K}_k\mathfrak{g}_k\mathfrak{g}_{-k} + \bar{z}\mathcal{K}_{l+1}\mathfrak{g}_k\mathfrak{g}_{l+1} + \mathfrak{g}_k\mathcal{G}_l\right)dt \\ &- \varepsilon\left(\frac{1}{\rho_k^{\mathfrak{c}} + \rho_l^{\mathfrak{c}}}\right)\left(z\mathcal{K}_l\mathfrak{g}_l\mathfrak{g}_{-l} + \bar{z}\mathcal{K}_{k+1}\mathfrak{g}_l\mathfrak{g}_{k+1} + \mathfrak{g}_l\mathcal{G}_k\right)dt \\ &- \varepsilon\left(\frac{1}{\rho_k^{\mathfrak{c}} + \rho_l^{\mathfrak{c}}}\right)\left(\frac{\mathfrak{a}\psi_c''\bar{z}^2\mathfrak{g}_3}{2 + 2\mathfrak{am}}\right)dt \\ &- \varepsilon\left[\frac{\mathfrak{g}_l\sqrt{q_k}(d\beta_{-k}(t) - id\beta_k(t)) + \mathfrak{g}_k\sqrt{q_l}(d\beta_l(t) + id\beta_{-l}(t))}{\rho_k^{\mathfrak{c}} + \rho_l^{\mathfrak{c}}}\right], \end{aligned}$$

and the result follows easily from a combination of the above.

□

Chapter 6

Multiscale Analysis for SPDEs with Multiplicative Noise Close to Hopf Bifurcation

The dynamical stochastic Hopf bifurcation was investigated in [17] for finite dimensional SDEs, driven by multiplicative noise, with coefficients dependent on some parameter γ. It has been shown that when γ varies in a way that the top Lyapunov exponent $\lambda(\gamma)$ changes sign from negative to positive, the trivial solution loses its almost-sure asymptotic stability and a non-trivial invariant measure is formed. In contrast to the finite dimensional cases, it is difficult to quantitatively describe the random invariant manifolds and stochastic bifurcations for SPDEs driven by multiplicative noise [79, 156, 107, 37].

On the other hand, when the separation of time-scales is naturally present in a neighbourhood of a deterministic bifurcation point, multiscale approximations of the dynamics using amplitude equations can describe the dynamics of the slowly-varying critical (dominating) modes of the SPDE. This dimension reduction technique has been used to study the transient dynamics of SPDEs driven by additive noise near the deterministic bifurcation [28][29]. More rigorous analysis on the approximation of the invariant measure was developed in [28].

In terms of dynamics driven by multiplicative noise, multiscale analysis and stochastic averaging/homogenization techniques have been applied to dimensional reduction problems of noisy nonlinear systems with rapidly oscillating and decaying components. For finite dimensional SDEs with small multiplicative noise, when many of the modes are 'heavily damped', reduced-order models were obtained using a martingale problem approach in [125]. The result verifies that as the noise becomes smaller, a lower dimensional Markov process characterizes the limiting behavior in the weak topology. In application, [150] derives a low-dimensional approximation of an 11-dimensional nonlinear stochastic aeroelastic problem near a deterministic Hopf bifurcation, with one critical mode and several stable modes. The reduced model performs well in terms of simulating the distribution, density, as well as the top Lyapunov exponent of the full system near the deterministic Hopf bifurcation.

The recent work [27] studies the impact of multiplicative noise in SPDEs near the bifurcation using amplitude equations. However, the hypothesis only guarantees a local existence of mild solution to the SPDEs up to a random explosion time. Under a stronger dissipative assumption in the critical modes, the derived amplitude equation can be used to approximate the solution to the original SPDE up to a fixed deterministic time with error converging in probability.

To investigate the long-term behavior of SPDEs near the deterministic Hopf bifurcation point and demonstrate the dynamical Hopf bifurcation under the impact of multiplicative noise, we impose proper conditions that guarantee the existence of invariant measures and study SPDEs of the following type

$$du(t) = \mathcal{A}(\gamma)u(t)dt + F(u(t))dt + \varepsilon G(u(t))dW(t), \tag{6.1}$$

where $u(t)$ takes value in an infinite-dimensional separable Hilbert space $\mathcal{H} = L^2(E)$ for some bounded $E \subseteq \mathbb{R}^n$. The self-adjoint unbounded linear $\mathcal{A}(\gamma)$ that also depends on a parameter $\gamma \in \mathbb{R}$ generates an analytic compact C_0 semigroup on $\mathcal{H}$. The nonlinearity F is a cubic mapping, and $G(u)$ is Hilbert-Schmidt operator with $G(0) = 0$. The noise W is a cylindrical Wiener process (see Definition D.2.2 for details).

We denote by γ_c the deterministic Hopf bifurcation point of (6.1) with the absence of noise ($\varepsilon = 0$), and aim to approximate the solution as well as the invariant measures around γ_c, i.e. for $\gamma = \gamma_c + \varepsilon^2\mathfrak{q}$ with some $\mathfrak{q} \in \mathbb{R}$, using multiscale technique. The purpose of this chapter, unlike [125, 150, 27], is not dimension reduction using the averaging/homogenization technique. We also do not intend to use averaging/homogenization results to study the structural change of the invariant measures. The approximation scheme only simplifies the dynamics with errors up to an acceptable scale for local analysis, which will be utilized later in Chapter 8 to connect with the almost-sure stability of the trivial solution and show the dynamical Hopf bifurcation of SPDEs of the form (6.1).

Note that the deterministic part that we consider is usually obtained by a local expansion and coordinate transformation of more general nonlinearilties, which captures the local dynamical behavior with topological equivalence. We do not involve any quadratic terms to avoid complicated interactions between the perturbed critical and stable modes. To better understand the impact of multiplicative noise, we focus on the given type (6.1).

Before proceeding, we introduce the notations and formulate necessary assumptions.

6.1 Notations and Main Assumptions

Given the separable Hilbert space $\mathcal{H}$, we denote by $\langle\cdot,\ \cdot\rangle$ the inner product in $\mathcal{H}$ and by $\|\cdot\|$ the norm. Due to the compactness of the associated semigroup, $\mathcal{A}(\gamma)$ has a pure point spectrum. We make the following assumptions on $\mathcal{A}(\gamma)$ about the point spectrum.

Assumption 6.1.1. *We assume that, for all γ,*

(1) the point spectrum $\{\rho_k(\gamma)\}_{k\in\mathbb{Z}_0}$ of $\mathcal{A}(\gamma)$ is complex, where $\rho_k(\gamma) := a_k(\gamma) + ib_k(\gamma) \in \mathbb{C}$ and $\rho_{-k}(\gamma) = \overline{\rho}_k(\gamma)$ for all $k \in \mathbb{Z}_0$;

(2) $\rho_k(\gamma)$ is analytic in γ for all $\gamma \in \mathbb{R}$ and all $k \in \mathbb{Z}_0$, and $a_1(\gamma) > a_2(\gamma) \geq \cdots \geq a_k(\gamma) \geq \ldots$;

(3) the corresponding eigenvectors $\{e_k\}_{k\in\mathbb{Z}_0}$ form a complete orthonormal basis of $\mathcal{H}$ such that $\mathcal{A}(\gamma)e_k = \rho_k(\gamma)e_k$, $\langle e_{-k}, e_k\rangle = 1$ for all $k \in \mathbb{Z}_0$, and $\langle e_i, e_j\rangle = 0$ for all $i + j \neq 0$.

Remark 6.1.2. *Note that by Assumption 6.1.1, $\mathcal{A}u = \sum_{k\in\mathbb{Z}_0} \rho_k\langle e_{-k}, u\rangle e_k$ for all $u \in \mathcal{H}$, and for all $u, v \in \mathcal{H}$ we have*

$$\langle \mathcal{A}u, v\rangle = \sum_{k\in\mathbb{Z}_0} \rho_k\langle e_{-k}, u\rangle\langle e_k, v\rangle = \sum_{k\in\mathbb{Z}_0} \rho_{-k}\langle e_k, u\rangle\langle e_{-k}, v\rangle = \sum_{k\in\mathbb{Z}_0} \rho_k\langle e_k, u\rangle\langle e_{-k}, v\rangle = \langle u, \mathcal{A}v\rangle,$$

which indicates the self-adjoint property of $\mathcal{A}$. The second to the last identity is in that, for each $k \neq 0$,

$$\rho_{-k}\langle e_k, u\rangle\langle e_{-k}, v\rangle + \rho_k\langle e_{-k}, u\rangle\langle e_k, v\rangle = \rho_k\langle e_k, u\rangle\langle e_{-k}, v\rangle + \rho_{-k}\langle e_{-k}, u\rangle\langle e_k, v\rangle \in \mathbb{R}.$$

Since γ_c is the deterministic Hopf bifurcation point, we have $a_{\pm 1}(\gamma_c) = 0$ as well as $a'_{\pm 1}(\gamma_c) \neq 0$, $b_{\pm 1}(\gamma_c) \neq 0$, whilst the rest of the spectrum stays in the left half-plane. We introduce the shorthand notation $\mathfrak{h} := e_1$ and $\bar{\mathfrak{h}} := e_{-1}$ to denote the critical eigenvectors. We denote by $\mathfrak{h}^*$ and $\bar{\mathfrak{h}}^*$ the associated adjoint eigenvectors , which satisfy

$$\langle \mathfrak{h}^*, \mathfrak{h}\rangle = 1,\ \langle \mathfrak{h}^*, \bar{\mathfrak{h}}\rangle = 0,\ \langle \bar{\mathfrak{h}}^*, \bar{\mathfrak{h}}\rangle = 1,\ \langle \bar{\mathfrak{h}}^*, \mathfrak{h}\rangle = 0.$$

Due to the existence of spectral gap, we also introduce the projections and basic properties of $\mathcal{A}(\gamma)$.

Definition 6.1.3. *The critical projection operator is defined as*

$$P_c(\cdot) := \langle \mathfrak{h}^*, \cdot\ \rangle\mathfrak{h} + \langle \bar{\mathfrak{h}}^*, \cdot\ \rangle\bar{\mathfrak{h}}, \tag{6.2}$$

The stable projection operator is $P_s = I - P_c$. For simplicity, we introduce shorthand notation $F_c := P_cF$. We define F_s, $\mathcal{A}_c(\gamma)$, $\mathcal{A}_s(\gamma)$, G_c, G_s, $\mathcal{H}_c$ and $\mathcal{H}_s$ in a similar way.

Definition 6.1.4 (Other notations for $\mathcal{A}(\gamma)$)**.** *To this end, we use*

(1) $\mathbb{Z}_c := \{\pm 1\}$, $\mathbb{Z}_s := \mathbb{Z}_0 \setminus \{\pm 1\}$.

(2) $\mathcal{A}_c^{\mathfrak{c}} := \mathcal{A}_c(\gamma_c)$, $\mathcal{A}_c^{\mathfrak{q}} := \mathfrak{q}\mathcal{A}_c'(\gamma_c)$*; the associated eigenvalues of* $\mathcal{A}_c^{\mathfrak{c}}$ *and* $\mathcal{A}_c^{\mathfrak{q}}$ *are respectively denoted by*

$$\rho_c^{\mathfrak{c}} = ib_c^{\mathfrak{c}} := ib_1(\gamma_c), \quad \bar{\rho}_c^{\mathfrak{c}} = -ib_c^{\mathfrak{c}} := -ib_1(\gamma_c),$$
$$\rho_c^{\mathfrak{q}} = a_c^{\mathfrak{q}} + ib_c^{\mathfrak{q}} := \mathfrak{q}(a_1'(\gamma_c) + ib_1'(\gamma_c))$$

and

$$\bar{\rho}_c^{\mathfrak{q}} = a_c^{\mathfrak{q}} - ib_c^{\mathfrak{q}} := \mathfrak{q}(a_1'(\gamma_c) - ib_1'(\gamma_c)).$$

(3) $\mathcal{A}_c^{\mathfrak{er}} := \varepsilon^{-2}[\mathcal{A}_c(\gamma_c + \varepsilon^2\mathfrak{q}) - \mathcal{A}_c^{\mathfrak{c}} - \varepsilon^2\mathcal{A}_c^{\mathfrak{q}}]$*, the associated eigenvalues of* $\mathcal{A}_c^{\mathfrak{er}}$ *are denoted as* $\rho_c^{\mathfrak{er}}$ *and* $\bar{\rho}_c^{\mathfrak{er}}$*.*

(4) $\mathcal{A}_s := \mathcal{A}_s(\gamma)$ *for all* γ *and* $\mathcal{A}^{\mathfrak{c}} := \mathcal{A}_s + \mathcal{A}_c^{\mathfrak{c}}$*.*

Remark 6.1.5. *We introduce the second-order expansion* $\mathcal{A}_c^{\mathfrak{er}}$ *of* $\mathcal{A}_c(\gamma)$ *around* γ_c *to better understand the effect in the multiscale expansion when the parameter of the linear operator is close to the deterministic Hopf bifurcation point.*

Assumption 6.1.6. *We assume that, for each* $\gamma \in \mathbb{R}$*,* $\mathcal{A}(\gamma)$ *generates an analytic compact* C_0 *semigroup* $\{e^{t\mathcal{A}(\gamma)}\}_{t\geq 0}$ *on* $\mathcal{H}$*, which also commute with the critical projection operator* P_c*. We further assume that*

(1) There exists some $M_s > 0$ *and* $c_s > 0$ *such that for all* $u \in \mathcal{H}_s$*,*

$$\|e^{t\mathcal{A}(\gamma)}u\| \leq Me^{-c_s t}\|u\|, \quad \forall t \geq 0.$$

(2) There exists $M > 0$ *and* $c \geq 0$ *such that for all* $u \in \mathcal{H}$ *and each* γ*,*

$$\|e^{t\mathcal{A}(\gamma)}u\| \leq Me^{ct}\|u\|, \quad \forall t \geq 0.$$

Proposition 6.1.7. *For each* $u \in \mathcal{H}$ *and for all* $\varepsilon \in (0,1)$*, there exist some* $C_{\mathfrak{q}} > 0$ *and* $C_{\mathfrak{er}} > 0$ *such that*

$$\langle \mathcal{A}_c^{\mathfrak{q}}u, u\rangle \leq C_{\mathfrak{q}}\|P_c u\|^2$$

and

$$\langle \mathcal{A}_c^{\mathfrak{er}}u, u\rangle \leq \varepsilon^2 C_{\mathfrak{er}}\|P_c u\|^2.$$

Proof. It is clear from the definition of projection that

$$\begin{aligned}\langle \mathcal{A}_c^{\mathfrak{q}} u, u\rangle &= \langle \mathcal{A}_c^{\mathfrak{q}}(P_c u + P_s u), P_c u + P_s u\rangle = \langle \mathcal{A}_c^{\mathfrak{q}}(P_c u), P_c u\rangle \\ &= \rho_c^{\mathfrak{q}}\langle \bar{v}, P_c u\rangle\langle v, P_c u\rangle + \bar{\rho}_c^{\mathfrak{q}}\langle v, P_c u\rangle\langle \bar{v}, P_c u\rangle \\ &\leq 2|a_c^{\mathfrak{q}}|^2 \cdot \|P_c u\|^2.\end{aligned} \tag{6.3}$$

The bound for $\langle \mathcal{A}_c^{\mathfrak{er}} u, u\rangle$ can be obtained in a similar way with by the Cauchy-Taylor expansion based on the additional analyticity of $\rho_{\pm 1}(\gamma)$ as in (2) of Assumption 6.1.1. □

For the same reason as in Chapter 5, we define the fractional power spaces w.r.t. $\mathrm{dom}(\mathcal{A}(\gamma))$. More details can be found in Appendix D.3.

Definition 6.1.8 (Fractional Power Space). *For $\alpha \in \mathbb{R}$, given the linear operator $\mathcal{A}(\gamma)$, define the interpolation fractional power (Hilbert) space [134] $\mathcal{H}_\alpha := \mathrm{dom}(\mathcal{A}^\alpha(\gamma))$ endowed with inner product $\langle u, v\rangle_\alpha = \langle \mathcal{A}^\alpha u, \mathcal{A}^\alpha v\rangle$ and corresponding induced norm $\|\cdot\|_\alpha := \|\mathcal{A}^\alpha \cdot\|$. Further more, we denote the dual space of $\mathcal{H}_\alpha$ by $\mathcal{H}_{-\alpha}$ w.r.t. the inner product in $\mathcal{H}$.*

Assumption 6.1.9. *We assume that given $\alpha \in (0, 1]$, there exists $\beta \in (\alpha - 1, \alpha]$ such that the mapping $F : (\mathcal{H}_\alpha)^3 \to \mathcal{H}_\beta$ is trilinear, symmetric, continuous. We use shorthand nations for $F(u) = F(u, u, u)$.*

We also assume that for all $u, v, w \in \mathcal{H}_c \setminus \{0\}$, we have

$$\langle F_c(u), u\rangle < 0, \tag{6.4}$$

$$\langle F_c(u, u, v), v\rangle < 0, \tag{6.5}$$

$$\langle F_c(u, v, w) - F_c(v), u\rangle \leq -C_0\|u\|^4 + C_1\|w\|^4 + C_2\|w\|^2\|v\|^2, \tag{6.6}$$

and that for all $\varsigma > 0$, there exists a $K > 0$ and $k \in [0, 1)$ such that for all $\varsigma > 0$,

$$\langle F(u + v), u\rangle_\alpha \leq -\varsigma\|u\|_\alpha^2 - k\langle \mathcal{A}(\gamma)u, u\rangle_\alpha + K\varsigma^2 + K\|v\|_\alpha^4, \quad \forall u, v \in \mathrm{dom}(\mathcal{A}(\gamma)). \tag{6.7}$$

Note that Assumption 6.1.9 implies that the nonlinear mapping behaves strongly dissipative in the critical subspace $\mathcal{H}_c$, where as the assumption on F in $\mathcal{H}$ is slightly weak, but still strong enough to guarantee nice properties of solutions to (6.1).

Now we move on to the assumptions on the last term in (6.1). Let $\mathcal{L}_2(E, K)$ denote the set of all Hilbert-Schmidt operators from E to K for any separable Hilbert spaces E and K. We write $\mathcal{L}_2(E)$ instead of $\mathcal{L}_2(E, E)$ for short. We denote by $\|\cdot\|_{\mathcal{L}_2(E,K)}$ the norm for Hilbert-Schmidt operators. If the spaces E and K are not emphasized, we also use the shorthand notation $\|\cdot\|_{\mathcal{L}_2}$.

Assumption 6.1.10. *Let $\mathcal{V}$ be a separable Hilbert space with orthonormal basis $\{\mathfrak{z}_k\}_{k\in\mathbb{Z}_0}$. We assume W is a $\mathcal{V}$-valued cylindrical Wiener process (see details in Definition D.2.2).*

Remark 6.1.11. *Given an orthonormal basis $\{\mathfrak{z}_k\}_{k\in\mathbb{Z}_0}$ of $\mathcal{V}$ and a set of i.i.d. Brownian motions $\{\beta_k\}$, in view of RKHS as in Appendix D, we can write*

$$W(t) = \sum_{k\in\mathbb{Z}^+} (\beta_k(t) + i\beta_{-k}(t))\mathfrak{z}_k + \sum_{k\in\mathbb{Z}^-} (\beta_{-k}(t) - i\beta_k(t))\mathfrak{z}_k. \tag{6.8}$$

Assumption 6.1.12. *Assume that, for each $\alpha \in (0, 1]$ and for all $u \in \mathcal{H}_\alpha$, $G(u)$ is a Hilbert-Schmidt operator from $\mathcal{V}$ to $\mathcal{H}_\alpha$, i.e., $G : \mathcal{H}_\alpha \to \mathcal{L}_2(\mathcal{V}, \mathcal{H}_\alpha)$. We further require that $G(u)$ is Fréchet-differentiable satisfying*

(1) there exists an $\ell_1 > 0$ s.t. $\|G(u)\|_{\mathcal{L}_2(\mathcal{V},\mathcal{H}_\alpha)} \le \ell_1\|u\|_\alpha$ for all $u \in \mathcal{H}_\alpha$;

(2) there exists an $\ell_2 > 0$ s.t. $\|G'(u)\cdot v\|_{\mathcal{L}_2(\mathcal{V},\mathcal{H}_\alpha)} \le \ell_2\|v\|_\alpha$ for all $v \in \mathcal{H}_\alpha$;

(3) $G''(u) = 0$ for all $u \in \mathcal{H}$.

Note that $G(u)$ is Hilbert-Schmidt operators from $\mathcal{V}$ to $\mathcal{H}_\alpha$, and we generally assume the commutative property between of $G(u)$ and id. Assuming W is a $\mathcal{V}$-valued cylindrical Wiener process is without loss of generality, we can view W as a $G(u)G^*(u)$-Wiener process in $\mathcal{H}_\alpha$, where $G(u)G^*(u)$ play the role of the covariance operator. If W is already a Q-Wiener process in $\mathcal{V}$, the correlation $G(u)QG(U)^*$ would make the process spatially smoother.

Remark 6.1.13. *Given an orthonormal basis $\{\mathfrak{z}_k\}_{k\in\mathbb{Z}_0}$ of $\mathcal{V}$, we can write*

$$G(u) = \sum_{j\in\mathbb{Z}_0}\sum_{k\in\mathbb{Z}_0} g_{jk}(u) e_j \otimes \mathfrak{z}_k,$$

where $g_{ik}(u) \in \mathbb{R}$ is the eigenvalue of the operator $G(u)$ for all $u \in \mathcal{H}_\alpha$ and for all $j, k \in \mathbb{Z}_0$.

Definition 6.1.14. *For any solution $u(t) \in \mathcal{H}$ to (6.1), we set $x(t) = \varepsilon^{-1}P_c u(\varepsilon^{-2}t)$ and $y(t) = \varepsilon^{-1}P_s u(\varepsilon^{-2}t)$. We further name $z(t) = \varepsilon^{-1}\langle\mathfrak{h}^*, u(\varepsilon^{-2}t)\rangle$, $\bar{z} = \varepsilon^{-1}\langle\bar{\mathfrak{h}}^*, u(\varepsilon^{-2}t)\rangle$ and $z_1 = \mathrm{Re}(z)$ as well as $z_2 = \mathrm{Im}(z)$.*

By the above notation, immediately we have $u(t) = \varepsilon x(\varepsilon^2 t) + \varepsilon y(\varepsilon^2 t) = \varepsilon z(\varepsilon^2 t)v + \varepsilon\bar{z}(\varepsilon^2)\bar{v} + \varepsilon y(\varepsilon^2 t)$. Due to the existence of the spectral gap, we decompose (6.1) in to the rescaled critical and fast-varying modes as follows:

$$dx = \mathcal{A}_c^{\mathfrak{q}} x dt + \mathcal{A}_c^{\mathfrak{er}} x dt + F_c(x+y)dt + G_c(x+y)dW_t, \tag{6.9a}$$
$$dy = \varepsilon^{-2}\mathcal{A}_s y dt + F_s(x+y)dt + G_s(x+y)dW_t, \tag{6.9b}$$

or equivalently,

$$dz = \rho_c^{\mathfrak{q}} z dt + \rho_c^{\mathfrak{er}} z dt + \langle \upsilon^*, F(x+y)\rangle dt + \langle \upsilon^*, G(x+y) dW_t\rangle, \tag{6.10a}$$
$$dy = \varepsilon^{-2}\mathcal{A}_s y dt + F_s(x+y)dt + G_s(x+y)dW_t, \tag{6.10b}$$

where the Wiener process W is obtained after being rescaled in both space and time, such that $\varepsilon W_{\varepsilon^{-2}t}$ has the same law with the old one. For simplicity, we keep the same notation.

6.2 Existence of Invariant Measures

Before proceeding to the approximation of solutions and invariant measures, we first show that the problem is well-posed under the assumptions.

Lemma 6.2.1. *Let Assumption 6.1.1, 6.1.6, 6.1.9 and 6.1.12 be satisfied. Let $u(0) = u_0 \in \mathcal{H}$ be the initial condition for (6.1) and $\varepsilon \in (0,1)$. Then there exists some $C_0 > 0$ such that*

$$\sup_{t\geq 0} \mathbf{E}\|u(t)\|_\alpha \leq C_0.$$

Moreover, if $\mathbf{E}\|u_0\|_\alpha$ is of order $\mathcal{O}(\varepsilon)$, then for any fixed $p > 0$ there exists some $C > 0$ such that

$$\sup_{t\geq 0} \mathbf{E}\|u(t)\|_\alpha^p \leq C\varepsilon^p.$$

Proof. Let $\mathcal{A}_n(\gamma) := n\mathcal{A}(\gamma)(nI - \mathcal{A}(\gamma))^{-1}$ be the Yosida approximation of $\mathcal{A}(\gamma)$ for each γ. By the property of F and G, the approximation equation

$$du_n = \mathcal{A}_n(\gamma)u_n dt + F(u_n)dt + \varepsilon G(u_n)dW_t, \quad u_n(0) = u_0.$$

Apply Itô's lemma to $\|u_n(t)\|_\alpha^2$ for sufficiently large n, we have

$$\begin{aligned} d\|u_n\|_\alpha^2 =& 2\langle \mathcal{A}_{s,n}u_n, u_n\rangle_\alpha dt + 2\langle \mathcal{A}_{c,n}^{\mathfrak{c}} u_n, u_n\rangle_\alpha dt + 2\langle F(u_n), u_n\rangle_\alpha dt + \varepsilon^2\|G(u_n)\|^2_{\mathcal{L}_2(\mathcal{V},\mathcal{H}_\alpha)}dt \\ &+ 2\varepsilon^2\langle \mathcal{A}_{c,n}^{\mathfrak{q}} u_n, u_n\rangle_\alpha dt + 2\varepsilon^2\langle \mathcal{A}_{c,n}^{\mathfrak{er}} u_n, u_n\rangle_\alpha dt + 2\varepsilon\langle u_n, G(u_n)dW_t\rangle_\alpha, \end{aligned}$$

which implies

$$\begin{aligned} \frac{d}{dt}\mathbf{E}\|u_n\|_\alpha^2 =& \mathbf{E}[2\langle \mathcal{A}_n^{\mathfrak{c}} u_n, u_n\rangle_\alpha + 2\varepsilon^2\langle \mathcal{A}_{c,n}^{\mathfrak{q}} u_n, u_n\rangle_\alpha + 2\varepsilon^2\langle \mathcal{A}_{c,n}^{\mathfrak{er}} u_n, u_n\rangle_\alpha] \\ &+ 2\mathbf{E}\langle F(u_n), u_n\rangle_\alpha + \varepsilon^2\mathbf{E}\|G(u_n)\|^2_{\mathcal{L}_2(\mathcal{V},\mathcal{H}_\alpha)}. \end{aligned} \tag{6.11}$$

Note that by Assumption 6.1.6 on the self-adjoint property of $\mathcal{A}$, using Definition 6.1.8 we are able to obtain the following bounds in a similar way as Proposition 6.1.7,

$$\langle \mathcal{A}_{c,n}^{\mathfrak{q}} u_n, u_n \rangle_\alpha \leq C_{\mathfrak{q}} \|u_n\|_\alpha^2, \quad \langle \mathcal{A}_{c,n}^{\mathfrak{er}} u_n, u_n \rangle_\alpha \leq \varepsilon^2 C_{\mathfrak{er}} \|u_n\|_\alpha^2 \tag{6.12}$$

for some $C_{\mathfrak{q}}, C_{\mathfrak{er}} > 0$. By Assumption 6.1.1, we have

$$\langle \mathcal{A}_n^{\mathfrak{c}} u_n, u_n \rangle_\alpha \leq 0. \tag{6.13}$$

Combining (6.11), (6.12), (6.13), Assumption 6.1.9 and 6.1.12, we have

$$\begin{aligned}
\frac{d}{dt}\mathbf{E}\|u_n\|_\alpha^2 \leq & \mathbf{E}[-\varsigma\|u_n\|_\alpha^2 + 2(1-k)\varepsilon^2\langle \mathcal{A}_{c,n}^{\mathfrak{q}} u_n, u_n\rangle_\alpha + K\varsigma^2 + \varepsilon^2\ell_1\|u_n\|_\alpha^2] \\
& + \varepsilon^2\mathbf{E}[2(1-k)\langle \mathcal{A}_{c,n}^{\mathfrak{er}} u_n, u_n\rangle_\alpha] \\
\leq & -\varsigma\mathbf{E}\|u_n\|_\alpha^2 + \varepsilon^2[2(1-k)C_{\mathfrak{q}} + \ell_1]\cdot\mathbf{E}\|u_n\|_\alpha^2 + K\varsigma^2 \\
& + 2\varepsilon^4(1-k)C_{\mathfrak{er}}\cdot\mathbf{E}\|u_n\|_\alpha^2
\end{aligned} \tag{6.14}$$

Now let $\tilde{\varsigma} := 2(1-k)C_{\mathfrak{q}} + \ell_1 + \varepsilon^2(1-k)C_{\mathfrak{er}}$, $\varsigma := 2\varepsilon^2\tilde{\varsigma}$, and $\tilde{C} := K\varsigma^2$, then $\varsigma > 0$ and $\tilde{C} > 0$. Therefore, (6.14) becomes

$$\frac{d}{dt}\mathbf{E}\|u_n\|_\alpha^2 \leq -\varepsilon^2\tilde{\varsigma}\|u_n\|_\alpha^2 + \varepsilon^4\tilde{C}.$$

Send n to infinity and rearrange, we have

$$\mathbf{E}\|u(t)\|_\alpha^2 \leq e^{-\varepsilon^2\tilde{\varsigma}t}\mathbf{E}\|u_0\|_\alpha^2 + \varepsilon^2\frac{\tilde{C}}{\tilde{\varsigma}}, \tag{6.15}$$

which completes the proof. □

Lemma 6.2.2. *Let all assumptions in Lemma 6.2.1 be satisfied. Fix any a time T and a $p \geq 2$, for any initial condition $u_0 \in \mathcal{H}_\alpha$, there exists some $C > 0$ such that*

$$\mathbf{E}\sup_{0\leq t\leq T}\|u(t)\|_\alpha^p \leq C.$$

Proof. For any $t \in [0, T]$, the mild solution is given as

$$u(t) = e^{t\mathcal{A}^{\mathfrak{c}}}u_0 + \int_0^t e^{(t-s)\mathcal{A}^{\mathfrak{c}}}[\varepsilon^2\mathcal{A}_c^{\mathfrak{q}}u(s) + \varepsilon^2\mathcal{A}_c^{\mathfrak{er}}u(s) + F(u(s))]ds + \varepsilon\int_0^t e^{(t-s)\mathcal{A}^{\mathfrak{c}}}G(u(s))dW_s. \tag{6.16}$$

Let $W^G_{\mathcal{A}^{\mathfrak{c}}}(t) := \varepsilon \int_0^t e^{(t-s)\mathcal{A}^{\mathfrak{c}}} G(u(s))dW_s$ denote the stochastic convolution. The bound for the stochastic convolution follows [43, Proposition 6.10, Proposition 7.3]. Indeed, by Lemma 6.2.1, there exists some $C > 0$ and $C' > 0$, such that

$$\int_0^T \mathbf{E}\|G(u(s))\|^p_{\mathcal{L}_2(\mathcal{V},\mathcal{H}_\alpha)} ds \le C \int_0^T \mathbf{E}\|u(s)\|^p_\alpha ds < C' < \infty.$$

Therefore $G(u(s))$ is $\mathcal{L}^2$-predictable and there exists constants $C_T > 0$ and $C'_T > 0$ such that

$$\mathbf{E} \sup_{0\le t\le T} \left\|W^G_{\mathcal{A}^{\mathfrak{c}}}(t)\right\|^p_\alpha \le \varepsilon^{2p} C_T \mathbf{E} \left(\int_0^t \|G(u(s))\|^p_{\mathcal{L}_2(\mathcal{V},\mathcal{H}_\alpha)} ds \right) \le \varepsilon^{2p} C'_T. \tag{6.17}$$

Let $v(t) := u(t) - W^G_{\mathcal{A}^{\mathfrak{c}}}(t) \in \mathcal{H}_\alpha$, then

$$\partial_t v = [\mathcal{A}^{\mathfrak{c}} v + \varepsilon^2 \mathcal{A}^{\mathfrak{q}}_c v + \varepsilon^2 \mathcal{A}^{\mathfrak{cr}}_c v + F(v + W^G_{\mathcal{A}^{\mathfrak{c}}}(t))]dt$$

and $\mathbf{E}\sup_{0\le t\le T} \|u(t)\|^p_\alpha \le \mathbf{E}\sup_{0\le t\le T} \|v(t)\|^p_\alpha + \mathbf{E}\sup_{0\le t\le T} \|W^G_{\mathcal{A}^{\mathfrak{c}}}(t)\|^p_\alpha$. Apply Itô' formula on $\|v\|^2_\alpha$ and use a similar trick as (6.11), we can show that there exist some $\tilde{\varsigma} > 0$ and $\tilde{C} > 0$ such that $\|v\|^2_\alpha \le e^{-2\varepsilon^2\tilde{\varsigma}t}\|u_0\|^2_\alpha + \tilde{C}(\left\|W^G_{\mathcal{A}^{\mathfrak{c}}}(t)\right\|^4_\alpha + \varepsilon^4)$. This together with the supreme bound for $W^G_{\mathcal{A}^{\mathfrak{c}}}(t)$ implies that $\mathbf{E}\sup_{0\le t\le T} \|u(t)\|^p_\alpha$ is bounded.

□

Corollary 6.2.3. *Let all assumptions in Lemma 6.2.1 be satisfied. For any $u_0 \in \mathcal{H}_\alpha$, the solution $u(t)$ to (6.1) is non-explosive.*

Proof. The conclusion follows by Chebychev inequality and Lemma 6.2.2. □

Proposition 6.2.4. *Let all assumptions in Lemma 6.2.1 be satisfied. The for any $u_0 \in \mathcal{H}_\alpha$, there exists at least one invariant measure to (6.1).*

Proof. Let $\Theta_t(u_0, \Gamma)$ be the transition function of (6.1), then by Lemma 6.2.2, there exists some sequence $T_n \to \infty$ such that the family of measures $\left\{\frac{1}{T_n}\int_0^{T_n} \Theta^*_t \delta_{u_0} dt\right\}_n$ is tight. The existence of invariant measure follows by Krylov–Bogoliubov's Theorem. □

Remark 6.2.5. *Note that due to the possible non-contraction of the linear operator $\mathcal{A}^{\mathfrak{c}}$, the system may not have a unique invariant measure, even though the nonlinearity has nice stability property. On the other hand, for $a^{\mathfrak{q}}_c \ll 0$, the system is strongly dissipative, which implies that the Dirac measure $\delta_{\mathbf{0}}$ is the unique invariant measure. The research focuses on the part where $a^{\mathfrak{q}}_c$ is in some neighborhood of 0 and the uniqueness of the invariant measure is uncertain.*

We next show some properties of the stable mode $P_c u(t)$ and the non-trivial (i.e., $u_0 \neq \mathbf{0}$) solution $u^\star(t)$ that is distributed by the invariant measure.

Lemma 6.2.6. *Let Assumption 6.1.1, 6.1.6, 6.1.9 and 6.1.12 be satisfied. Let $u(0) = u_0 \in \mathcal{H}_\alpha$ be the initial condition for (6.1) and $\varepsilon \in (0,1)$. Suppose additionally $\|P_s u_0\|_\alpha$ is of order $\mathcal{O}(\varepsilon^2)$, then for any fixed $p > 0$ there exists some $C > 0$ such that*

$$\sup_{t \geq 0} \mathbf{E}\|P_s u(t)\|_\alpha^p \leq C\varepsilon^{2p}.$$

Proof. Let $u_c(t) := P_c u(t)$ and $u_s(t) := P_s u(t)$ for all $t \geq 0$. Note that by the stability of $\mathcal{A}_s$, there exists some $c > 0$ such that $\langle \mathcal{A}_s u_s, u_s \rangle_\alpha \leq -c\|u_s\|_\alpha$. We use a similar trick in Lemma 6.2.1 and obtain

$$\frac{d}{dt}\mathbf{E}\|u_s\|_\alpha^2 = \mathbf{E}[2\langle \mathcal{A}_s u_s, u_s\rangle_\alpha + 2\langle F_s(u), u_s\rangle_\alpha] + \varepsilon^2 \mathbf{E}\|G_s(u)\|^2_{\mathcal{L}_2(\mathcal{V},\mathcal{H}_\alpha)}, \tag{6.18}$$

By Assumption 6.1.9,

$$\begin{aligned} \langle F_s(u_s + u_c), u_s\rangle_\alpha &\leq -\varsigma\|u_s\|_\alpha^2 - k\langle \mathcal{A}_s u_s, u_s\rangle_\alpha + K\varsigma^2 + K\|u_c\|_\alpha^4 \\ &\leq -\varsigma\|u_s\|_\alpha^2 - k\langle \mathcal{A}_s u_s, u_s\rangle_\alpha + K\varsigma^2 + C_1\varepsilon^4 \end{aligned} \tag{6.19}$$

for some $C_1 > 0$. By Assumption 6.1.12,

$$\varepsilon\mathbf{E}\|G_s(u)\|_{\mathcal{L}_2(\mathcal{V},\mathcal{H}_\alpha)} \leq \varepsilon\ell_1(\|u_s\|_\alpha + \|u_c\|_\alpha) \leq \varepsilon\ell_1\|u_s\|_\alpha + C_2\varepsilon^2 \tag{6.20}$$

for some $C_2 > 0$. Combine the above bounds and (6.18) and choose ς small enough to cancel the $\mathcal{O}(\varepsilon^2)$ terms, one can obtain some $\tilde{c}$ sufficiently close to c and some $\tilde{C} > 0$, such that

$$\frac{d}{dt}\mathbf{E}\|u_s\|_\alpha^2 \leq -2\tilde{c}\|u_n\|_\alpha^2 + \varepsilon^4\tilde{C}.$$

and thus,

$$\mathbf{E}\|u_s(t)\|_\alpha^2 \leq e^{-2\tilde{c}t}\mathbf{E}\|P_s u_0\|_\alpha^2 + \varepsilon^4\frac{\tilde{C}}{2\tilde{c}}. \tag{6.21}$$

The conclusion follows after this. □

Corollary 6.2.7. *Let all assumptions in Lemma 6.2.1 be satisfied. Then the solution $u^\star(t)$ that is distributed by the invariant measure has the following properties,*

(1) there exists some $C > 0$ such that $\mathbf{E}\|u^\star\|_\alpha^p \leq C\varepsilon^p$;

(2) there exists some $C_s > 0$ such that $\mathbf{E}\|P_s u^\star\|_\alpha^p \leq C_s \varepsilon^{2p}$.

Proof. The proof follows the second part of Lemma 6.2.1 as well as Lemma 6.2.6. One can use (6.15) and (6.21) and set $u_0 = u^\star$. The conclusion follows by sending t to infinity. □

Note that the above result is equivalent as $\mathbf{E}\|x^\star\|_\alpha^p \leq C$ and $\mathbf{E}\|y^\star\|_\alpha^p \leq C_s\varepsilon^p$, where $x^\star$ and $y^\star$ are stationary solutions for x and y in (6.9). Due to the effect of the multiplicative noise and the strong stability of $\mathcal{A}_s$, the stable mode $P_s u^\star$ (or $y^\star$) converges fast to a smaller neighborhood of $\mathbf{0}$ (recall that $u^\star = \varepsilon x^\star + \varepsilon y^\star$).

6.3 Primary Approximation of Solution

The ultimate goal is to build the connection between (6.9) and the top Lyapunov exponent $\lambda^{\mathfrak{q},\varepsilon} := \lambda(\gamma)$ for the linearized SPDE, where $\gamma = \gamma_c + \varepsilon^2\mathfrak{q}$. The purpose of this primary approximation in this section is to show that by dropping unnecessary terms, the errors would only contribute to the high order terms of the asymptotic expansion of $\lambda^{\mathfrak{q},\varepsilon}$. We intend to eventually show that as γ varies in the neighborhood of γ_c such that the corresponding $\lambda^{\mathfrak{q},\varepsilon}$ varies continuously from negative to positive, a new measure $\nu^{\mathfrak{q},\varepsilon}$ for (6.9) other than $\delta_{\mathbf{0}}$ will be generated, which is known as a dynamical bifurcation of (6.9). Since $\lambda^{\mathfrak{q},\varepsilon}$ will be approximated using the asymptotic expansion in Chapter 7, in this level of estimation, we do not drop too many terms in order to capture more nonlinear effects.

Remark 6.3.1. *Note that for numerical simulations of the density of $\nu^{\mathfrak{q},\varepsilon}$, we can conduct a simpler approximation of solutions to (6.9) by ignoring the nonlinear couplings through the cubic terms. However, this can only be used when the phenomenological bifurcation point of $\nu^{\mathfrak{q},\varepsilon}$ is determined such that the Fokker-Planck equation for this lower-dimensional approximation is well-posed.*

We aim to approximate the solution to (6.9) using the following truncated equation

$$d\tilde{x} = \mathcal{A}_c^{\mathfrak{q}}\tilde{x}dt + F_c(\tilde{x}+\tilde{y})dt + G_c(\tilde{x}+\tilde{y})dW_t, \tag{6.22a}$$

$$d\tilde{y} = \varepsilon^{-2}\mathcal{A}_s\tilde{y}dt + G_s(\tilde{x}+\tilde{y})dW_t \tag{6.22b}$$

with the same initial conditions. We denote by

$$\tilde{u}(t) := \varepsilon\tilde{x}(\varepsilon^2 t) + \varepsilon\tilde{y}(\varepsilon^2 t) \tag{6.23}$$

the approximation of $u(t)$. Note that the stable and critical modes of (6.22a) and (6.22b) are coupled in the noise terms as well as the critical projection of the nonlinearity. The approximated stable marginals are captured by a linear equation (6.22b) with a contraction semigroup, whereas the equation (6.22a) keep the same form as (6.9a) apart from the small error term $A_c^{\mathfrak{er}}\tilde{x}dt$. By a similar argument as in Section 6.2, the system (6.22) also has at least one invariant measure.

We claim that

$$\mathbf{E}\sup_{0\le t\le T}\|\tilde{y}(t)-y(t)\|_\alpha^p=\mathcal{O}(\varepsilon^{2p}),\quad \mathbf{E}\sup_{0\le t\le T}\|\tilde{x}(t)-x(t)\|^p=\mathcal{O}(\varepsilon^{2p}), \tag{6.24}$$

(which are of the same order as $\mathcal{A}_c^{\mathfrak{er}}x$, since $\mathbf{E}\sup_{0\le t\le T}\|\mathcal{A}_c^{\mathfrak{er}}x\|^p\le\varepsilon^{2p}C_{\mathfrak{er}}^p\sup_{0\le t\le T}\mathbf{E}\|x\|^p\le C\varepsilon^{2p}$) and consequently

$$\mathbf{E}\sup_{0\le t\le\varepsilon^{-2}T}\|\tilde{u}(t)-u(t)\|_\alpha^p=\mathcal{O}(\varepsilon^{3p}). \tag{6.25}$$

The following proposition provides the property for the approximated solution $\tilde{u}$, the proof of which follows the same procedure as in Lemma 6.2.1 and 6.2.2.

Proposition 6.3.2. *Let Assumption 6.1.1, 6.1.6, 6.1.9 and 6.1.12 be satisfied. Suppose $\|\tilde{u}_0\|_\alpha$ (or equivalently $\|u_0\|_\alpha$) is of order $\mathcal{O}(\varepsilon)$, then for any fixed $p>0$ there exists some $C_1>0$ such that*

$$\sup_{t\ge0}\mathbf{E}\|\tilde{u}(t)\|_\alpha^p\le C_1\varepsilon^p.$$

In addition, for any fixed time T and $p\ge2$, under the same initial condition as before, there exists some $C_2>0$ such that

$$\mathbf{E}\sup_{0\le t\le T}\|\tilde{u}(t)\|_\alpha^p\le C_2\varepsilon^p.$$

6.3.1 The Error Terms of the Approximation for the Stable Modes

To proceed, we first compare (6.9b) and (6.22b). Note that the solution to (6.9b) is given as

$$\begin{aligned}y(t)=&e^{\varepsilon^{-2}t\mathcal{A}_s}y(0)+\int_0^t e^{\varepsilon^{-2}(t-\sigma)\mathcal{A}_s}F_s(x(\sigma)+y(\sigma))d\sigma\\&+\int_0^t e^{\varepsilon^{-2}(t-\sigma)\mathcal{A}_s}G_s(x(\sigma)+y(\sigma))dW_\sigma\end{aligned} \tag{6.26}$$

We denote

$$I_F(t):=\int_0^t e^{\varepsilon^{-2}(t-\sigma)\mathcal{A}_s^{\mathfrak{c}}}F_s(x(\sigma)+y(\sigma))d\sigma$$

and

$$W_s^G(t):=\int_0^t e^{\varepsilon^{-2}(t-\sigma)\mathcal{A}_s^{\mathfrak{c}}}G_s(x(\sigma)+y(\sigma))dW_\sigma.$$

Furthermore, the solution to (6.22b) is given as

$$\tilde{y}(t) = e^{\varepsilon^{-2}t\mathcal{A}_s}\tilde{y}(0) + \int_0^t e^{\varepsilon^{-2}(t-\sigma)\mathcal{A}_s} G_s(\tilde{x}(\sigma) + \tilde{y}(\sigma))dW_\sigma. \tag{6.27}$$

Likewise, we denote $\widetilde{W_s^G}(t) := \int_0^t e^{\varepsilon^{-2}(t-\sigma)\mathcal{A}_s^\mathfrak{c}} G_s(\tilde{x}(\sigma) + \tilde{y}(\sigma))dW_\sigma$. Clearly, for $y(0) = \tilde{y}(0)$, we have

$$y(t) - \tilde{y}(t) = I_F(t) + W_s^G(t) - \widetilde{W_s^G}(t).$$

In particular, I_F has nothing to do with $\tilde{x}$ and $\tilde{y}$, whereas the comparison between $W_s^G(t)$ and $\widetilde{W_s^G}(t)$ is more subtle. Since the stochastic convolution terms involve mixed information of $\tilde{x}$ and $\tilde{y}$, which are slightly different from $x(t)$ and $y(t)$, we need an extra estimation of $\tilde{x}$ to obtain more precise error bounds for $\tilde{y}$.

The following lemma shows the property of I_F.

Lemma 6.3.3. *Given any $u(0) \in \mathcal{H}_\alpha$ with is of order $\mathcal{O}(\varepsilon)$, and some fixed time $T > 0$ and each $p > 0$, there exists a constant $C_F > 0$, such that*

$$\mathbf{E} \sup_{0 \le t \le T} \|I_F(t)\|_\alpha^p \le C_F \varepsilon^{2p}. \tag{6.28}$$

Proof. Note that $u(0) = \varepsilon(x(0) + y(0))$, by Lemma 6.2.1 and 6.2.6, we have $\mathbf{E}\sup_t \|x(t) + y(t)\|_\alpha^p \le C$ for some $C > 0$. Therefore,

$$\begin{aligned}
\mathbf{E} \sup_{0 \le t \le T} \|I_F(t)\|_\alpha^p &\le \mathbf{E} \sup_{0 \le t \le T} \left[\int_0^t \|e^{\varepsilon^{-2}(t-\sigma)\mathcal{A}_s} F_s(x(\sigma) + y(\sigma))\|_\alpha d\sigma \right]^p \\
&\le C_{\alpha,\beta}^p \varepsilon^{2p(\beta-\alpha)} \mathbf{E} \sup_{0 \le t \le T} \left[\int_0^t e^{-c\varepsilon^{-2}(t-\sigma)} (t-\sigma)^{\beta-\alpha} \|F_s(x(\sigma) + y(\sigma))\|_\beta d\sigma \right]^p \\
&\le C_{\alpha,\beta}^p \varepsilon^{2p(\beta-\alpha)} \mathbf{E} \sup_{0 \le t \le T} \left[\int_0^t e^{-c\varepsilon^{-2}(t-\sigma)} (t-\sigma)^{\beta-\alpha} \|(x(\sigma) + y(\sigma)\|_\alpha^3 d\sigma \right]^p \\
&\le C^{3p} \cdot C_{\alpha,\beta}^p \varepsilon^{2p(\beta-\alpha)} \sup_{0 \le t \le T} \left[\int_0^t e^{-c\varepsilon^{-2}(t-\sigma)} (t-\sigma)^{\beta-\alpha} d\sigma \right]^p \\
&\le C^{3p} \cdot C_{\alpha,\beta}^p \varepsilon^{2p(\beta-\alpha)} \sup_{0 \le t \le T} \left[\int_0^{-c\varepsilon^{-2}t} e^{-\sigma} \sigma^{\beta-\alpha} d\sigma \right]^p \le C_F \varepsilon^{2p},
\end{aligned} \tag{6.29}$$

where the second line is by the semigroup property given in Proposition D.3.3, the third line is by Assumption 6.1.9, and the last line is by change of variables. □

To better understand the effect of stochastic convolution in (6.9b) and (6.22b) under the contraction semigroup, we introduce the following lemma.

Lemma 6.3.4. *Given any $u(0) = \tilde{u}(0) \in \mathcal{H}_\alpha$ with is of order $\mathcal{O}(\varepsilon)$, and any fixed time $T > 0$ and fixed $p > 0$, there exist constants $C_w, \widetilde{C_w} > 0$ such that*

$$\mathbf{E} \sup_{0 \le t \le T} \|W_s^G(t)\|_\alpha^p \le C_w \varepsilon^p. \tag{6.30}$$

and

$$\mathbf{E} \sup_{0 \le t \le T} \|\widetilde{W_s^G(t)}\|_\alpha^p \le \widetilde{C_w} \varepsilon^p. \tag{6.31}$$

Proof. We show the proof for $\widetilde{W_s^G(t)}$ and the other should be similar. Note that there exists some $C > 0$ such that $\mathbf{E}\sup_{0 \le t \le T} \|\tilde{x}(t) + \tilde{y}(t)\|_\alpha \le C$ by the assumption and Proposition 6.3.2. Then,

$$\begin{aligned}
\mathbf{E} \sup_{0 \le t \le T} \|W_s^G(t)\|_\alpha^p &\le \mathbf{E} \sup_{0 \le t \le T} \left[\int_0^t \|e^{\varepsilon^{-2}(t-\sigma)\mathcal{A}_s} G_s(\tilde{x}(\sigma) + \tilde{y}(\sigma))\|_\alpha d\sigma \right]^p \\
&\le \ell_1^{2p} \mathbf{E} \sup_{0 \le t \le T} \left[\int_0^t e^{-2c\varepsilon^{-2}(t-\sigma)} \|\tilde{x}(\sigma) + \tilde{y}(\sigma)\|_\alpha^2 d\sigma \right]^{\frac{p}{2}} \\
&\le C^p \ell_1^{2p} \mathbf{E} \sup_{0 \le t \le T} \left[\int_0^t e^{-2c\varepsilon^{-2}(t-\sigma)} d\sigma \right]^{\frac{p}{2}} \le \widetilde{C_w} \varepsilon^p.
\end{aligned} \tag{6.32}$$

□

Corollary 6.3.5. *Given any $\tilde{u}(0) \in \mathcal{H}_\alpha$ with is of order $\mathcal{O}(\varepsilon)$ and $P_s\tilde{u}(0)$ of order $\mathcal{O}(\varepsilon^2)$. For some fixed time $T > 0$ and fixed p, there exist constants $C_c, C_s > 0$ such that*

$$\mathbf{E} \sup_{0 \le t \le T} \|\tilde{x}(t)\|^p \le C_c, \quad \mathbf{E} \sup_{0 \le t \le T} \|\tilde{y}(t)\|_\alpha^p \le C_s \varepsilon^p. \tag{6.33}$$

Proof. Since $\tilde{y}$ is driven by a linear equation, by the contraction property of $\mathcal{A}_s$, we have

$$\|\tilde{y}(t)\|_\alpha^p \le e^{-\varepsilon^{-2}ct} \|\tilde{y}(0)\|_\alpha^p + \|\widetilde{W_s^G(t)}\|_\alpha. \tag{6.34}$$

The conclusion follows by Lemma 6.3.4. The estimation for $\tilde{x}$ follows a similar way in Lemma 6.2.1 and 6.2.2. □

Note that the bounds given in Corollary 6.3.5 imply that the stochastic convolution generated by $\mathcal{A}_s$ only contributes a smaller order of value compared to the critical mode. Intuitively, if one can show that $\mathbf{E}\sup_{0\le t\le T}\|\tilde{x}(t)-x(t)+\tilde{y}(t)-y(t)\|_\alpha^p$ is of order $\mathcal{O}(\varepsilon^p)$, by the Lipschitz continuity of G as in (2) of Assumption 6.1.12 and by a similar argument as in (6.32), we can obtain that $\mathbf{E}\sup_{0\le t\le T}\|W_s^G(t)-\widetilde{W_s^G}(t)\|_\alpha^p$ is of order $\mathcal{O}(\varepsilon^{2p})$. We will formally complete this part of proof after we investigate the brief properties of $\tilde{x}$.

6.3.2 A Rough Estimation of the Error in the Critical Mode

We show in this subsection that the error between (6.9a) and (6.22a), i.e. $\mathbf{E}\sup_{0\le t\le T}\|\tilde{x}(t)-x(t)\|_\alpha^p$, is at most of order $\mathcal{O}(\varepsilon^p)$. To do this and to avoid complicated interactions between $\tilde{x}$, x, $\tilde{y}$ and y, we introduce an auxiliary equation

$$d\hat{x}=\mathcal{A}_c^{\mathfrak{q}}\hat{x}dt+F_c(\hat{x})dt+G_c(\hat{x})dW_t,\quad \hat{x}_0=x_0=\tilde{x}_0 \tag{6.35}$$

and show that $\mathbf{E}\sup_{0\le t\le T}\|\tilde{x}(t)-\hat{x}(t)\|^p$ and $\mathbf{E}\sup_{0\le t\le T}\|\hat{x}(t)-x(t)\|^p$ are both of order $\mathcal{O}(\varepsilon^p)$, where x is the solution to (6.9a). We provide the proof for $\mathbf{E}\sup_{0\le t\le T}\|\hat{x}(t)-x(t)\|^p$ and the other one should be similar.

Remark 6.3.6. *The over simplified critical model* (6.35) *provides error* $\mathbf{E}\sup_{0\le t\le T}\|\tilde{x}(t)-\hat{x}(t)\|^p$ *(resp.* $\mathbf{E}\sup_{0\le t\le T}\|\hat{x}(t)-x(t)\|^p$*) that is of same order as the stable modes* y *(resp.* $\tilde{y}$*) and of a smaller order than the linear perturbation* $\mathcal{A}_c^{\mathfrak{er}}x$ *(resp.* $\mathcal{A}_c^{\mathfrak{er}}\tilde{x}$*). As* ε *goes to* 0*, the error term affects the local behavior around the bifurcation parameter more than its own second order expansion. To see more accurately how the stable mode and the multiplicative noise influence the dynamical behavior near* γ_c*, we do not apply* (6.35) *to study the stochastic Hopf bifurcation. However, when the system is away from the dynamical bifurcation point, one can use* (6.35) *to simulate the invariant measure.*

We rewrite the solution to (6.9a) as

$$x(t)=x(0)+\int_0^t\mathcal{A}_c^{\mathfrak{q}}x(\sigma)d\sigma+\int_0^t F_c(x(\sigma))d\sigma+\int_0^t G_c(x(\sigma))dW_\sigma+R(t),$$

where the truncated error term is given as

$$\begin{aligned}R(t)=&\int_0^t\mathcal{A}_c^{\mathfrak{er}}x(\sigma)d\sigma+\int_0^t F_c(y(\sigma))d\sigma\\&+3\int_0^t F_c(x(\sigma),y(\sigma),y(\sigma))d\sigma+3\int_0^t F_c(x(\sigma),x(\sigma),y(\sigma))d\sigma\\&+\int_0^t[G_c(x(\sigma)+y(\sigma))-G_c(x(\sigma))]dW_\sigma.\end{aligned} \tag{6.36}$$

Let $\mathscr{R} = x - \hat{x}$. Then,

$$\mathscr{R}(t) = R(t) + \int_0^t \mathcal{A}_c^{\mathfrak{q}} \mathscr{R}(\sigma) d\sigma + \int_0^t F_c(x(\sigma)) d\sigma - \int_0^t F_c(\hat{x}(\sigma)) d\sigma + \int_0^t G_c(\mathscr{R}(\sigma)) dW_\sigma. \quad (6.37)$$

We first evaluate $\mathbf{E} \sup_{0 \le t \le T} \|R(t)\|^p$ and then $\mathbf{E} \sup_{0 \le t \le T} \|\mathscr{R}(t)\|^p$.

Lemma 6.3.7. *Given any $x(0) \in \mathcal{H}_\alpha$ of order $\mathcal{O}(1)$. For any fixed time $T > 0$ and fixed $p > 0$, there exist constants $C > 0$ such that*

$$\mathbf{E} \sup_{0 \le t \le T} \| \int_0^t \mathcal{A}_c^{\mathfrak{er}} x(\sigma) d\sigma \|_\alpha^p \le C \varepsilon^{2p}. \quad (6.38)$$

Proof. The above bound can be verified by the boundedness of the operator $\mathcal{A}_c^{\mathfrak{er}}$ as well as the property of $\mathbf{E} \sup_{0 \le t \le T} \|x(t)\|_\alpha^p$. □

Lemma 6.3.8. *Given any $x(0) \in \mathcal{H}_\alpha$ of order $\mathcal{O}(1)$ and $y(0) \in \mathcal{H}_\alpha$ of order $\mathcal{O}(\varepsilon)$. For any fixed time $T > 0$ and fixed $p > 0$, there exist some constant $C > 0$ such that*

$$\mathbf{E} \sup_{0 \le t \le T} \left\| \int_0^t F_c(x(\sigma), y(\sigma), y(\sigma)) d\sigma \right\|_\alpha^p \le C \varepsilon^{2p}. \quad (6.39)$$

Proof. The expansion is given as

$$\begin{aligned} \int_0^t F_c(x(\sigma), y(\sigma), y(\sigma) d\sigma = & \int_0^t F_c(x(\sigma), e^{\varepsilon^{-2}(t-\sigma)\mathcal{A}_s} y(0), e^{\varepsilon^{-2}(t-\sigma)\mathcal{A}_s} y(0)) d\sigma \\ & + \int_0^t F_c(x(\sigma), I_F(\sigma), I_F(\sigma)) d\sigma \\ & + \int_0^t F_c(x(\sigma), W_s^G(\sigma), W_s^G(\sigma)) d\sigma \\ & + 2 \int_0^t F_c(x(\sigma), e^{\varepsilon^{-2}(t-\sigma)\mathcal{A}_s} y(0), I_F(\sigma)) d\sigma \\ & + 2 \int_0^t F_c(x(\sigma), e^{\varepsilon^{-2}(t-\sigma)\mathcal{A}_s} y(0), W_s^G(\sigma)) d\sigma \\ & + 2 \int_0^t F_c(x(\sigma), I_F(\sigma), W_s^G(\sigma)) d\sigma \\ := & \sum_{i=1}^6 J^{(i)}(t). \end{aligned} \quad (6.40)$$

Note that,

$$\begin{aligned}\|J^{(1)}(t)\|_\alpha \leq \hat{C}_1\|J^{(1)}(t)\|_\beta &\leq \hat{C}_1\int_0^t \|F_c(x(\sigma), e^{\varepsilon^{-2}(t-\sigma)\mathcal{A}_s}y(0), e^{\varepsilon^{-2}(t-\sigma)\mathcal{A}_s}y(0))\|_\beta d\sigma \\ &\leq \hat{C}_1\|y(0)\|_\alpha^2 \sup_{0\leq t\leq T}\|x(t)\|_\alpha \int_0^t e^{2c_s\varepsilon^{-2}(t-\sigma)}d\sigma \\ &\leq \tilde{C}_1\varepsilon^2\|y(0)\|_\alpha^2 \sup_{0\leq t\leq T}\|x(t)\|_\alpha.\end{aligned} \tag{6.41}$$

Hence, by the property of $y(0)$ and $\mathbf{E}\sup_{0\leq t\leq T}\|x(t)\|_\alpha^p$, there exists somce $C_1 > 0$ such that

$$\mathbf{E}\sup_{0\leq t\leq T}\|J^{(1)}\|_\alpha^p \leq C_1\varepsilon^{4p}.$$

As for the bound for $J^{(2)}$, we have

$$\begin{aligned}\|J^{(2)}(t)\|_\alpha &\leq \hat{C}_2\|J^{(2)}(t)\|_\beta \\ &\leq \hat{C}_2\int_0^t \|F_c(x(\sigma), I_F(\sigma), I_F(\sigma))\|_\beta d\sigma \\ &\leq \tilde{C}_2\|I_F(t)\|_\alpha^2 \sup_{0\leq t\leq T}\|x(t)\|_\alpha.\end{aligned} \tag{6.42}$$

Note that

$$\begin{aligned}\|I_F(t)\|_\alpha &\leq \int_0^t \|e^{\varepsilon^{-2}(t-\sigma)\mathcal{A}_s}F_s(x(\sigma)+y(\sigma))\|_\alpha d\sigma \\ &\leq C_{\alpha,\beta}\varepsilon^{2(\beta-\alpha)}\int_0^t e^{-c\varepsilon^{-2}(t-\sigma)}(t-\sigma)^{\beta-\alpha}\|F_s(x(\sigma)+y(\sigma))\|_\beta d\sigma \\ &\leq C_{\alpha,\beta}\varepsilon^{2(\beta-\alpha)}\int_0^t e^{-c\varepsilon^{-2}(t-\sigma)}(t-\sigma)^{\beta-\alpha}\|(x(\sigma)+y(\sigma)\|_\alpha^3 d\sigma \\ &\leq \sup_{0\leq t\leq T}\|x(t)+y(t)\|_\alpha^3\cdot C_{\alpha,\beta}\varepsilon^{2(\beta-\alpha)}\int_0^t e^{-c\varepsilon^{-2}(t-\sigma)}(t-\sigma)^{\beta-\alpha}d\sigma \\ &\leq \varepsilon^2 C_{\alpha,\beta}\left(\sup_{0\leq t\leq T}\|x(t)\|_\alpha^3 + \sup_{0\leq t\leq T}\|y(t)\|_\alpha^3\right).\end{aligned} \tag{6.43}$$

Combining (6.41) and (6.43) we have

$$\|J^{(2)}(t)\|_\alpha \leq C_2\varepsilon^2\left(\sup_{0\leq t\leq T}\|x(t)\|_\alpha^3 + \sup_{0\leq t\leq T}\|y(t)\|_\alpha^3\right)^3 \tag{6.44}$$

Therefore, by Lemma 6.3.3 and the property of x and y, we obtain

$$\mathbf{E} \sup_{0\le t\le T} \|J^{(2)}\|_\alpha^p \le C_2\varepsilon^{4p}.$$

We proceed by a similar method for the rest of $J^{(i)}$'s and obtain

$$\mathbf{E} \sup_{0\le t\le T} \|J^{(3)}\|_\alpha^p \le C_3\varepsilon^{2p}, \quad \mathbf{E} \sup_{0\le t\le T} \|J^{(4)}\|_\alpha^p \le C_4\varepsilon^{4p},$$
$$\mathbf{E} \sup_{0\le t\le T} \|J^{(5)}\|_\alpha^p \le C_5\varepsilon^{2p}, \quad \mathbf{E} \sup_{0\le t\le T} \|J^{(6)}\|_\alpha^p \le C_6\varepsilon^{2p}.$$

□

By a similar approach, we obtain the following bounds.

Lemma 6.3.9. *Let the assumptions in Lemma 6.3.8 be satisfied. For any fixed time $T > 0$ and fixed $p > 0$, there exist some constant $C > 0$ such that*

$$\mathbf{E} \sup_{0\le t\le T} \left\| \int_0^t F_c(x(\sigma), x(\sigma), y(\sigma))d\sigma \right\|_\alpha^p \le C\varepsilon^p. \tag{6.45}$$

Lemma 6.3.10. *Let the assumptions in Lemma 6.3.8 be satisfied. For any fixed time $T > 0$ and fixed $p > 0$, there exist some constant $C > 0$ such that*

$$\mathbf{E} \sup_{0\le t\le T} \left\| \int_0^t F_c(y(\sigma))d\sigma \right\|_\alpha^p \le C\varepsilon^{3p}. \tag{6.46}$$

We now evaluate the last term in $R(t)$ (recall (6.36)).

Lemma 6.3.11. *Let the assumptions in Lemma 6.3.8 be satisfied. For any fixed time $T > 0$ and fixed $p > 0$, there exist some constant $C > 0$ such that*

$$\mathbf{E} \sup_{0\le t\le T} \left\| \int_0^t [G_c(x(\sigma) + y(\sigma)) - G_c(x(\sigma))]dW_\sigma \right\|_\alpha^p \le C\varepsilon^p. \tag{6.47}$$

Proof. By Burkholder–Davis–Gundy inequality,

$$
\begin{aligned}
&\mathbf{E}\sup_{0\le t\le T}\left\|\int_0^t [G_c(x(\sigma)+y(\sigma))-G_c(x(\sigma))]dW_\sigma\right\|_\alpha^p\\
\le& C'\mathbf{E}\left[\int_0^T \|G_c(x(\sigma)+y(\sigma))-G_c(x(\sigma))\|^2_{\mathcal{L}_2(\mathcal{V},\mathcal{H}_\alpha)}d\sigma\right]^{\frac{p}{2}}\\
\le& C'\mathbf{E}\left[\int_0^T \|G(x(\sigma)+y(\sigma))-G(x(\sigma))\|^2_{\mathcal{L}_2(\mathcal{V},\mathcal{H}_\alpha)}d\sigma\right]^{\frac{p}{2}}\\
\le& C'\mathbf{E}\left[\int_0^T \ell_2^2\|y(\sigma)\|_\alpha^2 d\sigma\right]^{\frac{p}{2}}\\
\le& C_p\mathbf{E}\left[\int_0^T (\|e^{\varepsilon^{-2}(t-\sigma)\mathcal{A}_s}y(0)\|_\alpha^2+\|I_F\|_\alpha^2+\|W_s^G\|_\alpha^2)d\sigma\right]^{\frac{p}{2}}\\
\le& C\mathbf{E}\left[(\varepsilon^p\|y(0)\|_\alpha^p+\sup_{0\le t\le T}\|I_F\|_\alpha^p+\sup_{0\le t\le T}\|W_s^G\|_\alpha^p)\right],
\end{aligned}
\tag{6.48}
$$

where we have used Hölder's inequality in the last line. The conclusion follows immediately by the properties verified in Section 6.3.1. □

Combining the bounds derived from Lemma 6.3.7 to 6.3.11, we obtain the bound for the truncation error $R(t)$.

Proposition 6.3.12. *Let the assumptions in Lemma 6.3.8 be satisfied. For any fixed time $T>0$ and fixed $p>0$, there exist some constant $C>0$ such that*

$$\mathbf{E}\sup_{0\le t\le T}\|R(t)\|_\alpha^p\le C\varepsilon^p. \tag{6.49}$$

We then proceed to derive the bound for $\mathscr{R}(t)$. However, since R is not Itô differentiable, we apply the method provided in [27, Lemma 4.9].

Proposition 6.3.13. *Let the assumptions in Lemma 6.3.8 be satisfied. For any fixed time $T>0$ and fixed $p>0$, there exist some constant $C>0$ such that*

$$\mathbf{E}\sup_{0\le t\le T}\|\mathscr{R}(t)\|_\alpha^p\le C\varepsilon^p. \tag{6.50}$$

Proof. Let $h = \mathscr{R} + R$, then

$$h(t) = \int_0^t \mathcal{A}_c^{\mathfrak{q}}(h(\sigma) - R(\sigma))d\sigma + \int_0^t F_c(\hat{x}(\sigma) - h(\sigma) + R(\sigma))d\sigma \\ - \int_0^t F_c(\hat{x}(\sigma))d\sigma + \int_0^t G_c(h(\sigma) - R(\sigma))dW_\sigma. \tag{6.51}$$

Let $f(\cdot) = \|\cdot\|^p$, then by [27, Lemma 4.9], we have

$$\operatorname{tr}[f''(h(\sigma))G_c(h(\sigma)-R(\sigma))G_c(h(\sigma)-R(\sigma))^*] \leq Cp(p-1)\|h(\sigma)\|^{p-2}\|h(\sigma)-R(\sigma)\|^2. \tag{6.52}$$

Applying Itô's formula to $\|\cdot\|^p$ and (6.52), we then have

$$\begin{aligned} \|h(t)\|^p \leq & p\int_0^t \|h(\sigma)\|^{p-2}\langle \mathcal{A}_c^{\mathfrak{q}}(h(\sigma) - R(\sigma)), h(\sigma)\rangle d\sigma \\ & + p\int_0^t \|h(\sigma)\|^{p-2}\langle F_c(\hat{x}(\sigma) - h(\sigma) + R(\sigma)) - F_c(\hat{x}(\sigma)), h(\sigma)\rangle d\sigma \\ & + p\int_0^t \|h(\sigma)\|^{p-2}\langle G_c(h(\sigma) - R(\sigma)), h(\sigma)\rangle d\sigma \\ & + \frac{1}{2}C'p(p-1)\int_0^t \|h(\sigma)\|^{p-2}\|h(\sigma) - R(\sigma)\|^2 d\sigma \\ \leq & C\int_0^t \|h(\sigma)\|^p d\sigma + C\int_0^t \|h(\sigma)\|^{p-1}\|R(\sigma)\| d\sigma \\ & + C\int_0^t \|h(\sigma)\|^{p-2}\|R(\sigma)\|^2 d\sigma + C\int_0^t \|h(\sigma)\|^{p-2}\|R(\sigma)\|^4 d\sigma \\ & + C\int_0^t \|h(\sigma)\|^{p-2}\|R(\sigma)\|^2\|y(\sigma)\|^2 d\sigma \\ & + p\int_0^t \|h(\sigma)\|^{p-2}\langle G_c(h(\sigma) - R(\sigma)), h(\sigma)\rangle d\sigma, \end{aligned} \tag{6.53}$$

where we have used Cauchy-Schwarz inequality and Assumption 6.1.9 for F_c in the second inequality. By Young's inequality (for products), we obtain

$$\begin{aligned} \|h(t)\|^p \leq & C\int_0^t \|h(\sigma)\|^p d\sigma + C\int_0^t \|R(\sigma)\|^p d\sigma + C\int_0^t \|R(\sigma)\|^{2p} d\sigma \\ \leq & C\int_0^t \|R(\sigma)\|^p\|y(\sigma)\|^p d\sigma + p\int_0^t \|h(\sigma)\|^{p-2}\langle G_c(h(\sigma) - R(\sigma)), h(\sigma)\rangle d\sigma, \end{aligned} \tag{6.54}$$

where the supreme bound for the last term can be obtained by using Burkholder–Davis–Gundy inequality and [27, Eq.(33)] on $[0, T]$,

$$
\begin{aligned}
&\mathbf{E} \sup_{0\le t\le T} \left| p\int_0^t \|h(\sigma)\|^{p-2}\langle G_c(h(\sigma)-R(\sigma)), h(\sigma)\rangle d\sigma \right| \\
\le & C\mathbf{E}\left[\int_0^T (\|h(\sigma)\|^{2p} + \|h(\sigma)\|^{2p-2}\|R(\sigma)\|^2)\right]^{\frac{1}{2}} \\
\le & \frac{1}{2}\mathbf{E} \sup_{0\le t\le T}\|h(t)\|^p + C\int_0^T \mathbf{E}\sup_{0\le\sigma\le T}\|h(\sigma)\|^p d\sigma + C\mathbf{E}\left[\int_0^T \|R(\sigma)\|^{2p}d\sigma\right]^{\frac{1}{2}}.
\end{aligned}
\tag{6.55}
$$

Combining (6.53) and (6.55), taking the supreme norm, we have

$$
\mathbf{E}\sup_{0\le t\le T}\|h(t)\|^p \le C\int_0^T \mathbf{E}\sup_{0\le\sigma\le T}\|h(\sigma)\|^p d\sigma + C\varepsilon^p. \tag{6.56}
$$

Therefore, by Grownwall's inequality, we have

$$
\mathbf{E}\sup_{0\le t\le T}\|h(t)\|^p \le C\varepsilon^p.
$$

The conclusion follows by a triangle inequality argument on $\mathscr{R} = h - R$. □

Remark 6.3.14. *As shown in the above proof, the bound for* $\mathbf{E}\sup_{0\le t\le T}\|h(t)\|^p$ *is dominated by the estimation of* $\mathbf{E}\sup_{0\le t\le T}\|R(t)\|^p$. *The order of error* $\mathscr{R}$ *should be consistent with the order of* R.

6.3.3 Final Estimation of Errors

We first show the estimation of $\mathbf{E}\sup_{0\le t\le T}\|\tilde{y}(t) - y(t)\|_\alpha^p$ in (6.24).

Proof. Let $\mathfrak{R}^y := y - \tilde{y}$, let $\mathfrak{R}^x := x - \tilde{x}$ then

$$
\mathfrak{R}^y(t) = I_F(t) + \int_0^t e^{\varepsilon^{-2}(t-\sigma)\mathcal{A}_s} G_s(\mathfrak{R}^x(\sigma) + \mathfrak{R}^y(\sigma))dW_\sigma. \tag{6.57}
$$

Note that

$$
\begin{aligned}
&\| \int_0^t e^{\varepsilon^{-2}(t-\sigma)\mathcal{A}_s} G_s(\mathfrak{R}^x(\sigma) + \mathfrak{R}^y(\sigma)) dW_\sigma \|_\alpha^p \\
\leq &\left[\int_0^t \| e^{\varepsilon^{-2}(t-\sigma)\mathcal{A}_s} G_s(\mathfrak{R}^x(\sigma) + \mathfrak{R}^y(\sigma)) \|_\alpha dW_\sigma \right]^p \\
\leq &\ell_1^p \sup_{0\leq t\leq T} \|\mathfrak{R}^x(t) + \mathfrak{R}^y(t)\|_\alpha^p \left[\int_0^t e^{-2c\varepsilon^{-2}(t-\sigma)} d\sigma \right]^{\frac{p}{2}} \\
\leq &C'\varepsilon^p (\sup_{0\leq t\leq T} \|\mathfrak{R}^x(t)\|_\alpha^p + \sup_{0\leq t\leq T} \|\mathfrak{R}^y(t)\|_\alpha^p) \\
\leq &C'\varepsilon^p (\sup_{0\leq t\leq T} \|\mathfrak{R}^x(t)\|_\alpha^p + \sup_{0\leq t\leq T} \|y(t)\|_\alpha^p) + \sup_{0\leq t\leq T} \|\tilde{y}(t)\|_\alpha^p).
\end{aligned}
\tag{6.58}
$$

We take the supreme norm on both sides of (6.57), use the results in Section 6.3.1, and the triangle inequality that $\sup_{0\leq t\leq T} \|\mathfrak{R}^x(t)\|_\alpha \leq \sup_{0\leq t\leq T} \|x - \hat{x}\|_\alpha + \sup_{0\leq t\leq T} \|x - \tilde{x}\|_\alpha$, as well as the result in Section 6.3.2, we then have

$$
\mathbf{E} \sup_{0\leq t\leq T} \|\mathfrak{R}^y(t)\|_\alpha^p \leq C\varepsilon^{2p}. \tag{6.59}
$$

□

We then show the improved bound for $\mathbf{E} \sup_{0\leq t\leq T} \|\tilde{x}(t) - x(t)\|^p$ given the information of $\mathbf{E} \sup_{0\leq t\leq T} \|\tilde{y}(t) - y(t)\|^p$ other than using the triangle inequality that was used above. Note that we can rewrite the solution to (6.9a) as

$$
x(t) = x(0) + \int_0^t \mathcal{A}_c^{\mathfrak{q}} x(\sigma) d\sigma + \int_0^t F_c(x(\sigma) + \tilde{y}(\sigma)) d\sigma + \int_0^t G_c(x(\sigma) + \tilde{y}(\sigma)) dW_\sigma + \tilde{R}(t),
$$

where the truncated error term is given as

$$
\begin{aligned}
\tilde{R}(t) = &\int_0^t \mathcal{A}_c^{\mathfrak{er}} x(\sigma) d\sigma + \int_0^t F_c(\mathfrak{R}^y(\sigma)) d\sigma \\
&+ 3 \int_0^t F_c(x(\sigma) + \tilde{y}(\sigma), \mathfrak{R}^y(\sigma), \mathfrak{R}^y(\sigma)) d\sigma \\
&+ 3 \int_0^t F_c(x(\sigma) + \tilde{y}(\sigma), x(\sigma) + \tilde{y}(\sigma), \mathfrak{R}^y(\sigma)) d\sigma \\
&+ \int_0^t [G_c(x(\sigma) + y(\sigma)) - G_c(x(\sigma) + \tilde{y}(\sigma))] dW_\sigma.
\end{aligned}
\tag{6.60}
$$

By the exact procedure from Lemma 6.3.7 to 6.3.11, we obtain the bound for the truncation error $\tilde{R}(t)$.

Proposition 6.3.15. *Let the assumptions in Lemma 6.3.8 be satisfied. For any fixed time $T > 0$ and fixed $p > 0$, there exist some constant $C > 0$ such that*

$$\mathbf{E} \sup_{0 \le t \le T} \|\tilde{R}(t)\|_{\alpha}^{p} \le C\varepsilon^{2p}. \tag{6.61}$$

The improved version of $\mathfrak{R}^x$ based on the truncated error $\tilde{R}$ is then

$$\begin{aligned} \mathfrak{R}^x(t) = \tilde{R}(t) &+ \int_0^t \mathcal{A}_c^{\mathfrak{q}} \mathfrak{R}^x(\sigma) d\sigma \\ &+ \int_0^t F_c(x(\sigma) + \tilde{y}(\sigma)) d\sigma - \int_0^t F_c(\tilde{x}(\sigma) + \tilde{y}(\sigma)) d\sigma \\ &+ \int_0^t G_c(\mathfrak{R}^x(\sigma)) dW_\sigma. \end{aligned} \tag{6.62}$$

We can define $\tilde{h} = \mathfrak{R}^x + \tilde{R}$ as in Proposition 6.3.13, then

$$\begin{aligned} \tilde{h}(t) = &\int_0^t \mathcal{A}_c^{\mathfrak{q}}(\tilde{h}(\sigma) - \tilde{R}(\sigma)) d\sigma + \int_0^t F_c(\tilde{x}(\sigma) + \tilde{y}(\sigma) + \tilde{h}(\sigma) - \tilde{R}(\sigma)) d\sigma \\ &- \int_0^t F_c(\tilde{x}(\sigma) + \tilde{y}(\sigma)) d\sigma + \int_0^t G_c(\tilde{h}(\sigma) - \tilde{R}(\sigma)) dW_\sigma. \end{aligned} \tag{6.63}$$

We proceed by the same argument as in Proposition 6.3.13 and obtain the second estimation in (6.24) as

$$\mathbf{E} \sup_{0 \le t \le T} \|\mathfrak{R}^x(t)\|_{\alpha}^{p} \le C\varepsilon^{2p}, \tag{6.64}$$

for each fixed $T > 0$ and $p > 0$.

6.4 Error Estimates for Invariant Measures

In this section, we estimate the error of the invariant measure for the approximated solutions to (6.22). We proceed by taking advantages of the results from Section 6.3.

We first reiterate the bounded Lipschitz distance and the total variation distance as the probability metrics on the space of probability measures. More details can be found in Appendix E. Given the state space $\mathcal{H}_\alpha$, we denote by $\mathfrak{P}(\mathcal{H}_\alpha)$ the space of probability measures.

Definition 6.4.1. *Let $\mu, \nu \in \mathfrak{P}(\mathcal{H}_\alpha)$, the bounded-Lipschitz distance is defined by*

$$\|\mu - \nu\|_{\mathrm{L}} := \sup_{\|h\|_{\mathrm{BL}} \le 1} \left\{ \left| \int_{\mathcal{H}_\alpha} h(x) d\mu(x) - \int_{\mathcal{H}_\alpha} h(x) d\nu(x) \right| \right\},$$

where the bounded Lipschitz metric is defined by

$$\|h\|_{\mathrm{BL}} = \|h\|_\infty + \sup_{\mathfrak{x} \neq \mathfrak{y}} \frac{|h(\mathfrak{y}) - h(\mathfrak{x})|}{\|\mathfrak{y} - \mathfrak{x}\|_\alpha}.$$

Definition 6.4.2. *Let $\mu, \nu \in \mathfrak{P}(\mathcal{H}_\alpha)$, the total variation distance is defined by*

$$\|\mu - \nu\|_{\mathrm{TV}} := \sup_{\|h\|_\infty \le 1} \left\{ \left| \int_{\mathcal{H}_\alpha} h(x) d\mu(x) - \int_{\mathcal{H}_\alpha} h(x) d\nu(x) \right| \right\},$$

where the uniform norm is defined by $\|h\|_\infty := \sup_{x \in \mathcal{H}_\alpha} |h(x)|$.

We also introduce bounded-Lipschitz distance with test functions being Lipschitz along the directions of stable modes.

Definition 6.4.3. *Let $\mu, \nu \in \mathfrak{P}(\mathcal{H}_\alpha)$, we define*

$$\|\mu - \nu\|_{\mathrm{L,\,s}} := \sup_{\|h\|_{\mathrm{BL,\,s}} \le 1} \left\{ \left| \int_{\mathcal{H}_\alpha} h(x) d\mu(x) - \int_{\mathcal{H}_\alpha} h(x) d\nu(x) \right| \right\},$$

where the Lipschitz norm along the stable direction is defined by

$$\|h\|_{\mathrm{BL,\,s}} := \|h\|_\infty + \sup_{\mathfrak{x} \neq \mathfrak{y} \in \mathcal{H}_\alpha \,|\, P_c \mathfrak{x} = P_c \mathfrak{y}} \frac{|h(\mathfrak{x}) - h(\mathfrak{y})|}{\|P_s \mathfrak{x} - P_s \mathfrak{y}\|_\alpha}.$$

Definition 6.4.4. *Given a measure $\mu \in \mathfrak{P}(\mathcal{H}_\alpha)$, we denote $P_c^* \mu$ by the marginal of μ on $P_c \mathcal{H}_\alpha$.*

Remark 6.4.5. *Based the above definitions, for any $\mu, \nu \in \mathfrak{P}(\mathcal{H}_\alpha)$, we can verify that*

$$\|\mu - \nu\|_{\mathrm{L}} \le \|\mu - \nu\|_{\mathrm{L,\,s}} \le \|\mu - \nu\|_{\mathrm{TV}}$$

and

$$\|P_c^* \mu - P_c^* \nu\|_{\mathrm{TV}} \le \|\mu - \nu\|_{\mathrm{L,\,s}}.$$

Note that by the assumptions in Section 6.1, both the solution u to the original equation and its approximation $\tilde{u}$ possess strong Markov property. We denote $\mathcal{T}_t$ by the transition semigroup generated by $\varepsilon^{-1}u(\varepsilon^{-2}t)$ and $\widetilde{\mathcal{T}}_t$ by the transition semigroup generated by $\varepsilon^{-1}\tilde{u}(\varepsilon^{-2}t)$.

Let non-trivial measures μ (resp. ν) on $\mathscr{B}(\mathcal{H}_\alpha)$ satisfy the bounds on moments given in Lemma 6.2.1, Lemma 6.2.6 and Proposition 6.3.2, i.e., for each $p \geq 1$, there exists some $C > 0$ such that

$$\int_{\mathcal{H}_\alpha} \|v\|_\alpha^p \, \mu(dv) \leq C\varepsilon^p, \quad \text{and} \quad \int_{\mathcal{H}_\alpha} \|P_s v\|_\alpha^p \, \mu(dv) \leq C\varepsilon^{2p}. \tag{6.65}$$

We aim to show a similar result as in [28, Theorem 5.2], i.e., for any t and $\varepsilon \in (0,1)$, there exist constants C, C' and $c > 0$ such that

$$\left\| \mathcal{T}_t^* \mu - \widetilde{\mathcal{T}}_t^* \nu \right\|_{\mathrm{L}} \leq C e^{-ct} \|\mu - \nu\|_{\mathrm{L}} + C'\varepsilon^2. \tag{6.66}$$

Consequently, the non-trivial invariant measures $\mu_\star, \nu_\star$ for $\mathcal{T}_t$ and $\widetilde{\mathcal{T}}_t$ are bounded by

$$\|\mu_\star - \nu_\star\|_{\mathrm{L}} \leq C'\varepsilon^2. \tag{6.67}$$

By Definition 6.4.1, we can rewrite the results in (6.24) as follows.

Proposition 6.4.6. *Given any $\mu \in \mathfrak{P}(\mathcal{H}_\alpha)$ satisfying (6.65). Fixed a $T > 0$. Then, there exists a constant C such that for all $t \in [0, T]$,*

$$\left\| \mathcal{T}_t^* \mu - \widetilde{\mathcal{T}}_t^* \mu \right\|_{\mathrm{L}} \leq C\varepsilon^2.$$

The following results are the dual version of Lemma 6.2.1, Lemma 6.2.6 and Proposition 6.3.2.

Proposition 6.4.7. *Let $\mu \in \mathfrak{P}(\mathcal{H}_\alpha)$ satisfy (6.65). Then, for each $p \geq 1$, there exists some $\tilde{C} > 0$ such that, for each $t > 0$,*

$$\int_{\mathcal{H}_\alpha} \|v\|_\alpha^p \, \mathcal{T}_t^* \mu(dv) \leq \tilde{C}\varepsilon^p, \quad \textit{and} \quad \int_{\mathcal{H}_\alpha} \|P_s v\|_\alpha^p \, \mathcal{T}_t^* \mu(dv) \leq \tilde{C}\varepsilon^{2p},$$

and

$$\int_{\mathcal{H}_\alpha} \|v\|_\alpha^p \, \widetilde{\mathcal{T}}_t^* \mu(dv) \leq \tilde{C}\varepsilon^p, \quad \textit{and} \quad \int_{\mathcal{H}_\alpha} \|P_s v\|_\alpha^p \, \widetilde{\mathcal{T}}_t^* \mu(dv) \leq \tilde{C}\varepsilon^{2p}.$$

We now collect other ingredients to prove (6.66). Apart from the multiplicative noise, the error estimate should fall in the same procedure as in the proof of [28, Theorem 5.2].

We show in the following lemma that transition semigroup $\mathcal{T}_t$ shows Feller property given fixed critical modes. Note that the result is same as [28, Lemma 5.4] except that we need to replace the noise by the multiplicative noise. We provide explicit proof for this case.

Lemma 6.4.8. *Let $\mu, \nu \in \mathfrak{P}(\mathcal{H}_\alpha)$. There exists a constant $C > 0$ such that, for each $t > 0$,*

$$\|\mathcal{T}_t^* \mu - \mathcal{T}_t^* \nu\|_{\mathrm{L,\,s}} \leq C(t^{-\frac{1}{2}} + 1) \|\mu - \nu\|_{\mathrm{L}} .$$

Proof. For any $v \in \mathcal{H}_\alpha$, we denote $\mathfrak{x} = P_c v$ and $\mathfrak{y} = P_s v$. Let $U(t, v) := \varepsilon^{-1} u(\varepsilon^{-2} t)$ denote the solution to (6.22) with initial condition $v = \mathfrak{x} + \mathfrak{y}$. We also define $X(t, \mathfrak{x})$ as the mild solution of $x(t)$ to (6.22a) with initial condition $\mathfrak{x}$; we define $Y(t, \mathfrak{y})$ in a similar way.

(1) We first show the growth rate of $\mathcal{T}_t \psi$ along the critical subspace, i.e. $D_c \mathcal{T}_t \psi$, given test functions satisfying $\|\psi\|_{\mathrm{BL,\,s}} \leq 1$, where D_c denotes the Fréchet derivative along the critical subspace.

By [43, Lemma 9.33], the identity

$$\psi(U(t, v)) = \mathcal{T}_t \psi(v) + \int_0^t \langle D_u \mathcal{T}_{t-s} \psi(U(s, v)), G(U(s, v)) dW_s \rangle \tag{6.68}$$

holds for any fixed $t > 0$, $\xi \in \mathcal{H}_\alpha$ and test function $\psi \in C_b^2(\mathcal{H}_\alpha)$.

Multiply both sides of (6.68) by

$$\int_0^t \langle G_c^{-1}(U(s, v)) D_{\mathfrak{x}} X(s, \mathfrak{x}) h,\ dW_s \rangle,$$

where $D_{\mathfrak{x}} X(s, \mathfrak{x}) h$ is the directional derivative w.r.t. $\mathfrak{x}$ along h. Then, we can show the following modified version of Bismut–Elworthy–Li formula,

$$\begin{aligned}
&\mathbf{E}\left[\psi(U(t, v)) \int_0^t \langle G_c^{-1}(U(s, v)) D_{\mathfrak{x}} X(s, \mathfrak{x}) h,\ dW_s \rangle\right] \\
=&\mathbf{E}\left[\int_0^t \langle G^*(U(s, v)) D_u \mathcal{T}_{t-s} \psi(U(s, v)),\ G_c^{-1}(U(s, v)) D_{\mathfrak{x}} X(s, \mathfrak{x}) h \rangle ds\right] \\
=&\mathbf{E}\left[\int_0^t \langle D_u \mathcal{T}_{t-s} \psi(U(s, v)),\ P_c^{-1} D_{\mathfrak{x}} X(s, \mathfrak{x}) h \rangle ds\right] \\
=&\int_0^t D_c \mathbf{E}[\mathcal{T}_{t-s} \psi(U(s, v))] h ds \\
=&\int_0^t D_c(\mathcal{T}_s \mathcal{T}_{t-s} \psi(v)) h ds = t D_c \mathcal{T}_t \psi(v) h.
\end{aligned} \tag{6.69}$$

For $\psi \in C_b^2(\mathcal{H}_\alpha)$, by (6.69), we have

$$\begin{aligned} |D_c \mathcal{T}_t \psi(v) h| &\leq \frac{1}{t} \|\psi\|_\infty \mathbf{E} \left| \int_0^t \langle G_c^{-1}(U(s,v)) D_{\mathfrak{x}} X(s,\mathfrak{x}) h,\ dW_s \rangle \right| \\ &\leq \frac{\hat{C}}{t} \left(\mathbf{E} \int_0^t \|D_{\mathfrak{x}} X(s,\mathfrak{x}) h\|_\alpha^2 ds \right)^{\frac{1}{2}}, \end{aligned} \tag{6.70}$$

where the constant $\hat{C}$ is obtained by $\|\psi\|_\infty$ and the Lipshitz constant of G.

By the differentiablity of $X(t,\mathfrak{x})$ w.r.t. the initial condition along the critical subspace, we can verify based on [43, Theorem 9.8] that the process $\zeta^h(t) := D_{\mathfrak{x}} X(t,\mathfrak{x}) h$ is the unique mild solution of

$$\begin{cases} d\zeta^h(t) = \mathcal{A}_c^{\mathfrak{q}} \zeta^h(t) dt + \mathcal{A}_c^{\mathfrak{cr}} \zeta^h(t) dt + D_{\mathfrak{x}} F_c(U(t,v)) \cdot \zeta^h dt + D_{\mathfrak{x}} G_c(U(t,v)) \cdot \zeta^h dW_t \\ \zeta^h(0) = h. \end{cases} \tag{6.71}$$

By the assumptions on F and G, using brute force as in the proof of [135, Lemma 2.5], we can verify that for t, there exists some $\tilde{C} > 0$ such that

$$\mathbf{E} \int_0^t \|\zeta^h(s)\|_\alpha^2 ds \leq \tilde{C} t e^{2\|\mathcal{A}_c\|_\alpha t}.$$

Combining with (6.70), then there exists some $C > 0$ such that for any $v \in \mathcal{H}_\alpha$, the operator norm

$$\|D_c \mathcal{T}_t(\psi(v))\| \leq \frac{C}{\sqrt{t}}.$$

Therefore,

$$\begin{aligned} |\mathcal{T}_t \psi(\mathfrak{x} + \mathfrak{y}) - \mathcal{T}_t \psi(\tilde{\mathfrak{x}} + \mathfrak{y})| &\leq \sup_{v \in \mathcal{H}_\alpha} |D_c(\mathcal{T}_t \psi(v))(\mathfrak{x} - \tilde{\mathfrak{x}})| \\ &\leq \frac{C}{\sqrt{t}} \|\mathfrak{x} - \tilde{\mathfrak{x}}\|_\alpha. \end{aligned} \tag{6.72}$$

By a similar argument as in [135, Lemma 2.2], the inequality in (6.72) also holds for all ψ such that $\|\psi\|_{\mathrm{BL,\,s}} \leq 1$.

(2) We then show the growth rate of $\mathcal{T}_t \psi$ along the stable subspace given test functions satisfying $\|\psi\|_{\mathrm{BL,\,s}} \leq 1$. Note that $\mathcal{A}_s$ is a strong contraction, then $\|D_{\mathfrak{y}} Y(t,\mathfrak{y})\| \leq 1$ in the operator norm for all $\mathfrak{y} \in \mathcal{H}_\alpha$. By the same procedure, we have

$$|\mathcal{T}_t \psi(\tilde{\mathfrak{x}} + \mathfrak{y}) - \mathcal{T}_t \psi(\tilde{\mathfrak{x}} + \tilde{\mathfrak{y}})| \leq C \|\mathfrak{y} - \tilde{\mathfrak{y}}\|_\alpha. \tag{6.73}$$

(3) Combining (6.72) and (6.73), by Definition 6.4.1, we have $\|\mathcal{T}_t\psi\|_{\mathrm{BL}} \leq C(\frac{1}{\sqrt{t}}+1)$ for test functions satisfying $\|\psi\|_{\mathrm{BL,\,s}} \leq 1$.

(4) By Definition 6.4.1, we have

$$\begin{aligned}\|\mathcal{T}_t^*\mu - \mathcal{T}_t^*\nu\|_{\mathrm{L,\,s}} &= \sup_{\|\psi\|_{\mathrm{BL,\,s}}\leq 1}\left|\int_{\mathcal{H}_\alpha}\mathcal{T}_t\psi(x)d\mu(x) - \int_{\mathcal{H}_\alpha}\mathcal{T}_t\psi(x)d\nu(x)\right| \\ &\leq \sup_{\|\varphi\|_{\mathrm{BL}}\leq 1}\sup_{\|\psi\|_{\mathrm{BL,\,s}}\leq 1}\left|\int_{\mathcal{H}_\alpha}\frac{\mathcal{T}_t\psi(x)}{\varphi(x)}\varphi(x)d\mu(x) - \int_{\mathcal{H}_\alpha}\frac{\mathcal{T}_t\psi(x)}{\varphi(x)}\varphi(x)d\nu(x)\right| \\ &\leq \sup_{\|\varphi\|_{\mathrm{BL}}\leq 1}\sup_{\|\psi\|_{\mathrm{BL,\,s}}\leq 1}\left\|\frac{\mathcal{T}_t\psi(x)}{\varphi(x)}\right\|_{\mathrm{Lip}}\left|\int_{\mathcal{H}_\alpha}\varphi(x)d\mu(x) - \int_{\mathcal{H}_\alpha}\varphi(x)d\nu(x)\right| \\ &\leq C(t^{-1/2}+1)\,\|\mu-\nu\|_{\mathrm{L}}\,. \end{aligned} \tag{6.74}$$

□

The following results repeat [28, Theorem 5.1]. We rephrase the statement and skip the proof due to the identical procedure.

Theorem 6.4.9. *[28, Thoerm 5.1]Let the assumptions in Section 6.1 be satisfied. There exists a $T > 0$ such that*

$$\|\mathcal{T}_T^*\mu - \mathcal{T}_T^*\nu\|_{\mathrm{L}} \leq \frac{1}{2}\|\mu-\nu\|_{\mathrm{L}} + \varepsilon^2\int_{\mathcal{H}_\alpha}(1+\|P_s v\|_\alpha)(\mu+\nu)dv$$

for any non-trivial $\mu, \nu \in \mathfrak{P}(\mathcal{H}_\alpha)$.

Combining Theorem 6.4.9, Proposition 6.4.6 and 6.4.7, as well as Lemma 6.4.8, we can verify (6.66) by the same procedure as in the proof of [28, Theorem 5.2]. In particular, we use the following triangle inequality

$$\begin{aligned}\left\|\mathcal{T}_t^*\mu - \widetilde{\mathcal{T}}_t^*\nu\right\|_{\mathrm{L}} &\leq \|\mathcal{T}_t^*\mu - \mathcal{T}_t^*\nu\|_{\mathrm{L}} + \left\|\mathcal{T}_t^*\nu - \widetilde{\mathcal{T}}_t^*\nu\right\|_{\mathrm{L}} \\ &\leq \|\mathcal{T}_t^*\mu - \mathcal{T}_t^*\nu\|_{\mathrm{L,\,s}} + \left\|\mathcal{T}_t^*\nu - \widetilde{\mathcal{T}}_t^*\nu\right\|_{\mathrm{L}}\,.\end{aligned}$$

6.5 Summary

In this chapter, we verified the regularities of solutions to (6.1) under proper assumptions given in Section 6.1. Then we proposed an approximation scheme (6.22) to the original system and investigated the error, including the error of transient solutions up to a fixed time T as well as the distance between the invariant measures.

Unlike the finite-dimensional amplitude equation as in [27, Eq. (16)] as well as the amplitude equation obtained in Chapter 5 that are the limiting equations for the corresponding infinite-dimensional dynamics, the purpose of the approximation scheme (6.22) is not to reduce dimension via homogenization. We do not cancel the coupling effects in the nonlinearies (in $\mathcal{H}_c$) and the multiplicative noise (in both $\mathcal{H}_c$ and $\mathcal{H}_s$) so as to keep as much interaction information between the critical and stable modes as in the original dynamics. This simplified scheme will finally be used along with the linearized equations to analyze the structural changes in random attractors as the trivial solution loses its stability.

In particular, the proposed approximation scheme in (6.22) has the same linearization as the original equation. By a careful multiscale analysis, we concluded that the errors of the transient critical and stable solutions up to a fixed time T are of the same order as the error generated by $\mathcal{A}_c^{\text{er}}$, which indicates a sufficient accuracy to capture the local bifurcation. By the same procedure as in [28], the invariant measure of (6.22) is also analyzed to be sufficient close to the one of the original system.

It is worth mentioning that a rough approximation (6.35) of the amplitude equation was used as an intermediate step to get the final error estimation. This equation cancels the random effect of the stable modes completely and provides a worst-case scenario of how large the deviation from the original solution could be. It will be shown in the next section that we cannot use (6.35) as a local approximation to investigate the dynamical Hopf bifurcation due to its oversimplified shape of the invariant measure for the stable modes, which is reduced to point mass around $\mathbf{0}$.

Chapter 7

Almost Sure Asymptotic Stability of Scalar SPDEs with Multiplicative Noise Close to Hopf Bifurcations

In this chapter, we investigate the almost-sure exponential asymptotic stability of the trivial solution of an SPDE driven by multiplicative noise near the deterministic Hopf bifurcation point. We show the existence and uniqueness of the invariant measure under proper assumptions, and approximate the exponential growth rate via asymptotic expansion, given the strength of the noise is small. We illustrate the results using a simplified stochastic Moore-Greitzer PDE model with multiplicative noise.

As introduced in the previous chapter, the almost-sure asymptotic stability/instability at the trivial solution is captured by the sign of the top Lyapunov exponent λ of the linearized system. In particular, under proper conditions, for finite dimensional SDEs with coefficients dependent on some parameter γ, if γ varies in a way that $\lambda(\gamma)$ changes sign from negative to positive, the trivial solution loses its almost-sure asymptotic stability and a nontrivial invariant measure is formed [17]. Despite the fact that it is difficult to quantitatively describe the random invariant manifolds and stochastic bifurcations for SPDEs driven by multiplicative noise [79, 156, 107, 37], in terms of stability for infinite-dimensional case, the almost-sure stability of scalar stochastic delay differential equation has been studied in [154].

As a necessary step to study the D-bifurcation (defined in [11]), we aim to investigate the almost-sure asymptotic stability of the trivial solutions of SPDEs driven by small multiplicative noise. In particular, we study the effect of multiplicative noise near the Hopf bifurcation point γ_c of the unperturbed system. While it remains difficult to obtain the exact expressions for top

Lyapunov exponents, we adopt a similar method as in [154] and find the asymptotic approximation of them. To be more precise, the multiscale analysis is conducted in the neighbourhood of the deterministic bifurcation point γ_c, as the parameter γ slowly passes through γ_c (quasi-static). For a fixed $\alpha \in (0, 1]$, we consider linear (linearized) SPDEs of the following general form

$$du(t) = \mathcal{A}(\gamma)u(t)dt + \varepsilon(G(u(t))dW(t),\ u(0) = u_0 \in \mathcal{H}_\alpha,\ \gamma \in \mathbb{R}, \tag{7.1}$$

where for each t, the state $u(t)$ takes value in the fractional power subspace $\mathcal{H}_\alpha$ of an infinite-dimensional separable Hilbert space $\mathcal{H} = L^2(E)$ for some bounded $E \subseteq \mathbb{R}^n$. The notations and assumptions keep the same as in Chapter 6.

Definition 7.0.1. *For future reference, we also introduce the space $\mathcal{H}_{\alpha,s} := P_s\mathcal{H}_\alpha$, where P_s is the stable projection as in Definition 6.1.3.*

7.1 Stability Analysis of the Trivial Solution

In this section, we investigate the almost-sure asymptotic stability of the trivial solution $\mathbf{0}$, or equivalently, the trivial invariant measure $\delta_{\mathbf{0}}$, using multiscale techniques. We set $\gamma = \gamma_c + \varepsilon^2\mathfrak{q}$ with some unfolding parameter $\mathfrak{q} \in \mathbb{R}$, and introduce the notion of solution under re-scaled space and time as follows, i.e. we introduce

$$z(t) = \varepsilon^{-1}\langle \mathfrak{h}^*, u(\varepsilon^{-2}t)\rangle$$

and

$$\bar{z}(t) = \varepsilon^{-1}\langle \bar{\mathfrak{h}}^*, u(\varepsilon^{-2}t)\rangle$$

as the complex amplitudes of the critical mode and $y(t) = \varepsilon^{-1}P_s u(\varepsilon^{-2}t)$. Then the solution u of (7.1) can be written as

$$u(t) = \varepsilon z(\varepsilon^2 t)\mathfrak{h} + \varepsilon\bar{z}(\varepsilon^2 t)\bar{\mathfrak{h}} + \varepsilon y(\varepsilon^2 t).$$

We denote the real part and imaginary part of z as $z_1 = \mathrm{Re}(z)$ and $z_2 = \mathrm{Im}(z)$, respectively.

Note that when the system is close to the critical point, due to the existence of the spectral gap, we decompose (7.1) into the re-scaled critical and fast-varying modes as follows:

$$\begin{aligned} dz &= \varepsilon^{-2}\rho_c^{\mathfrak{c}} z dt + \rho_c^{\mathfrak{q}} z dt + \langle \mathfrak{h}^*, G(z\mathfrak{h} + \bar{z}\bar{\mathfrak{h}} + y)dW_t\rangle + \rho_c^{\mathfrak{er}} z dt, \\ dy &= \varepsilon^{-2}\mathcal{A}_s y dt + G_s(z\mathfrak{h} + \bar{z}\bar{\mathfrak{h}} + y)dW_t, \end{aligned}$$

or equivalently,

$$
d\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} a_c^{\mathfrak{q}} & -\varepsilon^{-2}b_c^{\mathfrak{c}} - b_c^{\mathfrak{q}} \\ \varepsilon^{-2}b_c^{\mathfrak{c}} + b_c^{\mathfrak{q}} & a_c^{\mathfrak{q}} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} G_c^R(z\mathfrak{h} + \bar{z}\bar{\mathfrak{h}} + y) \\ G_c^I(z\mathfrak{h} + \bar{z}\bar{\mathfrak{h}} + y) \end{bmatrix} dW_t + \mathcal{O}(\varepsilon^2), \tag{7.2a}
$$

$$
dy = -\varepsilon^{-2}\mathcal{A}_s y dt + G_s(z\mathfrak{h} + \bar{z}\bar{\mathfrak{h}} + y)dW_t, \tag{7.2b}
$$

where $W(t) \sim \varepsilon^{-1}W(\varepsilon^{-2}t)$[1] and

$$
G_c^R(u) = \frac{\hat{G}_1(u) + \overline{\hat{G}_1(u)}}{2}, \; G_c^I(u) = \frac{\hat{G}_1(u) - \overline{\hat{G}_1(u)}}{2}; \tag{7.3a}
$$

$$
\hat{G}_1(u)w := \langle \mathfrak{h}^*, G(u)w \rangle, \quad \forall u \in \mathcal{H}_\alpha, \; \forall w \in \mathcal{V}. \tag{7.3b}
$$

We impose the initial condition to (7.2) as $z(0) = \varepsilon^{-1}\langle \mathfrak{h}^*, u_0 \rangle$ and $y(0) = \varepsilon^{-1}P_s u_0$.

The truncation error term $\mathcal{O}(\varepsilon^2)$ in (7.2) comes from the transformation of $\rho_c^{\mathfrak{er}} z dt$, whose property has been verified in Proposition 6.1.7 and Lemma 6.3.7. We use the notation $\mathcal{O}(\varepsilon^2)$ for short to indicate its order of error after integration. By dropping the high order term $\mathcal{O}(\varepsilon^2)$ in the asymptotic expansion, we are able to verify the error between the transient solutions and the error between the invariant measures in the same way as in Chapter 6. Due to the continuity of the term $\mathcal{O}(\varepsilon^2)$ in ε and the insignificant effect, to this end, we work on the following equation to derive the first order approximation of the top Lyapunov exponent.

$$
d\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} a_c^{\mathfrak{q}} & -\varepsilon^{-2}b_c^{\mathfrak{c}} - b_c^{\mathfrak{q}} \\ \varepsilon^{-2}b_c^{\mathfrak{c}} + b_c^{\mathfrak{q}} & a_c^{\mathfrak{q}} \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} G_c^R(z\mathfrak{h} + \bar{z}\bar{\mathfrak{h}} + y) \\ G_c^I(z\mathfrak{h} + \bar{z}\bar{\mathfrak{h}} + y) \end{bmatrix} dW_t, \tag{7.4a}
$$

$$
dy = -\varepsilon^{-2}\mathcal{A}_s y dt + G_s(z\mathfrak{h} + \bar{z}\bar{\mathfrak{h}} + y)dW_t. \tag{7.4b}
$$

7.1.1 The Furstenberg–Khasminskii Formula for the Top Lyapunov Exponent

Let

$$
\mathfrak{p}(t) = \frac{1}{2}\ln(z_1^2(t) + z_2^2(t)), \quad z_1(t) = e^{\mathfrak{p}(t)}\cos(\phi(t)), \quad z_2(t) = e^{\mathfrak{p}(t)}\sin(\phi(t)) \text{ and } \; \eta_t = e^{-\mathfrak{p}(t)}y_t,
$$

where ϕ is the phase angle in the unit sphere $\mathcal{S}^1$ satisfying

$$
z_1 = |z|\cos(\phi), \quad z_2 = |z|\sin(\phi), \quad \phi = \arctan\left(\frac{z_2}{z_1}\right). \tag{7.5}
$$

[1] We abuse the notation as what we did in Chapter 6 to avoid redundancy.

Therefore, by Itô's formula,

$$d\mathfrak{p} = a_c^{\mathfrak{q}} dt + \Xi(\phi, \eta) dt + [G_c^R(\phi, \eta) \cos \phi + G_c^I(\phi, \eta) \sin \phi] dW_t, \tag{7.6a}$$

$$d\phi = (\varepsilon^{-2} b_c^{\mathfrak{c}} + b_c^{\mathfrak{q}}) dt + \Gamma(\phi, \eta) dt - G_c^{\phi}(\phi, \eta) dW_t, \tag{7.6b}$$

$$d\eta = \varepsilon^{-2} \mathcal{A}_s \eta dt + G_s(\phi, \eta) dW_t, \tag{7.6c}$$

where[2]

$$G_c^R(\phi, \eta) := G_c^R(\cos(\phi)\mathfrak{h} + \sin(\phi)\bar{\mathfrak{h}} + \eta), \quad G_c^I(\phi, \eta) := G_c^I(\cos(\phi)\mathfrak{h} + \sin(\phi)\bar{\mathfrak{h}} + \eta),$$

$$G_s(\phi, \eta) := G_s(\cos(\phi)\mathfrak{h} + \sin(\phi)\bar{\mathfrak{h}} + \eta), \quad G_c^{\phi}(\phi, \eta) := G_c^R(\phi, \eta) \sin \phi - G_c^I(\phi, \eta) \cos \phi,$$

and

$$\Xi := -\frac{\cos(2\phi)}{2} \operatorname{tr}[G_c^R (G_c^R)^* - G_c^I (G_c^I)^*](\phi, \eta) - \frac{\sin(2\phi)}{2} \operatorname{tr}[G_c^R (G_c^I)^* + G_c^I (G_c^R)^*](\phi, \eta),$$

$$\Gamma := \frac{\sin(2\phi)}{2} \operatorname{tr}[G_c^R (G_c^R)^* - G_c^I (G_c^I)^*](\phi, \eta) - \frac{\cos(2\phi)}{2} \operatorname{tr}[G_c^R (G_c^I)^* + G_c^I (G_c^R)^*](\phi, \eta).$$

We also name $G_c^{\mathfrak{p}}(\phi, \eta) := G_c^R(\phi, \eta) \cos \phi + G_c^I(\phi, \eta) \sin \phi$ for future references.

The initial condition is such that $\mathfrak{p}_0 = \mathfrak{p}(0) = \ln |z(0)|$, $\phi_0 = \phi(0) = \arctan\left(\frac{z_2(0)}{z_1(0)}\right)$ and $\eta_0 = \eta(0) = e^{-\mathfrak{p}(0)} y(0)$.

We also denote the drift term of (7.6a) as

$$\mathcal{Q}^{\mathfrak{q}}(\phi, \eta) := a_c^{\mathfrak{q}} + \Xi(\phi, \eta). \tag{7.7}$$

Remark 7.1.1. *Let $a := \operatorname{tr}[G_c^R]$ and $b := \operatorname{tr}[G_c^I]$ for some fixed (ϕ, η), then $\Xi + i\Gamma = -e^{-2i\phi}(a + bi)^2$.*

Noticing that $\mathfrak{p}(t)$ only depends on $\phi(t)$ and $\eta(t)$, if there exists a unique invariant measure μ^{ε} for the product process $(\phi(t), \eta(t)) \in \mathcal{S}^1 \times \mathcal{H}_{\alpha,s}$, the top Lyapunov exponent of $\varepsilon^{-1} u$ can be determined by the Furstenberg–Khasminskii formula:

$$\begin{aligned} \lambda^{\mathfrak{q},\varepsilon} &= \lim_{t \to \infty} \frac{1}{t} \ln |z(t)| \\ &= \int_{\mathcal{S} \times \mathcal{H}_s} \mathcal{Q}^{\mathfrak{q}}(\phi, \eta) \mu^{\mathfrak{q},\varepsilon}(d\phi, d\eta) =: \langle \mathcal{Q}^{\mathfrak{q}}, \mu^{\mathfrak{q},\varepsilon} \rangle. \end{aligned} \tag{7.8}$$

Remark 7.1.2. *Since in this section we do not vary $\mathfrak{q}$ as what we do in the study of bifurcation theory, to simplify the notation, we use λ^{ε}, $\mathcal{Q}$ and μ^{ε} instead.*

[2] We abuse the notation G and recast the arguments as ϕ and η.

7.1.2 Existence of Invariant Measure

Note that (7.6b) and (7.6c) are coupled via the multiplicative noise. The mutual dependence of ϕ and η brings difficulty to study the explicit dependence of $\{\phi(t)\}_{t\geq 0}$ for the solution $\{\eta(t)\}_{t\geq 0}$ to (7.6c) pathwisely. On the other hand, we are able to take the advantage of the compactness of $\mathcal{S}^1$ and start with investigating the bounds for the stable marginals based on Assumptions 6.1.6 and 6.1.12.

Lemma 7.1.3. *Let Assumptions 6.1.1, 6.1.6 and 6.1.12 be satisfied. For each arbitrarily small $\varepsilon > 0$, there exists a $C > 0$ such that*

$$\sup_{t\geq 0} \mathbf{E}\|\eta(t)\|_\alpha^2 \leq C.$$

Proof. Consider Yosida approximation $\mathcal{A}_{s,n} := n\mathcal{A}_s(nI - \mathcal{A}_s)^{-1}$ of $\mathcal{A}_s$. We denote η_n by the solution to

$$d\eta_n = -\varepsilon^{-2}\mathcal{A}_{s,n}\eta_n dt + G_s(\phi, \eta_n)dW_t,\ \eta_n(0) = \eta(0).$$

Apply Itô's formula to $\|\eta_n(t)\|_\alpha^2$, then

$$\begin{aligned} d\|\eta_n(t)\|_\alpha^2 = &- 2\varepsilon^{-2}\langle \mathcal{A}_{s,n}\eta_n(t), \eta_n(t)\rangle_\alpha dt + \|G_s(\phi(t), \eta_n(t))\|_{\mathcal{L}_2}^2 dt \\ &+ 2\langle \eta_n(t),\ G_s(\phi(t), \eta_n(t))dW_t\rangle_\alpha. \end{aligned}$$

Taking the expectation, using the property of $\mathcal{A}_{s,n}$ and G_s, there exists an $w > 0$, $\tilde{w} > 0$ and $\tilde{C} > 0$ such that for all $t \geq 0$,

$$\begin{aligned} \frac{d\mathbf{E}\|\eta_n(t)\|_\alpha^2}{dt} &= \mathbf{E}\left\{-2\varepsilon^{-2}\langle \mathcal{A}_{s,n}\eta_n(t), \eta_n(t)\rangle_\alpha + \|G_s(\phi(t), \eta_n(t))\|_{\mathcal{L}_2}^2\right\} \\ &\leq -2\varepsilon^{-2}w\mathbf{E}\|\eta_n(t)\|_\alpha^2 + \ell_2\mathbf{E}|\cos(\phi(t)) + \sin(\phi(t))|^2 + \ell_2\mathbf{E}\|\eta_n(t)\|_\alpha^2 \\ &\leq -\tilde{w}\mathbf{E}\|\eta_n(t)\|_\alpha^2 + \tilde{C} \end{aligned}$$

It follows from Gronwall's inequality that $\mathbf{E}\|\eta_n(t)\|_\alpha^2 < C$ for every $t \geq 0$ and some $C > 0$. The conclusion follows by sending n to infinity. ∎

Lemma 7.1.4. *Let the assumptions in Lemma 7.1.3 be satisfied. Fix any $T > 0$ and any $p \geq 2$. For any initial condition $\eta(0) \in \mathcal{H}_{\alpha,s}$, there exists some $C > 0$ such that*

$$\mathbf{E} \sup_{0\leq t\leq T} \|\eta(t)\|_\alpha^p \leq \|\eta(0)\|_\alpha^p + C\varepsilon^p.$$

Proof. For any $t \in [0, T]$, the mild solution is given as

$$\eta(t) = e^{\varepsilon^{-2}t\mathcal{A}_s}\eta(0) + \int_0^t e^{\varepsilon^{-2}(t-s)\mathcal{A}_s} G_s(\phi(s), \eta(s))dW_s. \tag{7.9}$$

Let $P_s W_{\mathcal{A}}^G(t) := \int_0^t e^{\varepsilon^{-2}(t-s)\mathcal{A}_s} G_s(\phi(s), \eta(s))dW_s$ denote the stochastic convolution. The bound for the stochastic convolution follows [43, Proposition 7.3]. Indeed, by Lemma 7.1.3, there exists some $C > 0$ and $C' > 0$, such that

$$\int_0^T \mathbf{E}\|G_s(\phi(s), \eta(s))\|^p_{\mathcal{L}_2(\mathcal{V},\mathcal{H}_\alpha)} ds \le C \int_0^T \mathbf{E}\|\cos(\phi(s)) + \sin(\phi(s)) + \eta(s)\|^p_\alpha ds < C' < \infty.$$

Therefore $G_s(\phi(s), \eta(s))$ is $\mathcal{L}^2$ predictable and there exists constants $C_T > 0$ and $C'_T > 0$ such that

$$\mathbf{E} \sup_{0\le t\le T} \left\|P_s W_{\mathcal{A}}^G(t)\right\|^p_\alpha \le \varepsilon^p C_T \mathbf{E}\left(\int_0^t \|G_s(\phi(s), \eta(s)))\|^p_{\mathcal{L}_2(\mathcal{V},\mathcal{H}_\alpha)} ds\right) \le \varepsilon^p C'_T, \tag{7.10}$$

which implies that $\mathbf{E}\sup_{0\le t\le T} \|\eta(t)\|^p_\alpha \le \|\eta(0)\|_\alpha + C'_T\varepsilon^p$. □

For test functions $f \in C_b^2(\mathcal{S}^1 \times \mathcal{H}_{\alpha,s})$, the transition semigroup of (7.6b) and (7.6c) is such that $\mathcal{T}_t f = \mathbf{E}[f(\phi(t), \eta(t))|(\phi, \eta)]$. Based on the compactness of $\mathcal{S}^1$ and the above uniform bounds for $\{\eta(t)\}_{t\ge0}$, the existence of invariant measure is guaranteed.

Proposition 7.1.5. *Let the assumptions in Lemma 7.1.3 be satisfied. Then there exists an invariant measure for the transition semigroup* $\{\mathcal{T}_t\}_{t\ge0}$ *of* (7.6b) *and* (7.6c).

Proof. Let $\mathscr{L}(\cdot)$ denote the law of random variables on the canonical space generated by $\mathcal{S}^1 \times \mathcal{H}_{\alpha,s}$. By Lemma 7.1.4 and the compactness of $\mathcal{S}^1$, it is straightforward to show that

$$\left\{\frac{1}{t_n}\int_0^{t_n} \mathscr{L}(\phi(s), \eta(s))ds\right\}$$

forms a tight family of measure. The existence of invariant measure for $\{(\phi(t), \eta(t))\}_{t\ge0}$ under $\mathcal{T}_t$ follows by Krylov–Bogoliubov's Theorem (along the same time sequence). □

7.1.3 Transient Dissipativity of the Stable Modes

The following lemma shows the approximated dissipativity condition given any transient transitions.

Lemma 7.1.6. *Let the assumptions in Lemma 7.1.3 be satisfied. For each arbitrarily small $\varepsilon > 0$ and fixed $\phi \in \mathcal{S}^1$, there exists $w > 0$ such that, for all $\eta_1, \eta_2 \in \mathcal{H}_{\alpha,s}$,*

$$-2\varepsilon^{-2}\langle \mathcal{A}_{s,n}(\eta_1 - \eta_2), (\eta_1 - \eta_2)\rangle_\alpha + \|G_s(\phi, \eta_1) - G_s(\phi, \eta_2)\|^2_{\mathcal{L}_1} \leq -\varepsilon^{-2} w \|\eta_1 - \eta_2\|_\alpha,$$

where $\mathcal{A}_{s,n}$ is the Yosida approximation of $\mathcal{A}_s$.

Proof. The conclusion can be obtained under the assumptions considering sufficiently small $\varepsilon > 0$. □

Remark 7.1.7. *Note that by a similar approach as in [43, Theorem 11.30], we can verify that for each $\phi \in \mathcal{S}^1$, given initial condition $\eta(0) = \mathbf{0}$ a.s., there exists a unique random variable $\eta^\phi \in \mathcal{L}^2(\mathbf{\Omega}; \mathcal{H}_{\alpha,s})$ on some probability space $(\mathbf{\Omega}, \mathcal{F}, \mathbf{P})$ as $\varepsilon^{-2} t \to \infty$.*

Indeed, we can show it by shifting the time and letting $\eta^\phi(-\tau) = 0$ for some $\tau > 0$. We denote the solution for any $t \geq \tau$ by $\eta^\phi_\tau(t, 0)$, where the 0 is referred as the initial condition. Then it is clear that the probability law of $\eta^\phi_\tau(0, 0)$ is the same as $\eta^\phi(\tau)$. By Lemma 7.1.6, we have the mean square stability for each $\phi \in \mathcal{S}^1$:

$$\mathbf{E}\|\eta^\phi_\tau(0, 0) - \eta^\phi_\sigma(0, 0)\|^2_\alpha \leq MCe^{-\varepsilon^{-2} w\tau}, \quad \sigma > \tau.$$

The above inequality demonstrates that the Cauchy sequence in $\mathcal{L}^2$ is dominated by $Ce^{-\varepsilon^{-2}\omega\tau}$, and as $\varepsilon^{-2}\tau \to \infty$, the limit η^ϕ exits as a random variable in $\mathcal{L}^2(\Omega; \mathcal{H}_{\alpha,s})$.

The following proposition shows the transient behavior of the transition along the $\mathcal{H}_s$ subspace.

Proposition 7.1.8. *As $\varepsilon \to 0$, at each each $t > 0$, the marginal transition probability $H_t(\cdot \mid \phi, \eta)$ of (7.6c) behaves like a measure $\nu^\phi(d\eta)$ on $\mathcal{H}_{\alpha,s}$ that only depends on ϕ.*

Proof. By Remark 7.1.7, for $\eta(0) = 0$ and each $\phi \in \mathcal{S}^1$, there exists a unique limit η^ϕ with probability law $\nu^\phi(d\eta)$ as an $\mathcal{L}^2$ random variable as $\varepsilon^{-2} t \to \infty$. Note that as $\varepsilon \to 0$, at each $t > 0$, we have $\varepsilon^{-2} t \to \infty$. Therefore, the marginal transition of $\eta(t)$ is given as

$$\begin{aligned} H_t(d\eta_t \mid \phi, \eta) &= H_t(d\eta_t \mid \phi, \eta)\mathbb{1}_{\{\eta=0\}} \\ &\approx \nu^\phi(d\eta_t). \end{aligned} \tag{7.11}$$

We now consider random initial distribution and let $\mathscr{L}(\eta(0)) = \nu_0$. Since we have $\mathbf{E}[\eta^2(0)] < \infty$, by a similar argument as Lemma 7.1.3, we can show that

$$\lim_{\varepsilon^{-2} t \to \infty} \mathbf{E}\|\eta^{\phi,0}(t) - \eta^{\phi,\nu_0}(t)\|^2_\alpha = 0, \quad \phi \in \mathcal{S}^1, \tag{7.12}$$

where $\eta^{\phi,0}(t)$ denotes the solution of (7.6b) with $\eta(0) = 0$ a.s. and $\eta^{\phi,\nu_0}(t)$ denotes the solution of (7.6b) with $\mathscr{L}(\eta(0)) = \nu_0$ for some fixed ϕ. We aim to show that for any test function $f \in C_b(\mathcal{H}_{\alpha,s})$, each of $\{\eta^\phi(t)\}_\phi$ with an arbitrary initial distribution converges weakly to the same limit point with probability law in $\{\nu^\phi\}_\phi$ for $\nu^\phi(d\eta_t) := \lim_{\varepsilon^{-2}t\to\infty} H_t(d\eta_t \mid \phi, \eta)\nu_0(d\eta)$, i.e.,

$$\int_{\mathcal{H}_s} f(\eta(t))H_t(d\eta_t \mid \phi, \eta)\nu_0(d\eta) \xrightarrow{\varepsilon^{-2}t\to\infty} \int_{\mathcal{H}_s} f(\eta(t))\nu^\phi(d\eta_t), \quad \phi \in \mathcal{S}^1. \tag{7.13}$$

Following the approach in [42, Theorem 1], we can show that for each fixed ϕ,

$$\begin{aligned}
&\left|\mathbf{E}[f(\eta^{\phi,\nu_0}(t))] - \int_{\mathcal{H}_s} f(\eta)\nu^\phi(d\eta)\right| \\
\le&\mathbf{E}\left|f(\eta^{\phi,\nu_0}(t)) - f(\eta^{\phi,0}(t))\right| + \left|\mathbf{E}[f(\eta^{\phi,0}(t))] - \int_{\mathcal{H}_s} f(\eta)\nu^\phi(d\eta)\right| \\
=&:I_1(t) + I_2(t)
\end{aligned} \tag{7.14}$$

where $I_2(t) \to 0$ as discussed above, $I_1(t)$ is arbitrarily small. Indeed,

$$\begin{aligned}
I_1(t) &\le C \cdot \mathbf{P}[\|\eta_0\|_\alpha \ge R] + \mathbf{E}[\mathbb{1}_{\{\|\eta_0\|_\alpha \le R\}} \cdot f(\eta^{\phi,\nu_0}(t)) - f(\eta^{\phi,0}(t))] \\
&=: I_3(t) + I_4(t)
\end{aligned} \tag{7.15}$$

where the constant C in $I_3(t)$ is by the boundedness property of f. For arbitrary $\varsigma > 0$, since $\eta_0 \in \mathcal{L}^2(\Omega; \mathcal{H}_{\alpha,s})$, there exists $R > 0$ such that

$$\mathbf{P}[\|\eta_0\|_\alpha \ge R] < \varsigma.$$

We choose R based on an arbitrary small ς. Note that $I_4(t)$ is restricted in a compact subspace and f becomes uniformly continuous, by the property of f and (7.12), $I_4(t) \to 0$ as $\varepsilon^{-2}t \to \infty$.

We have seen that $\nu^\phi(d\eta)$ is unique w.r.t. each ϕ with arbitrary initial distribution ν_0, including δ_η. The statement hence follows. □

Remark 7.1.9. *Unlike the case in Chapter 5 where the linearized equations have no coupling effects, we are not able to explicitly solve the invariant measure for $\{\eta(t)\}_{t\ge 0}$ by considering the transitions separately.*

The above proposition only provides a view that the marginal transition along $\mathcal{H}_{\alpha,s}$ quickly forgets the initial point η for sufficiently small noise. In this view, we can represent an invariant measure by a disintegrated form

$$\begin{aligned}
\mu(d\phi \times d\eta) &= \int_{\mathcal{S}^1\times\mathcal{H}_s} R(d\phi \mid \phi, \eta)\nu^\phi(d\eta)\mu(d\phi \times d\eta) \\
&=: \tilde{\mu}(d\phi)\tilde{\nu}^\phi(d\eta).
\end{aligned}$$

Note that $\tilde{\mu}(d\phi)$ and $\tilde{\nu}^{\phi}(d\eta)$ are difficult to solve explicitly. This, however, motivates us to deliver an approximation in Section 7.2 of the invariant measure with the above disintegrated form.

7.1.4 Conditions on Uniqueness of Invariant Measure

We have seen in Section 7.1.3 that the disintegration measure $\nu^{\phi}(d\eta)$ along $\mathcal{H}_s$ uniquely exists given any ϕ. Similarly, under condition that $G_c^{\phi}(G_c^{\phi})^* \neq 0$ for all ϕ [91], for each η, the solution of

$$d\phi = (\varepsilon^{-2}b_c^{\mathfrak{c}} + b_c^{\mathfrak{q}})dt + \Gamma(\phi,\eta)dt - G_c^{\phi}(\phi,\eta)dW_t \tag{7.16}$$

admits a unique limit measure that is solved by the associated Fokker-Plank equation

$$\frac{dp}{d\phi}\big(\varepsilon^{-2}b_c^{\mathfrak{c}} + b_c^{\mathfrak{q}} + \Gamma(\phi,\eta)\big) + \frac{1}{2}\frac{d^2p}{d\phi^2}\left[G_c^{\phi}(G_c^{\phi})^*\right](\phi,\eta) = 0, \tag{7.17}$$

where p is the density function.

However, as discussed in Remark 7.1.9, it will not be enough to consider ergodicity or uniqueness of invariant measure for $\{\phi(t)\}_{t\geq 0}$ and $\{\eta(t)\}_{t\geq 0}$ separately. It is not sufficient to only suppose the full-rank property of G_c^{ϕ}. We hence impose a set of extra stronger conditions to guarantee the uniqueness of the invariant measure.

Assumption 7.1.10. *For any $\alpha \in (0,1]$, we assume that the operator $G(u)$ is invertible for each $u \in \mathcal{H}_\alpha \setminus \{\mathbf{0}\}$.*

The above assumption plays a role as the Lie algebra condition to guarantee the uniqueness of the $\{\mathcal{T}_t\}_{t\geq 0}$ for the coupled linearized system. It is equivalent to verify the non-singular condition, i.e., we need $G(u)$ to be bounded from below in the following sense:

$$\langle G(u)v, v\rangle_\alpha \geq m\|u\|_\alpha\|v\|,\ m > 0, v \in \mathcal{V}. \tag{7.18}$$

Combining with Assumption 6.1.12 and 6.1.6, it can be verified that $G(u)$ is holomorphic on $\mathcal{H}_\alpha \setminus \{\mathbf{0}\}$.

7.2 Asymptotic Approximation of the Top Lyapunov Exponent

Motivated by Remark 7.1.9, we derive the asymptotic expansion of the invariant measure μ^{ε} in this section for the approximation of the top Lyapunov exponent. By [43, Theorem 9.25], we can show that for any test function $f \in C_b^2(\mathcal{S}^1 \times \mathcal{H}_{\alpha,s})$, the quantity $\mathcal{T}_t f$ satisfies

$$\lim_{t \downarrow 0} \frac{\mathcal{T}_t f(\phi, \eta) - f(\phi, \eta)}{t} = \mathfrak{L}^{\mathfrak{q},\varepsilon} f(\phi, \eta)$$

where

$$\mathfrak{L}^{\mathfrak{q},\varepsilon} = \frac{1}{\varepsilon^2} \mathfrak{L}_0 + \mathfrak{L}_1^{\mathfrak{q}}, \tag{7.19}$$

and

$$\mathfrak{L}_0(\cdot) = \left[b_c^{\mathfrak{c}} \frac{\partial}{\partial \phi} + \mathcal{A}_s \eta \frac{\partial}{\partial \eta} \right] (\cdot)$$

$$\mathfrak{L}_1^{\mathfrak{q}}(\cdot) = \left[(b_c^{\mathfrak{q}} + \Gamma(\phi, \eta) \frac{\partial}{\partial \phi} \right] (\cdot) + \frac{1}{2} \operatorname{tr} \left[\frac{\partial^2 (\cdot)}{\partial \eta^2} G_s G_s^* + \frac{\partial^2 (\cdot)}{\partial \phi^2} G_c^\phi (G_c^\phi)^* \right] (\phi, \eta)$$

Remark 7.2.1. *To simplify the notation, we use $\mathfrak{L}^\varepsilon$ and $\mathfrak{L}_1$ in stead of $\mathfrak{L}^{\mathfrak{q},\varepsilon}$ and $\mathfrak{L}_1^{\mathfrak{q}}$ in this section.*

For any test function $f \in C^2(\mathcal{S}^1 \times \mathcal{H}_{\alpha,s})$ one should have $\langle \mathfrak{L}^\varepsilon f, \mu^\varepsilon \rangle = 0$. We expand μ^ε as

$$\mu^\varepsilon = \mu_0 + \varepsilon^2 \mu_1 + \mathcal{O}(\varepsilon^3), \tag{7.20}$$

then, respectively on the level $\mathcal{O}(\varepsilon^{-2})$ and $\mathcal{O}(1)$,

$$\langle \mathfrak{L}_0 f, \mu_0 \rangle = 0 \tag{7.21a}$$

$$\langle \mathfrak{L}_0 f, \mu_1 \rangle = -\langle \mathfrak{L}_1 f, \mu_0 \rangle \tag{7.21b}$$

We proceed to find the solutions to (7.21). Here we adopt a method similar to [154] to evaluate the first order asymptotic expansion of the top Lyapunov exponent.

Proposition 7.2.2. $\mu_0(d\phi \times d\eta) = \frac{d\phi}{2\pi} \delta_0(d\eta)$ *is an ergodic measure for Eq.(7.21a).*

Proof. Note that $\mathfrak{L}_0$ behaves like deterministic: $\eta(t) \to 0$ due to the stable semigroup generated by $\mathcal{A}_s$. Rigorously,

$$\begin{aligned} \langle \mathfrak{L}_0 f, \mu_0 \rangle &= \iint_{\mathcal{S} \times \mathcal{H}_s} b_c^{\mathfrak{c}} \frac{\partial f}{\partial \phi} \frac{d\phi}{2\pi} \delta_0(d\eta) + \iint_{\mathcal{S} \times \mathcal{H}_s} \mathcal{A}_s \eta \frac{\partial f}{\partial \eta}(\phi, \eta) \frac{d\phi}{2\pi} \delta_0(d\eta) \\ &= \int_{\mathcal{S}} b_c^{\mathfrak{c}} \frac{df(\phi, 0)}{2\pi} - 0 = 0. \end{aligned}$$

□

To solve (7.21b), we first calculate R.H.S. of (7.21b).

Lemma 7.2.3.

$$\begin{aligned}-\langle \mathfrak{L}_1 f, \mu_0\rangle = &\int_{\mathcal{S}} \frac{\partial \Gamma(\phi,\eta)}{\partial \phi}(\phi,0) f(\phi,0)\frac{d\phi}{2\pi} - \int_{\mathcal{S}} \left\{ f(\phi,\eta)\,\mathrm{tr}[(D_\phi G_c^\phi)(D_\phi G_c^\phi)^*]\right\}_{\eta=0} \frac{d\phi}{2\pi} \\ &- \frac{1}{2}\int_{\mathcal{S}} \left\{ \sum_{k,j\in\mathbb{Z}_s} [G_s G_s^*]_{kj} f''_\eta(\phi,\eta;e_k,e_j)\right\}_{\eta=0} \frac{d\phi}{2\pi},\end{aligned} \tag{7.22}$$

where $[G_s G_s^*]_{kj} := \langle G_s G_s^* e_k, e_j\rangle$, *and* $f''_\eta(\phi,\eta;e_k,e_j)$ *is the Fréchet derivative w.r.t.* η *along* e_k *and* e_j.

Proof.

$$\begin{aligned}-\langle \mathfrak{L}_1 f, \mu_0\rangle &= -\left\langle [b_c^{\mathfrak{q}} + \Gamma(\phi,\eta)]\frac{\partial}{\partial\phi} f,\ \frac{d\phi}{2\pi}\delta_0(d\eta)\right\rangle - \frac{1}{2}\left\langle \frac{\partial^2 f}{\partial \phi^2}\,\mathrm{tr}[G_c^\phi (G_c^\phi)^*],\ \frac{d\phi}{2\pi}\delta_0(d\eta)\right\rangle \\ &\quad - \frac{1}{2}\left\langle \mathrm{tr}\left[\frac{\partial^2 f}{\partial\eta^2} G_s G_s^*\right],\ \frac{d\phi}{2\pi}\delta_0(d\eta)\right\rangle \\ &= \int_{\mathcal{S}} \frac{\partial \Gamma(\phi,\eta)}{\partial \phi}(\phi,0) f(\phi,0)\frac{d\phi}{2\pi} - \int_{\mathcal{S}} \left\{ f(\phi,\eta)\,\mathrm{tr}[(D_\phi G_c^\phi)(D_\phi G_c^\phi)^*]\right\}_{\eta=0} \frac{d\phi}{2\pi} \\ &\quad - \frac{1}{2}\int_{\mathcal{S}} \left\{ \sum_{k,j\in\mathbb{Z}_s} [G_s G_s^*]_{kj} \left\langle \frac{\partial^2 f}{\partial\eta^2} e_k, e_j\right\rangle\right\}_{\eta=0} \frac{d\phi}{2\pi} \\ &= \int_{\mathcal{S}} \frac{\partial \Gamma(\phi,\eta)}{\partial \phi}(\phi,0) f(\phi,0)\frac{d\phi}{2\pi} - \int_{\mathcal{S}} \left\{ f(\phi,\eta)\,\mathrm{tr}[(D_\phi G_c^\phi)(D_\phi G_c^\phi)^*]\right\}_{\eta=0} \frac{d\phi}{2\pi} \\ &\quad - \frac{1}{2}\int_{\mathcal{S}} \left\{ \sum_{k,j\in\mathbb{Z}_s} [G_s G_s^*]_{kj} f''_\eta(\phi,\eta;e_k,e_j)\right\}_{\eta=0} \frac{d\phi}{2\pi}\end{aligned}$$

□

Observing the above, we try ansatz of the following form to match the R.H.S. of (7.21b):

$$\mu_1(d\phi\times d\eta) = \frac{d\phi}{2\pi}\kappa(\phi)\delta_0(d\eta) + \frac{d\phi}{2\pi}\frac{\partial^2\delta_0}{\partial\eta^2}(\chi(\phi),h)(d\eta), \tag{7.23}$$

where $\kappa : \mathcal{S}^1 \to \mathbb{R}$, $\chi : \mathcal{S}^1 \to \mathcal{H}_{\alpha,s}$, and an arbitrary $h = \sum_{k\in\mathbb{Z}_s}\langle h, e_k\rangle e_k \neq 0$ (recall notation $\mathbb{Z}_s$ in Definition 6.1.4) that can make the calculation simple.

Remark 7.2.4. *The expression* $\frac{\partial^2\delta_0}{\partial\eta^2}(a,b)(d\eta)$ *appears in the above ansatz is the measure on* $\mathcal{H}_s$ *in the sense of directional distribution, where* a *and* b *are the directions. For Fréchet differentiable test function* $f(\phi,\eta)$, *we define the Fréchet derivatives*

$$f'(\phi,\eta;a) := \frac{\partial f}{\partial\eta}(\phi,\eta)(a) \;\; \text{and} \;\; f''(\phi,\eta;a,b) := \frac{\partial^2 f}{\partial\eta^2}(\phi,\eta)(a,b).$$

Then the distributional measure should satisfy

$$\left\langle f(\phi,\eta), \frac{\partial\delta_0}{\partial\eta}(a)(d\eta)\right\rangle = \left\langle \frac{\partial f}{\partial\eta}(\phi,\eta)(a), \delta_0(d\eta)\right\rangle = -f'(\phi,0;a),$$

and, likewise,

$$\left\langle f(\phi,\eta), \frac{\partial^2\delta_0}{\partial\eta^2}(a,b)(d\eta)\right\rangle = f''(\phi,0;a,b).$$

Proposition 7.2.5. *Let* $\kappa(\phi)$ *and* $\chi(\phi)$ *be the notions in the ansatz (7.23). Let* $\chi_k(\phi) := \langle\chi(\phi), e_k\rangle$ *for* $k\in\mathbb{Z}_s$ *and choose* $h=\sum_{k\in\mathbb{Z}_s} h_k e_k \in \mathcal{H}_{\alpha,s}$, *where* $h_k = \frac{1}{2^{|k|+2}(1-\rho_k)}$. *Suppose* $\kappa(\phi)$ *solves*

$$-\omega_c\frac{\partial\kappa}{\partial\phi}(\phi) = \frac{\partial\Gamma}{\partial\phi}(\phi,0) - \operatorname{tr}[(D_\phi G_c^\phi)(D_\phi G_c^\phi)^*](\phi,0). \tag{7.24}$$

and $\chi_k(\phi)$ *solves*

$$\left(b_c^{\mathfrak{c}}\frac{\partial}{\partial\phi} - \rho_k + 1\right)\chi_k(\phi) = -\sum_{j\in\mathbb{Z}_s}\{[G_sG_s^*]_{kj}\}_{\eta=0} \tag{7.25}$$

for all $k\in\mathbb{Z}_s$. *Then*

$$\mu_1(d\phi,d\eta) = \frac{d\phi}{2\pi}\kappa(\phi)\delta_0(d\eta) + \frac{d\phi}{2\pi}\frac{\partial^2\delta_0}{\partial\eta^2}(\chi(\phi),h)(d\eta).$$

Proof. For test function $f\in C^{1,2}(\mathcal{S}^1\times\mathcal{H}_{\alpha,s})$, we have

$$\begin{aligned}
\left\langle \mathfrak{L}_0 f,\ \frac{d\phi}{2\pi}\kappa(\phi)\delta_0(d\eta)\right\rangle &= \left\langle \left[b_c^{\mathfrak{c}}\frac{\partial}{\partial\phi} + \mathcal{A}_s\eta\frac{\partial}{\partial\eta}\right](f),\ \frac{d\phi}{2\pi}\kappa(\phi)\delta_0(d\eta)\right\rangle \\
&= -\int_{\mathcal{S}}\left\{b_c^{\mathfrak{c}}\frac{\partial\kappa}{\partial\phi}(\phi)f(\phi,\eta)\right\}_{\eta=0}\frac{d\phi}{2\pi} \\
&= \int_{\mathcal{S}}\frac{\partial\Gamma(\phi,\eta)}{\partial\phi}(\phi,0)f(\phi,0)\frac{d\phi}{2\pi} \\
&\quad -\int_{\mathcal{S}} f(\phi,0)\operatorname{tr}[(D_\phi G_c^\phi)(D_\phi G_c^\phi)^*](\phi,0)\frac{d\phi}{2\pi}.
\end{aligned} \tag{7.26}$$

For each $k \in \mathbb{Z}_s$,

$$
\begin{aligned}
&\left\langle \mathfrak{L}_0 f,\ \frac{d\phi}{2\pi}\frac{\partial^2 \delta_0}{\partial \eta^2}(\chi_k(\phi), h)(d\eta) \right\rangle = \left\langle b_c^{\mathfrak{c}} \frac{\partial f}{\partial \phi} + \mathcal{A}_s \eta \frac{\partial f}{\partial \eta},\ \frac{d\phi}{2\pi}\frac{\partial^2 \delta_0}{\partial \eta^2}(\chi_k(\phi), h)(d\eta) \right\rangle \\
&= - \left\langle \frac{\partial \delta_0}{\partial \eta}(h)(d\eta),\ -f'_\eta \left(\phi, \eta; b_c^{\mathfrak{c}} \frac{\partial \chi_k}{\partial \phi}(\phi) \right) \frac{d\phi}{2\pi} \right\rangle \\
&\quad - \left\langle \frac{\partial \delta_0}{\partial \eta}(h)(d\eta),\ [f''_\eta(\phi, \eta; \mathcal{A}_s \eta, \chi_k(\phi)) + f'_\eta(\phi, \eta; \mathcal{A}_s \chi_k(\phi))] \frac{d\phi}{2\pi} \right\rangle \\
&= \left\langle \delta_0(d\eta),\ -f''_\eta \left(\phi, \eta;\ b_c^{\mathfrak{c}} \frac{\partial \chi_k}{\partial \phi}(\phi), h \right) \frac{d\phi}{2\pi} \right\rangle + \left\langle \delta_0(d\eta),\ f'''_\eta(\phi, \eta; \mathcal{A}_s \eta, \chi_k(\phi), h) \frac{d\phi}{2\pi} \right\rangle \\
&\quad + \left\langle \delta_0(d\eta),\ [f''_\eta(\phi, \eta; \mathcal{A}_s h, \chi_k(\phi)) + f''_\eta(\phi, \eta; \mathcal{A}_s \chi_k(\phi), h)] \frac{d\phi}{2\pi} \right\rangle \\
&= - \int_{\mathcal{S}} \left\{ f''_\eta \left(\phi, \eta;\ (b_c^{\mathfrak{c}} \frac{\partial}{\partial \phi} - \mathcal{A}_s) \chi_k(\phi), h \right) - f''_\eta(\phi, \eta; \chi_k(\phi), \mathcal{A}_s h) \right\}_{\eta=0} \frac{d\phi}{2\pi}.
\end{aligned} \tag{7.27}
$$

By the hypothesis on $\chi_k(\phi)$ and h, the last line of the above can be expanded as

$$
\begin{aligned}
&- \sum_{j \in \mathbb{Z}_s} (1 - \rho_j) h_j \int_{\mathcal{S}} \left\{ f''_\eta \left(\phi, \eta;\ (b_c^{\mathfrak{c}} \frac{\partial}{\partial \phi} - \mathcal{A}_s) \chi_k(\phi), e_j \right) + f''_\eta(\phi, \eta; \chi_k(\phi), e_j) \right\}_{\eta=0} \frac{d\phi}{2\pi} \\
&= - \frac{1}{2} \int_{\mathcal{S}} \sum_{j \in \mathbb{Z}_s} \{[G_s G_s^*]_{kj}\}_{\eta=0} f''_\eta(\phi, 0; e_k, e_j) \frac{d\phi}{2\pi}
\end{aligned}
$$

Combining this and (7.27), we have

$$
\left\langle \mathfrak{L}_0 f,\ \frac{d\phi}{2\pi}\frac{\partial^2 \delta_0}{\partial \eta^2}(\chi(\phi), h)(d\eta) \right\rangle = -\frac{1}{2} \int_{\mathcal{S}} \sum_{k,j \in \mathbb{Z}_s} \{[G_s G_s^*]_{kj}\}_{\eta=0} f''_\eta(\phi, 0; e_k, e_j) \frac{d\phi}{2\pi}.
$$

Thus, by Lemma 7.2.3, we have

$$
\langle \mathfrak{L}_0 f,\ \mu_1 \rangle = - \langle \mathfrak{L}_1 f,\ \mu_0 \rangle ,
$$

which completes the proof. □

By solving (7.24) and (7.25), we are able to obtain the exact form of μ_1 as in (7.23).

Given the assumptions on G, the terms $\frac{\partial \Gamma}{\partial \phi}$ and $\operatorname{tr}[(G_c^\phi)(G_c^\phi)^*]$, $\operatorname{tr}[(D_\phi G_c^\phi)(D_\phi (G_c^\phi)^*]$ are Lipschitz continuous in ϕ, the existence of solutions is guaranteed. Based on the differentiability assumption of G, the solution to (7.24) is

$$
\kappa(\phi) = -\Gamma'(\phi) + \frac{1}{2} \operatorname{tr}[G_c^\phi (D_\phi G_c^\phi)^* + (D_\phi G_c^\phi)(G_c^\phi)^*](\phi, 0) \tag{7.28}
$$

Note that (7.25) can also be represented as

$$\left(b_c^{\mathfrak{c}}\frac{\partial}{\partial\phi} - \mathcal{A}_s + 1\right)\chi(\phi) = -\sum_{k,j\in\mathbb{Z}_s}\{[G_sG_s^*]_{kj}\}_{\eta=0} < \infty,$$

since the operator on the L.H.S. is both unbounded in $\mathcal{S}$ and $\mathcal{H}_{\alpha,s}$, we have

$$\chi(\phi) = -\left(b_c^{\mathfrak{c}}\frac{\partial}{\partial\phi} - \mathcal{A}_s + 1\right)^{-1}\sum_{k,j\in\mathbb{Z}_s}\{[G_sG_s^*]_{kj}\}_{\eta=0}$$

well defined. For short, we let $G_k^R + iG_k^I = -\sum_{j\in\mathbb{Z}_s}\{[G_sG_s^*]_{kj}\}_{\eta=0}$, then for each $k\in\mathbb{Z}_s$,

$$b_c^{\mathfrak{c}}\frac{\partial}{\partial\phi}\begin{bmatrix}\chi_k^R\\ \chi_k^I\end{bmatrix}(\phi) - \begin{bmatrix}a_k-1 & -b_k\\ b_k & a_k-1\end{bmatrix}\chi_k(\phi) = \begin{bmatrix}G_k^R\\ G_k^I\end{bmatrix},$$

and

$$\begin{bmatrix}\chi_k^R\\ \chi_k^I\end{bmatrix}(\phi) = \begin{bmatrix}a_k-1 & -b_k\\ b_k & a_k-1\end{bmatrix}^{-1}\left(e^{\vartheta_k}\begin{bmatrix}\cos(w_k) & -\sin(w_k)\\ \sin(w_k) & \cos(w_k)\end{bmatrix} - \begin{bmatrix}G_k^R\\ G_k^I\end{bmatrix}\right) \tag{7.29}$$

is the solution, where $\vartheta_k := \frac{(a_k-1)\phi}{b_c^{\mathfrak{c}}}$ and $w_k := \frac{(b_k-1)\phi}{b_c^{\mathfrak{c}}}$. Then

$$\chi(\phi) = \sum_{k\in\mathbb{Z}_s}(\chi_k^R(\phi) + i\chi_k^I(\phi))e_k.$$

Now that $\lambda^\varepsilon = \langle\mathcal{Q},\mu^\varepsilon\rangle = \langle\mathcal{Q},\mu_0\rangle + \varepsilon^2\langle\mathcal{Q},\mu_1\rangle + r(\varepsilon)$, where $r(\varepsilon)$ represents the remainder. By a similar argument as [10, Section 3] and [154, Lemma 4.3], we show that $r(\varepsilon) = \mathcal{O}(\varepsilon^3)$.

Proposition 7.2.6. *For the generator $\mathfrak{L}_\varepsilon = \frac{1}{\varepsilon^2}\mathfrak{L}_0 + \mathfrak{L}_1$, there exists functions F_0, F_1 on $\mathcal{S}^1\times\mathcal{H}_{\alpha,s}$ and functions $\tilde{f}_0, \tilde{f}_1$ that are independent of $\mathcal{S}\times\mathcal{H}_{\alpha,s}$, such that the sequence of Poisson equations*

$$\begin{aligned}\mathfrak{L}_0F_0 \qquad &= \zeta - \tilde{f}_0\\ \mathfrak{L}_0F_1 + \mathfrak{L}_1F_0 &= \quad -\tilde{f}_1\end{aligned} \tag{7.30}$$

are satisfied. As a consequence,

$$r(\varepsilon) = -\varepsilon^3[\langle\mathfrak{L}_1F_1,\mu^\varepsilon\rangle + \langle\mathfrak{L}_1(F_0+\varepsilon^2F_1),\mu_1\rangle - \langle\mathfrak{L}_1F_1,\mu_0+\varepsilon^2\mu_1\rangle]. \tag{7.31}$$

Given the boundedness $\sup_{\phi,\eta}\{|\mathcal{L}_1 F_1|, |\mathcal{L}_1 F_0|\} = C$, we immediately have $|r(\varepsilon)| < \mathcal{O}(\varepsilon^3)$. Notice that in (7.24),

$$\frac{1}{2}\operatorname{tr}[G_c^\phi (D_\phi G_c^\phi)^* + (D_\phi G_c^\phi)(G_c^\phi)^*](\phi, 0) = \Gamma(\phi),$$

combining with μ^ε, we have

$$\begin{aligned}\lambda^\varepsilon =&\langle \mathcal{Q}, \mu_0\rangle + \varepsilon^2 \langle \mathcal{Q}, \mu_1\rangle + \mathcal{O}(\varepsilon^3)\\ =&\frac{1}{2\pi}\int_{\mathcal{S}} \mathcal{Q}(\phi,0)d\phi + \frac{\varepsilon^2}{2\pi}\int_{\mathcal{S}} \mathcal{Q}(\phi,0)\kappa(\phi)d\phi\\ &+ \frac{\varepsilon^2}{2\pi}\int_{\mathcal{S}} \mathcal{Q}''\left(\phi, 0; \sum_{k\in\mathbb{Z}_s}(\chi_k^R(\phi)+\chi_k^I(\phi))e_k, \sum_{k\in\mathbb{Z}_s}\frac{1}{2^{|k|+2}(1-\rho_k)}e_k\right)d\phi\\ &+ \mathcal{O}(\varepsilon^3).\end{aligned} \tag{7.32}$$

7.3 Example

We illustrate the main result in application to the simplified stochastic Moore-Greitzer PDE model in the subspace $\mathcal{H}$ rather than the full space $\mathscr{H}$. We replace the state variable $\mathfrak{d}$ in (1.2) by u to keep the notation consistent with this chapter.

To investigate the local exponential a.s. stability of $u \in \mathcal{H}$ under multiplicative noise, we linearize the system for $\gamma = \gamma_c + \varepsilon^2\mathfrak{q}$ with some $\mathfrak{q} \in \mathbb{R}$ and concern the perturbation, then we obtain

$$du(t) = [\mathcal{A}(\gamma_c) + \varepsilon^2\mathfrak{q}\mathcal{A}'(\gamma_c)]u(t)dt + \varepsilon G(u(t))\cdot dW_t. \tag{7.33}$$

We recall that

$$\mathcal{A}(\gamma_c) = [A + Df_{u_e}(\gamma_c)]|_{\mathcal{H}} = \left(\frac{\nu}{2}\frac{\partial^2}{\partial\theta^2} - \frac{1}{2}\frac{\partial}{\partial\theta} + a\psi_c'\right)\mathrm{K}^{-1},$$

and

$$\mathcal{A}'(\gamma_c) = (a\psi_c''\Phi_c)\,\mathrm{K}^{-1},$$

where $\psi_c' := \psi_c'(\Phi_e(\gamma_c)) = \frac{3\iota}{2\mathfrak{M}}\left[1 - \left(\frac{\Phi_e(\gamma_c)}{\mathfrak{M}} - 1\right)^2\right]$, $\psi_c'' := \psi_c''(\Phi_e(\gamma_c)) = -\frac{3\iota}{\mathfrak{M}^2}(\frac{\Phi_e(\gamma_c)}{\mathfrak{M}} - 1)$ and $\Phi_c = \Phi_e'(\gamma_c)$.

We consider a special case when $\mathcal{V} = \mathcal{H}$, then the periodic cylindrical Wiener process is such that

$$W = \sum_{k\in\mathbb{Z}^+}(\beta_k + i\beta_{-k})e_k + \sum_{k\in\mathbb{Z}^-}(\beta_{-k} - i\beta_k)e_k,$$

where $\{\beta_k\}$ are i.i.d. $\mathscr{F}_t$-Brownian motions. Let $G(u)$ be such that

$$G(u)h = \sum_{k\in\mathbb{Z}_0} \langle e_k^*, uQ^{1/2}h\rangle e_k,$$

where Q is a trace-class operator with eigenvalues $q_k = |k|^{-(4\alpha+1)}$ for $\alpha > 0$. It can be verified that $G(u)$ is a Hilbert-Schmidt operator satisfying Assumption 6.1.12.

Note that

$$a_c^{\mathfrak{q}} = \frac{a}{1+am}\psi_c''\Phi_c,\ b_c^{\mathfrak{c}} = \frac{1}{2a},\ b_c^{\mathfrak{q}} = 0,$$

and the eigenvalues of $\mathcal{A}_s(\gamma_c) + \varepsilon^2\mathfrak{q}\mathcal{A}_s(\gamma_c)$ are given below:

$$\rho_k = \frac{\mathfrak{a}|k|}{|k|+\mathfrak{am}}\left(\psi_c' - \frac{\mathfrak{v}k^2}{2\mathfrak{a}} + \varepsilon^2\mathfrak{q}\frac{\mathfrak{a}|k|}{|k|+\mathfrak{am}}\psi_c''\Phi_c\right) + i\frac{|k|}{2\mathfrak{a}},$$

and for $k\in\mathbb{Z}_s$, $\psi_c' = \frac{\mathfrak{v}k^2}{2a}$. We recast the critical mode of the abstract linear equation as in (7.2a), then

$$G_c^R(u)\cdot = \sum_{k\in\mathbb{Z}_+} \sqrt{q_{k+1}}\frac{u_{-k}\langle e_{-k-1},\ \cdot\rangle + u_k\langle e_{k+1},\ \cdot\rangle}{2}$$

and

$$G_c^I(u)\cdot = \sum_{k\in\mathbb{Z}_+} \sqrt{q_{k+1}}\frac{u_{-k}\langle e_{-k-1},\ \cdot\rangle - u_k\langle e_{k+1},\ \cdot\rangle}{2},$$

where $u_k = \langle e_{-k}, u\rangle$. Consequently,

$$\Xi(\phi,0) = -\frac{\sin(2\phi)}{2}q_2,\quad \Gamma(\phi,0) = -\frac{\cos(2\phi)}{2}q_2,\quad \mathcal{Q}(\phi,0) = a_c^{\mathfrak{q}} - \Xi(\phi,0).$$

We also have $\sum_{j\in\mathbb{Z}_s}\{[G_sG_s^*]_{kj}\}_{\eta=0} = \frac{q_k}{2},\quad \forall k\in\mathbb{Z}_s\setminus\{\pm 2\}$ and $\sum_{j\in\mathbb{Z}_s}\{[G_sG_s^*]_{kj}\}_{\eta=0} = 0$ for $k=\pm 2$. Therefore,

$$\lambda^\varepsilon = \frac{a_c^{\mathfrak{q}}}{2\pi} + \frac{\varepsilon^2}{2\pi}\int_{\mathcal{S}}\mathcal{Q}''\left(\phi, 0; \sum_{k\in\mathbb{Z}_s}(\chi_k^R(\phi)+\chi_k^I(\phi))e_k, \sum_{k\in\mathbb{Z}_s}\frac{1}{2^{|k|+2}(1-\rho_k)}e_k\right)d\phi + \mathcal{O}(\varepsilon^3), \tag{7.34}$$

The second term is negative by the above calculation of χ_k. However, since $\int_0^{2\pi}\Xi(\phi,0)\frac{d\phi}{2\pi} = 0$ in this special case, the multiplicative noise does not stabilize the system given small values of ε.

Remark 7.3.1. *Note that $\int_0^{2\pi}\Xi(\phi,0)\frac{d\phi}{2\pi}$ cannot generally be expected to be 0. However, in our special case, we have set such that the basis of $\mathcal{V}\ni W$ is exactly the same as $\mathcal{H}$, the term Ξ is averaged out to be 0 by the invariant measure.*

7.4 Summary

In this chapter, we provide for the first time a derivation of an asymptotic expansion for the top Lyapunov exponent for SPDEs with multiplicative noise when the parameter moves slowly through the deteriministic Hopf bifurcation point. Instead of obtaining a dimension reduction using homogenization, the formula of top Lyapunov exponent was provided explicitly. We prove the existence of invariant measure on the product space of the unit sphere and the stable mode, and show the conditions for ergodicity. The disintegrated form of invariant measure as in Remark 7.1.9 explains the long term dependence of the stable marginals on the unit sphere of the critical mode. However, since it is difficult to solve, we derive an asymptotic expansion of the invariant measure of the disintegrate form and apply it in the Furstenberg–Khasminskii formula for the top Lyapunov exponent. The derived formula is illustrated in an example of a simplified stochastic Moore-Greitzer PDE model with multiplicative noise.

Chapter 8

Stochastic Hopf Bifurcations of Semilinear SPDEs with Small Multiplicative Noise

We have seen in Chapter 6 that, under certain proper hypothesis, in the neighborhood of the critical point γ_c, a semilinear SPDE with a cubic nonlinearity can be approximated by an SPDE with linear (linearized) stable dynamics whilst the critical modes keep the same form. The approximated invariant measure generates a relatively small error of order $\mathcal{O}(\varepsilon^2)$. We will take advantages from this approximation to build a connection between the fully linearized equation (7.1) or (7.4) to see how the change of the almost-sure stability of $\delta_{\mathbf{0}}$ can affect the long-term behavior of the amplitude of the critical modes within $\mathbb{R}^2 \setminus \{\mathbf{0}\}$.

Due to the coupling effect in the linearized equation and the nonlinear equation, the critical mode itself is not a Markov process. Hence, it is not sufficient to consider an invariant measure only for the critical mode regardless of its dominant amount of mass. However, we can study the changes of marginal distribution of the critical mode by looking at the invariant measure for the joint process $(\tilde{z}, \tilde{y})$. In particular, we would like to use the information from the moment Lyapunov exponent and its derivatives to quantify the recurrence (resp. null-recurrence) property in $(\mathbb{R}^2 \setminus \{\mathbf{0}\}) \times \mathcal{H}_{\alpha,s}$ when the sign of $\lambda^{\mathfrak{q},\varepsilon}$ becomes positive (resp. 0).

The noncompactness of the support of the stable marginal distribution brings difficulties when constructing the invariant measure using the Lyapunov exponents and moment Lyapunov exponents for the joint process $(\tilde{z}, \tilde{y})$. However, based on the analysis in Chapter 6, given the strongly exponential stability of the stable semigroup, the marginal $\tilde{y}$ only possesses a petite amount (of order $\mathcal{O}(\varepsilon^2)$) of probability in the tail of $\mathcal{H}_{\alpha,s}$ (recall Definition 7.0.1). We take this advantage and consider an approximate result on $\mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s}$, where $\mathcal{H}^n_{\alpha,s}$ is a truncated bounded domain $\mathcal{H}^n_{\alpha,s} := \{y \in \mathcal{H}_{\alpha,s} : \|y\|_\alpha < n\} \subset \mathcal{H}_{\alpha,s}$, for some sufficiently large n.

As a consequence, the expected time spent within a deleted neighborhood of $\mathbf{0} \in \mathbb{R}^2$ before exiting some small ball with larger radius r is utilized to construct the invariant measure on $(\mathbb{R}^2 \setminus \{\mathbf{0}\}) \times \mathcal{H}^n_{\alpha,s}$. We show that, as $\varepsilon \to 0$, the regularity of the marginal measure on $\mathbb{R}^2 \setminus \{\mathbf{0}\}$ with reasonably good accuracy as well as the approximate bifurcation diagram can be developed.

The notations and assumptions keep the same as Chapter 6 and 7. We revisit (6.22) and consider the amplitude $\tilde{z}$ of $\tilde{x}$ in (6.22)

$$\begin{aligned} d\tilde{x} &= \mathcal{A}^{\mathfrak{q}}_c \tilde{x} dt + F_c(\tilde{x} + \tilde{y})dt + G_c(\tilde{x} + \tilde{y})dW_t, \\ d\tilde{y} &= \varepsilon^{-2}\mathcal{A}_s \tilde{y} dt + G_s(\tilde{x} + \tilde{y})dW_t, \end{aligned}$$

where $\tilde{z} = \langle \mathfrak{h}^*, x\rangle$ and, consequently, $\tilde{x} = \tilde{z}\mathfrak{h} + \bar{\tilde{z}}\bar{\mathfrak{h}}$. We denote the real part and imaginary part of $\tilde{z}$ as $\tilde{z}_1 = \mathrm{Re}(\tilde{z})$ and $\tilde{z}_2 = \mathrm{Im}(\tilde{z})$. Similar to (7.4), we convert (6.22) into dynamics of $\tilde{z}_1$, $\tilde{z}_2$, and $\tilde{y}$ as follows,

$$d\begin{bmatrix}\tilde{z}_1\\ \tilde{z}_2\end{bmatrix} = \begin{bmatrix} a^{\mathfrak{q}}_c & -\varepsilon^{-2}b^{\mathfrak{c}}_c - b^{\mathfrak{q}}_c \\ \varepsilon^{-2}b^{\mathfrak{c}}_c + b^{\mathfrak{q}}_c & a^{\mathfrak{q}}_c\end{bmatrix}\begin{bmatrix}\tilde{z}_1\\ \tilde{z}_2\end{bmatrix} + \begin{bmatrix} F^R_c(\tilde{z}\mathfrak{h} + \bar{\tilde{z}}\bar{\mathfrak{h}} + \tilde{y}) \\ F^I_c(\tilde{z}\mathfrak{h} + \bar{\tilde{z}}\bar{\mathfrak{h}} + \tilde{y})\end{bmatrix} + \begin{bmatrix} G^R_c(\tilde{z}\mathfrak{h} + \bar{z}\bar{\mathfrak{h}} + \tilde{y}) \\ G^I_c(\tilde{z}\mathfrak{h} + \bar{\tilde{z}}\bar{\mathfrak{h}} + \tilde{y})\end{bmatrix} dW_t, \tag{8.1a}$$

$$d\tilde{y} = -\varepsilon^{-2}\mathcal{A}_s \tilde{y} dt + G_s(\tilde{z}\mathfrak{h} + \bar{\tilde{z}}\bar{\mathfrak{h}} + \tilde{y})dW_t, \tag{8.1b}$$

where

$$F^R_c(u) = \frac{\hat{F}_1(u) + \overline{\hat{F}_1(u)}}{2},\ F^I_c(u) = \frac{\hat{F}_1(u) - \overline{\hat{F}_1(u)}}{2}; \tag{8.2a}$$

$$\hat{F}_1(u)w := \langle \mathfrak{h}^*, F(u)\rangle, \quad \forall u \in \mathcal{H}. \tag{8.2b}$$

We also name $\tilde{F}(\tilde{z}_1, \tilde{z}_2, \tilde{y}) := [F^R_c, F^I_c](\tilde{z}\mathfrak{h} + \bar{\tilde{z}}\bar{\mathfrak{h}} + \tilde{y})$ for simplicity. The linearized equation (7.4) is given as below with a new numbering.

$$d\begin{bmatrix}z_1\\ z_2\end{bmatrix} = \begin{bmatrix} a^{\mathfrak{q}}_c & -\varepsilon^{-2}b^{\mathfrak{c}}_c - b^{\mathfrak{q}}_c \\ \varepsilon^{-2}b^{\mathfrak{c}}_c + b^{\mathfrak{q}}_c & a^{\mathfrak{q}}_c\end{bmatrix}\begin{bmatrix}z_1\\ z_2\end{bmatrix} + \begin{bmatrix} G^R_c(z\mathfrak{h} + \bar{z}\bar{\mathfrak{h}} + y) \\ G^I_c(z\mathfrak{h} + \bar{z}\bar{\mathfrak{h}} + y)\end{bmatrix} dW_t, \tag{8.3a}$$

$$dy = -\varepsilon^{-2}\mathcal{A}_s y dt + G_s(z\mathfrak{h} + \bar{z}\bar{\mathfrak{h}} + y)dW_t. \tag{8.3b}$$

Definition 8.0.1 (Notation)**.** *We make the following clarifications.*

(1) Since we intend to investigate the bifurcation behaviors and particularly the change marginal distribution of the critical modes subjected to the changes of $\mathfrak{q}$ and ε, we denote the solutions to (8.1a) and (8.3a), respectively, as $\tilde{z}^{\mathfrak{q},\varepsilon}$ and $z^{\mathfrak{q},\varepsilon}$ to explicitly indicate the dependence.

(2) We denote the solutions to (8.1b) and (8.3b) respectively as $\tilde{y}^\varepsilon$ and y^ε due to their strong stability regardless of $\mathfrak{q}$.

(3) The top Lyapunov exponent will be denoted as $\lambda^{\mathfrak{q},\varepsilon}$ instead of the shorthand notation λ^ε.

(4) The associated generator of (8.1) and (8.3) are denoted by $\widetilde{\mathfrak{L}_u^{\mathfrak{q},\varepsilon}}$ and $\mathbf{T}_u\mathfrak{L}^{\mathfrak{q},\varepsilon}$, respectively.

(5) We recall that the generator for the $(\mathfrak{p},\phi,\eta)$-coordinate expression of the linearized system, as in (7.6), is denoted by $\mathfrak{L}^{\mathfrak{q},\varepsilon}$ and was defined in (7.19).

(6) If the dependence on $\mathfrak{q}$ or ε is not emphasized, we cancel the superscript accordingly.

The generator $\mathfrak{L}^{\mathfrak{q},\varepsilon}$ will be used as an intermediate step to study the properties of $\mathbf{T}_u\mathfrak{L}^{\mathfrak{q},\varepsilon}$, and hence the properties of $\widetilde{\mathfrak{L}_u^{\mathfrak{q},\varepsilon}}$.

The main proof in this chapter deals with construction of suitable Lyapunov type functions that are bounded away from 0 and ∞. In order to control the growth of these Lyapunov type functions, which are related to the eigenvalue problem associated with the moment Lyapunov exponent, we define a bounded subdomain below:

Definition 8.0.2. *For each n that is not dependent on ε, we define*

(1) A bounded subdomain of $\mathcal{H}^n_{\alpha,s}$ (w.r.t. $\|\cdot\|_\alpha$ for some fixed $\alpha\in(0,1]$) as

$$\mathcal{H}^n_{\alpha,s} := \{y\in\mathcal{H}_{\alpha,s} : \|y\|_\alpha < n\}, \tag{8.4}$$

where $\mathcal{H}_{\alpha,s}$ is given in Definition 7.0.1.

(2) The stopping time τ_n to be the corresponding first exit time of $\mathcal{H}^n_{\alpha,s}$.

(3) The stopped processes, for (8.1) and (8.3) respectively, as

$$(\tilde{z}^{\mathfrak{q},\varepsilon,\tau_n},\tilde{y}^{\varepsilon,\tau_n}) := \{(\tilde{z}^{\mathfrak{q},\varepsilon}(t\wedge\tau_n),\tilde{y}^{\varepsilon}(t\wedge\tau_n)\}_{t\geq 0}$$

and

$$(z^{\mathfrak{q},\varepsilon,\tau_n},y^{\varepsilon,\tau_n}) := \{(z^{\mathfrak{q},\varepsilon}(t\wedge\tau_n),y^{\varepsilon}(t\wedge\tau_n)\}_{t\geq 0}.$$

Given $y\in\mathcal{H}^n_{\alpha,s}$, it can be verified that $\mathbf{P}^{z,y}[\tau_n>0]\equiv 1$ and $\mathbf{P}^{z,y,\varepsilon}[\tau_n=\infty]\to 1$ as $\varepsilon\to 0$ for any $\mathfrak{q}\in\mathbb{R}$. Furthermore, to ensure that the above mentioned eigenvalue problem is well defined, we work with function spaces with nice properties.

Definition 8.0.3. *For continuous functions $h : \mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s} \to \mathbb{R}$, we define the bounded Liptshcitz metric as*

$$\|h\|_{\mathrm{BL}} = \|h\|_\infty + \sup_{(\phi,\mathfrak{x})\neq(\phi,\mathfrak{y})} \frac{|h(\phi,\mathfrak{y}) - h(\phi,\mathfrak{x})|}{\|\mathfrak{y}-\mathfrak{x}\|_\alpha}.$$

For continuous mapping $H : \mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s} \to \mathcal{L}(\mathcal{H}^n_{\alpha,s}; \mathbb{R})$, where $\mathcal{L}$ denotes the space of linear bounded operators, we define the bounded Liptshcitz metric as

$$\|H\|_{\mathrm{BL}} = \sup_{(\phi,\eta)\in\mathcal{S}^1\times\mathcal{H}^n_{\alpha,s}} \|H(\phi,\eta)\|_{\mathcal{L}(\mathcal{H}^n_{\alpha,s};\mathbb{R})} + \sup_{(\phi,\mathfrak{x})\neq(\phi,\mathfrak{y})} \frac{\|H(\phi,\mathfrak{y}) - H(\phi,\mathfrak{x})\|_{\mathcal{L}(\mathcal{H}^n_{\alpha,s};\mathbb{R})}}{\|\mathfrak{y}-\mathfrak{x}\|_\alpha}.$$

The $\|\cdot\|_{\mathrm{BL}}$ metrics are defined in a similar way for continuous mappings $H : \mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s} \to \mathcal{L}^k(\prod_k \mathcal{H}^n_{\alpha,s}; \mathbb{R})$ with $k \geq 2$.

Definition 8.0.4. *We define $C^k_{\mathrm{BL}}(\mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s}; \mathbb{R})$ as a subspace of $C^k(\mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s}; \mathbb{R})$ with norm*

$$\|h\|_{C^k} := \|h\|_{\mathrm{BL}} + \sum_{i=1}^{k} \|D^{(i)}_\phi h\|_\infty + \sum_{i=1}^{k} \left\|D^{(i)}_\eta h\right\|_{\mathrm{BL}}, \quad h \in C^k_{\mathrm{BL}}(\mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s}; \mathbb{R}).$$

Assumption 8.0.5. *To this end, we suppose Assumption 6.1.6, 6.1.12 and 7.1.10 are satisfied. In addition, we assume that for any $A \in \mathscr{B}((\mathbb{R}^2 \setminus \{\mathbf{0}\}) \times \mathcal{H}^n_{\alpha,s})$, and any $(z,y) \in (\mathbb{R}^2 \setminus \{\mathbf{0}\}) \times \mathcal{H}^n_{\alpha,s}$ there exists a $t \in (0,\infty)$, such that $\mathbf{P}^{z,y,\varepsilon}[(\tilde{z}^{\mathfrak{q},\varepsilon}(t\wedge\tau_n), \tilde{y}^{\varepsilon}(t\wedge\tau_n)) \in A] > 0$.*

8.1 Moment Lyapunov Exponents and Approximations

8.1.1 Moment Lyapunov Exponents and the Approximate Eigenvalue Problems

Let $(\mathbf{\Omega}, \mathcal{F}, \mathbf{P})$ be the probability space. The moment Lyapunov exponent, which controls the p^{th}-moment stability of (8.3), is given as

$$\Lambda^{\mathfrak{q},\varepsilon}(p) = \lim_{t\to\infty} \frac{1}{t} \log \mathbf{E}|z^{\mathfrak{q},\varepsilon}(t)|^p \tag{8.5}$$

Recall from (7.6) that

$$\log|z^{\mathfrak{q},\varepsilon}(t)| = \mathfrak{p}(t) = \mathfrak{p}(0) + \int_0^t \mathcal{Q}^{\mathfrak{q}}(\phi(s),\eta(s))ds + \int_0^t G^{\mathfrak{p}}_c(\phi(s),\eta(s))dW_s,$$

and hence, for $z^{\mathfrak{q},\varepsilon}(0) = z_0$,

$$|z^{\mathfrak{q},\varepsilon}(t)|^p = |z_0|^p \exp\left\{p\int_0^t \mathcal{Q}^{\mathfrak{q}}(\phi(s),\eta(s))ds + p\int_0^t G_c^{\mathfrak{p}}(\phi(s),\eta(s))dW_s\right\}.$$

We attempt to define a continuous semigroup $\{\mathcal{T}_t^{\mathfrak{q},\varepsilon}(p)\}_{t\geq 0}$ on $\mathcal{S}^1 \times \mathcal{H}_{\alpha,s}$ by

$$\begin{aligned}&\mathcal{T}_t^{\mathfrak{q},\varepsilon}(p)f(\phi,\eta)\\ :=&\mathbf{E}^{\phi,\eta}\left[f(\phi(t),\eta(t))\exp\left\{p\int_0^t \mathcal{Q}^{\mathfrak{q}}(\phi(s),\eta(s))ds + p\int_0^t G_c^{\mathfrak{p}}(\phi(s),\eta(s))dW_s\right\}\right],\end{aligned}$$

with test functions $f \in C_b^2(\mathcal{S}^1 \times \mathcal{H}_{\alpha,s})$.

However, the operator $G(u)$ is assumed to be linear in u and hence unbounded in the direction of $y \in \mathcal{H}_{\alpha,s}$. Consequently, $\{\mathcal{T}_t^{\mathfrak{q},\varepsilon}(p)\}_{t\geq 0}$ is not bounded w.r.t. η and hence not a well-defined semigroup. To fix this problem, we consider a bounded open domain $\mathcal{H}_{\alpha,s}^n$, as introduced in Definition 8.0.2, for arbitrarily large n that is not dependent on ε. Now we define the process $\{z^{\mathfrak{q},\varepsilon,\tau_n}(t)\}_{t\geq 0}$ as

$$\log|z^{\mathfrak{q},\varepsilon,\tau_n}(t)| = \mathfrak{p}(0)+\int_0^t \mathcal{Q}^{\mathfrak{q}}(\phi(s\wedge\tau_n),\eta(s\wedge\tau_n)))ds+\int_0^t G_c^{\mathfrak{p}}(\phi(s\wedge\tau_n)),\eta(s\wedge\tau_n)))dW_s, \tag{8.6}$$

define the moment Lyapunov exponent of $z^{\mathfrak{q},\varepsilon,\tau_n}$ as

$$\Lambda_n^{\mathfrak{q},\varepsilon}(p) = \lim_{t\to\infty}\frac{1}{t}\log \mathbf{E}|z^{\mathfrak{q},\varepsilon,\tau_n}(t)|^p, \tag{8.7}$$

and the semigroup

$$\begin{aligned}&\mathcal{T}_{t,n}^{\mathfrak{q},\varepsilon}(p)f(\phi,\eta) :=\\ &\mathbf{E}^{\phi,\eta}\left[f(\phi(t\wedge\tau_n),\eta(t\wedge\tau_n))\exp\left\{p\int_0^t \mathcal{Q}^{\mathfrak{q}}(\phi(s\wedge\tau_n),\eta(s\wedge\tau_n))ds\right.\right.\\ &\left.\left.\qquad +p\int_0^t G_c^{\mathfrak{p}}(\phi(s\wedge\tau_n),\eta(s\wedge\tau_n))dW_s\right\}\right],\end{aligned} \tag{8.8}$$

for $f \in C^2(\mathcal{S}^1 \times \mathcal{H}_{\alpha,s}^n)$. Then for each fixed n, we have $\lim\limits_{t\to\infty}\frac{1}{t}\log\mathcal{T}_{t,n}^{\mathfrak{q},\varepsilon}(p)1(\phi,\eta) = \Lambda_n^{\mathfrak{q},\varepsilon}(p)$ as $\varepsilon \to 0$ for all p.

Remark 8.1.1. *Note that for any $t \geq 0$ and $(\phi, \eta) \in \mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s}$, by Lemma 7.1.4, we have*

$$\begin{aligned} \mathbf{P}^{\phi,\eta}[\tau_n \leq t] &= \mathbf{P}^{\phi,\eta}\left[\sup_{0\leq s\leq t} \|\eta(s)\|_\alpha \geq n\right] \\ &\leq \frac{\mathbf{E}^{\phi,\eta}[\sup_{0\leq s\leq t}\|\eta(s)\|_\alpha^p]}{n^p} \\ &\leq \frac{\|\eta\|_\alpha^p + \mathcal{O}(\varepsilon^p)}{n^p} =: \vartheta_n^\varepsilon(p), \quad \forall p \geq 2. \end{aligned} \tag{8.9}$$

Therefore, as $\varepsilon \to 0$, for sufficiently large p, we have $\vartheta_n^\varepsilon := \vartheta_n^\varepsilon(p) \ll 1$ arbitrarily small. Consequently, the event $\{\tau_n > t\}$ carries a probability close to 1 for each t.

In addition, the probability law of $\{(\phi(t \wedge \tau_n), \eta(t \wedge \tau_n))\}_{t\geq 0}$ agrees with the probability law of $\{(\phi(t), \eta(t))\}_{t\geq 0}$ on $\mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s}$ until τ_n whenever $(\phi, \eta) \in \mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s}$. In view of Proposition 7.1.5, the invariant probability measure $\mu_n^{\mathfrak{q},\varepsilon}$ for $\{(\phi(t\wedge\tau_n), \eta(t\wedge\tau_n))\}_{t\geq 0}$ is the the weak limit of

$$\left\{\frac{1}{t_k}\int_0^{t_k} \mathscr{L}(\phi(s\wedge\tau_n), \eta(s\wedge\tau_n))ds\right\}$$

as $t_k \to \infty$, i.e.

$$\frac{1}{t_k}\int_0^{t_k} \mathscr{L}(\phi(s\wedge\tau_n), \eta(s\wedge\tau_n))ds \rightharpoonup \mu_n^{\mathfrak{q},\varepsilon}. \tag{8.10}$$

On the other hand, for each sufficiently large t_k and for all $A \in \mathscr{B}(\mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s})$,

$$\begin{aligned} \frac{1}{t_k}\int_0^{t_k} \mathscr{L}(\phi(s), \eta(s))(A)ds &= \frac{1}{t_k}\left[\int_0^{t_k} \mathscr{L}(\phi(s\wedge\tau_n), \eta(s\wedge\tau_n))(A)\cdot\mathbf{P}[\tau_n > t_k]ds\right] \\ &= \frac{1}{t_k}\left[\int_0^{t_k} \mathscr{L}(\phi(s\wedge\tau_n), \eta(s\wedge\tau_n))(A)\right](1 - \vartheta_n^\varepsilon) \end{aligned}$$

and hence

$$-C\vartheta_n^\varepsilon \leq \frac{1}{t_k}\int_0^{t_k} \mathscr{L}(\phi(s), \eta(s))(A)ds - \frac{1}{t_k}\int_0^{t_k} \mathscr{L}(\phi(s\wedge\tau_n), \eta(s\wedge\tau_n))(A) \leq C\vartheta_n^\varepsilon. \tag{8.11}$$

As $t_k \to \infty$, the above indicates that $|\mu^{\mathfrak{q},\varepsilon}(A) - \mu_n^{\mathfrak{q},\varepsilon}(A)| = \mathcal{O}(\vartheta_n^\varepsilon)$ for all $A \in \mathscr{B}(\mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s})$.

Lemma 8.1.2. *For all $t > 0$ and $p \in \mathbb{R}$, the infinitesimal generator of $\{\mathcal{T}^{\mathfrak{q},\varepsilon}_{t,n}(p)\}_{t\geq 0}$ is given by*

$$\mathfrak{L}_p^{\mathfrak{q},\varepsilon} = \mathfrak{L}^{\mathfrak{q},\varepsilon} + p\mathfrak{X} + p\mathcal{Q}^{\mathfrak{q}} + \frac{p^2}{2}\mathfrak{R}, \quad \mathrm{dom}(\mathfrak{L}_p^{\mathfrak{q},\varepsilon}) = C^2(\mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s}), \tag{8.12}$$

where $\mathfrak{L}^{\mathfrak{q},\varepsilon}$ is the generator of processes $\{(\phi(t), \eta(t))\}_{t\geq 0}$ (on the truncated domain $(\mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s})$) with $(\phi(0), \eta(0)) \in \mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s}$ in (7.6), the quantities $\mathfrak{X}$ and $\mathfrak{R}$ are continuous on $\mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s}$ satisfying $\mathfrak{X}(\phi,\eta) = \frac{1}{2}\mathrm{tr}[G_c^{\mathfrak{p}}(G_c^{\phi})^* + G_c^{\phi}(G_c^{\mathfrak{p}})^*](\phi,\eta)$ and $\mathfrak{R}(\phi,\eta) = \mathrm{tr}[G_c^{\mathfrak{p}}(G_c^{\mathfrak{p}})^*](\phi,\eta)$.

Proof. The proof falls in a standard procedure as in [7, Lemma 2.3]. We just show the sketch. We first define a new measure $\mathbf{Q}$ using Girsanov's theorem by

$$\left.\frac{d\mathbf{Q}}{d\mathbf{P}}\right|_{\mathcal{F}_t} = \exp\left\{p\int_0^t G_c^{\mathfrak{p}}(\phi(s\wedge\tau_n),\eta(s\wedge\tau_n))dW_s - \frac{p^2}{2}\int_0^t \mathfrak{R}(\phi(s\wedge\tau_n),\eta(s\wedge\tau_n))ds\right\},$$

Then, under the new measure $\mathbf{Q}$, for some valid test function $f\in C^2(\mathcal{S}^1\times\mathcal{H}_s^n)$, we have

$$\begin{aligned}
&\mathcal{T}_{t,n}^{\mathfrak{q},\varepsilon}(p)f(\phi,\eta)\\
=&\mathbf{E}_{\mathbf{Q}}^{\phi,\eta}\left[f(\phi(t\wedge\tau_n),\eta(t\wedge\tau_n))\exp\left\{\int_0^t [p\mathcal{Q}^{\mathfrak{q}}(\phi(s\wedge\tau_n),\eta(s\wedge\tau_n))ds\right.\right.\\
&\left.\left.\qquad\qquad + \frac{p^2}{2}\mathfrak{R}(\phi(s\wedge\tau_n),\eta(s\wedge\tau_n))\right]ds\right\}\Bigg].
\end{aligned}$$

Note that the generator determines a limit behavior when $t\to 0$. Since $\mathbf{P}[\tau_n>0]=1$, when $t\to 0$, by the Feynman-Kac formula [60], the generator w.r.t. the measure $\mathbf{Q}$ is given as $A^{\mathfrak{q},\varepsilon}+p\mathcal{Q}^{\mathfrak{q}}+\frac{p^2}{2}\mathfrak{R}$, where $A^{\mathfrak{q},\varepsilon}$ is the generator of the process $\{\phi(t),\eta(t)\}_{t\geq 0}$ for $(\phi(0),\eta(0))\in\mathcal{S}^1\times\mathcal{H}_{\alpha,s}^n$. By Girsanov's theorem, for $(\phi(0),\eta(0))\in\mathcal{S}^1\times\mathcal{H}_{\alpha,s}^n$, the generator $A^{\mathfrak{q},\varepsilon}$ of the process $\{\phi(t),\eta(t)\}_{t\geq 0}$ w.r.t. the original measure $\mathbf{P}$ is converted to $\mathfrak{L}^{\mathfrak{q},\varepsilon}+p\mathfrak{X}$. Combining the above, the result can be obtained. □

Remark 8.1.3. *Note that $\mathfrak{L}_p^{\mathfrak{q},\varepsilon}$ is analytic w.r.t. p on the subdomain $\mathcal{S}^1\times\mathcal{H}_{\alpha,s}^n$.*

Therefore, for each p, if $\{\mathcal{T}_{t,n}^{\mathfrak{q},\varepsilon}(p)\}_{t\geq 0}$ is irreducible, positive, and compact from $E\subseteq C^2(\mathcal{S}^1\times\mathcal{H}_{\alpha,s}^n)$ to E, then there exists a strictly positive $\mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon}\in E\subseteq C^2(\mathcal{S}^1\times\mathcal{H}_{\alpha,s}^n)$ such that $\mathcal{T}_{t,n}^{\mathfrak{q},\varepsilon}(p)1$ converges to $\langle\mu_{p,n}^{\mathfrak{q},\varepsilon},1\rangle\mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon}$ as $t\to\infty$, which is equal to $\exp\{t\Lambda_n^{\mathfrak{q},\varepsilon}(p)\}\mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon}$. The consequences are further summarized in Lemma 8.1.6.

Remark 8.1.4. *Note that, by letting $\mathcal{Y}_t=f(\phi(t\wedge\tau_n),\eta(t\wedge\tau_n))$ and*

$$\mathcal{Z}_t=\exp\left\{\int_0^t\left[p\mathcal{Q}^{\mathfrak{q}}(\phi(s\wedge\tau_n),\eta(s\wedge\tau_n))ds+\frac{p^2}{2}\mathfrak{R}(\phi(s\wedge\tau_n),\eta(s\wedge\tau_n))\right]ds\right\},$$

we have $\mathcal{T}_{t,n}^{\mathfrak{q},\varepsilon}(p)f(\phi,\eta)=\mathbf{E}_{\mathbf{Q}}^{\phi,\eta}[\mathcal{Y}_t\mathcal{Z}_t]$, where $\mathbf{Q}$ is the same measure generated by Girsanov's transformation as in the proof of Lemma 8.1.2. Given $f\in C_{\mathrm{BL}}^k(\mathcal{S}^1\times\mathcal{H}_{\alpha,s}^n)$ and the Lipschitz continuity of $\mathcal{Q}$ and $\mathfrak{R}$ in η, let $\mathfrak{Y}(t,\mathfrak{y})$ represent the solution of $\eta(t\wedge\tau_n)$ with initial condition $\eta(0)=\mathfrak{y}$, we have $\mathcal{T}_{t,n}^{\mathfrak{q},\varepsilon}(p)f(\phi,\eta)\leq L\mathbf{E}_{\mathbf{Q}}^{\phi,\eta}[D_{\mathfrak{y}}\mathfrak{Y}(t,\mathfrak{y})\eta\cdot\exp\{\int_0^t D_{\mathfrak{y}}\mathfrak{Y}(s,\mathfrak{y})\eta ds\}]$. Since $\zeta^\eta(t):=D_{\mathfrak{y}}\mathfrak{Y}(t,\mathfrak{y})\eta$ uniquely solves (6.71), applying Itô's formula to ζ^η and checking its boundness in moments by brute force in a similar way as [135, Lemma 2.5], we can verify that $\mathbf{E}_{\mathbf{Q}}^{\phi,\eta}[\zeta^\eta(t)\exp\{\int_0^t\zeta^\eta(s)ds\}]\leq C\|\eta\|_\alpha$, which implies the Lipschitz continuity of $\mathcal{T}_{t,n}^{\mathfrak{q},\varepsilon}(p)f(\phi,\eta)$ in η.

For future references, the $\mathfrak{z}^{\mathfrak{q},\varepsilon}_{p,n}$ is introduced to construct Lyapunov-like functions on the Cartesian coordinates (z, y) of the form $|z|^p \cdot \mathfrak{z}^{\mathfrak{q},\varepsilon}_{p,n}(\phi(z), y/|z|)$ and estimate the up-crossing behaviors of the radius $|\tilde{z}^{\mathfrak{q},\varepsilon,\tau_n}|$ near $\mathbf{0}$ regardless the value of $(\phi(\tilde{z}^{\mathfrak{q},\varepsilon,\tau_n}), \tilde{y}^{\varepsilon,\tau_n}/|\tilde{z}^{\mathfrak{q},\varepsilon,\tau_n}|)$. For this purpose, we directly work with the precompact subspace $E := C^k_{\mathrm{BL}}(\mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s})$ with sufficiently large $k \in \mathbb{N}$ to ensure uniform bounds for $\mathfrak{z}^{\mathfrak{q},\varepsilon}_{p,n}(\phi, \eta)$ and for its derivatives. It suffices to show the irreducibility, positiveness[1] of the semigroup $\{\mathcal{T}^{\mathfrak{q},\varepsilon}_{t,n}(p)\}_{t\geq 0}$.

Proposition 8.1.5. *For each p, the semigroup $\{\mathcal{T}^{\mathfrak{q},\varepsilon}_{t,n}(p)\}$ is irreducible and positive at each $t \geq 0$.*

Proof. Note that a linear operator $T : E \to E$ is positive if $Tf \geq 0$ whenever $f \in E$ and $f \geq 0$. It is clear from the definition of $\{\mathcal{T}^{\mathfrak{q},\varepsilon}_{t,n}(p)\}_{t\geq 0}$ that, for each $t \geq 0$ and p, given $f \in C^2(\mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s})$ and $f \geq 0$, we have $\mathcal{T}^{\mathfrak{q},\varepsilon}_{t,n}(p)f \geq 0$. The positiveness follows.

To show the irreducibility, we rely on the assumptions given in the statement. We aim to verify that for each $0 < f \in C^2\mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s})$ and each strictly positive μ in the dual space of $C^2(\mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s})$ (which is the space Borel measure on $\mathscr{B}(\mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s})$), there exists a $t \in (0, \infty)$ such that

$$\langle \mu, \mathcal{T}^{\mathfrak{q},\varepsilon}_{t,n}(p)f\rangle > 0.$$

It is clear that the irreducibility of $\{\mathcal{T}^{\mathfrak{q},\varepsilon}_{t,n}\}_{t\geq 0}$ (for $p = 0$) is guaranteed by the assumptions, i.e., given any μ-measurable set $\Gamma \in \mathscr{B}(\mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s})$, there exists a positive time t such that $\mathbf{P}^{\phi,\eta}[(\phi(t), \eta(t)) \in \Gamma] > 0$ for any $(\phi, \eta) \in \Gamma$. Since for $\varepsilon \ll 1$, $\mathbf{P}^{\phi,\eta}[\tau_n > t] \to 1$, we also have $\mathbf{P}^{\phi,\eta}[(\phi(t \wedge \tau_n), \eta(t \wedge \tau_n)) \in \Gamma] > 0$ for some t given any initial point.

It can be easily verified that, for each $\Gamma \in \mathscr{B}(\mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s})$ and each strictly positive μ, there exists a time $t \in (0, \infty)$ such that

$$\langle \mathbb{1}_\Gamma, \mathcal{T}^{\mathfrak{q},\varepsilon}_{t,n}(p)^*\mu\rangle > 0. \tag{8.13}$$

Since $f(\mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s})$ is bounded subset in $\mathbb{R}$, we can find a finite open convering $G_1, G_2, \cdots, G_n$ for the range of f in $\mathbb{R}$ with arbitrarily small diameter $\epsilon > 0$. We slice the set $\cup^n_i G_i$ into disjoint subsets by defining $A_1 = G_1$ and $A_k = G_k \setminus (\cup^{k-1}_{i=1} G_i)$. Let $\Gamma_k = f^{-1}(A_k)$, then we can approximate f as a simple functions with finite summation $f_\epsilon = \sum^n_{k=1} a_k \mathbb{1}_{\Gamma_k}$. Thus, by (8.13), we can immediately verify that, given an arbitrary valid f and $\epsilon > 0$, we have

$$\langle f, \mathcal{T}^{\mathfrak{q},\varepsilon}_{t,n}(p)^*\mu\rangle - \varsigma(\epsilon) < \langle f_\epsilon, \mathcal{T}^{\mathfrak{q},\varepsilon}_{t,n}(p)^*\mu\rangle < \langle f, \mathcal{T}^{\mathfrak{q},\varepsilon}_{t,n}(p)^*\mu\rangle + \varsigma(\epsilon),$$

where $\varsigma(\epsilon)$ is generated by the evaluation of inner product (integral) and is continuous w.r.t. ϵ. We set ϵ arbitrarily small such that $\langle f, \mathcal{T}^{\mathfrak{q},\varepsilon}_{t,n}(p)^*\mu\rangle - \varsigma(\vartheta) > 0$, which concludes the proof. □

[1] These standard definitions will be provided in the proofs of such properties.

We now summarize the result in view of the generator $\mathfrak{L}_p^{\mathfrak{q},\varepsilon}$ of $\{\mathcal{T}_{t,n}^{\mathfrak{q},\varepsilon}(p)\}$ as follows.

Lemma 8.1.6. *For any compact interval that contains p, there exists a constant $K \in (0,\infty)$, such that there exists a strictly positive sufficiently smooth function $\mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon} : \mathcal{S}^1 \times \mathcal{H}_{\alpha,s}^n \to \mathbb{R}$ with $\frac{1}{K} \leq \|\mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon}\|_{C^2} \leq K$ and $\mathfrak{z}_{0,n}^{\mathfrak{q},\varepsilon} \equiv 1$ for all $\varepsilon \ll 1$ sufficiently small, and satisfying*

(1)

$$\mathfrak{L}_p^{\mathfrak{q},\varepsilon}\mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon} = \Lambda_n^{\mathfrak{q},\varepsilon}(p) \cdot \mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon}, \quad \langle \mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon},\ \mu_{p,n}^{\mathfrak{q},\varepsilon}\rangle = 1. \tag{8.14}$$

Moreover, by differentiating both sides of the eigenvalue problem in (8.14), and taking the appropriate scalar product with $\mu_{p,n}^{\mathfrak{q},\varepsilon}$, we have the following set of equations and associated solvability conditions for $\varphi_n^{\mathfrak{q},\varepsilon} = \frac{\partial \mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon}}{\partial p}\Big|_{p=0}$ and $\psi_n^{\mathfrak{q},\varepsilon} = \frac{\partial^2 \mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon}}{\partial p^2}\Big|_{p=0}$:

(2) $\mathfrak{L}^{\mathfrak{q},\varepsilon}\varphi_n^{\mathfrak{q},\varepsilon} = \lambda_n^{\mathfrak{q},\varepsilon} - \mathcal{Q}^{\mathfrak{q}} + \vartheta_n^{\mathfrak{q},\varepsilon}$, *with* $\|\varphi_n^{\mathfrak{q},\varepsilon}\|_{C^2} \leq K$ *and*

$$\lambda_n^{\mathfrak{q},\varepsilon} = \langle \mu_n^{\mathfrak{q},\varepsilon}, \mathcal{Q}^{\mathfrak{q}}\rangle = \langle \mu^{\mathfrak{q},\varepsilon}, \mathcal{Q}^{\mathfrak{q}}\rangle \pm \vartheta_n^{\mathfrak{q},\varepsilon} = \lambda^{\mathfrak{q},\varepsilon} \pm \vartheta_n^{\mathfrak{q},\varepsilon}$$

for and some smallness $\vartheta_n^{\mathfrak{q},\varepsilon}$. The quantity $\mu_n^{\mathfrak{q},\varepsilon}$ is an approximate invariant probability measure with the weak limit $\mu^{\mathfrak{q},\varepsilon}$ as $\varepsilon \to 0$.

(3) $\mathfrak{L}^{\mathfrak{q},\varepsilon}\psi_n^{\mathfrak{q},\varepsilon} = V_n^{\mathfrak{q},\varepsilon} - 2(\mathfrak{X} + \mathcal{Q}^{\mathfrak{q}} - \lambda_n^{\mathfrak{q},\varepsilon})\varphi_n^{\mathfrak{q},\varepsilon} - \mathfrak{R}$, *where* $V_n^{\mathfrak{q},\varepsilon} = \frac{d^2\Lambda_n^{\mathfrak{q},\varepsilon}(p)}{dp^2}\Big|_{p=0}$, $\|\psi_n^{\mathfrak{q},\varepsilon}\|_{C^2} \leq K$ *and*

$$V_n^{\mathfrak{q},\varepsilon} = 2\langle \mu_n^{\mathfrak{q},\varepsilon}, (\mathfrak{X} + \mathcal{Q}^{\mathfrak{q}} - \lambda_n^{\mathfrak{q},\varepsilon})\varphi_n^{\mathfrak{q},\varepsilon}\rangle + \langle \mu_n^{\mathfrak{q},\varepsilon}, \mathfrak{R}\rangle.$$

Proof. We show a sketch of proof due to the similarity to [9, Theorem 2]. Equation (8.14) comes straightforward from the asymptotic property of the associated semigroup $\{\mathcal{T}_{t,n}^{\mathfrak{q},\varepsilon}(p)\}_{t\geq 0}$. For each $\varepsilon \ll 1$, we can obtain the bounds (denoted as K^ε and $1/K^\varepsilon$) for the eigenfunction $\mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon}$, which is achieved by utilizing the analyticity w.r.t. p and the particular choice $\mathfrak{z}_0^{\mathfrak{q},\varepsilon} = 1$. The result (2) and (3) for each $\varepsilon \ll 1$ follows easily given (8.14). We observe that the term $\mathfrak{X}$ is generated by an absolute continuous measure $\mathbf{Q}$ w.r.t. $\mathbf{P}$ in Lemma 8.1.2, which should be always 0 considering the invariant measure $\mu_n^{\mathfrak{q},\varepsilon}$ and the cancellation should not affect the almost-sure exponential stability. We also have used the fact that $\lambda_n^{\mathfrak{q},\varepsilon} = \frac{d\Lambda_n^{\mathfrak{q},\varepsilon}(p)}{dp}\Big|_{p=0}$. Note that the quantity $\vartheta_n^{\mathfrak{q},\varepsilon}$ is caused by the rare event $\{\tau_n \leq t\}$. In view of Remark 8.1.1, due to the smallness of the stable process $\tilde{y}^\varepsilon$ or y^ε in p^{th}-moment, the convergence follows.

In addition, since for each p and ε, we can write $\mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon} = \mathfrak{z}_{0,n}^{\mathfrak{q},\varepsilon} + p\varphi_n^{\mathfrak{q},\varepsilon} + \frac{p^2}{2}\psi_n^{\mathfrak{q},\varepsilon} + \mathcal{O}(p^3)$, and, by the expansion w.r.t. ε ([9]), we have $\left\|\mathfrak{f}_n^{\mathfrak{q},\varepsilon'} - \mathfrak{f}_n^{\mathfrak{q},\varepsilon}\right\|_{\text{BL}} \leq \mathcal{O}(\varepsilon^2) + \mathcal{O}((\varepsilon')^2)$ where $\mathfrak{f} = \mathfrak{z}_0, \varphi, \psi$.

For any $\varepsilon \ll 1$ sufficiently small, the uniform bounds K and $1/K$ can be obtained based on this perturbation.

□

Remark 8.1.7. *We abuse the notation of $\vartheta_n^{\mathfrak{q},\varepsilon}$ just to indicate the smallness regardless of the way of being generated. This quantity will eventually be utilized for the asymptotic analysis as $\varepsilon \to 0$.*

We have seen the eigenvalue problems of $\Lambda_n^{\mathfrak{q},\varepsilon}(p)$ and $\lambda^{\mathfrak{q},\varepsilon}$ w.r.t. the associated generators $\mathfrak{L}_p^{\mathfrak{q},\varepsilon}$ and $\mathfrak{L}^{\mathfrak{q},\varepsilon}$ for the converted $(\mathfrak{p}, \phi, \eta)$ coordinates. Now we consider the amplitude of the process $(z^{\mathfrak{q},\varepsilon,\tau_n}, y^{\varepsilon,\tau_n})$ given the original linearized equation (8.3) in the Cartesian coordinate. The following corollary provides direct connections between the generator $\mathbf{T}_u\mathfrak{L}^{\mathfrak{q},\varepsilon}$ of (8.3) and the quantities $\Lambda_n^{\mathfrak{q},\varepsilon}(p)$, $\lambda^{\mathfrak{q},\varepsilon}$ and $V_n^{\mathfrak{q},\varepsilon}$. Such relations will finally be utilized to estimate the amplitude $|\tilde{z}^{\mathfrak{q},\varepsilon,\tau_n}|$ of the (approximated) nonlinear system with the generator $\widetilde{\mathfrak{L}_u^{\mathfrak{q},\varepsilon}}$.

Corollary 8.1.8. *Let $\mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon}$, $\varphi_n^{\mathfrak{q},\varepsilon}$ and $\psi_n^{\mathfrak{q},\varepsilon}$ be given in Lemma 8.1.6. Then*

$$\mathbf{T}_u\mathfrak{L}^{\mathfrak{q},\varepsilon}\left(|z|^p \cdot \mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon}\left(\phi(z), \frac{y}{|z|}\right)\right) = \Lambda_n^{\mathfrak{q},\varepsilon}(p) \cdot |z|^p \cdot \mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon}\left(\phi(z), \frac{y}{|z|}\right), \tag{8.15}$$

$$\mathbf{T}_u\mathfrak{L}^{\mathfrak{q},\varepsilon}\left(\varphi_n^{\mathfrak{q},\varepsilon}\left(\phi(z), \frac{y}{|z|}\right) + \log|z|\right) = \lambda^{\mathfrak{q},\varepsilon} + \vartheta_n^{\mathfrak{q},\varepsilon}, \tag{8.16}$$

and, if $\lambda^{\mathfrak{q},\varepsilon} = 0$,

$$\mathbf{T}_u\mathfrak{L}^{\mathfrak{q},\varepsilon}\left((\log|z|)^2 + 2\varphi^{\mathfrak{q},\varepsilon}\left(\phi(z), \frac{y}{|z|}\right)\log|z| + \psi^{\mathfrak{q},\varepsilon}\left(\phi(z), \frac{y}{|z|}\right)\right) = V_n^{\mathfrak{q},\varepsilon}. \tag{8.17}$$

Proof. Let $r = |z^{\mathfrak{q},\varepsilon}|$, $\phi = \phi(z^{\mathfrak{q},\varepsilon})$ and $\eta = y/|z^{\mathfrak{q},\varepsilon}|$. Then

$$\mathbf{T}_u\mathfrak{L}^{\mathfrak{q},\varepsilon}(r^p\mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon}(\phi,\eta)) = r^p\mathfrak{L}_p^{\mathfrak{q},\varepsilon}\mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon}(\phi,\eta).$$

Similarly, we have

$$\mathbf{T}_u\mathfrak{L}^{\mathfrak{q},\varepsilon}(r^p\log(r)\mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon}(\phi,\eta)) = r^p(\log(r))\mathfrak{L}_p^{\mathfrak{q},\varepsilon}\mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon}(\phi,\eta) + r^p(\mathfrak{X} + \mathcal{Q}^{\mathfrak{q}} + p\mathfrak{R})\mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon}(\phi,\eta). \tag{8.18}$$

Equation (8.16) and (8.17) is by differentiating both sides of the eigenvalue problem in (8.18), taking the appropriate scalar product with $\mu_{p,n}^{\mathfrak{q},\varepsilon}$, and setting $p = 0$. □

8.1.2 Continuous Dependence on the Parameter

We have seen in Lemma 8.1.6 and Corollary 8.1.8 about the approximated eigenvalue problems regarding $\lambda^{\mathfrak{q},\varepsilon}$ and $\Lambda_n^{\mathfrak{q},\varepsilon}(p)$ for each fixed ε and $\mathfrak{q}$. We have also determined by asymptotic expansion in Chapter 7 that there exists a critical $w = w(\varepsilon)$ at which $\lambda^{w,\varepsilon} = 0$ for each fixed ε.

For bifurcation analysis, we are also interested in how the truncated measure $\mu_n^{\mathfrak{q},\varepsilon}$ on the product space $\mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s}$ continuously dependent on $\mathfrak{q}$ within some small neighborhood. For future references, this continuous dependence allows us to investigate the regularities of a family of solutions $(\tilde{z}^{\mathfrak{q},\varepsilon,\tau_n}, \tilde{y}^{\varepsilon,\tau_n})$ for $\mathfrak{q}$ residing in a small interval, including the expected time spent within some annulus.

Lemma 8.1.9. *For each $w \in \mathbb{R}$, there exists a neighborhood $\mathscr{N}$ of w such that*

(1) $\lambda^{\mathfrak{q},\varepsilon}$ is continuous in $\mathfrak{q}$ for $\mathfrak{q} \in \mathscr{N}$.

(2) For each $p \in \mathbb{R}$, $\Lambda_n^{\mathfrak{q},\varepsilon}(p)$ is continuous in $\mathfrak{q}$ for $\mathfrak{q} \in \mathscr{N}$.

Proof. Note that there exists an $\mathscr{N}$ such that the generators $\mathfrak{L}_p^{\mathfrak{q},\varepsilon}$ and $\mathfrak{L}^{\mathfrak{q},\varepsilon}$ are continuous in $\mathfrak{q}$, therefore, the family of solutions $\{\mu_n^{\mathfrak{q},\varepsilon}\}_{\mathfrak{q}}$ is continuous w.r.t. $\mathfrak{q}$ in the weak topology (see Appendix E for details about the weak topology). By Lemma 8.1.6, $\lambda^{\mathfrak{q},\varepsilon} = \langle \mathcal{Q}^{\mathfrak{q}}, \mu_n^{\mathfrak{q},\varepsilon}\rangle \pm \vartheta_n^{\mathfrak{q},\varepsilon}$. Since there exists a neighborhood $\mathscr{N}$ such that $\vartheta_n^{\mathfrak{q},\varepsilon}$ is uniformly bounded and continuously dependent on $\mathfrak{q}$, and $\mathcal{Q}^{\mathfrak{q}}$ is Lipschitz continuous in $\mathfrak{q}$ for all $(\phi,\eta) \in \mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s}$, the continuity of $\{\lambda^{\mathfrak{q},\varepsilon}\}_{\mathfrak{q}}$ for $\mathfrak{q} \in \mathscr{N}$ follows by the above facts.

To verify the continuity of $\Lambda_n^{\mathfrak{q},\varepsilon}(p)$ in $\mathfrak{q}$ for each p, we are able to show that, for each $\varsigma > 0$, there exists a small neighborhood $\mathscr{N}$ such that

$$|\langle \mathfrak{L}_p^{\mathfrak{q},\varepsilon}\mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon}, \mu_{p,n}^{\mathfrak{q},\varepsilon}\rangle - \langle \mathfrak{L}_p^{w,\varepsilon}\mathfrak{z}_{p,n}^{w,\varepsilon}, \mu_{p,n}^{w,\varepsilon}\rangle| \le \varsigma.$$

This is again by the continuity of $\mathfrak{L}_p^{\mathfrak{q},\varepsilon}$ w.r.t the parameter $\mathfrak{q}$. Therefore, by (8.14), we have

$$|(\Lambda_n^{\mathfrak{q},\varepsilon}(p) - \Lambda_n^{w,\varepsilon}(p))| \le \varsigma.$$

The conclusion follows after this. □

8.2 D-Bifurcation

Due to the the couplings in the multiplicative noise, to investigate the D-bifurcation of the critical marginal, we still need to determine the evolution w.r.t. the joint probability measure.

The problem is reduced to quantifying the expected time spent within some annulus using submartingale/supermartinale arguments. To do this, we first construct Lyapunov-like functions with the same procedure as in [17]. The difference should only be regarding the strongly damped stable modes.

Lemma 8.2.1. *Let* $r = |z|, \phi = \arctan(z_2/z_1), \eta = y/r$ *be the coordinate rather than a stochastic process* $\{r(t), \phi(t \wedge \tau_n), \eta(t \wedge \tau_n)\}_{t \geq 0}$ *as in* (7.6). *Then, for any* $\varsigma \in (0, 1)$*, any fixed* $r \in (0, \varsigma)$ *and* $(\phi, \eta) \in \mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s}$*, there exists a neighborhood* $\mathscr{N}$ *of* w *and a constant* K *such that, for all* $\mathfrak{q} \in \mathscr{N}$*,*

$$|(\widetilde{\mathfrak{L}^{\mathfrak{q},\varepsilon}_u} - \mathbf{T}_u \mathfrak{L}^{\mathfrak{q},\varepsilon}) f(r) h(\phi, \eta)| \leq K(r^2 f(r) + r^3 |f'(r)|) \cdot \|h\|_{C^2}, \tag{8.19}$$

where $f \in C^2(0, \varsigma)$ *(w.r.t.* $\|\cdot\|_\infty$*) and* $h \in C^2(\mathcal{S}^1 \times \mathcal{H}^n_{\alpha,s})$*.*

Proof. The proof follows directly by the comparison between the two generators at the same position of (r, ϕ, η) in a small neighborhood of r. Note that $\frac{df}{dz_i} = f'(r)\frac{z_i}{r}$ for $i = 1, 2$,

$$\frac{dh}{dz_1} = -h_\phi(\phi, \eta)\frac{\sin(\phi)}{r} - h_\eta(\phi, \eta)\frac{\eta\cos(\phi)}{r}, \quad \frac{dh}{dz_2} = h_\phi(\phi, \eta)\frac{\cos(\phi)}{r} - h_\eta(\phi, \eta)\frac{\eta\sin(\phi)}{r}$$

and

$$\frac{d(fh)}{dz_i} = h\frac{df}{dz_i} + f\frac{dh}{dz_i}, \;\; i = 1, 2.$$

Let $\mathfrak{L}_y h(\phi, \eta) = \mathcal{A}_s \eta h_\eta(\phi, \eta) + \frac{1}{2}\operatorname{tr}[\frac{\partial^2 h(\phi,\eta)}{\partial \eta^2} G_s(\phi + \eta) G_s^*(\phi + \eta)]$. Then,

$$\mathbf{T}_u \mathfrak{L}^{\mathfrak{q},\varepsilon} f(r) h(\phi, \eta) = \mathfrak{L}_y h(\phi, \eta) + \sum_i A_i \frac{d(fh)}{dz_i} + \frac{1}{2}\operatorname{tr}\left[G^{\mathfrak{p}}_c (G^{\mathfrak{p}}_c)^* \frac{\partial^2 (fh)}{\partial z^2}\right],$$

where $A_1 = a^{\mathfrak{q}}_c z_1 - (\varepsilon^{-2} b^{\mathfrak{c}}_c + b^{\mathfrak{q}}_c) z_2$ and $A_2 = a^{\mathfrak{q}}_c z_2 + (\varepsilon^{-2} b^{\mathfrak{c}}_c + b^{\mathfrak{q}}_c) z_1$. On the other hand,

$$\widetilde{\mathfrak{L}^{\mathfrak{q},\varepsilon}_u} f(r) h(\phi, \eta) = \mathfrak{L}_y h(\phi, \eta) + \sum_i \tilde{A}_i \frac{d(fh)}{dz_i} + \frac{1}{2}\operatorname{tr}\left[G^{\mathfrak{p}}_c (G^{\mathfrak{p}}_c)^* \frac{\partial^2 (fh)}{\partial z^2}\right],$$

where $\tilde{A}_1 = A_1 + F^R_c(r\cos(\phi)\mathfrak{h} + r\sin(\phi)\bar{\mathfrak{h}} + r\eta)$ and $\tilde{A}_2 = A_2 + F^I_c(r\cos(\phi)\mathfrak{h} + r\sin(\phi)\bar{\mathfrak{h}} + r\eta)$. Note that F^R_c and F^I_c are cubic nonlinearities with properties in Assumption 6.1.9, and for each

(r, ϕ, η), there exists a $K > 0$ such that both $\|F_c^R\|_\infty \le r^3 K$ and $\|F_c^I\|_\infty \le r^3 K$. Therefore,

$$\begin{aligned}
|(\widetilde{\mathfrak{L}_u^{\mathfrak{q},\varepsilon}} - \mathbf{T}_u \mathfrak{L}^{\mathfrak{q},\varepsilon}) f(r) h(\phi, \eta)| &= \left| F_c^R \cdot \frac{d(fh)}{dz_1} + F_c^I \cdot \frac{d(fh)}{dz_2} \right| \\
&\le K r^3 \left| \frac{d(fh)}{dz_1} + \frac{d(fh)}{dz_2} \right| \\
&= K r^3 \left| h(\phi, \eta) f'(r)(\cos(\phi) + \sin(\phi)) + f(r) \frac{dh}{dz_1} + f(r) \frac{dh}{dz_2} \right| \\
&\le K |r^2 f(r) + r^3 f'(r)| \|h\|_{C^2}.
\end{aligned}$$

We notice that, unlike the settings in [17], the only difference in the nonlinear equation and the linearized equation comes from F_c^R and F_c^I, whereas the multiplicative noise in the critical modes and the entire stable modes keep the same. Therefore, the second order derivatives $\frac{\partial^2 (fh)}{\partial z_i \partial z_j}$ for $i, j = 1, 2$ and the trace term in $\mathfrak{L}_y h$ should not contribute to the difference. □

Remark 8.2.2. *The small r and the boundedness of h are to match the conditions in application of Dynkin's formula (see Remark 3.1.13 for details) up to a first exit time of a bounded domain in $r\mathbb{B} \times \mathcal{H}_{\alpha,s}^n$. On the other hand, Lemma 8.2.1 is not intended to compare the difference of $f \otimes h$ acting on some processes generated by (8.3) and (8.1) at a deterministic time. The purpose is to quantify that, at each fixed position in the field, the quantity $\widetilde{\mathfrak{L}_u^{\mathfrak{q},\varepsilon}} f(r) h(\phi, \eta)$ provides a different evolution rate compared to the linear dynamics. Such a difference can be bounded by a term in proportion to r regardless of the input in $\mathcal{S}^1$ and $\mathcal{H}_{\alpha,s}^n$.*

Combining Lemma 8.2.1 and Corollary 8.1.8, we are able to construct Lyapunov-like functions which will be used for the sub/super-martingale arguments. The following results are rephrased from [17, Proposition 4.13].

Proposition 8.2.3. *Let Assumption 7.1.10 be satisfied. Then,*

(1) For each $p \neq 0$ and $\vartheta > 0$, there exist $\varsigma > 0$, a constant K (dependent on n), and a neighborhood $\mathscr{N}$ of w such that for every $\mathfrak{q} \in \mathscr{N}$ and $\varepsilon \ll 1$ sufficiently small, there exist smooth functions $\bar{\mathfrak{z}}_{p,\pm}^{\mathfrak{q},\varepsilon} : \mathbb{R}^2 \times \mathcal{H}_{\alpha,s}^n \to \mathbb{R}$ as perturbations of $\bar{\mathfrak{z}}_p^{\mathfrak{q},\varepsilon}(z, y) := |z|^p \cdot \mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon}(\phi(z), y/|z|)$[2] satisfying

$$(\widetilde{\mathfrak{L}_u^{\mathfrak{q},\varepsilon}} - \Lambda_n^{\mathfrak{q},\varepsilon}(p)) \bar{\mathfrak{z}}_{p,+}^{\mathfrak{q},\varepsilon} \ge 0 \ge (\widetilde{\mathfrak{L}_u^{\mathfrak{q},\varepsilon}} - \Lambda_n^{\mathfrak{q},\varepsilon}(p)) \bar{\mathfrak{z}}_{p,-}^{\mathfrak{q},\varepsilon} \tag{8.20}$$

[2] Recall that ϕ is the projection of z onto the $\mathcal{S}^1$ space. Therefore, the function $|z|^p \cdot \mathfrak{z}_p^{\mathfrak{q},\varepsilon}(\phi(z), y/|z|)$ is a function of z and y.

for all $(z,y) \in \mathbb{B}'(\varsigma)$*, where* $\mathbb{B}'(\varsigma) := \{(z,y) \in \mathbb{R}^2 \times \mathcal{H}^n_{\alpha,s} : 0 < |z| \le \varsigma\}$*. Furthermore, we have*

$$\frac{1}{K}(|z|)^p \le \bar{\mathfrak{z}}^{\mathfrak{q},\varepsilon}_{p,\pm}(z,y) \le K(|z|)^p, \quad \forall y \in \mathcal{H}^n_{\alpha,s}.$$

(2) There exist $\varsigma > 0$*, a constant* K*, and a neighborhood* $\mathcal{N}$ *of* w *such that, for every* $\mathfrak{q} \in \mathcal{N}$*, there exist smooth functions* $\varphi^{\mathfrak{q},\varepsilon}_{\pm} : \mathbb{B}'(\varsigma) \to \mathbb{R}$ *for all* $(z,y) \in \mathbb{B}'(\varsigma)$ *satisfying*

$$\widetilde{\mathfrak{L}^{\mathfrak{q},\varepsilon}_u}\varphi^{\mathfrak{q},\varepsilon}_{+}(z,y) \ge \lambda^{\mathfrak{q},\varepsilon} \ge \widetilde{\mathfrak{L}^{\mathfrak{q},\varepsilon}_u}\varphi^{\mathfrak{q},\varepsilon}_{-}(z,y) \tag{8.21}$$

(3) If $\lambda^{\mathfrak{q},\varepsilon} = 0$*, then for all* $\vartheta > 0$ *there exist an* $\varsigma > 0$*,* $K < \infty$ *and and smooth functions* $\psi^{\mathfrak{q},\varepsilon}_{\pm} : \mathbb{B}'(\varsigma) \to \mathbb{R}$ *such that*

$$\widetilde{\mathfrak{L}^{\mathfrak{q},\varepsilon}_u}\psi^{\mathfrak{q},\varepsilon}_{+}(z,y) \ge V^{\mathfrak{q},\varepsilon} \ge \widetilde{\mathfrak{L}^{\mathfrak{q},\varepsilon}_u}\psi^{\mathfrak{q},\varepsilon}_{-}(z,y) \tag{8.22}$$

and $|\psi^{\mathfrak{q},\varepsilon}_{\pm}(z,y) - (\log|z|)^2| \le K|\log|z||$ *for* $(z,y) \in \mathbb{B}'(\varsigma)$*.*

Remark 8.2.4. *The above proposition is intended for verifying the supper/sub-martingale properties of* $(\tilde{z}^{\mathfrak{q},\varepsilon,\tau_n}, \tilde{y}^{\varepsilon,\tau_n})$ *near* $\mathbf{0} \times \mathcal{H}^n_{\alpha,s}$*. The* y *argument plays a role as a 'dummy variable' that does not influence the estimation along* $z \in \mathbb{R}^2$*. This independence in* y *is inherited directly from Lemma 8.2.1.*

The proof for any fixed $\mathfrak{q} \in \mathcal{N}$ *is given in [18]. We do not show the details due to the similarity. The idea is to quantify a feasible range of* ς *based on Lemma 8.2.1 and Corollary 8.1.8 such that*

- ✧ *There exists a sufficiently small* k *and* $\bar{\mathfrak{z}}^{\mathfrak{q},\varepsilon}_{p,\pm} = \bar{\mathfrak{z}}^{\mathfrak{q},\varepsilon}_{p} \pm k|z|^{p'}\mathfrak{z}^{\mathfrak{q},\varepsilon}_{p',n}$ *satisfying the requirement as in (1) of the above, where* p' *can be carefully designed within a neighborhood of* p*. The associated moment Lyapunov exponents of* $\bar{\mathfrak{z}}^{\mathfrak{q},\varepsilon}_{p,\pm}$ *also only create small difference. The estimation relies on the following fact:*

$$\begin{aligned}
&(\widetilde{\mathfrak{L}^{\mathfrak{q},\varepsilon}_u} - \Lambda^{\mathfrak{q},\varepsilon}_n(p))\bar{\mathfrak{z}}^{\mathfrak{q},\varepsilon}_{p,\pm} \\
=&(\widetilde{\mathfrak{L}^{\mathfrak{q},\varepsilon}_u} - \Lambda^{\mathfrak{q},\varepsilon}_n(p))\bar{\mathfrak{z}}^{\mathfrak{q},\varepsilon}_{p} \pm (\widetilde{\mathfrak{L}^{\mathfrak{q},\varepsilon}_u} - \Lambda^{\mathfrak{q},\varepsilon}_n(p))(k|z|^{p'}\mathfrak{z}^{\mathfrak{q},\varepsilon}_{p',n}) \\
=&(\widetilde{\mathfrak{L}^{\mathfrak{q},\varepsilon}_u} - \mathbf{T}_u\mathfrak{L}^{\mathfrak{q},\varepsilon})\bar{\mathfrak{z}}^{\mathfrak{q},\varepsilon}_{p} \pm (\widetilde{\mathfrak{L}^{\mathfrak{q},\varepsilon}_u} - \Lambda^{\mathfrak{q},\varepsilon}_n(p') + \Lambda^{\mathfrak{q},\varepsilon}_n(p') - \Lambda^{\mathfrak{q},\varepsilon}_n(p))(k|z|^{p'}\mathfrak{z}^{\mathfrak{q},\varepsilon}_{p',n}) \\
=&(\widetilde{\mathfrak{L}^{\mathfrak{q},\varepsilon}_u} - \mathbf{T}_u\mathfrak{L}^{\mathfrak{q},\varepsilon})\bar{\mathfrak{z}}^{\mathfrak{q},\varepsilon}_{p} \pm (\widetilde{\mathfrak{L}^{\mathfrak{q},\varepsilon}_u} - \mathbf{T}_u\mathfrak{L}^{\mathfrak{q},\varepsilon})(k|z|^{p'}\mathfrak{z}^{\mathfrak{q},\varepsilon}_{p',n}) \\
&\pm (\Lambda^{\mathfrak{q},\varepsilon}_n(p') - \Lambda^{\mathfrak{q},\varepsilon}_n(p))(k|z|^{p'}\mathfrak{z}^{\mathfrak{q},\varepsilon}_{p',n})
\end{aligned} \tag{8.23}$$

Apart from the approximated quantities $\Lambda^{\mathfrak{q},\varepsilon}_n(p)$ *and* $\Lambda^{\mathfrak{q},\varepsilon}_n(p')$*, the rest of the above estimations are the same as in the proof of [18, Theorem 3.18]. In particular, by Lemma 8.2.1,* $\widetilde{\mathfrak{L}^{\mathfrak{q},\varepsilon}_u}$ *creates a distance with* $\mathbf{T}_u\mathfrak{L}^{\mathfrak{q},\varepsilon}$ *that is not dependent on* p *and* ε *given any valid test function. For a sufficiently small* $\varsigma > 0$ *and* $r \in (0,\varsigma)$*, the above construction is valid for (1).*

✧ *By differentiating both sides of the eigenvalue problems, and setting $p = 0$, the above construction procedure can be applied to (2) and (3) of the statements. In particular, $\varphi_{\pm}^{\mathfrak{q},\varepsilon}$ is selected based on*

$$\varphi_{\pm}^{\mathfrak{q},\varepsilon} = \log|z| + (\bar{\mathfrak{z}}_0^{\mathfrak{q},\varepsilon})' \pm k|z|^{p'} \mathfrak{z}_{p',n}^{\mathfrak{q},\varepsilon}$$

as well as a similar derivation as in (8.23). The effect of $\vartheta_n^{\mathfrak{q},\varepsilon}$ that appears in (2) can also be reduced by carefully choosing the parameters in this step.

The fact that constants ς and K can be selected locally uniform w.r.t. $\mathfrak{q}$ is by the continuous dependence on $\mathfrak{q}$ from Section 8.1.2. The proof follows [17, Proposition 4.13] without extra ingredients.

Taking the advantages of the above construction, the problem can be reduced to qualify the recurrence and null recurrence properties w.r.t. $(\mathbb{R}^2 \setminus \{\mathbf{0}\}) \times \mathcal{H}_{\alpha,s}^n$ of the joint process $\{(\tilde{z}^{\mathfrak{q},\varepsilon}(t \wedge \tau_n), \tilde{y}^{\varepsilon}(t \wedge \tau_n))\}_{t \geq 0}$ when $\lambda^{\mathfrak{q},\varepsilon} \geq 0$. The key is to use supper/sub-martingale arguments to estimate the expected occupation time near $\mathbf{0} \times \mathcal{H}_{\alpha,s}^n$, which will be utilized later to construct invariant measures on $\mathscr{B}(\mathbb{R}^2 \setminus \{\mathbf{0}\}) \times \mathcal{H}_{\alpha,s}^n$. The effect of $(\phi(\tilde{z}^{\mathfrak{q},\varepsilon,\tau_n}), \tilde{y}^{\varepsilon}/|\tilde{z}^{\mathfrak{q},\varepsilon,\tau_n}|)$ finally contribute to the top Lyapunov exponent $\lambda^{\mathfrak{q},\varepsilon}$, which only determines the evolution of $|\tilde{z}^{\mathfrak{q},\varepsilon,\tau_n}|$. We state the key ingredients and procedures for the final construction of invariant measures on $\mathscr{B}(\mathbb{R}^2 \setminus \{\mathbf{0}\}) \times \mathcal{H}_{\alpha,s}^n$, the proofs should be the same as [17].

Proposition 8.2.5. *On $\{\tau_n = \infty\}$, there exists $\varsigma > 0$, $K < \infty$ and a neighborhood $\mathscr{N}$ of w such that $\mathbf{P}^{z,y,\mathfrak{q}}[\tau_\vartheta \wedge \tau_R < \infty] = 1$ whenever $0 < \vartheta < |z| < R < \varsigma$[3], $y \in \mathcal{H}_{\alpha,s}^n$ and $\mathfrak{q} \in \mathscr{N}$, where*

$$\tau_\vartheta = \inf\{t > 0, (\tilde{z}^{\mathfrak{q},\varepsilon}(t \wedge \tau_n), \tilde{y}^{\varepsilon}(t \wedge \tau_n)) \in \partial\overline{\mathbb{B}}_\vartheta \times \mathcal{H}_{\alpha,s}^n\}$$

and

$$\tau_R = \inf\{t > 0, (\tilde{z}^{\mathfrak{q},\varepsilon}(t \wedge \tau_n), \tilde{y}^{\varepsilon}(t \wedge \tau_n)) \in \partial\overline{\mathbb{B}}_R \times \mathcal{H}_{\alpha,s}^n\}.$$

In particular, if $\lambda^{\mathfrak{q},\varepsilon} > 0$ for all $\mathfrak{q} \in \mathscr{N}$, then

$$\frac{1}{\lambda^{\mathfrak{q},\varepsilon}}\left[\log\frac{R}{|z|} - 2K\right] \leq \mathbf{E}^{z,y,\mathfrak{q}}[\tau_R] \leq \frac{1}{\lambda^{\mathfrak{q},\varepsilon}}\left[\log\frac{R}{|z|} + 2K\right].$$

In addition, for $\mathfrak{d} > 0$ such that $\Lambda_n^{\mathfrak{q},\vartheta}(-\mathfrak{d}) = 0$, there exists a $k \in (0,1)$ such that

$$\frac{1}{K}\left(\frac{\vartheta}{|z|}\right)^{\mathfrak{d}} \leq \mathbf{E}^{z,y,\mathfrak{q}}\left(\int_0^{\tau_R} \mathbb{1}_{\{\mathbb{B}_\vartheta \times \mathcal{H}_s^n\}}(\tilde{z}_t^{\mathfrak{q},\varepsilon,\tau_n}, \tilde{y}_t^{\varepsilon,\tau_n})dt\right) \leq K\left(\frac{\vartheta}{|z|}\right)^{\mathfrak{d}}.$$

whenever $0 < \vartheta < |z| < kR < k\varsigma$ and $y \in \mathcal{H}_{\alpha,s}^n$.

[3] Recall that $\tilde{z}_0^{\mathfrak{q},\varepsilon} = \langle \mathfrak{h}^*, x \rangle$.

Proposition 8.2.6. *Let w be such that $\lambda^{w,\varepsilon} = 0$ and $\Lambda_n^{w,\varepsilon}(p) \neq 0$. Let $\varsigma > 0$, $K < \infty$ and $\mathscr{N}$ be as in Lemma 8.2.5. Fix $0 < \vartheta < R < \varsigma$ with $\vartheta R < 1$. Then, on $\{\tau_n = \infty\}$,*

(1) $\mathbf{E}^{z,y,\varepsilon}[\tau_R] = \infty$ whenever $0 < |z| < R < \varsigma$ and $y \in \mathcal{H}^n_{\alpha,s}$.

(2) there exists $k \in (0,1)$ such that, whenever $0 < |z| < kR < k\varsigma$ and $y \in \mathcal{H}^n_{\alpha,s}$,

$$\begin{aligned}\frac{2}{V_n^{w,\varepsilon}}\left[\log\left(\frac{R}{|z|}\right) - 4K\right] &\leq \liminf_{\vartheta\to 0} \frac{1}{|log(\vartheta)|} \liminf_{\mathfrak{q}\to w} G^{\mathfrak{q},\varepsilon,-}_{\vartheta,R}(|z|) \\ &\leq \limsup_{\vartheta\to 0} \frac{1}{|log(\vartheta)|} \limsup_{\mathfrak{q}\to w} G^{\mathfrak{q},\varepsilon,+}_{\vartheta,R}(|z|) \\ &\leq \frac{2}{V_n^{w,\varepsilon}}\left[\log\left(\frac{R}{|z|}\right) + 4K\right],\end{aligned}$$

where

$$G^{\mathfrak{q},\varepsilon,+}_{\vartheta,R}(r) = \sup_{|z|=r} \mathbf{E}^{z,y,\mathfrak{q}}\left(\int_0^{\tau_R} \mathbb{1}_{\{(\mathbb{R}^2\setminus\vartheta\mathbb{B})\times\mathcal{H}^n_{\alpha,s}\}}(\tilde{z}_t^{\mathfrak{q},\varepsilon,\tau_n}, \tilde{y}_t^{\varepsilon,\tau_n})dt\right)$$

and

$$G^{\mathfrak{q},\varepsilon,-}_{\vartheta,R}(r) = \inf_{|z|=r} \mathbf{E}^{z,y,\mathfrak{q}}\left(\int_0^{\tau_R} \mathbb{1}_{\{(\mathbb{R}^2\setminus\vartheta\mathbb{B})\times\mathcal{H}^n_{\alpha,s}\}}(\tilde{z}_t^{\mathfrak{q},\varepsilon,\tau_n}, \tilde{y}_t^{\varepsilon,\tau_n})dt\right).$$

(3) for each $0 < R < \varsigma$ and $p \in (0,1]$, there exists a $\widetilde{K} < \infty$ such that $\mathbf{E}^{z,y,\mathfrak{q}} \int_0^{\tau_R} (|\tilde{z}_t^{\mathfrak{q},\varepsilon,\tau_n}|)^p dt \leq \widetilde{K}$ whenever $0 < |z| < R$ and $y \in \mathcal{H}^n_{\alpha,s}$.

Note that, based on Assumption 6.1.9 about the dissipativity of F in the critical mode, by a standard Lyapunov-like argument (using a quadrtic Lyapunov function), we can also show that there exists some K such that, on $\{\tau_n = \infty\}$,

$$\mathbf{E}^{z,y,\mathfrak{q}}[\tau_\vartheta] \leq K + z^2$$

whenever $|z| \geq \vartheta$ and $y \in \mathcal{H}^n_{\alpha,s}$. This in turn indicates a down crossing from the place away from $\mathbf{0} \times \mathcal{H}^n_{\alpha,s}$ for almost all sample paths on $\{\tau_n = \infty\}$.

The existence and construction of the new invariant measure on $(\mathbb{R}^2 \setminus \{\mathbf{0}\}) \times \mathcal{H}^n_{\alpha,s}$ when $\lambda^{\mathfrak{q},\varepsilon} \geq 0$ follows the procedure in [90, 17]. One can show that the up/down-crossings of the spheres with radius r and R are infinitely often for $\{(\tilde{z}^{\mathfrak{q},\varepsilon}(t\wedge\tau_n), \tilde{y}^{\varepsilon}(t\wedge\tau_n))\}_{t\geq 0}$ a.s. given $\tau_n = \infty$. It can be verified that, under Assumption 8.0.5, there exists a unique invariant measure $\hat{\nu}$ for the Markov chain on $\Gamma = \{(z,y) \in \mathbb{R}^2 \times \mathcal{H}^n_{\alpha,s} : |z| = r\}$ which is induced by $\{(\tilde{z}^{\mathfrak{q},\varepsilon}(t\wedge\tau_n), \tilde{y}^{\varepsilon}(t\wedge\tau_n))\}_{t\geq 0}$ (on $\{\tau_n = \infty\}$) stopping sequentially at the sphere. Let

$$\tau_0 = \inf\{t \geq 0, |\tilde{z}_t^{\mathfrak{q},\varepsilon}| = r\},$$

$$\tau_0' = \inf\{t \geq \tau_0, |\tilde{z}_t^{\mathfrak{q},\varepsilon}| = R\},$$

and

$$\tau_1 = \inf\{t \geq \tau_0', |\tilde{z}_t^{\mathfrak{q},\varepsilon}| = r\}.$$

Then,

(1) When $\lambda^{\mathfrak{q},\varepsilon} > 0$, the measure $\tilde{\nu}^{\mathfrak{q},\varepsilon} = \tilde{\nu}_\infty^{\mathfrak{q},\varepsilon} \cdot \mathbf{P}^{z,y,\mathfrak{q}}[\tau_n = \infty] + \mathscr{L}(\tilde{z}_{\tau_n}^{\mathfrak{q},\varepsilon}, \tilde{y}_{\tau_n}^{\varepsilon}) \cdot \mathbf{P}^{z,y,\mathfrak{q}}[\tau_n < \infty]$ where

$$\tilde{\nu}_\infty^{\mathfrak{q},\varepsilon}(A) = \frac{\int_\Gamma \mathbf{E}^{z,y,\mathfrak{q}} \left(\int_0^{\tau_1} \mathbb{1}_{\{A \times \mathcal{H}_s^n\}}(\tilde{z}_s^{\mathfrak{q},\varepsilon}, \tilde{y}_s^{\varepsilon}) ds\right) d\hat{\nu}(x,y)}{\mathbf{P}^{z,y,\mathfrak{q}}[\tau_n = \infty] \cdot \int_\Gamma \mathbf{E}^{z,y,\mathfrak{q}}[\tau_1] d\hat{\nu}(x,y)}, \quad A \in \mathscr{B}(\mathbb{R}^2 \setminus \{\mathbf{0}\}) \tag{8.24}$$

is a marginal invariant probability measure of $\tilde{z}^{\mathfrak{q},\varepsilon,\tau_n}$ for the stopped process $(\tilde{z}^{\mathfrak{q},\varepsilon,\tau_n}, \tilde{y}^{\varepsilon,\tau_n})$ w.r.t. $\hat{\nu}$, where the normalizer is finite based on Proposition 8.2.5. Such a measure already averages out the estimations along the $\mathcal{H}_{\alpha,s}^n$ direction due to the way of construction. Nonetheless, since the critical mode contributes more to the radius of the process $\{u(t)\}_{t \geq 0}$, the above result fulfills our goal.

Note that, since $\Lambda_n^{\mathfrak{q},\varepsilon}(p)$ is convex in p, there exists a unique $\mathfrak{d} = \mathfrak{d}^{\mathfrak{q},\varepsilon} > 0$ such that $\Lambda_n^{\mathfrak{q},\varepsilon}(-\mathfrak{d}) = 0$. Furthermore, using the results from Proposition 8.2.5, there exists $\widetilde{\varsigma} > 0$ and $\widetilde{K}^{\mathfrak{q}} < \infty$ such that, for all $\varepsilon \ll 1$, the (conditional) probability measure $\tilde{\nu}_\infty^{\mathfrak{q},\varepsilon}$ satisfies

$$\frac{r^{\mathfrak{d}}}{\widetilde{K}^{\mathfrak{q}}} \leq \tilde{\nu}_\infty^{\mathfrak{q},\varepsilon}(\{z \in \mathbb{R}^2 \setminus \{\mathbf{0}\} : \ |z| < r\}) \leq \widetilde{K}^{\mathfrak{q}} r^{\mathfrak{d}}, \quad r \in (0, \widetilde{\varsigma}). \tag{8.25}$$

Since for $\varepsilon \ll 1$, the quantity $\mathscr{L}(\tilde{z}_{\tau_n}^{\mathfrak{q},\varepsilon}, \tilde{y}_{\tau_n}^{\varepsilon}) \cdot \mathbf{P}^{z,y,\mathfrak{q}}[\tau_n < \infty]$, though may not be unique, can be arbitrarily small, a similar result can be obtained for $\tilde{\nu}^{\mathfrak{q},\varepsilon}$. Due to the uniqueness of $\tilde{\nu}_\infty^{\mathfrak{q},\varepsilon}$, the unique weak limit of $\tilde{\nu}^{\mathfrak{q},\varepsilon}$ as $\varepsilon \to 0$ is $\tilde{\nu}_\infty^{\mathfrak{q}}$.

(2) A quick corollary can be made on the marginal invariant measure $\nu^{\mathfrak{q},\varepsilon}$ of z for the process $\{z(t \wedge \tau_n), y(t \wedge \tau_n)\}_{t \geq 0}$ (the stopped solution to the original nonlinear equation (6.10) for $y(0) \in \mathcal{H}_{\alpha,s}^n$) based on (6.67) for arbitrarily small $\varepsilon > 0$, i.e., there exists $\varsigma > 0$ and $K^{\mathfrak{q}} < \infty$ such that the measure $\nu^{\mathfrak{q},\varepsilon}$ satisfies the marginal property

$$\frac{r^{\mathfrak{d}}}{K^{\mathfrak{q}}} \leq \nu^{\mathfrak{q},\varepsilon}(\{z \in \mathbb{R}^2 \setminus \{\mathbf{0}\} : \ |z| < r\}) \leq K^{\mathfrak{q}} r^{\mathfrak{d}}, \quad r \in (0, \varsigma). \tag{8.26}$$

(3) When $\lambda^{w,\varepsilon} = 0$, there exists a σ-finite measure on $(\mathbb{R}^2 \setminus \{\mathbf{0}\}) \times \mathcal{H}_{\alpha,s}^n$, such that the marginal on $\mathscr{B}(\mathbb{R}^2 \setminus \{\mathbf{0}\})$ is of the form $\tilde{\nu}^{w,\varepsilon} = \tilde{\nu}_\infty^{w,\varepsilon} \cdot \mathbf{P}^{z,y,w}[\tau_n = \infty] + \mathscr{L}(\tilde{z}_{\tau_n}^{\mathfrak{q},\varepsilon}, \tilde{y}_{\tau_n}^{\varepsilon}) \cdot \mathbf{P}^{z,y,w}[\tau_n < \infty]$

where

$$\tilde{\nu}_\infty^{w,\varepsilon}(A) = \frac{\int_\Gamma \mathbf{E}^{x,y,w}\left(\int_0^{\tau_1} \mathbb{1}_{\{A\times\mathcal{H}_s^n\}}(\tilde{z}_s^{w,\varepsilon}, \tilde{y}_s^\varepsilon)ds\right) d\hat{\nu}(x,y)}{\mathbf{P}^{x,y,w}[\tau_n = \infty]}, \quad A \in \mathscr{B}(\mathbb{R}^2 \setminus \{\mathbf{0}\}). \tag{8.27}$$

Similarly, as $\varepsilon \to 0$, $\tilde{\nu}_\infty^w$ captures the limit behavior.

It has been proved in [17, Theorem 2.12, 2.13] that, if $\lambda^{\mathfrak{q},\varepsilon} > 0$, $\lambda^{w,\varepsilon} = 0$, and as $\mathfrak{q} \to w$ continuously, the invariant measure $\tilde{\nu}^{\mathfrak{q},\varepsilon}$ in (8.24) converges weakly to the probability measure $\delta_{\mathbf{0}}$ rather than the σ-finite measure $\tilde{\nu}^{w,\varepsilon}$. The convergence rate is as follows.

Theorem 8.2.7. *[17, Theorem 2.13] Let $\mathfrak{q} \to w$ continuously for all $\mathfrak{q}$ satisfying $\lambda^{\mathfrak{q},\varepsilon} > 0$. Then,*

$$\frac{1}{\lambda^{\mathfrak{q},\varepsilon}} \int_{\mathbb{R}^2} h(x) d\tilde{\nu}^{\mathfrak{q},\varepsilon}(x) \to \int_{\mathbb{R}^2} h(x) d\bar{\nu}(x),$$

where $\bar{\nu}$ is the unique σ-finite invariant measure for $\tilde{z}^{w,\varepsilon}$ on $\mathbb{R}^2 \setminus \{0\}$ satisfying

$$\frac{\bar{\nu}(\{x \in \mathbb{R}^2 \setminus \{0\} : |x| > r\})}{|\log r|} \to \frac{2}{V_n^{w,\varepsilon}}, \quad r \to 0,$$

and $h : \mathbb{R}^2 \setminus \{\mathbf{0}\} \to \mathbb{R}$ satisfying $h(x)/|x|^p \to 0$ as $x \to 0$ for some $p > 0$.

Corollary 8.2.8. *Let $\nu^{\mathfrak{q},\varepsilon}$ be the invariant measure on $\mathbb{R}^2 \setminus \{\mathbf{0}\}$ for the amplitude z of the system (6.10). Then, as $\varepsilon \to 0$, and $\mathfrak{q} \to w = w^\varepsilon$, we have*

$$\frac{1}{\lambda^{\mathfrak{q},\varepsilon}} \int_{\mathbb{R}^2} h(x) d\nu^{\mathfrak{q},\varepsilon}(x) \to \int_{\mathbb{R}^2} h(x) d\bar{\nu}(x),$$

where $\bar{\nu}$ is the same as in Theorem 8.2.7.

Remark 8.2.9. *Note that the original system (6.10) has the same $\lambda^{\mathfrak{q},\varepsilon}$ for every fixed ε and $\mathfrak{q}$. We control ε in a way such that the $\mathcal{O}(\varepsilon^2)$ error term is absolutely continuous to $\lambda^{\mathfrak{q},\varepsilon}$ as $\mathfrak{q} \to w$.*

8.3 A Discussion on P-Bifurcation Point

We have seen in (8.26) that, if $\lambda^{\mathfrak{q},\varepsilon} > 0$, for each n, the new invariant measure for the process $\{(z(t\wedge\tau_n), y(t\wedge\tau_n))\}_{t\geq 0}$ with $y(0) \in \mathcal{H}_{\alpha,s}^n$ has marginal $\nu^{\mathfrak{q},\varepsilon}$ that possesses mass within a small neighborhood in a way that

$$\frac{\nu^{\mathfrak{q},\varepsilon}(r\overline{\mathbb{B}})}{\mathrm{vol}(r\overline{\mathbb{B}})} \sim Cr^{\mathfrak{d}-2}, \tag{8.28}$$

where $\mathfrak{d} = \mathfrak{d}^{\mathfrak{q},\varepsilon} > 0$ and $\Lambda_n^{\mathfrak{q},\varepsilon}(-\mathfrak{d}) = 0$.

Now, let $\nu_u^{\mathfrak{q},\varepsilon}$ be the invariant measure of the critical amplitude of u^{τ_n}, which is within the time scale $\varepsilon^2 t$ (recall Definition 6.1.14); let $\Lambda_{u,n}^{\mathfrak{q},\varepsilon}(p)$ denote the associated moment Lyapunov exponents. By the same argument, we can verify that

$$\frac{\nu_u^{\mathfrak{q},\varepsilon}(r\overline{\mathbb{B}})}{\mathrm{vol}(r\overline{\mathbb{B}})} \sim C r^{\mathfrak{d}-2}, \tag{8.29}$$

where $\mathfrak{d} = \mathfrak{d}^{\mathfrak{q},\varepsilon} > 0$ and $\Lambda_{u,n}^{\mathfrak{q},\varepsilon}(-\mathfrak{d}) = 0$ as well. The shape of the density of $\nu_u^{\mathfrak{q},\varepsilon}$ changes at some $\mathfrak{q}$ at which $\mathfrak{d}^{\mathfrak{q},\varepsilon} = 2$, i.e., the density has a pole at $x = 0$ for $\mathfrak{d}^{\mathfrak{q},\varepsilon} < 2$ and has a zero at $x = 0$ for $\mathfrak{d}^{\mathfrak{q},\varepsilon} > 2$. We need to decide the $\mathfrak{q}$ such that $\mathfrak{d}^{\mathfrak{q},\varepsilon} = 2$, which is known as the P-bifurcation point.

Proposition 8.3.1. *For any compact interval that contains* $\mathfrak{p}$*, there exists a strictly positive function* $\mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon} : \mathcal{S}^1 \times \mathcal{H}_{\alpha,s}^n \to \mathbb{R}$ *satisfying* $\|\mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon}\|_{C^2} \le K$, $\mathfrak{z}_{0,n}^{\mathfrak{q},\varepsilon} \equiv 1$, *such that*

$$\mathfrak{L}_{p,u}^{\mathfrak{q},\varepsilon} \mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon} = \Lambda_{u,n}^{\mathfrak{q},\varepsilon}(p) \cdot \mathfrak{z}_{p,n}^{\mathfrak{q},\varepsilon}, \quad \langle \mathfrak{z}_p^{\mathfrak{q},\varepsilon},\ \mu_{p,n}^{\mathfrak{q},\varepsilon} \rangle = 1, \tag{8.30}$$

where $\mathfrak{L}_{p,u}^{\mathfrak{q},\varepsilon} = \mathfrak{L}^{\mathfrak{q},\varepsilon} + \varepsilon^2 p \mathfrak{X} + \varepsilon^2 p \mathcal{Q} + \varepsilon^2 \frac{p^2}{2} \mathfrak{R}$, $\mathcal{Q}$ *and* $\mathfrak{R}$ *are the same as in Lemma (8.1.2).*

Proof. Due the change of time scale, the quantity $\Lambda_{u,n}^{\mathfrak{q},\varepsilon}(p)$ is driven by a continuous semigroup $\{\mathcal{T}_{t,n}^{\mathfrak{q},\varepsilon}(p)\}_{t \ge 0}$ on $\mathcal{S}^1 \times \mathcal{H}_s^n$ as

$$\begin{aligned} &\mathcal{T}_{t,u}^{\mathfrak{q},\varepsilon}(p) f(\phi,\eta) := \\ &\mathbf{E}\Big[f(\phi(t\wedge\tau_n), \eta(t\wedge\tau_n)) \\ &\qquad \cdot \exp\Big\{ \varepsilon^2 p \int_0^t \mathcal{Q}(\phi(s\wedge\tau_n), \eta(s\wedge\tau_n)) ds + \varepsilon p \int_0^t G_c^{\mathfrak{p}}(\phi(s\wedge\tau_n), \eta(s\wedge\tau_n)) dW_s \Big\} \Big], \end{aligned} \tag{8.31}$$

By the Girsanov's theorem as in the proof of Lemma 8.1.2, the generator $\mathfrak{L}_{p,u}^{\mathfrak{q},\varepsilon} = \mathfrak{L}^{\mathfrak{q},\varepsilon} + \varepsilon^2 p \mathfrak{X} + \varepsilon^2 p \mathcal{Q} + \varepsilon^2 \frac{p^2}{2} \mathfrak{R}$ can be obtained. The existence of the eigenvalue function is followed by the compactness of the transition semigroup. □

Corollary 8.3.2. *For* $\varepsilon \ll 1$*, the moment Lyapunov exponent* $\Lambda_{u,n}^{\mathfrak{q},\varepsilon}(p)$ *has the following asymptotic expansion:*

$$\Lambda_{u,n}^{\mathfrak{q},\varepsilon}(p) = p\varepsilon^2 \lambda^{\mathfrak{q},\varepsilon} + \frac{p^2\varepsilon^2}{2} V_n^{\mathfrak{q},\varepsilon} + \mathcal{O}(\varepsilon^3), \tag{8.32}$$

where $\lambda^{\mathfrak{q},\varepsilon}$ *and* $V_n^{\mathfrak{q},\varepsilon}$ *solve the eigenvalue problem of* $\mathfrak{L}^{\mathfrak{q},\varepsilon} \varphi_n^{\mathfrak{q},\varepsilon} = \lambda^{\mathfrak{q},\varepsilon} - \mathcal{Q}^{\mathfrak{q}} + \vartheta_n^{\varepsilon}$ *and* $\mathfrak{L}^{\mathfrak{q},\varepsilon} \psi_n^{\mathfrak{q},\varepsilon} = V_n^{\mathfrak{q},\varepsilon} - 2(\mathfrak{X} + \mathcal{Q}^{\mathfrak{q}} - \lambda^{\mathfrak{q},\varepsilon})\varphi^{\mathfrak{q},\varepsilon} - \mathfrak{R}$*, respectively.*

Proof. The expansion is done by consecutively differentiating both sides of the eigenvalue problem in (8.30), and taking the appropriate scalar product with $\mu_{p,n}^{\mathfrak{q},\varepsilon}$. We have also used the fact that $\Lambda_{u,n}^{\mathfrak{q},\varepsilon}(0) = 0$. □

Utilizing the fact that $\Lambda_{u,n}^{\mathfrak{q},\varepsilon}(-2) = 0$, we can approximate $\mathfrak{q}$ by eliminating the high order error term and equating

$$-2\varepsilon^2\lambda^{\mathfrak{q},\varepsilon} + 2\varepsilon^2 V_n^{\mathfrak{q},\varepsilon} \approx 0 \Leftrightarrow \lambda^{\mathfrak{q},\varepsilon} \approx V_n^{\mathfrak{q},\varepsilon}.$$

The problem is reduced to the approximation of $V_n^{\mathfrak{q},\varepsilon}$.

Recall that $\mathfrak{L}^{\mathfrak{q},\varepsilon} = \frac{1}{\varepsilon^2}\mathfrak{L}_0 + \mathfrak{L}_1$. Let $\varphi_n^{\mathfrak{q},\varepsilon} = \varphi_0 + \varepsilon^2\varphi_1 + \varepsilon^4\varphi_2$. Therefore,

$$\begin{aligned}&\frac{1}{\varepsilon^2}\mathfrak{L}_0\varphi_0 + \mathfrak{L}_0\varphi_1 + \mathfrak{L}_1\varphi_0 + \varepsilon^2\mathfrak{L}_1\varphi_1 + \varepsilon^2\mathfrak{L}_0\varphi_2 + \mathcal{O}(\varepsilon^3)\\ =&\lambda_0 - \mathcal{Q} + \varepsilon^2\lambda_1 + \mathcal{O}(\varepsilon^3).\end{aligned}$$

Then, $\mathfrak{L}_0\varphi_0 = \mathfrak{L}_1\varphi_0 = 0$ and φ_1 solves the first-order PDE

$$\mathfrak{L}_0\varphi_1 = \lambda_0 - \mathcal{Q}, \tag{8.33}$$

whereas φ_2 solves

$$\mathfrak{L}_0\varphi_2 = -\mathfrak{L}_1\varphi_1 + \lambda_1. \tag{8.34}$$

By the solvability condition of $\mathfrak{L}^{\mathfrak{q},\varepsilon}\psi_n^{\mathfrak{q},\varepsilon} = V_n^{\mathfrak{q},\varepsilon} - 2(\mathfrak{X} + \mathcal{Q} - \lambda^{\mathfrak{q},\varepsilon})\varphi_n^{\mathfrak{q},\varepsilon} - \mathfrak{R}$, we have

$$V_n^{\mathfrak{q},\varepsilon} = 2\langle\mu^{\mathfrak{q},\varepsilon}, (\mathfrak{X} + \mathcal{Q} - \lambda^{\mathfrak{q},\varepsilon})\varphi_n^{\mathfrak{q},\varepsilon}\rangle + \langle\mu^{\mathfrak{q},\varepsilon}, \mathfrak{R}\rangle \pm \vartheta_n^{\varepsilon}.$$

The expansion of $V_n^{\mathfrak{q},\varepsilon}$ is given as

$$V_n^{\mathfrak{q},\varepsilon} = \langle\mathfrak{R},\ \mu_0\rangle + 2\varepsilon^2\langle(\mathfrak{X} + \mathcal{Q} - \lambda_0)\varphi_1,\ \mu_0\rangle + \mathcal{O}(\varepsilon^3). \tag{8.35}$$

The approximated P-bifurcation point $\mathfrak{q}$ is determined by

$$(\lambda_0 + \varepsilon^2\lambda_1)(\mathfrak{q}) = (\langle\mathfrak{R},\ \mu_0\rangle + 2\varepsilon^2\langle(\mathfrak{X} + \mathcal{Q} - \lambda_0)\varphi_1,\ \mu_0\rangle)(\mathfrak{q}). \tag{8.36}$$

8.4 Summary

In this chapter, we have investigated the stochastic bifurcations of the approximation system (8.1) and hence the asymptotic bifurcation behaviors of the original system (6.10) when $\varepsilon \to$

0. The key point is to verify the approximate eigenvalue problems for $\Lambda_n^{\mathfrak{q},\varepsilon}$, $\lambda^{\mathfrak{q},\varepsilon}$ and $V_n^{\mathfrak{q},\varepsilon}$ by checking the compactness of the approximated semigroup $\mathcal{T}_{t,n}^{\mathfrak{q},\varepsilon}(p)$. Taking the advantages of the strong mixing effect of the stable modes, the problem is reduced to an $\mathbb{R}^2 \times \mathcal{H}_{\alpha,s}^n$ system with a small $\mathcal{L}^2$-stable stable marginal. Relying on the construction based on Lemma 8.2.1 and Proposition 8.2.3, the recurrence and null-recurrence property of $(\tilde{z}^{\mathfrak{q},\varepsilon,\tau_n}, \tilde{y}^{\varepsilon,\tau_n})$ on $\{\tau_n = \infty\}$ within the critical marginal for $\lambda^{\mathfrak{q},\varepsilon} > 0$ and $\lambda^{\mathfrak{q},\varepsilon} = 0$, respectively, can be verified and utilized for constructing the unique invariant marginal measure on $\mathbb{R}^2 \setminus \{\mathbf{0}\}$. The approximated D-bifurcation property of $(z^{\mathfrak{q},\varepsilon}, y^{\varepsilon})$ within the $\mathbb{R}^2$ projection is then justified. On the other hand, a P-bifurcation point can be approximated using the relation $\lambda^{\mathfrak{q},\varepsilon} \approx V_n^{\mathfrak{q},\varepsilon}$. The effect of the small stable marginals comes in with a second-order correction as in $\varepsilon^2\lambda_1$ and $\varepsilon^2\varphi_1$.

Compared to the approximation scheme in [27], where the critical modes and stable modes in the stochastic term are decoupled, we have investigated the mixing effect of the stable modes by an approximation up to $\mathcal{O}(\varepsilon^2)$. Such a coupling has a slight impact on the D-bifurcation point as well as on the P-bifurcation point via second-order corrections. Unlike the deterministic case where deterministic parametric perturbations do not affect the structural stability, even a small intensity of stochastic perturbations from the stable modes shifts both of the dynamical and phenomenological bifurcation points. Other than capturing the stochastic bifurcation points, the simplified scheme in [27] can be used to simulate the density of the critical amplitude z, which is solved by the approximated Fokker-Planck equation, with a reasonably small error when the system has passed the P-bifurcation point.

Chapter 9

Conclusions and Future Work

Within the scope of dynamical systems with uncertainties, this research is motivated to develop novel theories and methods of verification and control synthesis for finite-dimensional nonlinear systems with non-stochastic and stochastic perturbations, and to understand the impact of small Gaussian-type space-time noise on stochastic Hopf bifurcations of parabolic SPDEs. We summarize the main contributions in this final chapter and bring up some related future research direction.

Lyapunov-Barrier Approaches Safety and Stability Related Specifications

For deterministic systems with non-stochastic uncertainties, we showed that smooth Lyapunov barrier functions can be defined on the entire set of initial conditions from which stability with a safety constraint can be satisfied. We built the connection between reach-avoid-stay type specifications and stability with a safety constraints via a robustness argument, upon which a converse Lyapunov-barrier function theorem for reach-avoid-stay type specifications can be established. It was shown by an example that the statement cannot be strengthened without additional assumptions. We further extended the results to deterministic hybrid systems and establish converse control Lyapunov-barrier functions for deterministic systems with control inputs. The general topological structure of the initial sets, target sets and unsafe sets leaves us more flexibility to design control Lyapunov-barrier functions. In comparison with formal methods, the effectiveness of such an approach was investigated in a case study of jet engine compressor control problem using a simplified Moore-Greitzer ODE model.

Limitations exist in the current results. We only considered an additive measurable disturbance in the dynamical systems for the purpose of establishing converse Lyapunov-barrier results. Similar to other converse Lyapunov theorems, the existence results are not constructive.

In addition, the extension of the converse Lyapunov-barrier functions to the converse control Lyapunov-barrier functions was based on the existence of a Lipschitz continuous feedback law. It is also necessary to investigate the necessary and sufficient conditions for the existence of universal control laws for control-affine nonlinear systems. We would like to emphasize that the focus of this part is not on designing controllers. However, such theoretical results can be used to characterize control Lyapunov functions. Besides a deeper understanding of the feasibility of QP, it would also be necessary to investigate new algorithm framework other than QP to synthesize controllers utilizing the verified Lyapunov-barrier criteria. On the other hand, a crucial question to ask is whether the Lyapunov-barrier conditions can bring a finite number of feedback controllers to achieve a (robust) reach-avoid-stay specification and whether the resulting feedback controllers possess certain regularities.

Another interesting future direction is to explore computational techniques for constructing Lyapunov-barrier function that is defined on the whole set of initial conditions (or as large a subset as possible of this set) from which a stability with safety guarantee or reach-avoid-stay specification is achievable, for instance, learning techniques [142, 22, 183] or interval analysis [141, 47]. In this regard, the results, especially Theorems 2.3.7 and 2.3.18, can hopefully shed some light into the development of such computational techniques with completeness (or approximate completeness) guarantees.

It would be an promising application in using the current Lypunov-barrier characterizations, which provide the required building blocks of linear temporal logic specifications, as high-level abstractions for controller synthesis. It is also valuable to design algorithms based on this theoretical work to construct Lyapunov-like functions that improves the estimated region of attractions.

In the stochastic contexts, we also formulated stochastic Lyapunov-barrier functions to develop sufficient conditions on probabilistic reach-avoid-stay specifications for continuous-time stochastic systems with extra uncertainties. Robustness was also taken into account such that a worst-case scenario is guaranteed. Unlike solutions of deterministic systems, the diffusion effects renders more difficulties of selecting Lyapunov/barrier functions under the restrictive geometric requirements of the initial conditions and unsafe sets.

To improve the current stochastic Lyapunov-barrier approach, it would be interesting to establish converse stochastic Lyapunov-barrier function theorems. In addition, due to the difficulty of obtaining full observations, the existence of feasible control policies given the whole set of sample paths based on the stochastic Lyapunov-barrier scheme is not verified. We only delivered numerical examples given certain successfully controlled sample paths under constraints for control inputs. The existence of feasible controller and the conditional probability of satisfaction given such control inputs should be quantified mathematically for future work.

We finally proposed a data-driven approach embedded in the Lyapunov-barrier scheme dealing with safety-critical control of unknown stochastic systems. A first-moment convergence rather than convergence in probability is needed to deal with the potential numerical uncertainties. One straightforward future direction is to establish the convergence rate w.r.t. the size of the training data and the sampling time.

Stochastic Abstractions

We investigated the mathematical properties of formal abstractions for discrete-time control-free and controlled stochastic systems in view of metrizable space of probability measures. We proposed the concept of robust completeness in the stochastic context for the first time and constructed formal stochastic abstractions with both soundness and such a property. The philosophy of stochastic abstractions was discussed in comparison with numerical approximations.

For future work, it would be interesting to design algorithms to construct IMC (resp. BMDP) abstractions for more general robust stochastic (resp. control) systems with $\mathcal{L}^1$ perturbations based on metrizable space of measures and weak topology. The size of state discretization can be refined given more specific assumptions on system dynamics and LT objectives. Now that the stochastic abstractions are analyzed from the mathematical perspectives, it would be crucial to design more powerful robust verification and control synthesis algorithms based on abstractions. For verification or control synthesis w.r.t. probabilistic safety or reachability problems, comparisons can be made with stochastic Lyapunov-barrier function approaches.

Another important issue is in abstracting continuous-time stochastic systems by discrete-time stochastic systems with certain guarantees. The difficulties exist due to the conversion of measurability from the continuous-time canonical space to the discrete-time counterpart. The probabilistic behaviors in between sampling time need to be evaluated in a proper sense. We aim to show whether there exists some computable temporal space discretization and a decision procedure of a control policy based on the discretized sampling period that can realize a given probabilistic LTL specification.

Stochastic Hopf Bifurcation Analysis for SPDEs

We considered parabolic type SPDEs in the presence of small space-time stochastic perturbations near the deterministic bifurcation points. This setting renders us convenience to separate the slowly and fast varying modes and then conduct multiscale analysis. For the systems with additive noise, inspired by recent advances in stochastic PDEs given in [29], the bifurcation analysis was provided for the stochastic version of the Moore and Greitzer PDE model for an axial flow compressor. The homogenized evolution equation for the critical coordinates was derived and proved with relatively small error bound. Such a transient dynamic can be treated as the normal form.

For the case with multiplicative noise, we considered cubic nonlinearities with proper assumptions to guarantee the existence of invariant measures. Compared to the recent work in [27], where the stable modes and the critical modes are decoupled, we proposed a different simplification scheme for the original system and proved the error bounds. We keep the multiplicative noise in both stable and critical modes such that the linearization of the simplified scheme stay the same as that of the original system. This approximation scheme is readily allied with the almost-sure exponential stability of the trivial solution to analyze the stochastic bifurcation diagram as the noise becomes smaller. The analysis shows that the stochastic effects from the stable modes do have small impact on determining the stochastic bifurcation points.

Now that we have a clearer view of stochastic Hopf bifurcations of parabolic SPDEs with cubic nonlinearities and small multiplicative noise, a straightforward extension in the future work would be considering the presence of both bilinear and cubic nonlinearities. Since we have been taken the advantages of spectral gap and multiscale analysis due to the small intensity of stochastic perturbations, to fully understand the concept of stochastic bifurcations in SPDEs, we need to develop theories for more general type of noise with regular intensity.

As discovered in the stochastic Moore-Greitzer model with additive noise, the approximation depends on the spatial regularity of the solution. However, for parabolic SPDEs with more than one spatial dimension under a space-time white noise, the solutions only exit in a distribution sense (correspondingly with α-Hölder continuity for $\alpha < 0$). In this case, the traditional regularity analysis approaches fail to capture the irregularity. Instead, we will find regularity based pathwise arguments to make sense of equations. To look at the SPDEs in a 'rough path' perspective is a promising technique. Rough path allows stochastic integrals with less regular integrands that take values in a more general Banach space.

Formal Methods of Verification and Control Synthesis for PDEs

This would be a promising future direction to work with based on our current understanding of stochastic abstractions, whose solutions are closely related to Fokker-Planck parabolic PDE. Rather than LTL specifications, spatio-signal temporal logic would be a better fit for specifying space-time properties for PDE systems. Finite Element Method is preferred as the spatio-temporal discretization approach. It is possible to formulate optimization problems to improve the computational searching time. We can start the problem with considering linear parabolic PDEs such as heat equations, then it is natural to extend the result to a general class of semilinear parabolic PDEs.

References

[1] Alessandro Abate, Joost-Pieter Katoen, John Lygeros, and Maria Prandini. Approximate model checking of stochastic hybrid systems. *European Journal of Control*, 16(6):624–641, 2010.

[2] Ayush Agrawal and Koushil Sreenath. Discrete control barrier functions for safety-critical control of discrete systems with application to bipedal robot navigation. In *Robotics: Science and Systems*, 2017.

[3] Francisco Penedo Alvarez. *Formal Methods for Partial Differential Equations*. PhD thesis, Boston University, 2020.

[4] Aaron D Ames, Samuel Coogan, Magnus Egerstedt, Gennaro Notomista, Koushil Sreenath, and Paulo Tabuada. Control barrier functions: Theory and applications. In *2019 18th European Control Conference (ECC)*, pages 3420–3431. IEEE, 2019.

[5] Aaron D Ames, Xiangru Xu, Jessy W Grizzle, and Paulo Tabuada. Control barrier function based quadratic programs for safety critical systems. *IEEE Transactions on Automatic Control*, 62(8):3861–3876, 2016.

[6] Wolfgang Arendt and Jochen Glück. Positive irreducible semigroups and their long-time behaviour. *Philosophical Transactions of the Royal Society A*, 378(2185):20190611, 2020.

[7] L Arnold, E Oeljeklaus, and E Pardoux. Almost sure and moment stability for linear itô equations. In *Lyapunov exponents*, pages 129–159. Springer, 1986.

[8] Ludwig Arnold. Random dynamical systems. *Dynamical systems*, pages 1–43, 1995.

[9] Ludwig Arnold, MM Doyle, and N Sri Namachchivaya. Small noise expansion of moment lyapunov exponents for two-dimensional systems. *Dynamics and Stability of Systems*, 12(3):187–211, 1997.

[10] Ludwig Arnold, George Papanicolaou, and Volker Wihstutz. Asymptotic analysis of the lyapunov exponent and rotation number of the random oscillator and applications. *SIAM Journal on Applied Mathematics*, 46(3):427–450, 1986.

[11] Ludwig Arnold, N Sri Namachchivaya, and Klaus R Schenk-Hoppé. Toward an understanding of stochastic hopf bifurcation: a case study. *International Journal of Bifurcation and Chaos*, 6(11):1947–1975, 1996.

[12] Ebru Aydin Gol, Mircea Lazar, and Calin Belta. Language-guided controller synthesis for discrete-time linear systems. In *Proceedings of the 15th ACM international conference on Hybrid Systems: Computation and Control*, pages 95–104, 2012.

[13] OO Badmus, S Chowdhury, and CN Nett. Nonlinear control of surge in axial compression systems. *Automatica*, 32(1):59–70, 1996.

[14] Christel Baier and Joost-Pieter Katoen. *Principles of model checking*. MIT press, 2008.

[15] Andrzej Banaszuk, Höskuldur Ari Hauksson, and Igor Mezic. A backstepping controller for a nonlinear partial differential equation model of compression system instabilities. *SIAM journal on control and optimization*, 37(5):1503–1537, 1999.

[16] Martin T Barlow and Edwin A Perkins. Brownian motion on the sierpinski gasket. *Probability theory and related fields*, 79(4):543–623, 1988.

[17] Peter H Baxendale. A stochastic hopf bifurcation. *Probability Theory and Related Fields*, 99(4):581–616, 1994.

[18] PH Baxendale and DW Stroock. Large deviations and stochastic flows of diffeomorphisms. *Probability Theory and Related Fields*, 80(2):169–215, 1988.

[19] Ali Behzadan and M Holst. Multiplication in sobolev spaces, revisited. *arXiv preprint arXiv:1512.07379*, 2015.

[20] Calin Belta and Sadra Sadraddini. Formal methods for control synthesis: An optimization perspective. *Annual Review of Control, Robotics, and Autonomous Systems*, 2:115–140, 2019.

[21] Calin Belta, Boyan Yordanov, and Ebru Aydin Gol. *Formal Methods for Discrete-time Dynamical Systems*, volume 89. Springer, 2017.

[22] Felix Berkenkamp, Riccardo Moriconi, Angela P Schoellig, and Andreas Krause. Safe learning of regions of attraction for uncertain, nonlinear systems with gaussian processes. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 4661–4666. IEEE, 2016.

[23] Dimitri P Bertsekas and Ian B Rhodes. On the minimax reachability of target sets and target tubes. *Automatica*, 7(2):233–247, 1971.

[24] Patrick Billingsley. *Convergence of probability measures.* John Wiley & Sons, 2013.

[25] Björn Birnir and Höskuldur Ari Hauksson. Basic control for the viscous moore–greitzer partial differential equation. *SIAM Journal on Control and Optimization*, 38(5):1554–1580, 2000.

[26] Dirk Blömker. Amplitude equations for locally cubic nonautonomous nonlinearities. *SIAM Journal on Applied Dynamical Systems*, 2(3):464–486, 2003.

[27] Dirk Blömker and Hongbo Fu. The impact of multiplicative noise in spdes close to bifurcation via amplitude equations. *Nonlinearity*, 33(8):3905, 2020.

[28] Dirk Blömker and Martin Hairer. Multiscale expansion of invariant measures for spdes. *Communications in mathematical physics*, 251(3):515–555, 2004.

[29] Dirk Blömker, Martin Hairer, and GA Pavliotis. Multiscale analysis for stochastic partial differential equations with quadratic nonlinearities. *Nonlinearity*, 20(7):1721, 2007.

[30] Dirk Blömker and Fu Hongbo. The impact of multiplicative noise in SPDEs close to bifurcation via amplitude equations. *Nonlinearity*, 33(8):3905–3927, 2020.

[31] Philipp Braun and Christopher M Kellett. On (the existence of) control Lyapunov barrier functions. 2017.

[32] Philipp Braun and Christopher M Kellett. Comment on "Stabilization with guaranteed safety using control lyapunov–barrier function". *Automatica*, 122:109225, 2020.

[33] Philipp Braun, Christopher M Kellett, and Luca Zaccarian. Complete control Lyapunov functions: Stability under state constraints. *IFAC-PapersOnLine*, 52(16):358–363, 2019.

[34] Amarjit Budhiraja, Lingji Chen, and Chihoon Lee. A survey of numerical methods for nonlinear filtering problems. *Physica D: Nonlinear Phenomena*, 230(1-2):27–36, 2007.

[35] Nathalie Cauchi, Luca Laurenti, Morteza Lahijanian, Alessandro Abate, Marta Kwiatkowska, and Luca Cardelli. Efficiency through uncertainty: Scalable formal synthesis for stochastic hybrid systems. In *Proceedings of the 22nd ACM International Conference on Hybrid Systems: Computation and Control*, pages 240–251, 2019.

[36] Krishnendu Chatterjee, Laurent Doyen, Thomas A Henzinger, and Jean-François Raskin. Algorithms for omega-regular games with imperfect information. In *International Workshop on Computer Science Logic*, pages 287–302. Springer, 2006.

[37] Mickaël D Chekroun, Honghu Liu, and Shouhong Wang. *Stochastic parameterizing manifolds and non-Markovian reduced equations: stochastic manifolds for nonlinear SPDEs II*. Springer, 2014.

[38] Richard Cheng, Gábor Orosz, Richard M Murray, and Joel W Burdick. End-to-end safe reinforcement learning through barrier functions for safety-critical continuous control tasks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3387–3395, 2019.

[39] Panagiotis D Christofides and Prodromos Daoutidis. Feedback control of hyperbolic pde systems. *AIChE Journal*, 42(11):3063–3086, 1996.

[40] Andrew Clark. Control barrier functions for complete and incomplete information stochastic systems. In *2019 American Control Conference (ACC)*, pages 2928–2935. IEEE, 2019.

[41] Hans Crauel and Franco Flandoli. Additive noise destroys a pitchfork bifurcation. *Journal of Dynamics and Differential Equations*, 10(2):259–274, 1998.

[42] Giuseppe Da Prato, Dariusz Gatarek, and Jerzy Zabczyk. Invariant measures for semilinear stochastic equations. *Stochastic Analysis and Applications*, 10(4):387–408, 1992.

[43] Giuseppe Da Prato and Jerzy Zabczyk. *Stochastic equations in infinite dimensions*. Cambridge university press, 2014.

[44] Giannis Delimpaltadakis, Luca Laurenti, and Manuel Mazo Jr. Abstracting the sampling behaviour of stochastic linear periodic event-triggered control systems. *arXiv preprint arXiv:2103.13839*, 2021.

[45] Xu Chu Ding, Calin Belta, and Christos G Cassandras. Receding horizon surveillance with temporal logic specifications. In *49th IEEE Conference on Decision and Control (CDC)*, pages 256–261. IEEE, 2010.

[46] Xuchu Ding, Stephen L Smith, Calin Belta, and Daniela Rus. Optimal control of markov decision processes with linear temporal logic constraints. *IEEE Transactions on Automatic Control*, 59(5):1244–1257, 2014.

[47] Adel Djaballah, Alexandre Chapoutot, Michel Kieffer, and Olivier Bouissou. Construction of parametric barrier functions for dynamical systems using interval analysis. *Automatica*, 78:287–296, 2017.

[48] Alexandre Donzé, Thomas Ferrere, and Oded Maler. Efficient robust monitoring for stl. In *International Conference on Computer Aided Verification*, pages 264–279. Springer, 2013.

[49] Tommaso Dreossi. Sapo: Reachability computation and parameter synthesis of polynomial dynamical systems. In *Proceedings of the 20th International Conference on Hybrid Systems: Computation and Control*, pages 29–34, 2017.

[50] Maxence Dutreix and Samuel Coogan. Specification-guided verification and abstraction refinement of mixed monotone stochastic systems. *arXiv preprint arXiv:1903.02191*, 2019.

[51] Maxence Dutreix and Samuel Coogan. Specification-guided verification and abstraction refinement of mixed monotone stochastic systems. *IEEE Transactions on Automatic Control*, 66(7):2975–2990, 2020.

[52] Maxence Dominique Henri Dutreix. *Verification and synthesis for stochastic systems with temporal logic specifications*. PhD thesis, Georgia Institute of Technology, 2020.

[53] Peyman Mohajerin Esfahani, Debasish Chatterjee, and John Lygeros. The stochastic reach-avoid problem and set characterization for diffusions. *Automatica*, 70:43–56, 2016.

[54] Stewart N Ethier and Thomas G Kurtz. *Markov processes: characterization and convergence*, volume 282. John Wiley & Sons, 2009.

[55] Lawrence C Evans. Partial differential equations and monge-kantorovich mass transfer. *Current developments in mathematics*, 1997(1):65–126, 1997.

[56] Lawrence C. Evans. *Partial differential equations*. American Mathematical Society, Providence, R.I., 2010.

[57] Georgios E Fainekos, Antoine Girard, Hadas Kress-Gazit, and George J Pappas. Temporal logic motion planning for dynamic robots. *Automatica*, 45(2):343–352, 2009.

[58] Timm Faulwasser, Benjamin Kern, and Rolf Findeisen. Model predictive path-following for constrained nonlinear systems. In *Proceedings of the 48h IEEE Conference on Decision and Control (CDC) held jointly with 2009 28th Chinese Control Conference*, pages 8642–8647. IEEE, 2009.

[59] Aleksei Fedorovich Filippov. *Differential Equations with Discontinuous Righthand Sides*. Springer Science & Business Media, 1988.

[60] Franco Flandoli and KAY Uwe Schaumloffel. Stochastic parabolic equations in bounded domains: random evolution operator and lyapunov exponents. *Stochastics: An International Journal of Probability and Stochastic Processes*, 29(4):461–485, 1990.

[61] Emilio Frazzoli, Munther A Dahleh, and Eric Feron. Maneuver-based motion planning for nonlinear systems with symmetries. *IEEE transactions on robotics*, 21(6):1077–1091, 2005.

[62] Randy Freeman and Petar V Kokotovic. *Robust nonlinear control design: state-space and Lyapunov techniques*. Springer Science & Business Media, 2008.

[63] Laurent Fribourg and Romain Soulat. *Control of switching systems by invariance analysis: applcation to power electronics*. John Wiley & Sons, 2013.

[64] Masoumeh Ghanbarpour and Mohamed Maghenem. Barrier functions for robust safety in differential inclusions, part ii: The converse problem. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 5344–5349. IEEE, 2021.

[65] Alison L Gibbs and Francis Edward Su. On choosing and bounding probability metrics. *International statistical review*, 70(3):419–435, 2002.

[66] Iosif Il'ich Gihman and Anatolij Vladimirovič Skorohod. *Controlled stochastic processes*. Springer Science & Business Media, 2012.

[67] Iosif Ilyich Gikhman and Anatoli Vladimirovich Skorokhod. Stochastic differential equations. In *The theory of stochastic processes III*, pages 113–219. Springer, 2007.

[68] Antoine Girard, Gregor Gossler, and Sebti Mouelhi. Safety controller synthesis for incrementally stable switched systems using multiscale symbolic models. *IEEE Transactions on Automatic Control*, 61(6):1537–1549, jun 2016.

[69] Antoine Girard, Giordano Pola, and Paulo Tabuada. Approximately bisimilar symbolic models for incrementally stable switched systems. *IEEE Transactions on Automatic Control*, 55(1):116–126, 2009.

[70] Paul Glotfelter, Jorge Cortés, and Magnus Egerstedt. Nonsmooth barrier functions with applications to multi-robot systems. *IEEE Control Systems Letters*, 1(2):310–315, 2017.

[71] Rafal Goebel, Ricardo G Sanfelice, and Andrew R Teel. *Hybrid dynamical systems*. Princeton University Press, 2012.

[72] Michel Goossens, Frank Mittelbach, and Alexander Samarin. *The LaTeX Companion*. Addison-Wesley, Reading, Massachusetts, 1994.

[73] Nicolas Gourdain, Frédéric Sicot, Florent Duchaine, and L Gicquel. Large eddy simulation of flows in industrial compressors: a path from 2015 to 2035. *Philosophical Transactions of the Royal society A: Mathematical, Physical and engineering sciences*, 372(2022):20130323, 2014.

[74] Jan Tommy Gravdahl. Modeling and control of surge and rotating stall in compressors. *Journal of Turbomachinery*, (152):544–615, 1998.

[75] J.T. Gravdahl and O. Egeland. Two results on compressor surge control with disturbance rejection. In *Proceedings of the 37th IEEE Conference on Decision and Control (Cat. No.98CH36171)*, volume 3, pages 2563–2568 vol.3, 1998.

[76] G Greiner. Spektrum und asymptotic stark stetiger halbgruppen positiver operatoren. *Sitzungsber. Heidelberger Akad. d. Wiss. (math.-naturw. Klasse)*, 230(1-2):27–36, 1982.

[77] EM Greitzer and FK Moore. A theory of post-stall transients in axial compression systems: part ii—application. 1986.

[78] John Guckenheimer and Philip Holmes. *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*, volume 42. Springer Science & Business Media, 2013.

[79] Yanfeng Guo, Jinqiao Duan, and Donglong Li. Approximation of random invariant manifolds for a stochastic swift-hohenberg equation. *Discrete & Continuous Dynamical Systems-S*, 9(6):1701, 2016.

[80] Iman Haghighi, Austin Jones, Zhaodan Kong, Ezio Bartocci, Radu Gros, and Calin Belta. Spatel: a novel spatial-temporal logic and its applications to networked systems. In *Proceedings of the 18th International Conference on Hybrid Systems: Computation and Control*, pages 189–198, 2015.

[81] Martin Hairer. An introduction to stochastic pdes. *arXiv preprint arXiv:0907.4178*, 2009.

[82] Martin Hairer and Xue-Mei Li. Markov processes. *Course Notes, Imperial College London*, 2020.

[83] Darald J Hartfiel. *Markov Set-Chains*. Springer, 2006.

[84] L'ubica Holá and Dušan Holỳ. Compactness in function spaces. *Topology and its Applications*, 262:20–29, 2019.

[85] Kyle Hsu, Rupak Majumdar, Kaushik Mallik, and Anne-Kathrin Schmuck. Multi-layered abstraction-based controller synthesis for continuous-time systems. In *Proc. of HSCC*, pages 120–129, 2018.

[86] Shao-Chen Hsu, Xiangru Xu, and Aaron D Ames. Control barrier function based quadratic programs with application to bipedal robotic walking. In *2015 American Control Conference (ACC)*, pages 4542–4548. IEEE, 2015.

[87] Mrdjan Jankovic. Combining control Lyapunov and barrier functions for constrained stabilization of nonlinear systems. In *2017 American Control Conference (ACC)*, pages 1916–1922. IEEE, 2017.

[88] Nikolaos Kariotoglou, Sean Summers, Tyler Summers, Maryam Kamgarpour, and John Lygeros. Approximate dynamic programming for stochastic reachability. In *2013 European Control Conference (ECC)*, pages 584–589. IEEE, 2013.

[89] Hassan K Khalil and Jessy W Grizzle. *Nonlinear systems*, volume 3. Prentice hall Upper Saddle River, NJ, 2002.

[90] Rafail Z Khas' minskii. Ergodic properties of recurrent diffusion processes and stabilization of the solution to the cauchy problem for parabolic equations. *Theory of Probability & Its Applications*, 5(2):179–196, 1960.

[91] Rafail Zalmanovich Khas' minskii. Necessary and sufficient conditions for the asymptotic stability of linear stochastic systems. *Theory of Probability & Its Applications*, 12(1):144–147, 1967.

[92] Taihyun Kim and Eyad H Abed. Closed-loop stability monitoring of axial flow compression systems. *Nonlinear Dynamics*, 20(2):181–196, 1999.

[93] Michał Kisielewicz. Stochastic differential inclusions and diffusion processes. *Journal of mathematical analysis and applications*, 334(2):1039–1054, 2007.

[94] Michał Kisielewicz et al. *Stochastic differential inclusions and applications*. Springer, 2013.

[95] Peter E Kloeden and Eckhard Platen. Stochastic differential equations. In *Numerical solution of stochastic differential equations*, pages 103–160. Springer, 1992.

[96] Marius Kloetzer and Calin Belta. A fully automated framework for control of linear systems from temporal logic specifications. *IEEE Transactions on Automatic Control*, 53(1):287–297, 2008.

[97] Donald Knuth. *The TEXbook*. Addison-Wesley, Reading, Massachusetts, 1986.

[98] Miroslav Krstic and Petar V Kokotovic. Lean backstepping design for a jet engine compressor model. In *Proceedings of International Conference on Control Applications*, pages 1047–1052. IEEE, 1995.

[99] Y Kurzweil. On the inversion of the second theorem of Lyapunov on stability of motion. *Czechoslovak Mathematical Journal*, 81(6):217–259, 1956.

[100] Harold J Kushner. Stochastic stability and control. Technical report, Brown Univ Providence RI, 1967.

[101] Yuri A Kuznetsov. *Elements of applied bifurcation theory*, volume 112. Springer Science & Business Media, 2013.

[102] Morteza Lahijanian, Sean B Andersson, and Calin Belta. Formal verification and synthesis for discrete-time stochastic systems. *IEEE Transactions on Automatic Control*, 60(8):2031–2045, 2015.

[103] Vangipuram Lakshmikantham and Srinivasa Leela. *Differential and Integral Inequalities: Theory and Applications: Volume I: Ordinary Differential Equations*. Academic Press, 1969.

[104] Leslie Lamport. *LATEX — A Document Preparation System*. Addison-Wesley, Reading, Massachusetts, second edition, 1994.

[105] Steven M LaValle. *Planning Algorithms*. Cambridge University Press, 2006.

[106] John M Lee. *Smooth Manifolds*. Springer, 2013.

[107] Limei Li, Marco Hernandez, and Kiah Wah Ong. Stochastic attractor bifurcation for the two-dimensional swift-hohenberg equation. *Mathematical Methods in the Applied Sciences*, 41(5):2105–2118, 2018.

[108] Yinan Li. Robustly complete temporal logic control synthesis for nonlinear systems. 2019.

[109] Yinan Li and Jun Liu. ROCS: A robustly complete control synthesis tool for nonlinear dynamical systems. In *Proc. of HSCC*, pages 130–135, 2018.

[110] Yinan Li and Jun Liu. Robustly complete synthesis of memoryless controllers for nonlinear systems with reach-and-stay specifications. *IEEE Transactions on Automatic Control*, 2020.

[111] Yuandan Lin, Eduardo D Sontag, and Yuan Wang. A smooth converse lyapunov theorem for robust stability. *SIAM Journal on Control and Optimization*, 34(1):124–160, 1996.

[112] Jun Liu. Robust abstractions for control synthesis: Completeness via robustness for linear-time properties. In *Proceedings of the 20th International Conference on Hybrid Systems: Computation and Control*, pages 101–110, 2017.

[113] Jun Liu. Closing the gap between discrete abstractions and continuous control: Completeness via robustness and controllability. In *International Conference on Formal Modeling and Analysis of Timed Systems*, pages 67–83. Springer, 2021.

[114] Jun Liu. Converse barrier functions via lyapunov functions. *IEEE Transactions on Automatic Control*, 67(1):497–503, 2022.

[115] Jun Liu, Yiming Meng, Yinan Li, and Maxwell Fitzsimmons. Smooth converse lyapunov-barrier theorems for asymptotic stability with safety constraints and reach-avoid-stay specifications. *arXiv preprint arXiv:2009.04432*, 2020.

[116] Gabriel J Lord, Catherine E Powell, and Tony Shardlow. *An introduction to computational stochastic PDEs*, volume 50. Cambridge University Press, 2014.

[117] Rupak Majumdar, Kaushik Mallik, and Sadegh Soudjani. Symbolic controller synthesis for büchi specifications on stochastic systems. In *Proceedings of the 23rd International Conference on Hybrid Systems: Computation and Control*, pages 1–11, 2020.

[118] Marek T Malinowski and Mariusz Michta. The interrelation between stochastic differential inclusions and set-valued stochastic differential equations. *Journal of Mathematical Analysis and Applications*, 408(2):733–743, 2013.

[119] Xuerong Mao. *Stochastic differential equations and applications*. Elsevier, 2007.

[120] James D Meiss. *Differential Dynamical Systems*. SIAM, 2007.

[121] Yiming Meng, Yinan Li, Maxwell Fitzsimmons, and Jun Liu. Smooth converse lyapunov-barrier theorems for asymptotic stability with safety constraints and reach-avoid-stay specifications. *Automatica*, 144:110478, 2022.

[122] Yiming Meng, Yinan Li, and Jun Liu. Control of nonlinear systems with reach-avoid-stay specifications: A lyapunov-barrier approach with an application to the moore-greizer model. In *2021 American Control Conference (ACC)*, pages 2284–2291, 2021.

[123] Thomas Meurer and Andreas Kugi. Tracking control for boundary controlled parabolic pdes with varying parameters: Combining backstepping and differential flatness. *Automatica*, 45(5):1182–1194, 2009.

[124] Franklin K Moore and Edward Marc Greitzer. A theory of post-stall transients in axial compression systems: Part i—development of equations. 1986.

[125] N Namachchivaya and Henry J Van Roessel. Averaging of noisy nonlinear systems with rapidly oscillating and decaying components. *Nonlinear Dynamics*, 36(2):329–347, 2004.

[126] Ameneh Nejati, Abolfazl Lavaei, Sadegh Soudjani, and Majid Zamani. Data-driven estimation of infinitesimal generators of stochastic systems. *IFAC-PapersOnLine*, 54(5):277–282, 2021.

[127] Quan Nguyen, Xingye Da, JW Grizzle, and Koushil Sreenath. Dynamic walking on stepping stones with gait library and control barrier functions. In *Algorithmic Foundations of Robotics XII*, pages 384–399. Springer, 2020.

[128] Quan Nguyen and Koushil Sreenath. Exponential control barrier functions for enforcing high relative-degree safety-critical constraints. In *Proc. of ACC*, pages 322–328. IEEE, 2016.

[129] Petter Nilsson, Necmiye Ozay, and Jun Liu. Augmented finite transition systems as abstractions for control synthesis. *Discrete Event Dynamic Systems*, 27(2):301–340, 2017.

[130] Ben Niu and Jun Zhao. Barrier Lyapunov functions for the output tracking control of constrained nonlinear switched systems. *Systems & Control Letters*, 62(10):963–971, 2013.

[131] Bernt Øksendal. Stochastic differential equations. In *Stochastic differential equations*, pages 65–84. Springer, 2003.

[132] Ivan Papusha, Jie Fu, Ufuk Topcu, and Richard M Murray. Automata theory meets approximate dynamic programming: Optimal control with temporal logic constraints. In *2016 IEEE 55th Conference on Decision and Control (CDC)*, pages 434–440. IEEE, 2016.

[133] Grigoris Pavliotis and Andrew Stuart. *Multiscale methods: averaging and homogenization.* Springer Science & Business Media, 2008.

[134] Amnon Pazy. *Semigroups of linear operators and applications to partial differential equations*, volume 44. Springer Science & Business Media, 2012.

[135] Szymon Peszat and Jerzy Zabczyk. Strong feller property and irreducibility for diffusions on hilbert spaces. *The Annals of Probability*, pages 157–172, 1995.

[136] Giordano Pola, Antoine Girard, and Paulo Tabuada. Approximately bisimilar symbolic models for nonlinear control systems. *Automatica*, 44(10):2508–2516, 2008.

[137] Jorge I Poveda, Andrew R Teel, and Dragan Nešić. Flexible nash seeking using stochastic difference inclusions. In *2015 American Control Conference (ACC)*, pages 2236–2241. IEEE, 2015.

[138] Stephen Prajna and Ali Jadbabaie. Safety verification of hybrid systems using barrier certificates. In *International Workshop on Hybrid Systems: Computation and Control*, pages 477–492. Springer, 2004.

[139] Stephen Prajna, Ali Jadbabaie, and George J Pappas. A framework for worst-case and stochastic safety verification using barrier certificates. *IEEE Transactions on Automatic Control*, 52(8):1415–1428, 2007.

[140] Stefan Ratschan. Converse theorems for safety and barrier certificates. *IEEE Transactions on Automatic Control*, 63(8):2628–2632, 2018.

[141] Stefan Ratschan and Zhikun She. Providing a basin of attraction to a target region of polynomial systems by computation of Lyapunov-like functions. *SIAM Journal on Control and Optimization*, 48(7):4377–4394, 2010.

[142] Hadi Ravanbakhsh and Sriram Sankaranarayanan. Learning Lyapunov (potential) functions from counterexamples and demonstrations. *arXiv preprint arXiv:1705.09619*, 2017.

[143] Gunther Reissig and Matthias Rungger. Symbolic optimal control. *IEEE Transactions on Automatic Control*, 64(6):2224–2239, 2018.

[144] Muhammad Zakiyullah Romdlony and Bayu Jayawardhana. Stabilization with guaranteed safety using control lyapunov–barrier function. *Automatica*, 66:39–47, 2016.

[145] Matthias Rungger and Majid Zamani. SCOTS: a tool for the synthesis of symbolic controllers. In *Proc. of HSCC*, pages 99–104, 2016.

[146] Naci Saldi, Serdar Yüksel, and Tamás Linder. Finite model approximations for partially observed markov decision processes with discounted cost. *arXiv preprint arXiv:1710.07009*, 2017.

[147] Cesar Santoyo, Maxence Dutreix, and Samuel Coogan. A barrier function approach to finite-time stochastic system verification and control. *Automatica*, 125:109439, 2021.

[148] Meenakshi Sarkar, Debasish Ghose, and Evangelos A Theodorou. High-relative degree stochastic control Lyapunov and barrier functions. *arXiv preprint arXiv:2004.03856*, 2020.

[149] Yasser Shoukry, Pierluigi Nuzzo, Alberto L Sangiovanni-Vincentelli, Sanjit A Seshia, George J Pappas, and Paulo Tabuada. Smc: Satisfiability modulo convex optimization. In *Proceedings of the 20th International Conference on Hybrid Systems: Computation and Control*, pages 19–28, 2017.

[150] Prince Singh, Hoong Chieh Yeong, Huiqing Zhang, Zoi Rapti, and Navaratnam Sri Namachchivaya. Stochastic stability and dynamics of a two-dimensional structurally nonlinear airfoil in turbulent flow. *Meccanica*, 51(11):2665–2688, 2016.

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

[152] Eduardo D Sontag. Comments on integral variants of ISS. *Systems & Control Letters*, 34(1-2):93–100, 1998.

[153] Eduardo D Sontag. *Mathematical Control Theory: Deterministic Finite Dimensional Systems.* Springer Science & Business Media, 1998.

[154] N Sri Namachchivaya and Volker Wihstutz. Almost sure asymptotic stability of scalar stochastic delay equations: Finite state markov process. *Stochastics and Dynamics*, 12(01):1150010, 2012.

[155] Anantharaman Subbaraman. *Robust stability theory for stochastic dynamical systems.* University of California, Santa Barbara, 2015.

[156] Xu Sun, Jinqiao Duan, and Xiaofan Li. An impact of noise on invariant manifolds in nonlinear dynamical systems. *Journal of Mathematical physics*, 51(4):042702, 2010.

[157] Zhibing Sun. *Control of non-deterministic transition systems for linear temporal logic specifications.* PhD thesis, University of Waterloo, 2021.

[158] Paulo Tabuada and George J Pappas. Linear time logic control of discrete-time linear systems. *IEEE Transactions on Automatic Control*, 51(12):1862–1877, 2006.

[159] Keng Peng Tee, Shuzhi Sam Ge, and Eng Hock Tay. Barrier Lyapunov functions for the control of output-constrained nonlinear systems. *Automatica*, 45(4):918–927, 2009.

[160] Andrew R Teel and Laurent Praly. A smooth Lyapunov function from a class-$\mathcal{KL}$ estimate involving two positive semidefinite functions. *ESAIM: Control, Optimisation and Calculus of Variations*, 5:313–367, 2000.

[161] Ilya Tkachev, Alexandru Mereacre, Joost-Pieter Katoen, and Alessandro Abate. Quantitative model-checking of controlled discrete-time markov processes. *Information and Computation*, 253:1–35, 2017.

[162] Moshe Y Vardi. Automatic verification of probabilistic concurrent finite state programs. In *26th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 327–338. IEEE, 1985.

[163] P-CG Vassiliou. Non-homogeneous markov set systems. *Mathematics*, 9(5):471, 2021.

[164] Abraham P Vinod and Meeko MK Oishi. Scalable underapproximation for the stochastic reach-avoid problem for high-dimensional lti systems using fourier transforms. *IEEE control systems letters*, 1(2):316–321, 2017.

[165] John B Walsh. An introduction to stochastic partial differential equations. In *École d'Été de Probabilités de Saint Flour XIV-1984*, pages 265–439. Springer, 1986.

[166] Chuanzheng Wang, Yiming Meng, Yinan Li, Stephen L Smith, and Jun Liu. Learning control barrier functions with high relative degree for safety-critical control. In *2021 European Control Conference (ECC)*, pages 1459–1464. IEEE, 2021.

[167] Chuanzheng Wang, Yiming Meng, Stephen L Smith, and Jun Liu. Safety-critical control of stochastic systems using stochastic control barrier functions. In *2021 IEEE Conference on Decision and Control (CDC)*, pages 5924–5931, 2021.

[168] Chuanzheng Wang, Yiming Meng, Stephen L Smith, and Jun Liu. Data-driven learning of safety-critical control with stochastic control barrier functions. *arXiv preprint arXiv:2205.11513*, 2022.

[169] Chuanzheng Wang†, Yiming Meng†, Stephen L. Smith, and Jun Liu. Safety-critical control of stochastic systems usingstochastic control barrier functions. In *Proceedings of the Conference on Decisions and Control*, 2021.

[170] Jiechang Wen, Mingqing Xiao, and Jianhong Xu. Control of hopf bifurcations of nonlinear infinite-dimensional systems: application to axial flow engine compressor. *Applicable Analysis*, 91(11):1959–1980, 2012.

[171] Peter Wieland and Frank Allgöwer. Constructive safety using control barrier functions. *IFAC Proceedings Volumes*, 40(12):462–467, 2007.

[172] Rafael Wisniewski and Christoffer Sloth. Converse barrier certificate theorems. *IEEE Transactions on Automatic Control*, 61(5):1356–1361, 2015.

[173] Eric M Wolff, Ufuk Topcu, and Richard M Murray. Optimal control with weighted average costs and temporal logic specifications. *Proc. of Robotics: Science and Systems VIII*, (2012), 2012.

[174] Zhe Wu, Fahad Albalawi, Zhihao Zhang, Junfeng Zhang, Helen Durand, and Panagiotis D Christofides. Control Lyapunov-barrier function-based model predictive control of nonlinear systems. *Automatica*, 109:108508, 2019.

[175] MingQing Xiao. Quantitative characteristic of rotating stall and surge for moore–greitzer pde model of an axial flow compressor. *SIAM Journal on Applied Dynamical Systems*, 7(1):39–62, 2008.

[176] Mingqing Xiao and Tamer Basar. Center manifold of the viscous moore–greitzer pde model. *SIAM Journal on Applied Mathematics*, 61(3):855–869, 2000.

[177] Wei Xiao and Calin Belta. Control barrier functions for systems with high relative degree. In *Proc. of CDC*, pages 474–479. IEEE, 2019.

[178] Xiangru Xu. Constrained control of input–output linearizable systems using control sharing barrier functions. *Automatica*, 87:195–201, 2018.

[179] Xiangru Xu, Paulo Tabuada, Jessy W Grizzle, and Aaron D Ames. Robustness of control barrier functions for safety critical control. *IFAC-PapersOnLine*, 48(27):54–61, 2015.

[180] Guang Yang, Calin Belta, and Roberto Tron. Self-triggered control for safety critical systems using control barrier functions. In *2019 American Control Conference (ACC)*, pages 4454–4459. IEEE, 2019.

[181] Serdar Yüksel. Control of stochastic systems. *Course Notes, Queen's University*, 2022.

[182] Majid Zamani, Giordano Pola, Manuel Mazo, and Paulo Tabuada. Symbolic models for nonlinear control systems without stability assumptions. *IEEE Transactions on Automatic Control*, 57(7):1804–1809, 2011.

[183] Hengjun Zhao, Xia Zeng, Taolue Chen, and Zhiming Liu. Synthesizing barrier certificates using neural networks. In *Proceedings of the 23rd International Conference on Hybrid Systems: Computation and Control*, pages 1–11, 2020.

APPENDICES

Appendix A

Linear Temporal Logic (LTL)

As the need of regulating a transition progress or performing surveillance in a dynamical system is growing, controllers are designed to guarantee the satisfaction of complex task specifications with LT properties. A particular class of LT properties can be conveniently specified by LTL. This appendix provides a brief introduction to transition systems, LT properties, LTL and automation methods. We follow the references from [14, 112, 108, 157] for succinct definitions.

Transition systems

Definition A.0.1. *A transition system is a tuple*

$$T = (\mathcal{X}, \mathcal{U}, R, \mathrm{AP}, L),$$

where $\mathcal{X}$ is the set of states; $\mathcal{U}$ is the set of actions (or control inputs); $R \subseteq \mathcal{X} \times \mathcal{U} \times \mathcal{X}$ is the transition relation; AP is the set of atomic propositions; $L : \mathcal{X} \to 2^{\mathrm{AP}}$ is the labelling function.

An execution of T is an infinite sequence of state and actions, i.e.,

$$\mathcal{Z} = x_0 u_0 x_1 u_1 x_2 u_2 \cdots ,$$

where $x_i \in \mathcal{X}$ and $u_i \in \mathcal{U}$ for all i. The path of $\mathcal{Z}$ is the extraction of the sequence of state evolution, i.e.,

$$\mathrm{Path}(\mathcal{Z}) = x_0 x_1 x_2 \cdots .$$

The trace of $\mathcal{Z}$ is defined by the sequence of observations generated by the labelling function,

$$\mathrm{Trace}(\mathcal{Z}) = L(x_0) L(x_1) L(x_2) \cdots .$$

LT properties

An LT property over a set of atomic proposition AP is a subset of the power set of $(2^{\mathrm{AP}})^\omega$, which is defined by

$$(2^{\mathrm{AP}})^\omega = \{l_0 l_1, \cdots : l_i \in 2^{\mathrm{AP}},\ i \geq 0\}.$$

We also call $(2^{\mathrm{AP}})^\omega$ the set of all infinite words (or infinite repetition) over the alphabet 2^{AP}.

Remark A.0.2. *Here, we use ω to denote the infinity.*

Remark A.0.3. *LT properties of a transition system are evaluated over its traces.*

Definition A.0.4 (ω-regular properties)**.** *LT property P over AP is called ω-regular if P is an ω-regular language over 2^{AP}, which admits the inductive form*

- ✧ *A^ω where A is a nonempty regular language*[1] *not containing the empty string;*
- ✧ *AB, the concatenation of a regular language A and an ω-regular language B;*
- ✧ *$A \cup B$ where A and B are ω-regular languages.*

We now introduce LTL for specifying a particular class of LT properties. All LTL formulas are ω-regular. An LTL formula consists of propositional logic operators (e.g., true ($\top$), negation ($\neg$), and conjunction $\wedge$)), and temporal operators (e.g., next ($\bigcirc$) and until (**U**)). The syntax of LTL over AP is defined inductively:

$$\varphi ::= \top \mid p \mid \varphi_1 \wedge \varphi_2 \mid \neg\varphi \mid \bigcirc \varphi \mid \mathbf{U}\varphi,$$

and reads as

- ✧ $\varphi = \top$ is an LTL formula;
- ✧ $\varphi = p \in \mathrm{AP}$ is an LTL formula;
- ✧ if $\varphi, \varphi_1, \varphi_2$ are LTL formulas, then $\neg\varphi$, $\varphi_1 \wedge \varphi_2$, $\bigcirc\varphi$ and $\mathbf{U}\varphi$ are also LTL formulas.

Other commonly used logical operators can be defined based on the syntax, for example,

1. $\varphi_1 \vee \varphi_1 := \neg(\neg\varphi_1 \wedge \neg\varphi_2)$;

[1] We omit the definition for regular language to prevent extra confusion. As an analogy, regular languages are generated in a similar way as generating algebras based on union and concatenation operations.

2. $(\varphi_1 \implies \varphi_2) := \neg\varphi_2 \wedge \varphi_2$;

3. (eventually) $\Diamond\varphi := \top\mathbf{U}\varphi$;

4. (always) $\Box\varphi := \neg\Diamond\neg\varphi$.

The semantics of LTL are defined over an abstract system model with a run of words (observations) $w = w_0 w_1 w_2 ...$ in 2^{AP}. The satisfaction of an LTL formula φ by w at position i is written as $w_i \models \varphi$, which is inductively defined as:

1. $w_i \models \top$ if and only if $w_i = \top$;

2. $w_i \models p$ if and only if $p \in w_i$;

3. $w_i \not\models \neg\varphi$ if and only if $w_i \models p$;

4. $w_i \models \bigcirc\varphi$ if and only if $w_{i+1} \models \varphi$;

5. $w_i \models \varphi_1 \vee \varphi_2$ if and only if $w_i \models \varphi_1 \ \vee w_i \models \varphi_2$;

6. $w_i \models \varphi_1 \mathbf{U} \varphi_2$ if and only if $\exists j \geq i$ s.t. $w_j \models \varphi_2$ and $w_n \models \varphi_1 \ \forall i \leq n < j$.

A run of words $w \models \varphi$ if $w_0 \models \varphi$. An execution $\mathcal{Z}$ of a transition system T is said to satisfy an LTL formula φ, written as $\mathcal{Z} \models \varphi$, if and only if its trace $\mathrm{Trace}(\mathcal{Z}) \models \varphi$.

Automation

The underneath idea of an automation is to keep track of an infinite word using a small finite graph.

Definition A.0.5. *A DA is a tuple $DA = (Q, L, R, q_0, Acc)$, where Q is the set of automation states; $L \subseteq 2^{\mathrm{AP}}$ is the set of labels; $R : L \times Q \to Q$ is a deterministic transition function; $q_0 \in Q$ is the unique initial state; $Acc \subseteq 2^{\mathrm{Q}}$ is the set of final states in a finite automaton or the set of accepting states in an ω-automation.*

Remark A.0.6. *A NA generalizes DA by considering a set of initial states $Q_0 \subseteq Q$, and a non-deterministic transition $R : L \times Q \to 2^{\mathrm{Q}}$. Given a label l, e.g. from some LTL formula, to an NA state q_i, there may be more than one transitions that cause us to lose track of the status of a token. An NA cannot be used to synthesize controllers or to solve a game, but can be used in a posteriori explanation or verification of LTL formulas, i.e., the reason why the run of the token satisfies the accepting condition of the NA or not.*

Remark A.0.7. *Compared to NA, a DA can be considered as a graph, where the automaton states Q are the nodes and the transition function R is hidden in the edges.*

We compare NBA and DRA due to their equivalent expressive power in terms of recognizing the complete ω-regular languages. Apart from the general differences between an NA and a DA, the accepting conditions are different.

(1) Büchi Condition. An infinite run of automation states is accepted by NBA if and only if one of the states in Acc occurs infinitely many often in the run.

(2) Rabin Condition. Acc $= \{(G_1, B_1), \cdots, (G_k, B_k)\}$, where $G_i, B_i \subseteq Q$, $i = 1, \cdots, k$. G_i's are referred as the set of 'Good' states, whereas B_i's are the set of 'Bad' states. An infinite run of automation states is accepted by DRA if and only if at least one of the states in G_i's occurs infinitely often and the states in B_i's occur only finitely many times in the run.

We prefer to use the systematic automatic conversion from an LTL formula φ to an automation to increase the efficiency. After a φ is converted into an NBA, we can always determinize it into a DRA. The trade-off is, however, the complexity issue.

Remark A.0.8. *In the control or verification problems, a trajectory satisfies the accepting condition of an LTL formula φ if its input word satisfies the accepting condition of the DA converted from φ. We find the winning set and the winning strategy such that all the controlled (or automatic) trajectories initialized in the winning set subjected to the winning strategy satisfy the accepting condition.*

Note that, by connecting non-deterministic transition systems with some DA generated from som LTL formula, the control and verification problems w.r.t. such an LTL specification can be automatically solved based on the product graph as a graph searching problem. The size of the non-deterministic transition systems is the dominating factor of the size of the product graph.

Appendix B

Martingales, Markov Processes, and Martingale Problem

This appendix provides a brief introduction to martingales, Markov processes, and the martingale problem. Details and more rigorous descriptions can be found in .

Stochastic processes, filtrations, and martingales

A stochastic process is a collection $X := \{X_t\}_{t\in\mathbb{T}}$ indexed by a time parameter $\mathbb{T}$, where $\mathbb{T}$ falls into the following three cases: 1) discrete time $\mathbb{T} = \mathbb{N}$; 2) continuous time finite horizon $\mathbb{T} = [0, T]$ for some $T > 0$; and 3) continuous time infinite horizon $\mathbb{T} = [0, \infty)$. On the canonical spaces $(\mathbf{\Omega}, \mathcal{F}, \mathbf{P})$, the mapping $t \mapsto X_t(\varpi)$ for some fixed $\varpi \in \mathbf{\Omega}$ is called a sample path.

A filtration is a collection of σ-field $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ with the property $\mathcal{F}_s \subseteq \mathcal{F}_t$ for $s, t \in \mathbb{T}$ and $s \leq t$. Given a stochastic process X, the natural filtration up to time t generated by X is the family of σ-fields $\mathcal{F}_t := \sigma\{X_s : s \in \mathbb{T}, s \leq t\}$, which contains the information about the evolution of X up to time t.

Now consider a given filtration $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$, we say that a stochastic process X is adapted to $\{\mathcal{F}_t\}_{t\in\mathbb{T}}$ if X is $\mathcal{F}_t$-measurable for each t. Roughly speaking, this means that X does not look into the information available in future.

Definition B.0.1 (Martingale)**.** *For a fixed filtered probability space* $(\mathbf{\Omega}, \mathcal{F}, \{\mathcal{F}_t\}_{t\in\mathbb{T}}, \mathbf{P})$*, a process* X *is called a martingale if* X *is adapted,* $\mathbf{E}[X_t|\mathcal{F}_s] = X_s$ *for all* $s < t$*, and* $\mathbf{E}[X_t] < \infty$ *for all* t*.*

We can also consider a decreasing filtration and backward martingales, based on which the $\mathcal{L}^1$ convergence of LLN. For simplicity, we only consider the situation when $\mathbb{T} = \mathbb{N}$.

Definition B.0.2 (Backward martingale)**.** *A backward martingale is a stochastic process $\{X_{-n}\}_{n=1,2,\cdots}$ such that, for each n, X_{-n} is $\mathcal{L}_1$ integrable and $\mathcal{F}_{-n}$-measurable, and satisfies*

$$\mathbf{E}[X_{-n-1} \mid \mathcal{F}_{-n}] = X_{-n}. \tag{B.1}$$

Theorem B.0.3 (Backward martingale convergence theorem)**.** *For every backward maringale, as $n \to \infty$,*

$$X_{-n} \to \mathbf{E}[X_{-1} \mid \mathcal{F}_{-\infty}] \ \ \mathbf{P}\text{-a.s. and in } \mathcal{L}^1. \tag{B.2}$$

Theorem B.0.4 (Kolmogorov's 0-1-law)**.** *Let $\mathcal{F}_1, \mathcal{F}_2, \cdots$ be independent σ-fields and denote by $\mathcal{F}_\infty = \cap_{n=1}^\infty \sigma\left(\cup_{k=n}^\infty \mathcal{F}_k\right)$ the corresponding tail field. Then*

$$\mathbf{P}[A] \in \{0, 1\}, \ \ \forall A \in \mathcal{F}_\infty.$$

Proof of LLN: Let Y_i be $\mathcal{L}^1$ integrable and i.i.d. w.r.t. $\mathbf{P}$. Let $S_n = \sum_{i=1}^n Y_i$ be the finite sum and let $X_{-n} = \frac{S_n}{n}$ be the average. Then the σ-field $\mathcal{F}_{-n} = \sigma\{S_n, S_{n+1}, \cdots\}$ is a decreasing filtration. Due to the independence of $\{Y_i\}$, we have

$$\begin{aligned}\mathbf{E}[X_{-1} \mid \mathcal{F}_{-n}] &= \mathbf{E}[Y_1 \mid S_n, S_{n+1}, \cdots] \\ &= \mathbf{E}[Y_1 \mid S_n, Y_{n+1}, Y_{n+2}, \cdots] \\ &= \mathbf{E}[Y_1 \mid S_n].\end{aligned} \tag{B.3}$$

Notice that $\mathbf{E}[Y_i \mid S_n] = \mathbf{E}[Y_j \mid S_n]$ by symmetry for $i, j \in \{1, \cdots, n\}$, then

$$n\mathbf{E}[Y_i \mid S_n] = \sum_{i=1}^n \mathbf{E}[Y_i \mid S_n] = \mathbf{E}[S_n \mid S_n] = S_n. \tag{B.4}$$

Combining the above, we have $\mathbf{E}[X_{-1} \mid \mathcal{F}_{-n}] = \frac{S_n}{n} = X_{-n}$, which verifies that $\{X_{-n}\}$ is a backward martingale. By the backward martingale convergence theorem, we immediately have

$$\frac{S_n}{n} \to \mathbf{E}[Y_1 \mid \mathcal{F}_{-\infty}], \ \ \mathbf{P}\text{-a.s. and in } \mathcal{L}^1.$$

By Kolmogorov's 0-1 law, we have that all A in the tail field $\mathscr{F}_{-\infty}$ have probability either 0 or 1, which in turn implies that the conditional expectation $\mathbf{E}[Y_1 \mid \mathscr{F}_{-\infty}]$ must be a constant (by the definition of conditional expectation) and should be equal to the average $\mathbf{E}[Y_1]$. In words, $\frac{\sum_{i=1}^n Y_i}{n} \to \mathbf{E}[Y_1]$ $\mathbf{P}$-a.s. and in $\mathcal{L}^1$.

Markov Processes

Consider a complete and separable state space $\mathcal{X}$.

Definition B.0.5 (Markov process)**.** *A stochastic process X is said to be a Markov process if each X is adapted and, for any $\Gamma \in \mathscr{B}(\mathcal{X})$ and $t > s$, we have*

$$\mathbf{P}[X_t \in \Gamma \mid \mathcal{F}_s] = \mathbf{P}[X_t \in \Gamma \mid X_s], \quad \text{a.s.} \tag{B.5}$$

Correspondingly, for every t, we define the transition operator $\mathcal{T}_{s,t} : C_b(\mathcal{X}) \to C_b(\mathcal{X})$ as

$$(\mathcal{T}_{s,t} f)(x) := \mathbf{E}[f(X_t) \mid X_s = x], \; s \leq t. \tag{B.6}$$

The function $\Theta_{s,t}(x, \Gamma) = \mathcal{T}_{s,t}\mathbb{1}_\Gamma(x)$, $\Gamma \in \mathscr{B}(\mathcal{X})$, is called transition function (probability). We denote $\Theta_{s,t} := \{\Theta_{s,t}(x, \Gamma) : \; x \in \mathcal{X}, \; \Gamma \in \mathscr{B}(\mathcal{X})\}$ as the family of transition probabilities from s to t.

Remark B.0.6. *Note that $\{\mathcal{T}_{s,t}\}_{s,t \in \mathbb{T}}$ is a semigroup. Homogeneous (autonomous) Markov processes are such that $\mathcal{T}_{s,t} = \mathcal{T}_{0,t-s}$ for all $t \neq s$.*

Remark B.0.7. *For discrete time Markov processes, i.e. $\mathbb{T} = \mathbb{N}$, we specifically consider one-step transition function at every $t \in \mathbb{T}$, which is defined as*

$$\Theta_t(x, \Gamma) := \mathbf{P}[X_{t+1} \in \Gamma \mid X_t = x], \;\; \Gamma \in \mathscr{B}(\mathcal{X}). \tag{B.7}$$

We denote correspondingly $\Theta_t := \{\Theta_t(x, \Gamma) : \;\; x \in \mathcal{X}, \;\; \Gamma \in \mathscr{B}(\mathcal{X})\}$ as the family of one-step transition probabilities at time t. Homogeneous (autonomous) Markov processes are such that $\Theta_t = \Theta_s$ for all $t \neq s$, and the n-step transition can be recursively defined by $\Theta^{n+1}(x, \cdot) = \int_{\mathcal{X}} \Theta(y, \cdot)\Theta^n(x, dy)$ with any initial distribution $\Theta^0(x, \cdot) = \delta_x$.

The generator of a time-homogenous Markov process is defined as in Definition 3.1.8. Any finite dimensional distribution can be constructed based on the transition function and the Chapman-Kolmogorov property. We use a discrete-time homogenous Markov process X with initial distribution $\mu(dx)$ as an example, the finite-dimension distribution is given as

$$\begin{aligned} &\mathbf{P}[X_0 \in A_0, X_1 \in A_1, \cdots, X_n \in A_n] \\ =& \int_{A_0} \int_{A_1} \cdots \int_{A_{n-1}} \Theta(x_{n-1}, A_n)\Theta(x_{n-2}, dx_{n-1}) \cdots \Theta(x_0, dx_1)\mu_0(dx_0) \end{aligned} \tag{B.8}$$

The probability law $\mathbf{P}$ of X on the canonical space can be guaranteed by the famous Kolmogorov extension theorem.

Martingale problem

Consider system $\mathscr{S}$ (see Equation (3.1)) with $X_0 = x$ a.s.. For any test function $f \in C_b^2(\mathcal{X})$, we define

$$M_t^f = f(X_t) - f(x) - \int_0^t \mathfrak{L} f(X_s) ds. \tag{B.9}$$

It can be verified that the process M^f is a martingale.

Now consider an inverse problem. Suppose X is a continuous process such that the process M^f defined in (B.9) is a martingale, then it can be shown that there exists a Wiener process W such that the process X satisfies the SDE (3.1) (with $\vartheta = 0$). In other words, X becomes a weak solution of $\mathscr{S}$ if the above M^f is a martingale. We say that X solves (or the probability law $\mathbf{P}$ of X solves) the martingale problem for operator $\mathfrak{L}$ if M^f is a martingale under $\mathbf{P}$ for valid test functions f. The martingale problem for $\mathfrak{L}$ is called well posed if there is a unique probability law solving the martingale problem.

Remark B.0.8. *Note that the martingale problem for finite-dimension system can be applied to construct weak (in terms of probability) solutions for SPDEs (see procedures in [43, Chapter 8]). Since weak solution in the PDE context has a different meaning, we also use the term martingale solutions to refer the notion of weak solutions in Definition 3.1.3 (2).*

Appendix C

Controlled Stochastic Processes

Let the set of states $\mathcal{X}$ and set of control $\mathcal{U}$ be given. Suppose that $\mathcal{X}$ and $\mathcal{U}$ are complete and separable metric spaces.

C.1 Canonical Setup for Discrete-Time Controlled Processes

Let $(\Omega^\dagger, \mathscr{F}^\dagger, \mathbb{P}^\dagger)$ be a probability space (most likely unknown). Without loss of generality, we assume that a stochastic process X and a process of control values $\mathfrak{u}$ are defined on $(\Omega^\dagger, \mathscr{F}^\dagger, \mathbb{P}^\dagger)$, and that $\mathfrak{u}_t$ is provided (according to some rule) at each instant of time. To describe the controlled process, it is necessary to define the probability law of X and the rule that $\mathfrak{u}$ is selected.

Definition C.1.1. *Given processes X and $\mathfrak{u}$, for any fixed $T > 0$, we define the following short hand notations:*

$$X_{[0,T]} := \{X_t\}_{t\in[0,T]}, \text{ and } \mathfrak{u}_{[0,T]} := \{\mathfrak{u}_t\}_{t\in[0,T]} \tag{C.1}$$

Suppose X is fully observed, it is natural to assume that the marginal distribution of X_T for each T is completely determined by $X_0, \cdots, X_{T-1}$ and the values of the controls $\mathfrak{u}_0, \cdots, \mathfrak{u}_{T-1}$. Let $\mu_T(\cdot \mid \cdot)$ be such that

$$\mu_T(\Gamma \mid X_{[0,T-1]}; \mathfrak{u}_{[0,T-1]}) = \mathbb{P}^\dagger[X_t \in \Gamma \mid X_{[0,T-1]}; \mathfrak{u}_{[0,T-1]}], \quad \Gamma \in \mathscr{B}(\mathcal{X}), \tag{C.2}$$

then for each realization, $\mu_T(\cdot \mid X_0 = x_0, \cdots, X_{T-1} = x_{T-1}; \mathfrak{u}_0 = u_0, \cdots, \mathfrak{u}_{T-1} = u_{T-1})$ defines the (transitional) conditional distribution of X_T, and for each $\Gamma \in \mathscr{B}(\mathcal{X})$, the random variable $\mu_T(\Gamma \mid X_{[0,T-1]}; \mathfrak{u}_{[0,T-1]})$ is jointly measurable w.r.t. the product σ-algebra $\mathscr{B}(\mathcal{X}^T) \otimes \mathscr{B}(\mathcal{U}^T)$.

For any initial distribution $\mu_0 \in \mathfrak{P}(\mathcal{X})$, given any realization of $\mathfrak{u}$ and fixed $T > 0$, the finite-dimensional distribution of $X_{[0,T]}$ is

$$\begin{aligned}&\int_{\Gamma_0} \mu_0(dx_0) \int_{\Gamma_1} \mu_1(dx_1 \,|X_0 = x_0; \mathfrak{u}_0 = u_0) \\ &\cdots \int_{\Gamma_T} \mu_T(dx_T \,|X_0 = x_0, \cdots X_{T-1} = x_{T-1}; \mathfrak{u}_0 = u_0, \cdots, \mathfrak{u}_{T-1} = u_{T-1}) \\ =:&\mathfrak{p}^T \left(\prod_{i=0}^{T} \Gamma_i \mid \mathfrak{u}_0 = u_0, \cdots, \mathfrak{u}_{T-1} = u_{T-1} \right)\end{aligned} \tag{C.3}$$

where $dx_i = [x_i, x_i + dx]$ for $i \in \{1, \cdots, T\}$. It is clear that the finite-dimensional distribution $\mathfrak{p}^T$ of $X_{[0,T]}$ on the cylinder set $\prod_{i=0}^{T} \Gamma_i \in \mathscr{B}(\mathcal{X}^T)$ is only dependent on the realization of $\mathfrak{u}_{[0,T-1]}$. Note that by Kolmogorov's extension theorem, for any known process of control values, there exists a unique $\mathfrak{p}(\cdot \mid \mathfrak{u})$ on $\mathcal{X}^\infty$.

Now we determine how $\mathfrak{u}$ is generated. It is natural to suppose that the selection of a control at time T is based on $X_{[0,T]}$ and $\mathfrak{u}_{[0,T-1]}$. For each fixed $T > 0$, let $\kappa_T(\cdot \mid \cdot)$ be such that

$$\kappa_T(\mathfrak{C} \mid X_{[0,T]}; \mathfrak{u}_{[0,T-1]}) = \mathbb{P}^\dagger[\mathfrak{u}_T \in \mathfrak{C} \mid X_{[0,T]}; \mathfrak{u}_{[0,T-1]}], \quad \mathfrak{C} \in \mathscr{B}(\mathcal{U}), \tag{C.4}$$

then for each realization, $\kappa_T(\cdot \mid X_0 = x_0, \cdots, X_T = x_T; \mathfrak{u}_0 = u_0, \cdots, \mathfrak{u}_{T-1} = u_{T-1})$ defines the conditional distribution of $\mathfrak{u}_T$, and for each $\mathfrak{C} \in \mathscr{B}(\mathcal{U})$, the random variable $\kappa_T(\Gamma \mid X_{[0,T]}; \mathfrak{u}_{[0,T-1]})$ is jointly measurable w.r.t. the product σ-algebra $\mathscr{B}(\mathcal{X}^{T+1}) \otimes \mathscr{B}(\mathcal{U}^T)$.

Definition C.1.2. *A randomized control policy is the sequence*

$$\kappa = \{\kappa_t, \ t \geq 0\},$$

where, for each $t \geq 0$, κ_t is given in the form of (C.4).

Similarly, given any realization of $\mathcal{X}$ and fixed $T > 0$, the finite-dimensional distribution of $\mathfrak{u}_{[0,T]}$ under the policy κ is

$$\begin{aligned}&\int_{\mathfrak{C}_0} \kappa_0(du_0 \mid X_0 = x_0) \int_{\mathfrak{C}_1} \kappa_1(du_1 \,|X_0 = x_0, X_1 = x_1; \mathfrak{u}_0 = u_0) \\ &\cdots \int_{\mathfrak{C}_T} \kappa_T(dx_T \,|X_0 = x_0, \cdots X_T = x_T; \mathfrak{u}_0 = u_0, \cdots, \mathfrak{u}_{T-1} = u_{T-1}) \\ =:&\mathfrak{d}^T \left(\prod_{i=0}^{T} \mathfrak{C}_i \mid \mathcal{X}_0 = x_0, \cdots, \mathcal{X}_T = x_T \right)\end{aligned} \tag{C.5}$$

where $du_i = [u_i, u_i + du]$ for $i \in \{1, \cdots, T\}$. Given the control policy κ, it is clear that the finite-dimensional distribution $\mathfrak{d}^T$ of $\mathfrak{u}_{[0,T]}$ on the cylinder set $\prod_{i=0}^{T} \mathfrak{C}_i \in \mathscr{B}(\mathcal{U}^T)$ is only dependent on the realization of $\mathcal{X}_{[0,T]}$. Again, by Kolmogorov's extension theorem, suppose the process X is known, there exists a unique law $\mathfrak{d}(\cdot \mid \mathcal{X})$ on $\mathcal{U}^\infty$.

Suppose the initial distribution $\mu_0 \in \mathfrak{P}(\mathcal{X})$ and a control policy is given, one can construct a product process $(X, \mathfrak{u}) := \{(X_t, \mathfrak{u}_t)\}_{t \geq 0}$ with law $\mathbf{P}^{\mu_0,\kappa}$. We also alternatively denote $X^\mathfrak{u}$ by the controlled process if we emphasize on the state-space marginal of $(X, \mathfrak{u})$. Note that every finite distribution of $(X, \mathfrak{u})$ is given as

$$\begin{aligned}
&\mathbf{P}^{\mu_0,\kappa}[X_0 \in \Gamma_0, \mathfrak{u}_0 \in \mathfrak{C}_0, \cdots, X_T \in \Gamma_T, \mathfrak{u}_T \in \mathfrak{C}_T] \\
=&\int_{\Gamma_0} \mu_0(dx_0) \int_{\mathfrak{C}_0} \kappa_0(du_0 \mid X_0 = x_0) \cdots \\
&\times \int_{\Gamma_T} \mu_T(dx_T \,|X_0 = x_0, \cdots X_{T-1} = x_{T-1}; \mathfrak{u}_0 = u_0, \cdots, \mathfrak{u}_{T-1} = u_{T-1}) \\
&\times \int_{\mathfrak{C}_T} \kappa_T(d\mathfrak{u}_T \,|X_0 = x_0, \cdots X_T = x_T; \mathfrak{u}_0 = u_0, \cdots, \mathfrak{u}_{T-1} = u_{T-1}).
\end{aligned} \tag{C.6}$$

In particular, the transitional distribution of the controlled process X at each T is such that

$$\begin{aligned}
&\mathbf{P}^{\mu_0,\kappa}[X_T \in \Gamma_T \mid X_0 = x_0, \mathfrak{u}_0 = u_0, \cdots X_{T-1} = x_{T-1}, \mathfrak{u}_{T-1} = u_{T-1}] \\
=&\mu_T(\Gamma_T \,|X_0 = x_0, \cdots X_{T-1} = x_{T-1}; \mathfrak{u}_0 = u_0, \cdots, \mathfrak{u}_{T-1} = u_{T-1});
\end{aligned} \tag{C.7}$$

whereas

$$\begin{aligned}
&\mathbf{P}^{\mu_0,\kappa}[\mathfrak{u}_T \in \mathfrak{C}_T \mid X_0 = x_0, \mathfrak{u}_0 = u_0, \cdots X_{T-1} = x_{T-1}, \mathfrak{u}_{T-1} = u_{T-1}, X_T = x_T] \\
=&\kappa_T(\mathfrak{C}_T \,|X_0 = x_0, \cdots X_{T-1} = x_{T-1}; \mathfrak{u}_0 = u_0, \cdots, \mathfrak{u}_{T-1} = u_{T-1}, X_T = x_T).
\end{aligned} \tag{C.8}$$

Let

$$\mathcal{F} := \sigma\{(X_t, \mathfrak{u}_t) \in (\Gamma, \mathfrak{C}),\ (\Gamma, \mathfrak{C}) \in \mathscr{B}(\mathcal{X}) \otimes \mathscr{B}(\mathcal{U}),\ \ t \in \mathbb{N}\}.$$

We then work on the space $((\mathcal{X} \times \mathcal{U})^\infty, \mathcal{F}, \mathbf{P})$ rather than the original probability space $(\Omega^\dagger, \mathscr{F}^\dagger, \mathbb{P}^\dagger)$.

C.2 Canonical Setup for Continuous-Time Controlled Processes

Without loss of generality, we consider non-randomized control. Note that, by [66], control policy can not always be established if we attempt a sequential definition as in the discrete-time cases. However, if we apply a step control strategy, the difficulty can be conquered.

Definition C.2.1. *[66] A control $\mathfrak{u}(t)$ is called a step control if $\mathfrak{u}(t)$ is piecewise constant on some given time horizon $[0, T]$, i.e., for some n and $0 = t_0 < t_1 < \cdots < t_n = T$,*

$$\mathfrak{u}(t) = \mathfrak{u}(t_k) = u_k, \ t \in [t_k, t_{k+1}).$$

In fact, a constant control (within some time horizon $[0, T]$) uniquely determines the joint distributions of $(X, \mathfrak{u})$ on the given horizon $[0, T]$. At the jump points t_k's, the new control values are decided. One can verify that the controlled object and the step control uniquely determine the joint distributions of the controlled process in a similar manner as discrete-time controlled processes.

We can define a step control policy κ similar to Definition C.1.2, where the non-random decision is made at the jump points. We denote by $\mathbf{P}^{\mu_0,\kappa}$ the law of $(X, \mathfrak{u})$ given the initial distribution $\mu_0 \in \mathfrak{P}(\mathcal{X})$ and the step control policy κ.

C.3 Controlled Markov Models and Classes of Control Policies

We focus on discrete-time systems. Consider

$$X_{t+1} = f(X_t, \mathfrak{u}_t, w_t), \tag{C.9}$$

where f is a measurable function, the state $X_t(\varpi) \in \mathcal{X}$ for all $t \in \mathbb{N}$, $\mathfrak{u}$ is a $\mathcal{U}$-valued control signal, $\{w_t\}_{t\in\mathbb{N}}$ are i.i.d. Gaussian random variables. We have seen in Section C.1 that, given an initial distribution μ_0 and a control policy κ, (C.9) generated processes $(X, \mathfrak{u}) := \{(X_t, \mathfrak{u}_t)\}_{t\geq 0}$ with the probability law $\mathbf{P}^{\mu_0,\kappa}$. The model (C.9) also possesses a Markov transition property in the sense that

$$\mathbf{P}^{\mu_0,\kappa}[X_{t+1} \in \Gamma \mid X_{[0,t]}, \mathfrak{u}_{[0,t]}] = \mathbf{P}^{\mu_0,\kappa}[X_{t+1} \in \Gamma \mid X_t, \mathfrak{u}_t]. \tag{C.10}$$

We further define the Markov transition function as

$$\Theta_t^u(x, \Gamma) = \mathbf{P}^{\mu_0,\kappa}[X_{t+1} \in \Gamma \mid X_t = x, \mathfrak{u}_t = u]. \tag{C.11}$$

Classes of control policies

We suppose that X is fully observed. Admissible control policies $\mathfrak{K}_A$ is a class of control policies such that $\kappa_t \in \mathfrak{P}(\mathcal{U})$ for all $t \in \mathbb{N}$, i.e., $\kappa_t(\,\cdot \mid X_{[0,t]}, \mathfrak{u}_{[0,t-1]})$ is a (random) measure on $\mathcal{U}$. Given any realization of the history $(X_{[0,t]}, \mathfrak{u}_{[0,t-1]})$, if κ_t is a Dirac measure $\delta_{\{u\}}$ for some $u \in \mathcal{U}$ and for all $t \in \mathbb{N}$, we call it a deterministic admissible control policy. It is equivalent to write $\mathfrak{u}_t = \kappa_t(X_{[0,t]}, \mathfrak{u}_{[0,t-1]})$ in this case. Note that the notion of 'deterministic' is in term of the output based on the history realization $(X_{[0,t]}, \mathfrak{u}_{[0,t-1]})$.

We focus on Markov control policies $\mathfrak{K}_M$, which is a subclass of $\mathfrak{K}_A$.

Definition C.3.1 (Markov policies). *A policy is Markov if, for all $t \in \mathbb{N}$,*

$$\kappa_t(\mathfrak{u}_t \in \mathfrak{C} \mid X_{[0,t]}, \mathfrak{u}_{[0,t-1]}) = \kappa_t(\mathfrak{u}_t \in \mathfrak{C} \mid X_t), \ \forall \mathfrak{C} \in \mathscr{B}(\mathcal{U}). \tag{C.12}$$

A deterministic Markov policy is such that $\mathfrak{u}_t = \kappa_t(X_t)$ for all $t \in \mathbb{N}$.

By (C.8), a straightforward consequence of κ being Markov is that

$$\mathbf{P}^{\mu_0,\kappa}[\mathfrak{u}_t \in \mathfrak{C} \mid X_{[0,t]}, \mathfrak{u}_{[0,t-1]}] = \kappa_t(\mathfrak{u}_t \in \mathfrak{C} \mid X_t), \ \forall \mathfrak{C} \in \mathscr{B}(\mathcal{U}).$$

Definition C.3.2. *A class $\mathfrak{K}_S \subseteq \mathfrak{K}_M$ is called stationary if for all $\kappa \in \mathfrak{K}_S$, we have $\kappa_t = \kappa_s$ for all $t \neq s$. The deterministic stationary policy can be defined accordingly.*

The following proposition shows a nice property of controlled Markov model under a Markov policy κ.

Proposition C.3.3. *Let $\kappa \in \mathfrak{K}_M$. Then the state process X (or $X^{\mathfrak{u}}$) becomes a Markov process, i.e., for every $t \in \mathbb{N}$, we have*

$$\mathbf{P}^{\mu_0,\kappa}(X_t \in \Gamma \mid X_{[0,t]}, \mathfrak{u}_{[0,t-1]}] = \mathbf{Q}^{\kappa}[X_{t+1} \in \Gamma \mid X_t], \quad \Gamma \in \mathscr{B}(\mathcal{X}),$$

where $\mathbf{Q}^{\kappa}$ is generally a transition kernel defining a Markov chain (recall (B.7)).

Proof.

$$\begin{aligned}
&\mathbf{P}^{\mu_0,\kappa}[X_{t+1} \in \Gamma \mid X_0 = x_0, \cdots X_t = x_t] \\
&= \int_{\mathcal{U}} \mathbf{P}^{\mu_0,\kappa}[X_{t+1} \in \Gamma, \mathfrak{u}_t \in du \mid X_0 = x_0, \cdots X_t = x_t] \\
&= \int_{\mathcal{U}} \mathbf{P}^{\mu_0,\kappa}[X_{t+1} \in \Gamma \mid \mathfrak{u}_t = u, X_0 = x_0, \cdots X_t = x_t] \mathbf{P}^{\mu_0,\kappa}[\mathfrak{u}_t \in du \mid X_0 = x_0, \cdots X_t = x_t] \\
&= \int_{\mathcal{U}} \Theta_t(x, u, \Gamma) \kappa_t(\mathfrak{u}_t = u \mid X_t = x_t) \\
&= \int_{\mathcal{U}} \Theta_t(x, u, \Gamma) \kappa_t(\mathfrak{u}_t = u \mid X_t = x_t) \\
&= \int_{\mathcal{U}} \mathbf{Q}^{\kappa}[X_{t+1} \in \Gamma, \mathfrak{u}_t \in du \mid X_t = x_t] = \mathbf{Q}^{\kappa}[X_{t+1} \in \Gamma \mid X_t = x_t]
\end{aligned} \tag{C.13}$$

Note that if $\kappa \in \mathfrak{K}_S$, the transition kernel becomes time-independent, i.e., the controlled process X (or $X^{\mathfrak{u}}$) is a homogeneous Markov process. □

In terms of optimal control, the following rephrased statements allow us to restrict the policies to be deterministic Markov.

Theorem C.3.4. *[66, Theorem 1.2] For any $\kappa \in \mathfrak{K}_A$, there exists a non-randomized control $\bar{\kappa}$ such that*

$$\mathbf{E}^{\mu_0,\bar{\kappa}}\left[\sum_{t=0}^{N-1} c(X_t,\mathfrak{u}_t) + C_N(X_N)\right] \leq \mathbf{E}^{\mu_0,\kappa}\left[\sum_{t=0}^{N-1} c(X_t,\mathfrak{u}_t) + C_N(X_N)\right]$$

Theorem C.3.5. *[181, Theorem 5.12] Let $(X,\mathfrak{u})$ be a process generated by a Markov model with initial distribution μ_0. Consider the minimization of*

$$J := \inf_{\kappa\in\mathfrak{K}_A} \mathbf{E}^{\mu_0,\kappa}\left[\sum_{t=0}^{N-1} c(X_t,\mathfrak{u}_t) + C_N(X_N)\right]$$

for some Borel measurable and bounded cost function c. Then, any $\kappa \in \mathfrak{K}_A$ can be replaced with one deterministic $\kappa^\star \in \mathfrak{K}_M$ which is at least as good as the original policy. In particular, if an optimal control policy exists, there is no loss in restricting policies to be Markov.

The above deterministic (or non-randomized) control policy is guaranteed by [66, Theorem 1.2], which is, 'if $\mathfrak{K}$ is a class of control policies, then for any $\kappa \in \mathfrak{K}$, there exists a non-randomized control $\kappa^\star \in \mathfrak{K}$ such that

$$\mathbf{E}^{\mu_0,\kappa^\star}\left[\sum_{t=0}^{N-1} c(X_t,\mathfrak{u}_t) + C_N(X_N)\right] \leq \mathbf{E}^{\mu_0,\kappa}\left[\sum_{t=0}^{N-1} c(X_t,\mathfrak{u}_t) + C_N(X_N)\right].$$

Appendix D

A Brief Introduction to SPDEs

This appendix provides a brief introduction to Gaussian measure theory, model of space-time Gaussian-type noise, and semilinear parabolic SPDEs driven by space-time Wiener processes.

D.1 Gaussian Measure Theory

Gaussian measures on separable Banach spaces

We first consider Gaussian measure for more general separable Banach spaces. Let E be a separable Banach space w.r.t. norm $\|\cdot\|$.

Gaussian measures on $\mathbb{R}^n$ can be characterised by prescribing that the projections of the measure onto any one-dimensional subspace of $\mathbb{R}^n$ are all Gaussian. This is a property that can readily be generalised to infinite-dimensional spaces [81].

Definition D.1.1 (Gaussian measure on a Banach space)**.** *A probability measure μ on $(E, \mathscr{B}(E))$ is said to be a Gaussian measure if and only if the law of an arbitrary linear functional $h \in E^*$, considered as a random variable on $(E, \mathscr{B}(E), \mu)$, is a Gaussian measure on $(\mathbb{R}, \mathscr{R}(\mathbb{R}))$ [43].*

If the law of each $h \in E^*$ additionally a mean-zero Gaussian on $\mathbb{R}$, then μ is called a symmetric Gaussian measure on $(E, \mathscr{B}(E))$.

Given a symmetric Gaussian measure we define the covariance operator $C_\mu : E^* \times E^* \to \mathbb{R}$ by

$$C_\mu(h, k) = \int_E h(x)k(x)\mu(dx). \tag{D.1}$$

The following theorem identifies the property of Gaussian tails.

Theorem D.1.2 (Fernique). *Let μ be an arbitrary symmetric Gaussian measure on E. Let $r > 0$ and $\alpha > 0$ be such that*

$$\log\left(\frac{1-\mu(\overline{\mathbb{B}}_r)}{\mu(\overline{\mathbb{B}}_r)}\right) + 32\alpha r^2 \leq -1.$$

Then $\int_E e^{\alpha\|x\|^2}\mu(dx) < \infty$.

As an immediate corollary of Fernique's theorem, μ has all moments finite. One can also identify that the linear operator $\hat{C}_\mu : E^* \to E$

$$\hat{C}_\mu h := \int_E x h(x)\mu(dx) \tag{D.2}$$

is one-to-one and continuous by applying Fernique's theorem. As a consequence, for any $h, k \in E^*$, we have $C_\mu(h,k) = k(\hat{C}_\mu h)$.

Reproducing kernel Hilbert spaces

Let μ be a symmetric Gaussian measure on a Banach space E. A linear subspace $H \subseteq E$ with a Hilbert norm $\|\cdot\|_H$ is said to be a RKHS for μ if H is complete, continuously embedded in E and such that for arbitrary $h \in E^*$, we have $h(x) \sim \mathcal{N}(0, \|h\|_H^2)$, where $\|h\|_H = \sup_{\|x\|_H \leq 1} |h(x)|$ [43, Section 2.2.2].

It has been verified in [43, Theorem 2.9] that, every arbitrary symmetric Gaussian measure μ on E admits a unique RKHS $(H, \|\cdot\|_H)$ with scalar product

$$\langle \hat{C}_\mu(h), \hat{C}_\mu(k)\rangle_H = \int_E h(x)k(x)\mu(dx). \tag{D.3}$$

Within H, we also have the following reproducing kernel formula:

$$\int_E \langle h, x\rangle_H \langle g, x\rangle_H \mu(dx) = \langle h, g\rangle_H, \quad h, g \in H. \tag{D.4}$$

Remark D.1.3. *RKHS is isomorphic to the Cameron-Martin space (see [81, Section 3.2] for more details) in a natural way. There are authors who use a slightly different terminology.*

The RKHS characterizes those directions (denoted by $h \in H$) in E in which translations leave the measure μ 'quasi-invariant', i.e., the measure $\mu^h(\Gamma) := \mu(\Gamma - h)$ for $\Gamma \in \mathscr{B}(E)$ is absolutely continuous w.r.t. μ. Moreover,

Proposition D.1.4. *Assume that a Banach space E_1 is continuously and as a Borel set embedded in E. If the measure μ is symmetric and Gaussian on E and E_1, then the RKHS calculated with respect to E or E_1 is the same [43, Proposition 2.10].*

Based on the notion of RKHS, we are able to construct an arbitrary symmetric Gaussian measure μ on E via expansions. Let H_μ be the associated RKHS of μ, let $\{\chi_k\}$ be an orthonormal and complete basis in H_μ, and $\{\xi_n\}$ be a sequence of independent $\mathbb{R}$-valued r.v. such that $\xi_n \sim \mathcal{N}(0,1)$ for all $n \in \mathbb{N}$. Then $\sum_{n=1}^\infty \xi_n \chi_n$ converges a.s. in E and the probability law of $\sum_{n=1}^\infty \xi_n \chi_n$ is μ [43, Theorem 2.12].

Gaussian measures on Hilbert spaces

We are more interested in Gaussian measures on Hilbert spaces. By Definition D.1.1, a probability measure μ on $(H, \mathscr{B}(H))$ is called Gaussian if for arbitrary $h \in H$, there exist $m \in \mathbb{R}$ and $q \geq 0$ such that

$$\mu(\{x \in H : \langle h, x\rangle \in A\}) = \mathcal{N}(m,q)(A), \quad \forall A \in \mathscr{B}(\mathbb{R}). \tag{D.5}$$

It turns out that if μ is Gaussian on H, then there exist an $m \in H$ and a symmetric nonnegative linear operator Q, such that

$$\int_H \langle h, x\rangle \mu(dx) = \langle m, h\rangle, \quad \forall h \in H, \tag{D.6}$$

and

$$\int_H \langle h, x-m\rangle \langle k, x-m\rangle \mu(dx) = \langle Qh, k\rangle, \quad \forall h,k \in H. \tag{D.7}$$

The operator Q is called the covariance operator of μ, which is also verified to be of trace class in the sense of Definition D.1.5. A Gaussian measure μ on H is uniquely determined by the mean m and the covariance Q, which is denoted by $\mathcal{N}(m, Q)$.

Definition D.1.5. *Let $\{\chi_k\}$ be a complete orthonormal basis in H. A linear operator $T : H \to H$ is said to be a trace class operator, denoted by $\mathcal{L}_1(H)$, if*

$$\|T\|_{\mathcal{L}_1} := \operatorname{tr} T = \sum_k \langle T\chi_k, \chi_k\rangle < \infty.$$

Let G be another separable Hilbert space. A linear operator $T : H \to G$ is said to be Hilbert-Schmidt if

$$\|T\|_{\mathcal{L}_2(H;G)} := \left(\sum_k \|T\chi_k\|^2\right)^{1/2} < \infty.$$

D.2 Q-Wiener Processes

In this section, we focus on Hilbert space valued Wiener processes before we get into SPDEs. Suppose we are given a separable Hilbert space H with a complete orthonormal basis $\{\chi_k\}$, and W is a H-valued Wiener process. Then, the probability law of $W(t)$ for each t is a Gaussian measure on H with a trace class covariance Q satisfying

$$\mathbf{E}[\langle W(t), a\rangle\langle W(s), b\rangle] = t\wedge s\mathbf{E}[\langle W(1), a\rangle\langle W(1), b\rangle] = t\wedge s\langle Qa, b\rangle, \quad a, b \in H, \ t, s \geq 0. \quad \text{(D.8)}$$

Conversely, given a trace class nonnegative operator Q (Q has all nonnegative real part of eigenvalues) on H as the spatial covariance, we are able to define an H-valued Q-Wiener process in a similar way as the finite-dimensional cases.

Definition D.2.1 (Q-Wiener process)**.** *An H-valued stochastic process W is called a Q-Wiener process if*

(1) $W(0) = 0$,

(2) W *has continuous trajectories (in* t*)*

(3) W *has independent increments,*

(4) $W(t) - W(s) \sim \mathcal{N}(0, (t-s)Q), \ 0 \leq s < t.$

In view of RKHS, we are able to define W as an expansion

$$W(t) = \sum_k \sqrt{q_k}\beta_k(t)\chi_k,$$

where β_k are real valued Brownian motions mutually independent, q_k are the eigenvalues of Q. The operator Q characterizes the spatial regularity of Q-Wiener processes.

Now suppose that Q is only self-adjoint and positive definite without finite trace, e.g. $Q =$ id, we are not able define spatially Gaussian processes in H as in Definition D.2.1. Instead, we consider generalized Wiener processes in the following sense.

Definition D.2.2 (Generalized Q-Wiener process)**.** *Let H' be a larger Hilbert space containing H as a dense subset and such that the inclusion map $\mathcal{I} : H \to H'$ is Hilbert-Schmidt. Let Q be self-adjoint and positive definite without finite trace. Suppose W is a process with*

$$\mathbf{E}[\langle W(t), a\rangle\langle W(s), b\rangle] = t \wedge s\langle Qa, b\rangle, \quad a, b \in H, \ t, s \geq 0.$$

We then call W a generalized Q-Wiener process on H if it is an H'-valued $\mathcal{I}\mathcal{I}^*Q$-Wiener process such that

$$\mathbf{E}[\langle W(t), a\rangle_{H'}\langle W(s), b\rangle_{H'}] = t \wedge s\langle \mathcal{I}\mathcal{I}^*Qa, b\rangle_{H'}, \quad a, b \in H', \ t, s \geq 0.$$

In particular, if $Q = \mathrm{id}$, we simply call W a cylindrical Wiener process on H.

Proposition D.2.3. *The Gaussian measure μ on H' with covariance $\mathcal{I}\mathcal{I}^*Q$ has H as its RKHS. Furthermore, for every $h, k \in H$, we have $\langle h, k\rangle_\mu = \langle \mathcal{I}h, \mathcal{I}k\rangle = \langle \mathcal{I}\mathcal{I}^*h, k\rangle_{H'}$.*

The above proposition implies that we can still use the same basis as H to construct cylindrical Wiener processes on H, that is $\mathcal{I}\mathcal{I}^*$-Wiener process on H'.

Remark D.2.4. *For conventional Wiener processes on $\mathbb{R}^n$, the Gaussian measure μ on the sample space $E = C([0, T]; \mathbb{R}^n)$ is such that the (temporal) convariance operator is $s \wedge t$. The RKHS of E is the Hilbert space of all absolutely continuous functions h with $h(0) = 0$ and $\int_0^T \dot{h}^2(t)dt < \infty$[1]. Similarly, we can define a the cylindrical Wiener process on H in an alternative way as a space-time white noise, where the Gaussian measure on the spatial projection has the same property as the temporal domain.*

D.3 Semilinear Parabolic SPDEs

Given separable Hilbert spaces H, V. In particular, we denote by $\langle \cdot, \cdot\rangle$ and $\|\cdot\|$ the asscociated inner product and norm, respectively. For simplicity, we assume that $\{\mathfrak{h}_n := e^{in\theta}\}_{n\in\mathbb{Z}_0}$ is the complete orthonormal basis of H and denote by $\{\rho_n\}_{n\in\mathbb{Z}_0}$ the eigenvalues. Semilinear parabolic SPDEs have the form of

$$du(t) = \mathcal{A}u(t)dt + f(u(t))dt + G(u(t))dW(t) \tag{D.9}$$

where $\mathcal{A}$ is a self-adjoint elliptic linear operator, f is a nonlinear function, W is a generalized Q-Wiener process on V, and $G(u) : V \to H$. The solutions should be in H. For the special case when $G(u) = G \in R$, we should require that the space V has the same RKHS as H. The famous stochastic reaction-diffusion equations fall in the category of semilinear parabolic SPDEs.

The solution of (D.9) with $u(0) = u_0$ is given as

$$u(t) = S(t)u_0 + \int_0^t S(t-s)f(u(s))ds + \int_0^t S(t-s)G(u(s))dW_s. \tag{D.10}$$

[1] $\dot{h}(t)$ is the derivative w.r.t. t in a distributional sense.

We have seen the solution subspace $\mathbf{H}^2_{\text{per}}$ (see Remark 1.2.1) of the deterministic Moore-Greitzer model restricted to the $\mathcal{H}$ subspace, which coincides with the domain of its linear operator $\mathcal{A}(\gamma)$. However, with the appearance of noise that is white in time, either white or colored in space, the regularity especially the differentiability of the solution varies.

Now we assume that $\mathrm{dom}(\mathcal{A})$ in (D.9) equals to $\mathbf{H}^2_{\text{per}}$, and define the fractional spaces w.r.t. $\mathrm{dom}(\mathcal{A})$ as well as the Sobolev space in order to render a more flexible scale of regularity.

Definition D.3.1 (Fractional Power Space). *For $\alpha \in \mathbb{R}$, define the interpolation fractional power (Hilbert) space [134] $\mathcal{H}_\alpha := \mathrm{dom}(\mathcal{A}^\alpha)$ endowed with inner product $\langle u, v\rangle_\alpha = \langle \mathcal{A}^\alpha u, \mathcal{A}^\alpha v\rangle$ and corresponding induced norm $\|\cdot\|_\alpha := \|\mathcal{A}^\alpha \cdot\|$. Further more, we denote the dual space of $\mathcal{H}_\alpha$ by $\mathcal{H}_{-\alpha}$ w.r.t. the inner product in $\mathcal{H}$.*

Remark D.3.2. *We list other properties [81] of the fractional power spaces and the semigroup $e^{t\mathcal{A}}$:*

1. *$\mathcal{H}_\alpha \subset \mathcal{H}_\beta$ for $\alpha \geq \beta$.*
2. *For any $\alpha > 0$, we have $\mathcal{H}_\alpha \subset \mathcal{H} \subset \mathcal{H}_{-\alpha}$;*
3. *The quantity $e^{t\mathcal{A}}$ commutes with any power of its generator;*
4. *$\|\mathcal{A}^\alpha e^{t\mathcal{A}}\| \leq \frac{C_\alpha}{t^\alpha} e^{-ct}$ for all $t > 0$ and for some $c_\alpha > 0$. In particular, $\|\mathcal{A}^\alpha e^{t\mathcal{A}}\| \leq \frac{C_\alpha}{t^\alpha}$ when $t \in (0, 1]$.*

The following proposition is straightforward from the above remark.

Proposition D.3.3. *For any $\alpha > \beta$, there exists a constant $C_{\alpha,\beta} > 0$ such that for $t \in (0, 1]$, $\|e^{t\mathcal{A}}x\|_\alpha \leq C_{\alpha,\beta}\|x\|_\beta t^{\beta-\alpha}$. Moreover, if all eigenvalues of $\mathcal{A}$ only have negative real part, for all $t > 0$ and $u \in \mathcal{H}$, there exists a constant $C'_{\alpha,\beta} > 0$ and a $c_s > 0$ such that $\|e^{t\mathcal{A}}u\|_\alpha \leq C'_{\alpha,\beta}\|x\|_\beta t^{\beta-\alpha}e^{-c_s t}$.*

Now we look at the same problem from the Sobolev space point of view. We specify that the spatial domain is $\mathcal{D} = [0, 2\pi]$ as in the Moore-Greitzer model.

Definition D.3.4 (Sobolev Space). *For $r \in \mathbb{N}_0$, denote by $H^r(\mathcal{D})$ as the Hilbert Sobolev space with weak derivative up to r, the norm on $H^r(0, 2\pi)$ is*

$$\|u\|_{H^r} := \left(\sum_{0\leq|k|\leq r} \|D^k u\|_2^2 \right)^{1/2}. \tag{D.11}$$

Definition D.3.5 ((Fractional) Sobolev Space)**.** *Denote $W^{r,p}(\mathcal{D})$ by the (fractional) Sobolev space. We particularly define the following cases:*

(a) for $r \in \mathbb{N}$, $p \in [1, \infty)$

$$\mathcal{W}^{r,p}(\mathcal{D}) = \left\{ u \in L^p(\mathcal{D}) : \|u\|_{\mathcal{W}^{r,p}} := \sum_{0 \leq |k| \leq r} \|\partial^k u\|_p < \infty \right\}; \tag{D.12}$$

(b) for $r \in (0, 1)$, $p \in [1, \infty)$

$$\mathcal{W}^{r,p}(\mathcal{D}) := \left\{ u \in L^p(\mathcal{D}) : \|u\|_{\mathcal{W}^{r,p}} := \left(\iint_{\mathcal{D} \times \mathcal{D}} \frac{|u(x) - u(y)|^p}{|x - y|^{1+rp}} \right)^{1/p} < \infty \right\}; \tag{D.13}$$

(c) for $r = k + q$, $p \in [1, \infty)$ $k \in \mathbb{N}$, $q \in (0, 1)$,

$$\mathcal{W}^{r,p}(\mathcal{D}) := \left\{ u \in L^p(\mathcal{D}) : \|u\|_{\mathcal{W}^{r,p}} := \|u\|_{\mathcal{W}^{k,p}} + \sum_{|\nu|=k} \|\partial^\nu u\|_{\mathcal{W}^{q,p}} < \infty \right\}; \tag{D.14}$$

(d) for $r \in (-\infty, 0)$, $p \in (1, \infty)$, $\mathcal{W}^{r,p}(\mathcal{D}) = (\mathcal{W}^{r,p}(\mathcal{D}))^*$.

If $p = 2$, we use $\mathbf{H}^r(\mathcal{D})$ for short. Correspondingly, we can also define $\mathbf{H}^r_{\text{per}} := \{u \in \mathbf{H}^r : u(0) = u(2\pi), u_\theta(0) = u_\theta(2\pi), \int_0^{2\pi} u(\theta) d\theta = 0\}$ for all $r \in \mathbb{R}$.

It is clear that in the special case where $r = 1$, $\text{dom}(\mathcal{A}^r) = \mathbf{H}^{2r}_{\text{per}}$. In the following lemma, we show that such relation holds for any $r \in \mathbb{N}$. The results can be further extended when $\mathbf{H}^{2r}_{\text{per}}[0, 2\pi]$ is defined for negative and non-integer r.

Lemma D.3.6. *On the spatial domain $\mathcal{D} = [0, 2\pi]$, the Sobolev norm $\|\cdot\|_{\mathbf{H}^r}$ is equivalent as the fractional power norm $\|\cdot\|_{r/2}$ for $r \in \mathbb{N}$.*

Proof. The proof easily follows [116, Proposition 1.93]. Initially, we have

$$\text{dom}(\mathcal{A}) = \mathbf{H}^2_{\text{per}}(\mathcal{D}) \subset U.$$

For $u \in \mathbf{H}^r$, we can write $u = \sum_{n \in \mathbb{Z} \setminus \{0\}} u_n e^{in\theta}$, then the Sobolev norm can be expanded as

$$\|u\|^2_{\mathbf{H}^r} = \sum_{n \in \mathbb{Z} \setminus \{0\}} (1 + n^2 + ... + n^{2r}) |u_n|^2$$

However, as the discrete spectrum of $\mathcal{A}$ are $\rho_n \in \mathbb{C}$ $(\forall n \in \mathbb{Z}_0)$, we can explicitly express $\|u\|^2_{r/2}$ by

$$\|u\|^2_{r/2} = \langle \mathcal{A}^{r/2}u, \mathcal{A}^{r/2}u \rangle = C \sum (\rho_n \rho_{-n})^{r/2} |u_n|^2,$$

where C is a normalizer. By the definition of ρ_n, we can find $C_1, C_2 > 0$ such that $C_1(1+n^2)^r \leq C(\rho_n \rho_{-n})^{r/2} \leq C_2(1+n^2)^r$. We also have

$$\frac{1}{2^r}(1+n^2)^r \leq (1 + n^2 + ... + n^{2r}) \leq (1+n^2)^r$$

Combine the above two sets of inequalities,

$$\frac{C}{C_2 2^r}(\rho_n \rho_{-n})^{r/2} \leq (1 + n^2 + ... + n^{2r}) \leq \frac{C}{C_1}(\rho_n \rho_{-n})^{r/2}$$

Then, by the definition of the two norms, it is not hard to see that $\|u\|^2_{r/2} \sim \|u\|^2_{\mathbf{H}^r}$. □

For more regularity results, we kindly refer readers to [43].

Appendix E

Metric and Topological Spaces of Probability Measures

We briefly review some of the useful facts concerning metric spaces of probability measures. Note that, 'if a space is metrisable, the topology is determined by convergences of sequences, which explains we sometimes only define the concept of convergence, without explicitly mention the topology.'[82] We introduce two concepts of convergence of sequence of probability measures as well as probability metrics.

Consider any separable and complete state space (Polish space) $\mathcal{X}$, we denote by $\mathfrak{P}(\mathcal{X})$ by the space of probability measures on $(\mathcal{X}, \mathscr{B}(\mathcal{X}))$.

Convergence concepts

Definition E.0.1 (Setwise (strong) convergence). *A sequence $\{\mu_n\}_{n=0}^{\infty} \subseteq \mathfrak{P}(\mathcal{X})$ is said to converge setwisely to a probability measure μ, denoted by $\mu_n \to \mu$, if*

$$\int_{\mathcal{X}} h(x)\mu_n(dx) \to \int_{\mathcal{X}} h(x)\mu(dx) \tag{E.1}$$

for all measurable and bounded test functions h.

The notion of setwise convergence is a stringent notion for convergence.

Example E.0.2. *It is interesting to note that $x_n \to x$ in $\mathcal{X}$ does not imply the strong convergence of the associated Dirac measures. A classical counterexample is to let $x_n = 1/n$ and $x = 0$, and we do not have $\lim_{n\to\infty} \delta_{1/n} = \delta_0$ in the strong sense since, i.e., $0 = \lim_{n\to\infty} \delta_{1/n}(\{0\}) \neq \delta_0(\{0\}) = 1$.* □

Definition E.0.3 (Weak convergence)**.** *A sequence $\{\mu_n\}_{n=0}^{\infty} \subseteq \mathfrak{P}(\mathcal{X})$ is said to converge weakly to a probability measure μ, denoted by $\mu_n \rightharpoonup \mu$, if*

$$\int_{\mathcal{X}} h(x)\mu_n(dx) \to \int_{\mathcal{X}} h(x)\mu(dx), \quad \forall h \in C_b(\mathcal{X}). \tag{E.2}$$

We frequently use the following alternative condition [43, Proposition 2.2]:

$$\mu_n(A) \to \mu(A), \quad \forall A \in \mathscr{B}(\mathcal{X}) \text{ s.t. } \mu(\partial A) = 0. \tag{E.3}$$

Remark E.0.4. *Weak convergence describes the weak topology[1]. The meaning of the weak topology is to extend the normal convergence in deterministic settings. Note that, in Example E.0.2, $x_n \to x$ in $\mathcal{X}$ is equivalent to the weak convergence of Dirac measures $\delta_{x_n} \rightharpoonup \delta_x$.*

To describe the convergence (in probability law) of more general random variables $\{X_n\}$ in $\mathcal{X}$, it is equivalent to investigate the weak convergence of their associated measures $\{\mu_n\}$. It is also straightforward from Definition E.0.3 that weak convergence also describes the convergence of probabilistic properties related to $\{\mu_n\}$.

Weak compactness

We aim to give a compactness theorem that provides us with a very useful criteria to verify whether a given sequence of probability measures has a weak convergent subsequence. We first introduce the notion of tightness.

Definition E.0.5 (Tightness of set of measures)**.** *Let $\mathcal{X}$ be any topological state space and $M \subseteq \mathfrak{P}(\mathcal{X})$ be a set of probability measures on $\mathcal{X}$. We say that M is tight if, for every $\epsilon > 0$ there exists a compact set $K \subset \mathcal{X}$ such that $\mu(K) \geq 1 - \epsilon$ for every $\mu \in M$.*

Once can show that, for each sequence $\{\mu_n\}$ of tight Λ, there exists a $\mu \in \bar{\Lambda}$ and a subsequence $\{\mu_{n_k}\}$ such that $\mu_{n_k} \rightharpoonup \mu$.

[1] The weak topology in this case is actually the weak* topology. However, as the term 'weak' is commonly accepted under the clear context of spaces of probability measures, we use weak topology for simplicity unless otherwise noted.

Theorem E.0.6 (Prokhorov). *Let $\mathcal{X}$ be a complete separable metric space. A family $\Lambda \subseteq \mathfrak{P}(\mathcal{X})$ is relatively compact if an only if it is tight.*

Remark E.0.7. *Prokhorov's theorem provides an alternative criterion for verifying the compactness of family of measures w.r.t. the corresponding metric space using tightness. In addition, on a compact metric space $\mathcal{X}$, every family of probability measures is automatically tight.*

Probability metrics

The space of probability measures on a complete, separable, metric (metrisable) space endowed with the topology of weak convergence is itself a complete, separable, metric (metrisable) space [24]. While not easy to compute, the Prohorov metric can be used to metrize weak topology. We introduce two other frequently used metrics that can implies weak convergence.

$$\begin{aligned}
&\text{Total Variation} \implies \textbf{setwise convergence}\\
&\Downarrow\\
&\text{Wasserstein}(\dim(\mathcal{X}) < \infty)\\
&\Downarrow\\
&\text{Prokhorov} \iff \textbf{weak topology}
\end{aligned} \tag{E.4}$$

Definition E.0.8 (Total variation distance). *Given two probability measures μ and ν on $\mathscr{B}(\mathcal{X})$, the total variation distance is defined as*

$$\|\mu - \nu\|_{\mathrm{TV}} = 2 \sup_{\Gamma \in \mathscr{B}(\mathcal{X})} |\mu(\Gamma) - \nu(\Gamma)|. \tag{E.5}$$

In particular, if $\mathcal{X}$ is a discrete space,

$$\|\mu - \nu\|^d_{\mathrm{TV}} = \|\mu - \nu\|_1 = \sum_{q \in \mathcal{X}} |\mu(q) - \nu(q)|. \tag{E.6}$$

Remark E.0.9. *It is equivalent to use the dual representation*

$$\|\mu - \nu\|_{\mathrm{TV}} = \sup_{\|h\|_\infty \le 1} \left| \int_{\mathcal{X}} h(x)\mu(dx) - \int_{\mathcal{X}} h(x)\nu(dx) \right|. \tag{E.7}$$

Definition E.0.10 (Wasserstein distance). *Let $\mu, \nu \in \mathfrak{P}(\mathcal{X})$ for $(\mathcal{X}, |\cdot|)$, the Wasserstein distance[2] is defined by $\|\mu - \nu\|_{\mathrm{W}} = \inf \mathbf{E}|X - Y|$, where the infimum is is taken over all joint distributions of the random variables X and Y with marginals μ and ν respectively.*

[2] This is formally termed as 1^{st}-Wasserstein metric. We choose 1^{st}-Wasserstein metric due to the convexity and nice property of test functions.

We frequently use the following duality form of definition[3],

$$\|\mu-\nu\|_{\mathrm{W}} := \sup\left\{\left|\int_{\mathcal{X}} h(x)d\mu(x) - \int_{\mathcal{X}} h(x)d\nu(x)\right|,\ h\in C(\mathcal{X}), \mathrm{Lip}(h)\le 1\right\}.$$

The discrete case, $\|\cdot\|_{\mathrm{W}}^d$, *is nothing but to change the integral to summation. Let* $\mathcal{B}_W = \{\mu \in \mathfrak{P}(\mathcal{X}) : \|\mu-\delta_0\|_{\mathrm{W}} < 1\}$. *Given a set* $\mathfrak{G} \subseteq \mathfrak{P}(\mathcal{X})$, *we denote* $\|\mu\|_{\mathfrak{G}} = \inf_{\nu\in\mathfrak{G}} \|\mu-\nu\|_{\mathrm{W}}$ *by the distance from* μ *to* $\mathfrak{G}$, *and* $\mathfrak{G} + r\mathcal{B}_W := \{\mu :\ \|\mu\|_{\mathfrak{G}} < r\}$[4] *by the* r*-neighborhood of* $\mathfrak{G}$.

Remark E.0.11. *Note that* $\mathcal{B}_W$ *is dual to* $\mathcal{B}$. *For any* $\mu\in\mathcal{B}_W$, *the associated random variable* X *should satisfy* $\mathbf{E}|X|\le 1$, *and vice versa.*

The following well-known result estimates the Wasserstein distance between two Gaussians. Note that the R.H.S. of (E.8) is the 2[nd]-Wasserstein metric, which, intuitively, captures the second moment deviation.

Proposition E.0.12. *Let* $\mu\sim\mathcal{N}(m_1,\Sigma_1)$ *and* $\nu\sim\mathcal{N}(m_2,\Sigma_2)$ *be two Gaussian measures on* $\mathbb{R}^n$. *Then*

$$|m_1-m_2| \le \|\mu-\nu\|_{\mathrm{W}} \le \left(\|m_1-m_2\|_2^2 + \|\Sigma_1^{1/2}-\Sigma_2^{1/2}\|_F^2\right)^{1/2}, \tag{E.8}$$

where $\|\cdot\|_F$ *is the Frobenius norm.*

On finite state spaces, total variation and Wasserstein distances manifest equivalence [65, Theorem 4]. We only show the following side of inequality in favor of our needs.

Proposition E.0.13. *For any* μ,ν *on some discrete and finite space* Q, *we have*

$$\|\mu-\nu\|_{\mathrm{W}}^d \le \mathrm{diam}(Q)\cdot\|\mu-\nu\|_{\mathrm{TV}}^d. \tag{E.9}$$

Definition E.0.14. *(Bounded-Lipschitz metric) Let* $\mu,\nu\in\mathfrak{P}(\mathcal{X})$ *for* $(\mathcal{X},\|\cdot\|)$, *the bounded-Lipschitz metric is defined by*

$$\|\mu-\nu\|_{\mathrm{L}} := \sup\left\{\left|\int_{\mathcal{X}} h(x)d\mu(x) - \int_{\mathcal{X}} h(x)d\nu(x)\right|,\ h\in C(\mathcal{X}), \|h\|_{\mathrm{BL}}\le 1\right\},$$

where

$$\|h\|_{\mathrm{BL}} := \sup_{x,y\in\mathcal{X}}\left\{|h(x)|,\ \frac{|h(x)-h(y)|}{\|x-y\|}\right\}.$$

[3] $\mathrm{Lip}(h)$ is the Lipschitz constant of h such that $|h(x_2)-h(x_1)|\le \mathrm{Lip}(h)|x_2-x_1|$.

[4] This is valid by definition.