Oliver Schön

Correct-by-Design Control of Cyber-Physical Systems

Robustness and Learning under Uncertainty

Thesis for the degree of Doctor of Philosophy

Newcastle upon Tyne, July 2025



Abstract

Artificial intelligence (AI) has paved the way for transformative new applications — from autonomous vehicles and robotic-assisted surgery to smart energy solutions; yet, entrusting these systems to operate without human oversight raises urgent questions about ensuring their safe and reliable behaviour. Developing control methods that formally guarantee intended system functionality rigorously by design remains a fundamental challenge, especially in the presence of randomness (aleatoric uncertainty) and incomplete model knowledge (epistemic uncertainty).

A key difficulty with formal approaches for designing controllers arises from the fact that they typically establish guarantees based on an assumed model that needs to accurately reflect the real system. This thesis addresses this core issue by developing a continuum of formal control synthesis and verification frameworks for stochastic systems under uncertainty — evolving from classical, robustness-focused techniques to highly expressive, learning-based methods that retain meaningful guarantees even under increasing uncertainty.

The first part explores robustness under model parametric uncertainty by developing a new type of probabilistic simulation relation that links a nominal model to a target system with uncertain parameters. Combined with Bayesian inference to extract credible parameter sets from measurement data, the resulting data-driven approach enables correct-by-design control synthesis for nonlinear systems capable of satisfying complex temporal logic specifications with quantified probabilistic guarantees. The approach is further extended to networked systems with inherent structure arising from interacting components.

In a narrative transition toward learning under increasing levels of uncertainty, the subsequent chapters develop structured kernel methods for data-driven verification. Gaussian processes based on binary-tree kernels are used to learn formal abstractions directly from data without requiring an explicit system model and dramatically reducing the associated computational cost compared to existing approaches. Finally, the thesis culminates in the verification of fully black-box systems with AI-based controllers in the loop, leveraging spectral control barrier certificates derived via conditional mean embeddings and Fourier series expansions.

Collectively, these contributions trace a clear trajectory from robustness to learning under uncertainty, demonstrating that formal guarantees and practical applicability can advance hand in hand to realise trustworthy autonomy.

Zusammenfassung

Künstliche Intelligenz (KI) ermöglicht neuartige Anwendungen mit weitreichendem Potenzial, von autonomen Fahrzeugen und robotergestützter Chirurgie bis hin zu intelligenten Energielösungen wie selbstorganisierenden Stromnetzen. Doch der Betrieb solcher Systeme ohne menschliche Aufsicht wirft grundlegende Fragen über ihre Sicherheit und Verlässlichkeit auf. Die Entwicklung von Regelungsverfahren, die die gewünschte Systemfunktionalität *per Design* zuverlässig gewährleistet, stellt dabei eine zentrale Herausforderung dar, insbesondere angesichts intrinsischer Zufallseinflüsse (aleatorische Unsicherheit) und unvollständigen Modellwissens (epistemische Unsicherheit).

Eine besondere Schwierigkeit formaler Regelungsansätze besteht darin, dass deren Garantien nur dann gelten, wenn das zugrunde gelegte Modell die Realität hinreichend genau abbildet. Diese Dissertation addressiert dieses Grundproblem durch die Entwicklung eines Spektrums formaler Synthese- und Verifikationsmethoden für stochastische Systeme unter Unsicherheit, ausgehend von klassischen, robustheitsorientierten Verfahren bis hin zu hochflexiblen, lernbasierten Ansätzen, die auch bei wachsender Unsicherheit belastbare Garantien ermöglichen.

Im ersten Teil wird eine probabilistische Simulationsrelation eingeführt, die ein nominales Modell mit einem unsicher parametrisierten Zielsystem verknüpft. In Kombination mit Bayesscher Inferenz auf Basis von Messdaten ermöglicht diese datengetriebene Methode die korrektheitsgarantierende Synthese nichtlinearer Regelungen unter temporalen Logikspezifikationen mit quantifizierten Wahrscheinlichkeitsgarantien. Die Methodik wird auf vernetzte Systeme mit inhärenter Struktur, die durch interagierende Komponenten entsteht, erweitert.

Im weiteren Verlauf werden strukturierte Kernel-Methoden zur datengestützten Verifikation eingesetzt. Mithilfe Binärbaum-Gaußprozesse lassen sich formale Abstraktionen effizient direkt aus Daten lernen, ohne explizites Systemmodell. Abschließend wird die Verifikation vollständig unbekannter Systeme mit KI-basierten Regelungen durch spektrale Control Barrier Certificates ermöglicht, die auf Conditional Mean Embeddings und Fourierreihen basieren.

Die Beiträge dieser Arbeit zeichnen eine klare Entwicklung nach, von modellbasierter Robustheit bis hin zu lernbasierter Flexibilität, und zeigen, dass formale Garantien und praktische Anwendbarkeit Hand in Hand gehen können, um vertrauenswürdige autonome Systeme zu realisieren.

To the crazy ones. The dreamers.
To those who see things differently
and change the world.

And to my parents,
for their everlasting love and support.

Acknowledgements

On paper, this thesis marks the culmination of three and a half years of research that shaped me into an independent scientist. Behind the scenes, it marks the end of an incredibly challenging and rewarding journey — one that shaped me into the person I am today. A period of growth, resilience, and discovery that extends far beyond the academic achievements. Moving to a new country and building a new place to call home, while leaving behind loved ones I appreciate endlessly; venturing into unfamiliar fields and trying to run as fast as possible; often pushing myself to the limit — and, at times, beyond. It has been a journey of sweat and tears, but also joy. A journey filled with unexpected encounters, unforgettable places, and experiences I will always carry with me. Submitting this thesis feels like the end of an era — like closing a chapter that, looking back, leaves me proud of what I have achieved and profoundly grateful that I was allowed to experience it. I want to take this moment to thank the people who have walked alongside me, supported me, and made this journey everything it turned out to be.

First and foremost, I want to thank my supervisor, Sadegh Soudjani. Throughout my PhD, you have been both a calm anchor and a steady compass in a sea full of uncertainty, with a remarkable talent for breaking down even the most complicated things into something understandable. I am grateful that you believed in me and my ideas early on — that you entrusted me with the freedom to shape my own research, manage my own time, and that you enabled me to present my work in so many places around the world. Even with a packed plate, you always made time when I needed it and encouraged me to grow — both as an independent researcher and in becoming a better leader. I especially admire your calm and easy-going nature, your kindness, and your unwavering dedication to good science — all balanced with a healthy taste for adrenaline when the deadline gets precariously tight.

Next, I would like to thank Sofie Haesaert. You co-supervised me from the very beginning of my PhD, and from the outset, working with you and your PhD student Birgit was both intense and incredibly formative. You have a brilliant mind and a fascinating talent for always thinking two steps ahead — no matter how much I prepared — which made it a real challenge to keep up initially, but ultimately pushed me to learn faster than I ever thought I could. I am really grateful that I got to work

with you and Birgit for such a long stretch of my PhD. I especially appreciated the one-on-one sessions, where we dove into the depths of highly specific mathematics or code — I always found them extremely helpful and efficient.

I would also like to thank a few people I have had the pleasure to collaborate with over the years. Birgit, my PhD would not have been the same without you. You were there from the very beginning — offering advice and being someone I could look up to. You were always available for a chat, to give support, or to patiently answer all of my (more or less intelligent) questions. Your kind and laid-back nature, paired with your dry humour, made working with you truly a pleasure. Zhengang, I am so grateful we met at ACC. We instantly clicked, and meeting you gave me someone to dive deep into kernel methods with — I learned so much from you, and we had such a fun time together. I like to think I did not just find a collaborator in you, but a friend. Our visit to Sea Life in San Diego is a day I will never forget.

Next, I would like to thank everyone in the HyCoDeV group, as well as the wider AMBER research group at Newcastle. In particular, I want to thank Ben, Milad, and Ernesto, with whom I shared especially close bonds throughout my time here. Ben and Milad, you were part of the reason I decided to move to Newcastle and pursue this PhD in the first place — and I am so glad I did. You made arriving and settling in so much easier, and for that I will always be grateful. Milad, thank you for being a great office mate and friend. You are such a kind person. You always answered my questions with 150% confidence, and you were incredibly generous with your time. I especially enjoyed our hiking trips with the mountaineering society. Ben, I do not think you know how much your friendship has meant to me. I will always be in your debt for welcoming me with open arms, introducing me to your friends, and just being someone I could always share a coffee with when times were tough or celebrate the wins. I like to think we walked much of this journey side by side, and it has been a privilege to see you grow into such a talented researcher. Ernesto, our connection came about a bit unexpectedly, but I can not even count how many hours we spent together in office 6.030 — probably about a million Dynabook charges? You are such a kind, fun person to be around, and a truly gifted programmer. I learned a lot from you and can not thank you enough for all the coding help when I was once again stuck on some debugging nightmare. I also really enjoyed our long days in Newcastle and our trips to Durham and Birmingham together.

I would also like to thank other colleagues from both HyCoDeV and AMBER who contributed to my time in Newcastle. Alex, thank you for being such a great listener and friend — you always had an open ear when times were tough, and our chats and breakfasts gave me both focus and something to look forward to. Marco, thank you for the countless squash matches and whiteboard sessions. Behrad, Mahdieh, Ayse, Yasemin, Omid, Negar, and Andreas — thank you for being fantastic office mates and for the many shared moments: trips into town, movie nights, cooking together, visits to Durham, Birmingham, and the beach. To everyone else who passed through office 6.030 or the sixth floor, or whom I worked with more broadly in Newcastle — thank you for being part of this journey. I especially enjoyed all the interesting lunch debates and hallway chats with John, Zhi, Ivan, Dasha, Anna, Zhichao, Ziqi, Abdel, Adnane, Andrea, Odette, and Dovydas. I am also grateful for the support and advice

I received from staff and faculty, including Ani, Cliff Jones, Steve Riddle, John Fitzgerald, Ken Pierce, Ben Morris, Jaume Bacardit, and Leo Freitas — especially those that joined the annual Computing Lab trips. And a big thanks to everyone I demonstrated with, and to the admin and support staff for fostering such a welcoming working environment. Finally, thanks to those I met at conferences, especially my colleagues from Eindhoven University of Technology — Chris, Max, Peter, and Yorick. Those are memories I will never forget. It has truly been a pleasure working and growing alongside you all.

I would also like to recognise my friends who supported me throughout this journey, and all those I met during my time in Newcastle. Emma, Josi, Vanessa — I am deeply grateful for meeting you — for making my arrival so easy and for making me fall in love with Newcastle. The memories we made will stay with me forever. I also especially want to thank Antonin, Kary, Josh, and all the close friendships I formed during my life in Newcastle — including those I met during my time in the Erasmus Student Network. You were cornerstones of my new life, and you will always hold a special place in my heart. I will always look back and be grateful for the time I have been given.

At last, I want to thank my parents, my sister, and all my dear friends back in Germany. Thank you for sticking with me throughout this journey — for being my backbone, for believing in me, and for always being there to listen. My PhD has been a rollercoaster of emotions, and you have been my anchor through it all.

Oliver Schön

Newcastle upon Tyne, July 2025

Contents

1 Introduction 1
- 1.1 Background and Motivation 1
  - 1.1.1 Autonomy in High-Risk Applications 3
  - 1.1.2 Formal Methods for Cyber-Physical Systems 7
- 1.2 Central Research Question and Thesis Scope 14
- 1.3 Claims and Key Takeaways 15
- 1.4 Overview of the Thesis 17

PART I ROBUSTNESS UNDER UNCERTAINTY

2 Bayesian Correct-by-Design Synthesis of Uncertain Systems 23
- 2.1 Introduction 23
- 2.2 Preliminaries and Problem Statement 26
  - 2.2.1 Probability Theory 26
  - 2.2.2 Stochastic Control System 27
  - 2.2.3 Temporal Logic Specifications 29
  - 2.2.4 Problem Statement 30
- 2.3 Data-Driven System Identification via Bayesian Linear Regression . 32
  - 2.3.1 Parameter Estimate 33
  - 2.3.2 Credible Set 34
- 2.4 Control Refinement via Sub-Simulation Relations 34
  - 2.4.1 Control Refinement 36
  - 2.4.2 Valid Control Refinement and Sub-Simulation Relations . . . 38
  - 2.4.3 Temporal Logic Control via Sub-Simulation Relations 42
- 2.5 Data-Driven Simulation Relations 44
- 2.6 Case Studies 49
  - 2.6.1 Linear System with Complex Specification 49
  - 2.6.2 7-Dimensional Building Automation System 53
  - 2.6.3 Van der Pol Oscillator 56
- 2.7 Conclusion 56

2.7.1 Temporal Logic Extensions 57
2.7.2 System Extensions 59

3 Correct-by-Design Synthesis for Networks of Uncertain Systems 61
3.1 Introduction 61
3.2 Preliminaries and Problem Statement 63
3.2.1 Networks of Uncertain Stochastic Systems 63
3.2.2 Gaussian Mixture Models 64
3.2.3 Local Controllers 66
3.2.4 Local Specifications 66
3.2.5 Problem Statement 66
3.3 Sub-Simulation Relations for Networks of Uncertain Systems 68
3.3.1 Models and Interconnections 69
3.3.2 Relation on Subsystem Level 69
3.3.3 Relation on Network Level 70
3.3.4 Global Guarantees 72
3.4 Sub-Simulation Relations for Gaussian Mixture Models 74
3.5 Case Studies 75
3.5.1 Package Delivery 75
3.5.2 Car Platoon 76
3.6 Conclusion 79

PART II LEARNING UNDER UNCERTAINTY

4 Learning Formal Abstractions via Binary-Tree Gaussian Processes . . 83
4.1 Introduction 83
4.2 Related Work 85
4.3 Preliminaries and Problem Statement 86
4.3.1 Reproducing Kernel Hilbert Spaces 86
4.3.2 Problem Statement 88
4.4 System Identification via Binary-Tree Gaussian Processes 90
4.4.1 Gaussian Process Regression 90
4.4.2 Binary-Tree Gaussian Processes 92
4.4.3 Statistical Error Bound 95
4.5 Verification Approach 97
4.5.1 BTGP-Based IMC Construction 98
4.5.2 Verification via Interval Iteration Algorithm 99
4.6 Case Study 102
4.7 Conclusion 105

5 Kernel-Based Learning of Black-Box Safety Certificates 107
5.1 Introduction 107
5.2 Related Work 109
5.3 Preliminaries and Problem Statement 110
5.3.1 System and Specification 111

5.3.2 Control Barrier Certificates . . . 111
5.3.3 Problem Statement . . . 113
5.4 Kernel Mean Embeddings . . . 114
5.4.1 Embedding Probability Measures . . . 115
5.4.2 Embedding Probability Kernels . . . 115
5.5 Data-Driven Barrier Formulation . . . 117
5.6 Extension to General Classes of Temporal Logic Specifications . . . 121
5.6.1 Linear Temporal Logic over Finite Traces . . . 121
5.6.2 Specifications Modelled via Co-Büchi Automata . . . 122
5.6.3 ω-Regular Specifications . . . 122
5.7 Problem Characterisation and Comparison of Possible Solutions . . . 124
5.7.1 Monolithic Solutions . . . 125
5.7.2 Divide-and-Conquer Solutions . . . 127
5.8 Fourier Barrier Solution . . . 130
5.8.1 Finite Fourier Feature Expansion . . . 131
5.8.2 Finite-Constraint Relaxation . . . 134
5.8.3 Linear Program . . . 135
5.9 Case Studies . . . 136
5.9.1 Complex Safety Specification . . . 136
5.9.2 Overtaking Scenario Featuring a Neural Network Controller 138
5.10 Conclusion . . . 141

6 Conclusion . . . 143
6.1 Summary . . . 143
6.2 Tool Development . . . 144

7 Future Directions . . . 147
7.1 Methodological Extensions . . . 147
7.2 Abstraction and Sparsity . . . 149
7.3 Expressiveness, Integration, and Deployment . . . 149

References . . . 153

A Additional Proofs of Statements . . . 171
A.1 Proof of Theorem 2.3 . . . 171
A.2 Proof of Theorem 3.1 . . . 174
A.3 Proof of Theorem 3.4 . . . 176
A.4 Proof of Proposition 4.1 . . . 177
A.5 Proof of Proposition 4.2 . . . 178
A.6 Proof of Theorem 4.1 . . . 179

B SILP to LP Relaxation . . . 181

List of Symbols . . . 183

List of Abbreviations . . . 185

Curriculum Vitae 189

List of Publications 191

Chapter 1
Introduction

This thesis deals with the challenges of designing trustworthy artificial intelligence (AI) that will one day drive our cars, fly our aircraft, perform robotic surgeries, and orchestrate our entire energy infrastructure. Such high-risk applications demand accountable forms of AI, which can learn and act robustly, even in the face of uncertainty. The motivation is clear — to make technology safer, smarter, and more cost-effective — and ultimately, to overcome existing technological limitations to enable entirely new applications previously beyond reach.

1.1 Background and Motivation

Voice assistants, facial recognition, seamless language translation, personalised content recommendations, and increasingly omniscient chatbots: undoubtedly, AI has already proven transformative, powering technologies that we can no longer imagine living without. At the same time, the media continues to fuel our imagination with the promise of emerging innovations, reminding us that AI wields the power to advance technology far beyond this. Echoing this promise, cyber-physical systems (CPSs) — systems in which software and physical components such as sensors and actuators are tightly interwoven — are increasingly evolving into forms of so-called *embodied AI*, referring to agentic forms of AI operating within the physical world. For CPSs, the impact of AI agency is particularly profound and poised to reshape entire industries. Examples include autonomous vehicles, robotic systems, implantable medical devices, and intelligent infrastructure. Autonomous vehicles (AVs), for instance, have recently been reported to have surpassed human-level safety whilst driving under controlled conditions — an achievement supported largely by extensive and costly testing (Kalra and Paddock, 2016). Yet, achieving the level of safety required to earn societal trust in AI-driven systems remains elusive. In fact, in a world governed by non-determinism, it is fundamentally infeasible to exhaustively test systems against every conceivable edge case. Crucially, it cannot be the goal to

Embodied AI

eliminate hazardous behaviour entirely, but to reduce the risk to as low as reasonably practicable.

This limitation is not merely theoretical. Unlike conventional applications of AI, premature deployment in safety-critical applications has already resulted in severe accidents (Axelrod, 2013; Amodei et al., 2016). For AVs alone, there have been numerous documented failures[1] leading to serious injuries and loss of life (NHTSA, 2025). Notably, the reasons for such incidents often remain opaque, owing to the complexity and black-box nature of AI components. With a rapid adoption of embodied AI across various high-risk domains such as robotics and advanced medical technologies, there is an unprecedented demand for tools that can provide guarantees for their safe operation. This demand is further reinforced by the dawn of new regulations targeting so-called *high-risk AI*, signalling that the evolving legal landscape will shape the path of automation and AI. This marks a pivotal shift in the global governance of intelligent systems toward enforcing a certifiable demonstration of robust reliability under increasingly stringent requirements (Veale and Zuiderveen Borgesius, 2021). Examples include China's Guidelines for Next-Generation AI (January 2022), the US Algorithmic Accountability Act (April 2022), the EU AI Act (April 2023), and the Colorado Artificial Intelligence Act (May 2024) — among others. Although intended as safeguards, these regulatory efforts have recently become a major roadblock for innovation, demanding evidence of reliability and compliance in the absence of mature tools to assert these requirements. As a result, many companies have exited the market, leaving a market increasingly concentrated around a handful of large, AI-focused corporations.

In response to these developments, a new technical imperative has emerged: to design new frameworks that provide rigorous safety guarantees, even in the face of uncertainty (Bengio, 2025; Bengio et al., 2024; Dalrymple et al., 2024; Annaswamy et al., 2024). Central to this endeavour is the ability to enforce complex logical properties — such as safety, reachability, and invariance — on intelligent systems. This thesis contributes to this broader agenda by developing algorithmic tools that blend data-driven learning with formal verification and control synthesis to provide robust probabilistic guarantees of the *correct* system functionality *by design*. As such, this work lies at the intersection of AI, control theory, and computer science, drawing particularly on formal methods and statistical learning.

The following two subsections provide the technical and conceptual background that motivates this thesis. Subsection 1.1.1 explores the role of autonomy in high-risk domains, outlining concrete application areas and their specific challenges. Subsection 1.1.2 then turns to the formal methods that underpin the systematic design and verification of such systems, highlighting their current limitations and motivating the need for uncertainty-aware, data-driven approaches.

[1] https://www.avcrashes.net/

1.1.1 Autonomy in High-Risk Applications

A scene from the TV show "Friends" (Season 3, Episode 19, aired in 1997) captures a prescient outlook:

> *Pete:* [...] within a few years, that voice recognition is gonna be pretty much standard on any computer you buy. Y'know, so you could be like-like, "Wash my car." "Clean my room." It's not gonna be able to do any of those things, but it'll understand what you're saying.

More than 25 years later, this quip remains surprisingly accurate: many AI breakthroughs have been concentrated around systems that understand, but rarely act — remaining disembodied and advisory. Overcoming this lack of *agency* in the physical world calls for systems that are not only intelligent but autonomous. Autonomy refers to the ability of a system to execute tasks without human oversight (Provan, 2024). This thesis focuses in particular on *dynamical* systems, whose behaviour unfolds over time, driven by the evolution of an underlying state[2]. Unlike conventional control applications, the growing complexity of the dynamics and task specifications of modern systems increasingly necessitates the integration of AI-based components. Additionally, they usually operate in tightly coupled feedback loops with their environment and its inhabitants. For example, in practise, autonomous systems often operate in collaboration with humans at different levels of autonomy. Equipped with only incomplete or imprecise model knowledge, they must repeatedly make safety-critical decisions under uncertain dynamics and in the face of inherently unpredictable human behaviour. To navigate these challenges, autonomous systems are typically structured around a set of core components. Various sensors provide high-dimensional, multimodal data streams. AI-driven perception and localisation extract and update abstract, uncertainty-aware representations of the world they inhabit. These inform the planning and control unit, which acts according to specified objectives and constraints. The resulting actions are executed via actuators, enabling the system to interact with the real world (Luckcuck, 2023; Wongpiromsarn et al., 2023). This thesis focuses on the part of constructing abstract system representations and subsequent control, especially, developing formal techniques that perform these tasks robustly and with quantified guarantees of producing the desired behaviour in the actual real-world system.

AI Agency

Dynamical Systems

Let us now consider some examples of autonomous systems, loosely categorised by application domain.

Autonomous Vehicles Whilst "autonomous vehicle" (AV) is used as an umbrella term for systems deployed across various sectors — including automotive (Provan, 2024), aviation (Emanuilov, 2017), maritime (Ramos et al., 2020), and agriculture (Eaton et al., 2010) — it is most commonly associated with self-driving cars. Technologically, these systems are particularly mature and represent the first use case

[2] The *state* of a system encompasses all quantities — such as position, velocity, and internal variables — needed to uniquely describe its current configuration. Dynamical systems theory formalises this behaviour; control theory provides tools for analysis and synthesis (Slotine and Li, 1991; Åström and Murray, 2021).

of embodied AI (Pfeifer and Iida, 2004). The primary motivation of automation in this area is simple: 94% of serious crashes are due to human error (NHTSA, 2022). The development of AVs — concerning both hardware and software — centres largely around the challenges of two key subsystems: perception & localisation, and planning & control (Wongpiromsarn et al., 2023). Whilst navigating through urban traffic, an AV must process high-dimensional sensor data (e.g., LiDAR, radar, cameras) in real-time, comprehend the environment and its agents, and act in accordance with a comprehensive set of rules (Esterle et al., 2020; Rizaldi et al., 2017; Mehdipour et al., 2023). However, complying with rules alone is not sufficient: rare but critical edge cases — such as sudden pedestrian crossings or erratic driver behaviour — pose particular challenges and demand robust behaviour under uncertainty.

As perhaps the most prominent commercial use case, robotaxis deployed by companies such as Waymo and Baidu operate at level 4 autonomy (high driving automation) (ISO 21448, 2022, (3.15)), functioning without human input in constrained environments. They are remotely monitored, with to date no commercially available robotaxi having reached level 5 autonomy (full automation under all conditions). Apart from their technological maturity, AVs are also notable for the relatively advanced regulatory landscape they inhabit. Central among relevant standards is ISO 21448 (SOTIF), which defines safety as the absence of unreasonable risk due to hazards arising from functional insufficiencies or their implementations — even in the absence of technical faults (ISO 21448, 2022, (3.25)). This highlights the core challenge in the development of systems that *do not behave as intended, despite functioning as designed*. Notably, the notion of *intended behaviour* (ISO 21448, 2022, (3.13)) makes clear that the requirements imposed on autonomous systems often go far beyond simple safety constraints, such as maintaining a safe distance or avoiding collisions. Instead, many specifications describe rich, temporally extended patterns of behaviour: sequences of actions, context-sensitive responses, and interactions that must unfold in a certain way. These are often abstract, high-level objectives — grounded in traffic rules, ethical considerations, or operational norms — and can be numerous and interdependent (Rizaldi et al., 2017; Jobin et al., 2019). Specifying, reasoning about, and ensuring such behaviours in a rigorous and traceable manner demands formal specification languages, such as temporal logic (see Subsection 2.2.3). Developing methods that translate such specifications into correct-by-design control strategies lies at the core of trustworthy autonomy — and forms a central theme of this thesis.

Intended Behaviour

Robotics Robotic systems are moving beyond controlled settings, such as factory floors and logistics hubs, into unstructured, human-populated environments (Penders et al., 2018). In doing so, robotics generalises the challenge of embodied AI to more diverse and less standardised domains (Pfeifer and Iida, 2004). Applications range from delivery robots and drones, assistive systems in healthcare and elderly care, to robotic surgery. these systems begin to interact more closely with humans, they must contend with non-deterministic, context-dependent human behaviour — often reactive in response to their own actions — while maintaining safe and interpretable operation. This transition raises profound challenges for guaranteeing trustworthy

behaviour under uncertainty, especially in scenarios involving rich interactions and dynamic decision-making (Barajas et al., 2025). Several surveys have begun to explore how formal methods can contribute to this goal (Luckcuck et al., 2019; Araujo et al., 2023).

Aviation Compared to other domains, aviation is anchored in a deeply conservative safety culture, where liability concerns outweigh innovation due to the catastrophic potential of failure (Emanuilov, 2017). As a result, autonomous capabilities are introduced only incrementally. While developments span drones, air taxis, and even commercial aircraft, true autonomy remains extremely limited. Traditional avionics rely on high-integrity sensors (e.g., GPS, IMU, altimeters), well-understood models, and sufficiently limited automation to enable the formal verification of critical functions such as autopilot control. For instance, semi-autonomous aircraft operate within rigorously defined safety envelopes (Annaswamy et al., 2024, Chapter 4.B). However, increasing functional requirements and the gradual integration of machine learning components — especially in perception — challenge existing frameworks and call for new verification approaches (Provan, 2024). As a case in point, in most countries, stringent safety regulations mandate commercial aircraft to carry certified collision avoidance systems. The long-standing *Traffic Alert and Collision Avoidance System* (TCAS) (Kuchar and Drumm, 2007) provided rule-based advisories, but has been superseded by the more adaptive *Airborne Collision Avoidance System X* (ACAS X) (Kochenderfer et al., 2012), designed to accommodate the increasing density of air traffic and evolving safety requirements. Notably, ACAS X remains purely advisory — the pilot retains full control. This underscores the prevailing regulatory ethos: even when autonomy is introduced, agency is not relinquished — instead, additional challenges arise from keeping the human in the loop.

Personalised/Smart Healthcare Advances in sensing and data analytics have enabled systems that monitor human health through digital histories of multimodal measurements — from wearable insulin and heart rate monitors to MRI scans and blood tests. These systems often aim to support clinical decision-making by identifying temporal patterns, classifying patient trajectories, or suggesting personalised interventions (Corral-Acero et al., 2020). Applications under development increasingly place patients in closed-loop interaction with medical devices, ranging from remote patient monitoring that informs treatment decisions, to automatic volume control in ventilators (Hewing et al., 2021), and predictive vagus nerve stimulation for epilepsy (Ingham et al., 2024). Despite these advancements in AI-based monitoring and diagnostic tools, clinical adoption remains limited, as they often fail to map multimodal symptoms onto actionable, coherent disease profiles, undermining explainability and trust (Tjoa and Guan, 2020). This challenge is compounded by the fact that many biological processes remain only partially understood, introducing significant epistemic uncertainty. Regulatory frameworks further restrict innovation: medical devices typically require full re-certification for any software change, severely restricting the deployment of adaptive, learning-enabled systems — especially in the absence of accepted verification procedures and tools. As a result,

most systems remain strictly advisory. Comprehensive accounts of these systemic barriers and emerging trends are given in Thimbleby (2021); Orwat et al. (2008).

Smart Grids and Critical Infrastructure Modern critical infrastructure — including power grids, communication networks, and transportation systems — is becoming increasingly decentralised, data-driven, and interconnected. The power grid, in particular, is undergoing a profound transition in response to global climate imperatives: shifting from centralised, dispatchable generation to distributed energy resources (DERs) such as rooftop photovoltaics, wind farms, and electric vehicle fleets. These systems must continuously balance fluctuating demand and variable supply, especially under intermittent renewable inputs. High penetration of renewables reduces grid inertia, rendering the system more sensitive to disturbances and raising the risk of cascading failures (Wooding et al., 2021). Blackouts have demonstrated how insufficient automation and inadequate coordination can trigger widespread societal disruption — affecting hospitals, public transport, payment services, factories, and communication networks — with severe economic consequences and fatalities. Such network-wide failures are rarely the result of isolated technical faults; rather, they emerge from the inability of the network to maintain or recover safe and functional operation under unforeseen conditions. Formal methods offer a promising path toward provably safe coordination of DERs, enabling the automatic arrest of frequency drops and prevention of load-shedding events. Techniques for synthesising distributed controllers with network-level guarantees could significantly reduce reliance on manual SCADA[3] interventions, which often fail under large-scale disturbances (Wooding et al., 2021). Yet, most existing methods rely on accurate network models — a limitation that motivates the uncertainty-aware, data-driven approaches developed in this thesis.

Military and Defence Fuelled by resurging defence spendings, autonomous systems are increasingly explored in military contexts (Reis et al., 2021). While many previously mentioned technologies allow dual use, notable applications include autonomous drones and vehicles used for surveillance, tactical missions, recovery, and transportation — reducing human exposure in high-risk zones. Systems such as Palantir's AI-enabled decision-support platforms assist with real-time situational analysis and strategy formulation. While these tools remain advisory, they operate over dynamic data streams and asset networks, underscoring the role of AI in managing complex, time-sensitive military operations.

This overview is not exhaustive. Numerous other domains also demand certified, trustworthy autonomy — each with their own technical and societal challenges. Building on this, the next section examines the formal foundations that enable reasoning about the behaviour of such systems and outlines the principal methodological challenges they face.

[3] SCADA – supervisory control and data acquisition

1.1.2 Formal Methods for Cyber-Physical Systems

Formal Verification & Synthesis

Formal methods encompass a broad spectrum of mathematically grounded techniques for specifying the desired behaviour of a system and rigorously reasoning about its adherence (Clarke and Wing, 1996; Belta et al., 2017). In this regard, *formal verification* refers to determining the compliance of a system with a given specification via formal methods; *formal synthesis* extends this concept to the algorithmic design of controllers that render the controlled system compliant. Unlike testing-based solutions, formal methods provide rigorous correctness guarantees by systematically accounting for all possible behaviours permitted by a given model — typically by construction or formal proof. This level of guarantee is especially compelling for autonomous CPSs, where embedded controllers operate without human oversight (Wongpiromsarn et al., 2023; Provan, 2024; Luckcuck, 2023) — aiming at asserting compliance before actual deployment.

Many formal methods are designed for nonlinear system dynamics and abstract high-level tasks, which classical control methods struggle to represent or reason over (Khalil and Grizzle, 2002). Yet, the application of these methods to real-world CPSs poses fundamental challenges: many techniques rely on accurate models of the system and its environment, and struggle to cope with high-dimensional dynamics and uncertainty. In fact, establishing scalable yet formal synthesis methods was highlighted as a major goal in the control community's 50-year road map (Annaswamy et al., 2024).

Challenges Despite their appeal, formal methods face fundamental obstacles when applied to real-world CPSs (Luckcuck et al., 2019). Most approaches presuppose access to accurate and tractable models of both the system and its environment — an assumption that often breaks down in complex physical domains, where dynamics are often nonlinear, stochastic, and only partially known. Even when such models exist, formal reasoning typically incurs prohibitive computational costs, especially in high-dimensional or partially observed systems, where the interplay of discrete and continuous variables leads to combinatorial explosion (Norman et al., 2017). To remain tractable, these methods often rely on structural simplifications — such as assuming low-dimensional or linear dynamics, or Gaussian noise (Park et al., 2012) — but any guarantees derived under such assumptions may no longer hold if the real system violates them. In practice, autonomous systems often fail due to discrepancies between design-time models and the actual operating environment (Luckcuck, 2023). This demands for methods that can bridge the simulation–reality gap between increasingly abstract models and the actual physical system they aim to represent (see Fig. 1.1). As systems grow in complexity and autonomy, addressing these challenges calls for methods that can confront uncertainty, abstraction, and learning, without relinquishing formal rigour.

Abstraction

Simulation vs Reality

To navigate these limitations, different strands of formal methods have emerged (Lavaei et al., 2022a; Yin et al., 2024) — each adopting a distinct formal reasoning approach, resting on specific system assumptions and supporting different specification types.

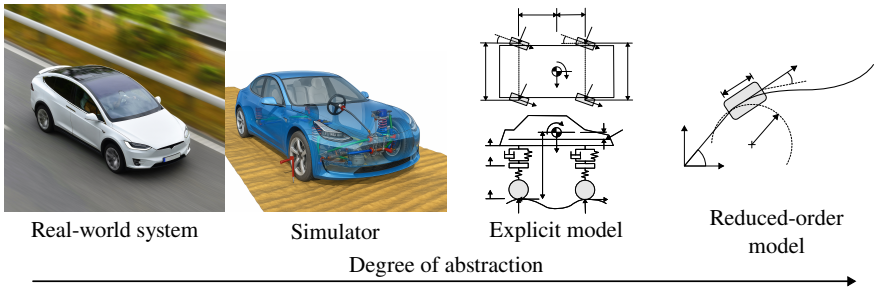


Fig. 1.1: Levels of abstraction: From the real-world system, via simulation models and mathematical descriptions, to reduced-order models.

Symbolic Abstraction

Abstraction-Based Approaches A prominent class of approaches enables control synthesis by abstracting the continuous dynamics of CPSs into finite-state models (see Fig. 1.2). These *symbolic abstractions*, typically constructed through space discretisation, enable control synthesis from high-level specifications — often expressed using temporal logic languages — via automata-based techniques (Vardi, 1985; Baier and Katoen, 2008; Tabuada, 2009; Belta et al., 2017; Forejt et al., 2011a). That is, specifications are translated into equivalent automata, and synthesis is performed on the product of the symbolic abstraction and the specification automaton. Note that the number of discrete states grows exponentially with the system dimension and complexity of the automaton. To mitigate this *curse of dimensionality* inherent in abstraction-based approaches, several techniques have been proposed to reduce the abstraction size while preserving formal guarantees:

- Adaptive partitioning and refinement seeks to reduce the total number of partitions on a dynamics manifold of given dimensionality by exploiting its regularity (Soudjani and Abate, 2013), the inflicted non-determinism (Kattenbelt et al., 2010), or its emanating behaviour with respect to a given specification (Banse et al., 2023).
- Compositional techniques break down large-scale networked systems via dissipativity or small-gain reasoning (Arcak et al., 2016; Alur et al., 2001; Hahn et al., 2013; Lavaei et al., 2023). They often rely on appropriate assumptions, including input-to-state stability, bounded subsystem interaction, and decomposable task specifications.
- Model order reduction (MOR) methods approximate high-dimensional systems by lower-dimensional surrogate models that preserve key behavioural or input-output properties (Chou et al., 2017). For stochastic systems, typical MOR methods providing computable error bounds include balanced truncation (Gugercin and Antoulas, 2004) and tensor-based approaches (Wang et al., 2025; Mathiesen et al., 2025).

While such additional steps can facilitate more scalable formal reasoning (Miner and Parker, 2004), every layer of abstraction introduces a behavioural mismatch between

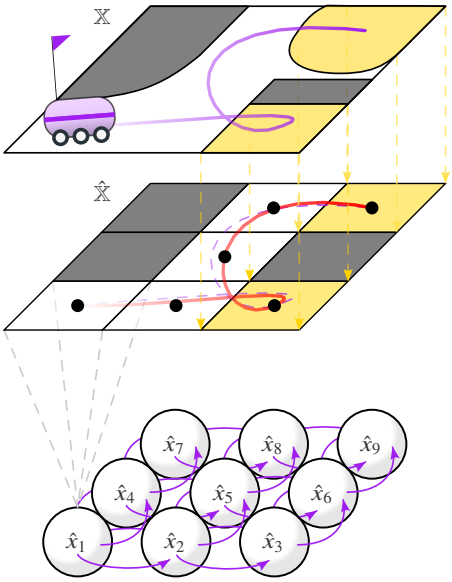


Fig. 1.2 To reason formally about the behaviour of a delivery robot in the continuous physical world (top), the state space $\mathbb{X}$ is discretised (middle) to construct a symbolic model (bottom), suitable for synthesis. Target regions (yellow) are underapproximated and obstacles (grey) overapproximated to ensure correctness despite model error. The discrete trajectory (black dots) induces a predicted continuous trajectory (red), whilst the resulting actual trajectory of the robot (purple) may diverge due to dynamics mismatch or uncertainty.

Simulation Relations

the symbolic model and the physical system. *Simulation relations* are used to formally quantify the behavioural similarity between the two systems and guide the synthesis of robust controllers for which the guarantees transfer to the original system (Segala and Lynch, 1995; Baier and Katoen, 2008; Girard and Pappas, 2011). However, constructing tractable abstractions and establishing simulation relations for nonlinear stochastic systems is extremely computationally demanding (Van Huijgevoort et al., 2023), and ensuring robustness becomes even more challenging when the underlying model is uncertain or incomplete. These challenges have motivated the development of alternative approaches that reason directly about the continuous-state system dynamics, circumventing the need for fully symbolic abstractions.

Abstraction-Free Approaches Rather than constructing symbolic abstractions, an alternative strand of methods operates directly on the continuous dynamics, typically by trying to find witness functions or performing set-based computations. These so-called *abstraction-free* approaches avoid the combinatorial blow-up associated with discretisation, but come with their own trade-offs: they often require specific structural assumptions to remain computationally feasible and the introduced conservatism tends to be less controllable. *Control barrier certificates*[4] (CBCs) are among the most widely used techniques in this category (see Fig. 1.3). They certify safety or

Barrier Certificates

[4] In this thesis, CBCs and discrete-time control barrier functions (CBFs) (Cosner et al., 2024) are jointly referred to as CBCs since their definitions are equivalent.

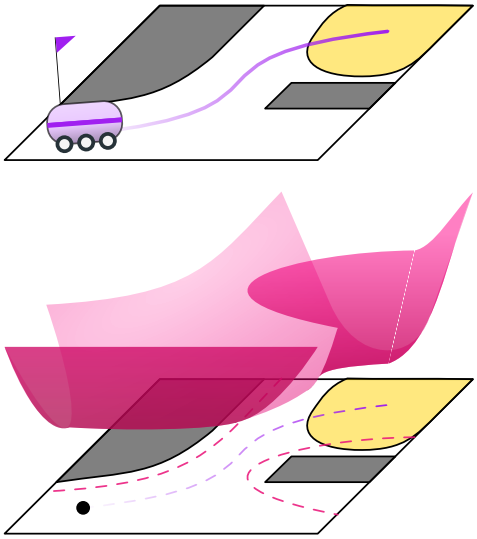


Fig. 1.3 The delivery robot is steered toward a target region (in yellow) without constructing a symbolic abstraction of the continuous system (top), by identifying a barrier function (red surface) that exceeds a safety threshold in the avoid regions (in grey) — indicated by the red dashed level set — and decreases towards the target.

reachability properties by proving the existence of a (barrier) function that satisfies a given set of inequalities with respect to the expected behaviour of the system (Prajna et al., 2007). The resulting global optimisation problem is in general a robust optimisation problem featuring infinitely many constraints. Thus, most approaches either impose strong assumptions — such as control-affine or polynomial dynamics and constraints to permit sum-of-squares (SOS) relaxations — or resort to spatial abstractions to render the problem tractable. For specifications beyond reachability, the problems become sequential reachability tasks (Jagtap et al., 2020b), greatly impeding their applicability. Chapter 5 explores these approaches in more detail.

Reachability Analysis

Reachability analysis forms a second major class of abstraction-free techniques, which computes over-/underapproximations of the state space reachable under a given control law by iteratively propagating sets of states through the system dynamics (Althoff et al., 2021). For linear dynamics, these sets remain convex, enabling highly efficient computation and placing such methods among the most practically applied formal tools. While extensions to nonlinear systems exist, most approaches resort to abstraction techniques or rely on specific system properties to handle nonconvexity. For stochastic systems, viable approaches are scarce, and often either depend on linearity (Arcari et al., 2023) or isolate the stochastic influence from the deterministic dynamics (Jafarpour et al., 2025) — a decomposition that may not hold in more complex systems.

Both reachability analysis and CBCs struggle to handle richer temporal or logical specifications beyond basic reach-avoid properties. Consequently, abstraction-free methods offer a promising direction for formal methods, but remain sensitive to structural assumptions. With a focus on autonomous CPSs, these assumptions might

not hold as systems become increasingly complex, uncertain, or opaque. As a result, even supposedly abstraction-free approaches regularly resort to some form of abstraction when moving towards more complex systems or specifications — placing them in subtle tension with abstraction-based methods, which, despite their discretisation overhead, naturally support broader classes of dynamics and specifications (see Section 5.7).

Formal Synthesis Pipeline Powered by the formal approaches discussed above, one can construct end-to-end formal synthesis pipelines that take high-level task specifications and system representations as inputs and produce controllers with probabilistic correctness guarantees as outputs (see Fig. 1.4). This systematic application of formal methods is commonly referred to as *temporal logic control* (Belta et al., 2017), where the specification is typically provided in a temporal logic language — such as linear temporal logic (LTL), signal temporal logic (STL), or ω-regular expressions — whilst possible system representations range from discrete-time stochastic models, such as Markov decision processes (MDPs) and partially observed MDPs (POMDPs), to data-driven approximations. Based on these inputs, a suitable formal

Temporal Logic Control

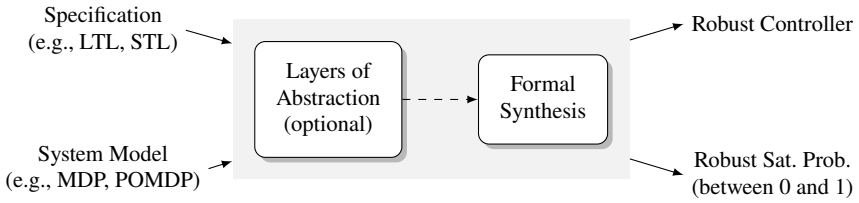


Fig. 1.4: Formal synthesis pipeline.

synthesis framework computes a controller that ensures that the controlled system satisfies the specification with mathematically guaranteed probability. The resulting controllers are typically deployed within layered architectures, with low-level feedback controllers interfacing with the physical sensors and actuators of the system in real-time (Matni et al., 2024). This provides a powerful alternative to classical control frameworks such as model predictive control (MPC), which is effective for convex objectives and constraints but struggles to address rich logical combinations of events or long-horizon sequential tasks. In contrast, the temporal logic interface allows users to specify rich desired behaviours directly and unambiguously in symbolic form, enabling a principled, machine-verifiable synthesis process for the design of autonomous systems, that can account for various system dynamics when combined with a suitable formal synthesis approach (Luckcuck, 2023).

However, the algorithmic approaches used in this pipeline — whether abstraction-based or abstraction-free — still inherit the same bottlenecks: assumptions about model structure and the availability of sufficiently accurate but low-dimensional dynamics models seldom hold in practice. As such, whilst formal synthesis is promising to the control of autonomous CPSs, these limitations necessitate the development of

scalable approaches that better accommodate system uncertainty and model opacity — setting the stage for the discussion that follows.

The Role of Uncertainty When interacting with the real world, autonomous CPSs must operate in open, dynamic environments, where uncertainty is the norm rather than the exception. As highlighted in the formal synthesis pipeline (Fig. 1.4), the effectiveness of any controller depends critically on the accuracy of the underlying system model. If this model is incomplete or not updated when the environment changes, the resulting behaviour may diverge from the intended specification — even if the synthesis procedure is formally correct (Luckcuck, 2023). Hence, the reliability of formal methods hinges not just on their algorithmic rigour, but on their capacity to model and reason under uncertainty.

Aleatoric vs Epistemic Uncertainty

Two principal types of uncertainty arise in this context. *Aleatoric uncertainty* captures randomness inherent in system behaviour — such as environmental influences and stochastic transitions — and is typically modelled probabilistically, e.g., through Markov processes. *Epistemic uncertainty*, by contrast, stems from incomplete or imprecise knowledge about the system, whether due to coarse abstractions, inaccurate or sparse sensor measurements, or unobserved dynamics — thus also referred to as *systematic* or *model uncertainty*. While aleatoric uncertainty is often seen as irreducible, epistemic uncertainty can, in principle, be reduced through data collection, improved sensing, or better modelling. In practice, the boundary between these types is blurry: what appears stochastic may reflect unmodelled structure or sensor limitations, and what seems systematic may resist precise characterisation. Nonetheless, both forms of uncertainty challenge the assumptions underpinning standard synthesis methods. Addressing them requires techniques that either explicitly model uncertainty or ensure robust performance despite it. These challenges motivate the need for synthesis frameworks that not only tolerate uncertainty but embed it as a first-class modelling concern. The next paragraphs discuss how robustness and learning can serve this goal. For an in-depth discussion on the epistemological nuances of uncertainty types, see Kirchhof et al. (2025); Der Kiureghian and Ditlevsen (2009); Smith (2024); Hüllermeier and Waegeman (2021).

Robustness under Uncertainty Designing controllers for systems under uncertainty requires methods that can cope with both aleatoric and epistemic uncertainty. For the former, stochastic models such as Markov decision processes (MDPs) allow formal verification against task specifications via probabilistic model checking (Vardi, 1985; Kwiatkowska et al., 2007). Recall that unbounded stochasticity generally prevents binary satisfaction of specifications; instead, one quantifies a *robust* satisfaction probability, i.e., a lower bound on the probability of satisfying a given specification that accounts for modelling errors and abstraction steps in the formal pipeline (Fig. 1.4).

Robust Satisfaction Probability

Robust MDPs

To robustify against epistemic uncertainty, models such as robust MDPs (RMDPs) — including interval and parametrised MDPs (IMDPs, pMDPs) — are widely used (Suilen et al., 2025; Puterman, 2014; Hahn et al., 2019). Parametrised models can capture variability in system configurations and tolerances, such as deterioration over time or manufacturing deviations. Crucially, such models avoid the need to

design a new controller for each system instance, making them essential for scalable, real-world deployment. Despite their promise, formal synthesis under epistemic uncertainty remains computationally challenging. Many approaches require reasoning over a set of models, i.e., performing worst-case analysis with respect to every state-specific model instantiation. This becomes intractable for high-dimensional systems, especially in the absence of MOR. For instance, MOR-compatible methods for IMDPs are extremely limited (Mathiesen et al., 2025), and distributionally robust RMDPs based on ambiguity sets are computationally tractable only under fixed partially-known dynamics (Gracia et al., 2023), with recent MOR techniques limited to linear systems (Pulch, 2021). These challenges underscore a critical limitation: sufficiently abstract system representations must often be constructed or adapted for each setting — a process that is costly and subject to statistical error. In light of these limitations — especially in high-dimensional or poorly understood settings — incorporating data-driven frameworks into the formal pipeline to move from known models to abstractions of unknown systems learned directly from observations is becoming essential.

Learning under Uncertainty In practice, constructing accurate models of complex systems is often infeasible or expensive and time-consuming. This is especially true for AI-enabled CPSs, where components may be black-box or only observable through data (Corso et al., 2021). This tension has sparked interest in statistical data-driven learning frameworks — including non-parametric methods such as Gaussian processes (Jackson et al., 2020; Jiang et al., 2022; Jagtap et al., 2020a; Wajid et al., 2022), non-parametric estimation (Zhang et al., 2024), conditional mean embeddings (Thorpe et al., 2022; Romao et al., 2023), neural network dynamics models (Adams et al., 2022), and scenario-based methods (Lavaei et al., 2022b; Salamati et al., 2024; Kazemi et al., 2024a; Ajeleye et al., 2023) — where models with quantified statistical errors are inferred from data, or even more, controllers are learned directly from data, for example using model-free reinforcement learning (Kazemi and Soudjani, 2020; Lavaei et al., 2020a; Hasanbeig et al., 2019; Kazemi et al., 2024b).

Statistical Learning Frameworks

However, this is notoriously difficult: approaches often suffer from high sample complexity and data must typically be supplemented with information on the regularity of the system dynamics to establish rigorous error bounds. This often forces assumptions on the system structure or intermediate representations — such as linearity, Lipschitz continuity, or kernel-related complexity bounds — that are difficult to justify in uncertain or partially known systems. In fact, few works provide practical approaches to obtaining such quantities (Zhang et al., 2024).

A key challenge in this setting is not only to learn accurate dynamics, but to learn system abstractions that are amenable to efficient formal reasoning. More specifically, most learning-enabled models are continuous — just like the system dynamics they aim to represent — and thus not natively amenable to formal reasoning. Abstracting them to, e.g., IMDPs, is extremely computationally intensive, as the conversion typically requires optimisation over all possible transitions to determine worst-case probability bounds (Wooding and Lavaei, 2024). Even more, if more data becomes available, most models have to be recomputed entirely, with few existing

Compatibility with Formal Reasoning

approaches relating robust satisfaction probabilities directly to data availability. For abstraction-free approaches, learning controllers via CBCs without a given model concentrates around establishing a chance constraint on the latent dynamics of the system based solely on data (Wajid et al., 2022; Jagtap et al., 2020a). To obtain the required guarantees, existing work is often limited to control-affine systems (Lopez and Slotine, 2023; Cohen et al., 2022a), or requires exponentially large data sets and knowledge of bounds on the regularity of the system (Salamati et al., 2024). Adjacent work explores model-free reinforcement learning (Kazemi and Soudjani, 2020; Lavaei et al., 2020a) or reachability analysis via sampling-based schemes (Lew and Pavone, 2021), typically relying on either Lipschitz constants or providing only asymptotic guarantees.

As such, learning-based formal methods for CPSs remain in early stages, and a key open challenge is to develop frameworks that integrate efficient inference of abstract models and controllers directly from data, whilst retaining formal guarantees under uncertainty and relaxing the reliance on restrictive assumptions.

1.2 Central Research Question and Thesis Scope

The previous sections have outlined the vast challenges and opportunities that formal methods offer for developing trustworthy autonomy. This thesis addresses key limitations of existing formal control synthesis and verification approaches for CPSs operating under both aleatoric and epistemic uncertainty. The focus is on robust, data-driven frameworks that relax unrealistic assumptions prevalent in the literature.

> **Central Research Question** How can correct-by-design controllers be constructed for uncertain stochastic systems to ensure satisfaction of temporal logic specifications based on data?

Rather than layering post-hoc verification tools onto existing pipelines, this thesis builds from the ground up — engaging deeply with the theoretical foundations of formal methods and data-driven learning to develop learning-enabled architectures that naturally embed uncertainty handling and correctness by design. To make progress on this question, a scoped view of control under uncertainty is adopted. The following assumptions define the class of systems and specifications considered, balancing expressiveness and tractability:

- **Discrete-Time Control:** Most engineered systems operate in discrete-time feedback loops, with sensing and actuation occurring at fixed sampling rates. This assumption aligns with practical implementations of embedded controllers and supports common data-driven learning pipelines. Extensions to asynchronous sensing and continuous-time guarantees are discussed in Chapter 7.
- **Nonlinear Stochastic Dynamics:** Randomness and limited model knowledge are key features of real-world CPSs. Thus, uncertainty is treated as a first-class

modelling element. The thesis builds on general stochastic nonlinear systems, in contrast to many existing methods that rely on deterministic or linear dynamics to obtain guarantees.

- **Observability and Markovianity:** The systems are assumed to have Markovian dynamics with full-state observability. This avoids the additional computational burden of online estimation or POMDP[5] reasoning and isolates the core challenge of synthesising controllers under model uncertainty. Extensions to partially observable systems are discussed in Chapter 7.
- **Time Invariance:** The system behaviour is assumed static with respect to time. This abstracts away the additional task of modelling the evolution of the dynamics themselves to focus on designing robust controllers.
- **Static Specifications:** Temporal logic specifications are assumed to be known and time-invariant. This excludes adaptive tasks and keeps the focus on the development of scalable synthesis techniques. Future work discusses possible extensions to dynamic settings.

These assumptions define a tractable yet challenging problem setting, centred around uncertainty-aware models such as pMDPs, IMDPs, and kernel-based models. These models capture both aleatoric and epistemic uncertainty while enabling policy synthesis with formal guarantees. Subsequent chapters may focus on fragments of temporal logic, whilst pointing towards possible extensions, wherever applicable.

By drawing clear boundaries around what is modelled and controlled, the thesis concentrates on the synthesis of data-driven controllers that satisfy rich specifications under uncertainty — without being encumbered by orthogonal challenges such as partial observability, non-stationary environments, or evolving task logic. These aspects are acknowledged and revisited in Chapter 7, which outlines pathways for lifting or relaxing the present assumptions.

1.3 Claims and Key Takeaways

The overarching contributions of this work fall into five main areas, which are reflected in the structure and technical progression of the thesis.

I Addressing Aleatoric and Epistemic Uncertainty in Formal Synthesis For continuous-state systems, existing formal approaches typically consider stochasticity (aleatoric uncertainty) or incomplete model knowledge (epistemic uncertainty) in isolation (see related work in Subsection 1.1.2). This thesis develops methods that explicitly incorporate both types of uncertainty, addressing the second-order problem of uncertainty about the stochastic law itself. This is especially relevant for CPSs, as an exact stochastic model of the system is rarely available and guarantees only hold for the specific model based on which they were obtained. The chapters follow a progression from model-rich settings to minimal prior knowledge, show-

[5] POMDP — partially observable Markov decision process

ing how formal guarantees can be achieved under increasing levels of compounded uncertainty.

II Exploring the Trade-Off Between Structure and Guarantees Dual to uncertainty, the strength and tractability of formal guarantees are often tied to the degree of structural assumptions (e.g., linearity, parametrised forms, Lipschitzness) imposed about the system. This thesis spans a continuum from model-rich to model-free paradigms, relaxing assumptions toward fully black-box systems whilst retaining a quantifiable notion of correctness. The chapters explore how imposing structure in progressively more abstract form — from spatial to spectral symbolisms — enables more flexible guarantees and more efficient synthesis. The underlying spectrum of structural assumptions is central to understanding the trade-offs in practical verification and synthesis workflows.

III Property-Centric Learning for Verification As systems grow in complexity and uncertainty, constructing full models becomes both computationally expensive and often unnecessary. Instead, this thesis advocates for learning structured representations designed to support formal synthesis directly. Two complementary developments are pursued: on one hand, structure is systematically imposed in learned models — including parametrised dynamics, hierarchical tree structures, and spectral representations — to ensure compatibility with abstraction and model checking procedures. On the other hand, explicit model construction is increasingly bypassed in favour of learning predictors that assess the property of interest directly from data. This shift reflects a broader trend toward integrating property-centric reasoning deep into the learning process itself.

IV Correct-by-Design Synthesis over Post-Hoc Verification A central theme in this thesis is the emphasis on synthesising controllers that satisfy temporal logic specifications with formal guarantees established during the design phase. This stands in contrast to runtime safety monitoring or statistical falsification techniques, and reflects a principled shift in how safety is approached for autonomous CPSs. A key challenge addressed in this context is the ability to provide such guarantees even when system knowledge is incomplete or when configurations may vary at deployment. Recall that the ability to synthesise policies that generalise across varying configurations or improve with accumulating data is a recurring objective.

V Selective Abstraction and Robust Refinement Several contributions point toward the importance of selective abstraction and refinement as mechanisms for scalable formal synthesis. In particular, abstraction is used to construct reduced-order models that preserve critical system behaviour, while refinement techniques ensure that guarantees obtained at the abstract level carry over to the original system. Although only partially explored in this thesis, these directions offer promising avenues for building scalable, compositional synthesis pipelines in future work.

These claims align with the technical progression of modelling paradigms across the thesis chapters, as illustrated in Fig. 1.5. Throughout, the thesis highlights that identifying structure conducive to efficient model checking and synthesis whilst

quantifying and compensating for the resulting abstraction error is central to advancing the practical scalability of formal methods.

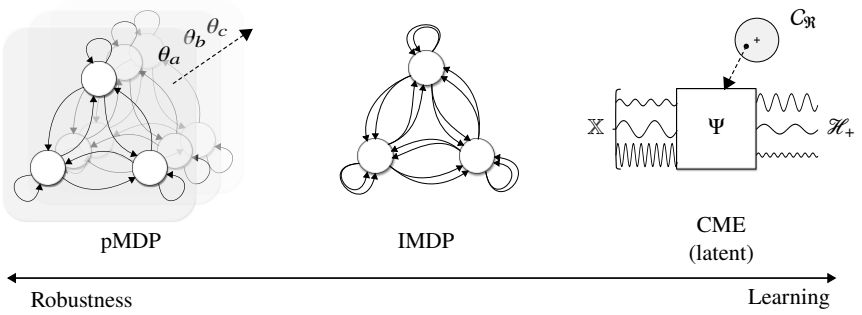


Fig. 1.5: Qualitative comparison of dynamics modelling paradigms by their robustness and expressive learning capacity — from parametric discrete-state models, and interval-based abstractions, to latent function-space representations (left to right).

1.4 Overview of the Thesis

This section maps the thesis chapters to the claims and main takeaways outlined previously. The contents of this thesis are based on original research published or under review in peer-reviewed international conferences and journals. As the exclusive first author, I have written all works in their entirety. Across all chapters, co-authors supported the work through feedback or assistance with the implementation. While many sections are adapted from the original papers, all writing has been revised to ensure consistency and coherence throughout the thesis.

PART I: ROBUSTNESS UNDER UNCERTAINTY

Chapter 2: Bayesian Correct-by-Design Synthesis of Uncertain Systems This chapter presents a correct-by-design synthesis approach for stochastic systems, integrating robustness to uncertain model parameters via Bayesian learning and probabilistic simulation relations. This is especially relevant as deployed systems can rarely be modelled exactly and parameters can change due to manufacturing-related or operational factors. To this end, continuous-space nonlinear dynamics are abstracted into finite-state (reduced-order) models, used to synthesise controllers for the original system that satisfy complex temporal logic specifications upon refinement. A known model structure is imposed, with unknown parameters treated as random variables. Bayesian linear regression is applied to derive credible parameter sets from data. Probabilistic simulation relations — formalised through new *sub*-probability cou-

plings — then link the unknown data-generating system to its symbolic abstraction, capturing both aleatoric and parametric uncertainty. The approach achieves rigorous robustness guarantees through comprehensive error bound computations, highlighting that scalability hinges on the availability of reduced-order models or additional structural information.

Associated Publications:

↪ **Oliver Schön**, Birgit van Huijgevoort, Sofie Haesaert, and Sadegh Soudjani (2025). *"Bayesian Formal Synthesis of Unknown Systems via Robust Simulation Relations"*. In *IEEE Transactions on Automatic Control (TAC)*, **Full Paper**, 70(3):1531–1546, 16 pages, DOI: 10.1109/TAC.2024.3459308.

↪ **Oliver Schön**, Birgit van Huijgevoort, Sofie Haesaert, and Sadegh Soudjani (2023). *"Poster Abstract: Data-Driven Correct-by-Design Control of Parametric Stochastic Systems"*. In Proceedings of the *26th ACM International Conference on Hybrid Systems: Computation and Control (HSCC)*, San Antonio, TX, USA, 2023. ACM, 2 pages. DOI: 10.1145/3575870.3589547.

↪ **Oliver Schön**, Birgit van Huijgevoort, Sofie Haesaert, and Sadegh Soudjani (2022). *"Correct-by-Design Control of Parametric Stochastic Systems"*. In *2022 IEEE 61st Conference on Decision and Control (CDC)*, Cancun, Mexico, 2022, pp. 5580–5587, DOI: 10.1109/CDC51059.2022.9992723.

Chapter 3: Correct-by-Design Synthesis for Networks of Uncertain Systems
CPSs are often composed of multiple dynamically interconnected subsystems operating under uncertainty. This chapter builds on the previous results by extending correct-by-design synthesis to such networked settings, addressing key challenges of scalability and heterogeneous disturbances. Local probabilistic simulation relations are constructed to handle parametric uncertainty from data and are composed into a global relation, enabling the synthesis of local controllers that collectively satisfy temporal logic specifications for the entire networked system. Arbitrary noise distributions are approximated via learnable Gaussian mixture models. The overall approach scales linearly with the number of subsystems, significantly improving the tractability of correct-by-design synthesis for systems with inherent network structure.

Associated Publication:

↪ **Oliver Schön**, Birgit van Huijgevoort, Sofie Haesaert, and Sadegh Soudjani (2023). *"Verifying the Unknown: Correct-by-Design Control Synthesis for Networks of Stochastic Uncertain Systems"*. In *2023 62nd IEEE Conference on Decision and Control (CDC)*, Singapore, Singapore, 2023, pp. 7035–7042, DOI: 10.1109/CDC49753.2023.10383307.

PART II: LEARNING UNDER UNCERTAINTY

Chapter 4: Learning Formal Abstractions via Binary-Tree Gaussian Processes This chapter marks the shift towards a focus on learning formal abstractions of stochastic systems directly from data. Gaussian process (GP) regression is employed to learn system dynamics with additive Gaussian stochasticity from data without relying on a given parametrised model structure. While IMDPs provide a verification-friendly model representation often used in conjunction with GPs, constructing them from learned dynamics is typically computationally expensive. This chapter demonstrates that GPs with structural binary-tree kernels — known as *binary-tree Gaussian processes* (BTGPs) — can be used to generate IMDPs with orders-of-magnitude speed-up, by learning representations that are naturally discontinuous and thus inherently amenable to formal analysis. In doing so, BTGPs avoid computationally intensive abstraction and error quantification procedures. By showing how kernel-based learning can be integrated directly into the modelling pipeline, this chapter illustrates the synergy between machine learning and formal methods, offering a scalable solution for the robust verification of reachability specifications under both aleatoric and epistemic uncertainty.

Associated Publication:

↪ **Oliver Schön**, Shammakh Naseer, Ben Wooding, and Sadegh Soudjani (2024). *"Data-Driven Abstractions via Binary-Tree Gaussian Processes for Formal Verification"*. *IFAC-PapersOnLine*, 58(11):115–122, In: *The 8th IFAC Conference on Analysis and Design of Hybrid Systems (ADHS)*, Boulder, CO, USA, 2024, DOI 10.1016/j.ifacol.2024.07.434.

Chapter 5: Kernel-Based Learning of Black-Box Safety Certificates This chapter turns to the challenge of verifying safety for fully black-box stochastic systems using data alone. Control barrier certificates are learned directly from observed transitions via *conditional mean embeddings* (CMEs), capturing uncertainty and enabling distributionally robust guarantees. A spectral abstraction based on a finite Fourier expansion casts the verification problem as a linear program, allowing efficient computation using the fast Fourier transform. The approach is extremely flexible and extends conceptually to broader temporal logic specifications, including ω-regular properties.

Associated Publications:

↪ **Oliver Schön**, Zhengang Zhong, and Sadegh Soudjani (TBA). *"Kernel-Based Learning of Safety Barriers"*. Manuscript submitted for publication in *Journal of Artificial Intelligence Research (JAIR)*, 37 pages.

↪ **Oliver Schön**, Zhengang Zhong, and Sadegh Soudjani (2024). *"Data-Driven Distributionally Robust Safety Verification Using Barrier Certificates and Conditional Mean Embeddings"*. In *American Control Conference (ACC)*, Toronto, ON, Canada, 2024, pp. 3417–3423, DOI: 10.23919/ACC60939.2024.10644349.

Chapter 6: Conclusion This chapter summarises the main results and accompanying software tools developed throughout the thesis, reflecting on their contributions to advancing the state of the art in formal synthesis under uncertainty.

Chapter 7: Future Directions This chapter concludes the thesis with an outline of promising directions for future work.

PART I
ROBUSTNESS UNDER UNCERTAINTY

In an uncertain world, robustness is key. A robust strategy expects uncertainty, plans for the worst, and guarantees success despite it.

Chapter 2

Bayesian Correct-by-Design Synthesis of Uncertain Systems via Robust Simulation Relations

Abstract This chapter proposes a data-driven approach to synthesising controllers for continuous-space stochastic systems with parametric uncertainty that are correct by design and can provably satisfy complex temporal logic specifications. The first stage utilises Bayesian regression to infer credible parameter sets from data, which contain the true parameters of the data-generating system with a specified level of confidence. In the second stage, a new class of simulation relations based on sub-probability couplings is introduced, which enable correct-by-design control refinement under both stochastic and parametric uncertainty. These relations are crucial for constructing abstract systems that are formally related to not just a single system, but an entire set of ambiguously parametrised systems. The approach is demonstrated on three case studies, including a nonlinear and a high-dimensional system.



2.1 Introduction

Many existing formal control synthesis approaches are based on the assumption that the *exact* stochastic model of the target system is known (Lavaei et al., 2022a; Yin et al., 2024). This constitutes a significant limitation, as any guarantees obtained during synthesis apply only to the specific model used. To address this issue, this chapter explores how classical data-driven system identification techniques can be integrated into formal synthesis frameworks.

System identification of stochastic systems is a well-established field (Keesman, 2011; Van Den Hof and Schrama, 1995). Typically, a parametrised model structure is imposed, whose parameters are estimated from data by minimising an appropriate distance of *closeness* — either in the time or frequency domain — between observed output data and the output trajectories generated by the model. This approach aligns

Closeness

naturally with engineering practice, where the general structure of a system is often modelled from first principles, with parameters determined empirically.

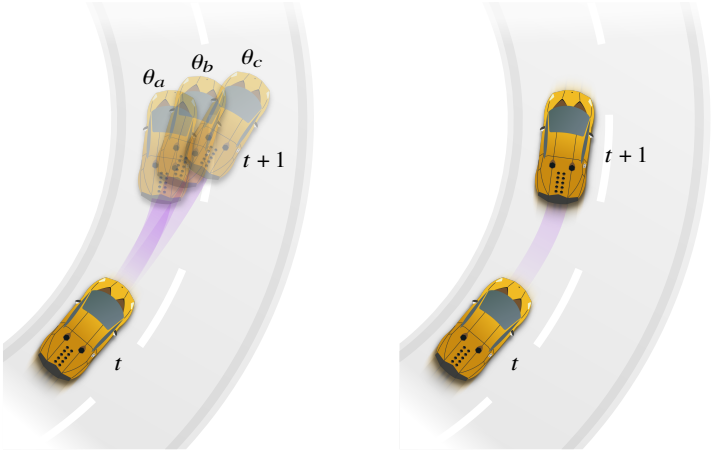


Fig. 2.1: The goal of system identification is to approximate the uncertain system $\mathcal{M}(\theta)$ (parametrised by uncertain parameters $\theta \in \{\theta_a, \theta_b, \theta_c\}$) with a model $\mathcal{M}(\hat{\theta})$ characterised by a concrete parametrisation $\hat{\theta}$. Note that the dynamics under each concrete parametrisation $\theta_a, \theta_b, \theta_c, \hat{\theta}$ are still stochastic.

Building on this foundation, this chapter assumes a given parametrised stochastic model and uses system identification tools to compute credible parameter sets that capture the uncertainty induced by limited data. This externalises epistemic uncertainty as an additional parametric dimension overarching a set of coherent stochastic models indexed by concrete parameter realisations, as illustrated in Fig. 2.1. The approach then builds on the theoretical foundations established by Haesaert et al. (2017); Haesaert and Soudjani (2021) to design robust controllers using abstraction-based methods via probabilistic simulation relations.

The problem studied in this chapter can be summarised as follows.

Problem 2.1 (Data-Driven Synthesis) Given sample data and a parametrised model of the uncertain true system, design a controller such that the controlled system satisfies a given temporal logic specification with probability at least $p_\psi \in (0, 1)$ and confidence $(1 - \alpha) \in (0, 1)$.[1]

[1] Two-layered probabilities are typical for problems involving both aleatoric and epistemic uncertainty. Whilst they can be merged (see, e.g., Corollary 2.1), keeping them separate helps attribute conservatism more transparently. The outer probability $(1 - \alpha)$ reflects the confidence in the identified model — including its stochastic law — used to compute the inner probability p_ψ, which quantifies the probability of specification satisfaction under the worst-case behaviour consistent

Unlike existing work on data-driven synthesis — mainly focused around more flexible machine learning techniques — this principled parametric approach facilitates rigorous yet tractable robustness guarantees in a practical form. For instance, GP-based IMDP construction (Jackson et al., 2020; Jiang et al., 2022) is extremely computationally intensive and requires bounds on the complexity of the latent dynamics in a reproducing kernel Hilbert space (RKHS) — an impractical assumption. Alternative, scenario-based approaches (Lavaei et al., 2022b) demand known Lipschitz constants to obtain formal guarantees, which are hard to obtain in practice. Moreover, IMDPs capture uncertainty in a state-wise independent fashion, leading to often overly conservative models (Schnitzer et al., 2024). In contrast, the approach taken here treats epistemic uncertainty globally and structurally via a parametrised model. While this may reduce expressiveness, it is better aligned with engineering intuition and compatible with established tools like MOR techniques (van Huijgevoort and Haesaert, 2022). It will be shown that this latter property is crucial to scale to high-dimensional systems (see Subsection 2.6.2).

In the context of formal synthesis for stochastic systems, few results are available to relate satisfaction of temporal requirements by the learned model to that of the true, unknown system in terms of the size of the dataset. In fact, existing IMDP-based approaches typically require full model reconstruction to incorporate new data, rendering them unsuitable for real-time learning applications such as autonomous systems.

This chapter presents an abstraction-based solution to Problem 2.1 for discrete-time stochastic systems and co-safe LTL specifications (see Definition 2.1). The proposed scheme proceeds in two stages:

1. **System Identification:** Available system identification techniques are utilised to learn credible parameter sets from data such that the true parameters lie within them with high confidence. As each concrete parametrisation associates a corresponding parametrised model (see Fig. 2.1), the result is a set of credible models.
2. **Robust Controller Synthesis:** A novel type of simulation relations is developed to quantify behavioural closeness between all models in the identified credible set and a chosen nominal model. As one of the main contributions, this type of simulation relation enables a control refinement[2] that is *independent* of the concrete valuation of the uncertain parameters — thus relaxing the need for exact parameter knowledge.

Theoretical results for establishing this new relation and the associated closeness guarantees for nonlinear parametrised systems with additive Gaussian uncertainty are provided, enabling the design of controllers with probabilistic specification satisfaction uniformly across the entire parameter set. The trade-off between the size of

with the identified law. See Hüllermeier and Waegeman (2021) for a detailed discussion on this interplay.

[2] Control refinement refers to adapting a controller synthesised for a surrogate model to a suitable controller for the target system, typically via an interface function and state mapping. See Subsection 2.4.1 for details.

the dataset and the conservatism introduced by the robust treatment of the parametric uncertainty is illustrated through numerical experiments (see Section 2.6). Whilst the proposed framework is generally compatible with any estimation method providing robust credible parameter sets (Jaynes and Kempthorne, 1976), this chapter illustrates the approach using Bayesian linear regression (Bishop, 2006).

Organisation This chapter unfolds as follows. After briefly recalling essential notation and foundational concepts from probability and measure theory, Section 2.2 presents the system and temporal logic specifications needed to formalise Problem 2.1 in Subsection 2.2.4, revealing two key subproblems. The first subproblem is tackled by the data-driven system identification framework presented in Section 2.3. Section 2.4 introduces the novel notion of sub-simulation relations and control refinement via partial coupling, demonstrating how to design robust controllers and compute rigorous bounds on specification satisfaction probabilities. Building on this, Section 2.5 leverages the identification approach from Section 2.3 to learn credible parameter sets from data and establish the connection between parametrised nonlinear models and their abstractions. The framework is validated across various systems in Section 2.6. The chapter closes with Section 2.7, which explores potential extensions.

2.2 Preliminaries and Problem Statement

Below is a collection of basic notation used throughout the thesis.

Notation Let $\mathbb{R}$, $\mathbb{R}_{\geq 0}$, $\mathbb{N}$, $\mathbb{N}_0$, be the sets of reals, non-negative reals, and natural numbers excluding and including zero, respectively. The transpose of a matrix A is indicated by $A^\top$. Given variables a_1, a_2, a_3 of suitable dimension, $[a_1; a_2; a_3]$ denotes the column vector $[a_1^\top, a_2^\top, a_3^\top]^\top$. I_n denotes the identity matrix of dimension $n \times n$, where $n \in \mathbb{N}$. For two sets A, B, their Cartesian product is written as $A \times B$. The indicator function $\mathbb{1}_A(x)$ for a set A is defined as $\mathbb{1}_A(x) = 1$ if $x \in A$ and $\mathbb{1}_A(x) = 0$ otherwise.

2.2.1 Probability Theory

The contents of this thesis require a basic understanding of some notions from probability and measure theory, summarised in this section. For further details and reading, please refer to the monograph Papoulis (1965) for a general introduction to probability theory and Bogachev (2007) for modern measure theory.

Probability Measures Let $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$ be a *Borel measurable space*, with a *Polish* sample space $\mathbb{X}$ and a Borel σ-algebra $\mathscr{B}(\mathbb{X})$ defined over $\mathbb{X}$, i.e., the smallest σ-algebra containing all open subsets of $\mathbb{X}$ (Bogachev, 2007). A positive *measure* ν

on $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$ is a non-negative map $\nu\colon \mathscr{B}(\mathbb{X}) \to \mathbb{R}_{\geq 0}$ such that for all countable collections $\{A_i\}_{i=1}^{\infty}$ of pairwise disjoint sets in $\mathscr{B}(\mathbb{X})$ it holds that $\nu(\bigcup_i A_i) = \sum_i \nu(A_i)$. A measure $p\colon \mathscr{B}(\mathbb{X}) \to [0,1]$ is a *probability measure* if $p(\mathbb{X}) = 1$ and a *sub-probability measure* if $p(\mathbb{X}) \leq 1$. A *probability space* is given by $(\mathbb{X}, \mathscr{B}(\mathbb{X}), p)$, with realisations $x \sim p$. The set of all probability measures for a given measurable space $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$ is denoted as $\mathscr{P}(\mathbb{X})$.

In this context, the following concepts are used throughout the thesis:

- The *Dirac delta* measure $\delta_a\colon \mathscr{B}(\mathbb{X}) \to [0,1]$ concentrated at a point $a \in \mathbb{X}$ is defined as $\delta_a(A) = 1$ if $a \in A$ and $\delta_a(A) = 0$ otherwise, for any measurable set $A \in \mathscr{B}(\mathbb{X})$.
- The n-dimensional *Gaussian* measure with mean $\mu \in \mathbb{R}^n$ and covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$ is defined by

$$\mathcal{N}(dx|\mu,\Sigma) := \frac{dx}{\sqrt{(2\pi)^n\,|\Sigma|}} \exp\left[-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right],$$

  with $|\Sigma|$ being the determinant of Σ.
- The *cumulative distribution function* (CDF) of a 1-dim. Gaussian distribution[3],

$$\text{cdf}\,(x|\mu,\Sigma) := \int_{-\infty}^{x} \mathcal{N}(d\xi|\mu,\Sigma)\; d\xi,$$

  is abbreviated to cdf (x) for the standard case ($\mu = 0$, $\Sigma = 1$).

Probability Kernels For two measurable spaces $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$ and $(\mathbb{Y}, \mathscr{B}(\mathbb{Y}))$, a *(sub-)probability kernel* is a mapping $\mathbf{p}\colon \mathbb{X}\times\mathscr{B}(\mathbb{Y}) \to [0,1]$ such that $\mathbf{p}(x, \cdot)\colon \mathscr{B}(\mathbb{Y}) \to [0,1]$ is a (sub-)probability measure for each $x \in \mathbb{X}$, and $\mathbf{p}(\cdot, Y)\colon \mathbb{X} \to [0,1]$ is measurable for all $Y \in \mathscr{B}(\mathbb{Y})$; subsequently, the (sub-)probability measure associated to or *conditioned on* $x \in \mathbb{X}$ is denoted by $\mathbf{p}(\cdot\,|x)$.

2.2.2 Stochastic Control System

This section presents the mathematical details used to describe the dynamical behaviour of a target system and its corresponding controller.

Stochastic Dynamics Model Consider a system whose dynamical behaviour can be modelled by a discrete-time stochastic nonlinear model, perturbed by additive stochastic noise and model-parametric uncertainty:

$$\boldsymbol{\mathcal{M}}(\theta)\colon \begin{cases} x_{t+1} = f_\theta(x_t, u_t) + w_t, & w_t \sim \mathcal{N}(\cdot\,|0,\Sigma) \\ y_t \;\;\;\; = h(x_t) \end{cases}, \tag{2.1}$$

[3] The probability measure of a Gaussian distribution is a Gaussian measure.

with state $x_t \in \mathbb{X} \subset \mathbb{R}^{n_x}$, input $u_t \in \mathbb{U} \subset \mathbb{R}^{n_u}$, and observation $y_t \in \mathbb{Y} \subset \mathbb{R}^{n_y}$ at time step $t \in \mathbb{N}_0$. The functions $f\colon \mathbb{X} \times \mathbb{U} \times \Omega \to \mathbb{X}$ and $h\colon \mathbb{X} \to \mathbb{Y}$ specify, respectively, the parametrised state evolution — subject to an uncertain parametrisation $\theta \in \Omega$ from some parameter space Ω — and the observation map of the system. The additive process noise $w_t \in \mathbb{W}$ is an independent, identically distributed (i.i.d.) noise sequence with Gaussian distribution $w_t \sim \mathcal{N}(\cdot\,|0, \Sigma)$. Consequently, the noise is generally unbounded — unlike many existing approaches assuming noise with bounded support.

System Evolution The goal of this chapter is to design a controller for the system $\boldsymbol{\mathcal{M}}(\theta)$, which generates input sequences of the form $\mathbf{u} := u_0, u_1, u_2, \ldots$. Starting from an initial state $x_0 \in \mathbb{X}_0 \subset \mathbb{X}$, the system evolves along (finite) *executions*, i.e, sequences of states $\mathbf{x} = x_0, x_1, x_2, \ldots$ (respectively, $\mathbf{x}_N = x_0, x_1, x_2, \ldots, x_N$). In each execution, the consecutive state $x_{t+1} \in \mathbb{X}$ is obtained as a realisation $x_{t+1} = f_\theta(x_t, u_t) + w_t$, $w_t \sim \mathcal{N}(\cdot\,|0, \Sigma)$. The resulting *execution history* $(x_0, u_0, x_1, \ldots, u_{N-1}, x_N)$ grows with the number of steps N and takes values in the *history space* $\mathbb{H}_N := (\mathbb{X} \times \mathbb{U})^N \times \mathbb{X}$.

Control Policies

Controller A controller determines which inputs to apply, based on the evolution history of the system. Formally, a *control policy* $\boldsymbol{\eta} = (\eta_0, \eta_1, \eta_2, \ldots)$ is a sequence of universally measurable maps $\eta_t\colon \mathbb{H}_t \to \mathscr{P}(\mathbb{U})$, $t \in \mathbb{N}_0$ from execution histories to distributions over inputs. Two common subclasses of control policies are differentiated:

- A *Markov policy* $\boldsymbol{\rho} = (\rho_0, \rho_1, \rho_2, \ldots)$ restricts this mapping to depend only on the current state, i.e., $\rho_t\colon \mathbb{X} \to \mathscr{P}(\mathbb{U})$. A Markov policy is called *stationary*, if $\boldsymbol{\rho} = (\rho, \rho, \rho, \ldots)$ for some fixed map ρ.
- A *finite memory policy* augments the state with a memory state that evolves with the system history, e.g., tracking task-relevant information (e.g., whether a delivery robot has picked up a pizza or already delivered it). To this end, it first maps the finite state execution of the system to a finite set (memory), and then chooses the input similar to the Markov policy as a function of the system state and the memory state. This class of policies is needed for satisfying temporal specifications on the system executions.

In the remainder of this chapter, a controller for the model $\boldsymbol{\mathcal{M}}(\theta)$ in (2.1) is denoted by $\boldsymbol{C}$ — typically the deterministic implementation of a policy above — acting as a state-feedback law,

$$u_t = \boldsymbol{C}(x_t).$$

The closed-loop composition of $\boldsymbol{C}$ and $\boldsymbol{\mathcal{M}}(\theta)$ is written as $\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}}(\theta)$. Additional formal definitions and technical details can be found in the work of Haesaert and Soudjani (2021); Tabuada (2009).

2.2.3 Temporal Logic Specifications

Temporal logic specifications provide a natural and unambiguous semantic for specifying the desired system behaviour — used as an input to the control synthesis task. Such specifications are defined as formulas composed of atomic propositions and operators.

Atomic Propositions Let $AP := \{p_1, \ldots, p_L\}$ be a set of *atomic propositions* that defines an *alphabet* $\boldsymbol{\Sigma} := 2^{AP}$, where any *letter* $l \in \boldsymbol{\Sigma}$ is composed of a set of atomic propositions. An infinite string of letters forms a *word* $\boldsymbol{\omega} = l_0 l_1 l_2 \ldots \in \boldsymbol{\Sigma}^{\mathbb{N}_0}$. For any $i \in \mathbb{N}_0$, $\boldsymbol{\omega}_i = l_i l_{i+1} l_{i+2} \ldots$ denotes a suffix of $\boldsymbol{\omega}$.

This chapter considers the *co-safe* subset of *linear-time temporal logic* properties (scLTL) (Kupferman and Vardi, 2001). This subset consists of temporal logic formulas for which violation can be determined from a finite bad prefix of a word (cf. Fig. 2.2). Possible extensions are discussed in Section 2.7.

Definition 2.1 (Syntactically Co-Safe Linear Temporal Logic (scLTL)) scLTL formulas can be constructed according to the syntax Co-Safe LTL

$$\psi ::= \text{true} \mid p \mid \neg p \mid \psi_1 \vee \psi_2 \mid \psi_1 \wedge \psi_2 \mid \psi_1 \,\mathsf{U}\, \psi_2 \mid \bigcirc \psi,$$

where $p \in \text{AP}$ is an atomic proposition. The semantics of scLTL are defined recursively over $\boldsymbol{\omega}_i$ as follows:

- **Proposition:** $\boldsymbol{\omega}_i \models p$ iff $p \in l_i$;
- **And:** $\boldsymbol{\omega}_i \models \psi_1 \wedge \psi_2$ iff $(\boldsymbol{\omega}_i \models \psi_1) \wedge (\boldsymbol{\omega}_i \models \psi_2)$;
- **Or:** $\boldsymbol{\omega}_i \models \psi_1 \vee \psi_2$ iff $(\boldsymbol{\omega}_i \models \psi_1) \vee (\boldsymbol{\omega}_i \models \psi_2)$;
- **Until:** $\boldsymbol{\omega}_i \models (\psi_1 \,\mathsf{U}\, \psi_2)$ iff $\exists j \geq i$ s.t. $(\boldsymbol{\omega}_j \models \psi_2) \wedge (\boldsymbol{\omega}_t \models \psi_1), \forall t = i, \ldots, j-1$;
- **Next:** $\boldsymbol{\omega}_i \models \bigcirc\psi$ iff $\boldsymbol{\omega}_{i+1} \models \psi$.

The eventually operator $\Diamond\psi$ is used as a shorthand for $\text{true}\,\mathsf{U}\psi$. A word $\boldsymbol{\omega}$ *satisfies* a specification ψ if $\boldsymbol{\omega}_0$ does, i.e., $\boldsymbol{\omega} \models \psi$ iff $\boldsymbol{\omega}_0 \models \psi$.

Temporal logic specifications are defined over the output traces of a system. To this end, the system $\boldsymbol{\mathcal{M}}$ is equipped with a labelling function $\mathcal{L}\colon \mathbb{Y} \to \Sigma$ assigning a letter l to each system output y, thus itself producing an abstraction of the output domain of the system. Consequently, each output trace $\mathbf{y} = y_0, y_1, y_2, \ldots$ can be translated to a word $\boldsymbol{\omega} = \mathcal{L}(\mathbf{y}) = l_0, l_1, l_2, \ldots$. Fig. 2.2 depicts some basic example formulas.

A system satisfies the specification ψ with probability at least p_ψ if $\mathbb{E}[\mathcal{L}(\mathbf{y}) \models \psi] \geq p_\psi$, where $\mathbf{y}$ is any output trace emitted by $\boldsymbol{\mathcal{M}}$. When the labelling function $\mathcal{L}$ is known from the context, writing the equivalent $\mathbb{P}(\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}} \models \psi)$ emphasises that the output traces of $\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}}$ are used for checking the satisfaction. Similarly, $\mathbb{P}(\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}}(\theta) \models \psi)$ denotes the probability that the parametrised system $\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}}(\theta)$ satisfies ψ.

Remark 2.1 (Infinite-Horizon Formulas) The *formula horizon* of a specification ψ Formula Horizon

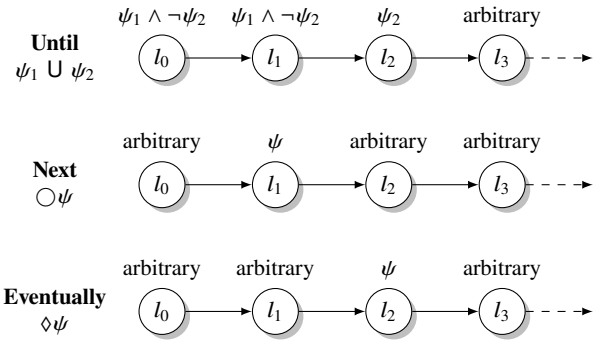


Fig. 2.2: Examples of satisfying words for basic scLTL formulas.

is the shortest prefix $\mathbf{y}_i := y_0, y_1, \ldots, y_i$ ($i \in \mathbb{N}_0 \cup \{\infty\}$) needed to assess the satisfaction of *any* output trace $\mathbf{y} = y_0, y_1, y_2, \ldots$ generated by the system (Farahani et al., 2019). Thus, any formula containing an eventually operator is by definition an *infinite-horizon* specification.

2.2.4 Problem Statement

The results of this chapter are built on the assumption that the unknown data-generating target system $\boldsymbol{\mathcal{M}}^*$ is of the parametrised form (2.1), i.e., $\boldsymbol{\mathcal{M}}^* := \boldsymbol{\mathcal{M}}(\theta^*)$, where the functions f and h are known but the underlying concrete parametrisation $\theta^* \in \Omega$ is unknown. The goal of Problem 2.1 is to design a controller $\boldsymbol{\mathcal{C}}$ such that the controlled system $\boldsymbol{\mathcal{C}} \bowtie \boldsymbol{\mathcal{M}}^*$ satisfies a given scLTL specification ψ, i.e, $\boldsymbol{\mathcal{C}} \bowtie \boldsymbol{\mathcal{M}}^* \models \psi$.

Data Collection To estimate the unknown parameters θ^* — and thus by extension the model $\boldsymbol{\mathcal{M}}^*$ —, data is collected from $\boldsymbol{\mathcal{M}}^*$ by sampling a finite number of state-input pairs $(x^i, u^i) \in \mathbb{X} \times \mathbb{U}$, $i \in \{1, \ldots, N\}$, and recording the corresponding successor $x_+^i \in \mathbb{X}$. Thus, the resulting data $\mathcal{D}_N$ is of the form

$$\begin{aligned} \mathcal{D}_N &:= \left\{(x^1, u^1, x_+^1), \ldots, (x^N, u^N, x_+^N)\right\}, \quad \text{where} \\ x_+^i &= f_{\theta^*}(x^i, u^i) + w^i, \quad w^i \sim \mathcal{N}(\,\cdot\,|0, \Sigma), \quad i = 1, \ldots, N. \end{aligned} \tag{2.2}$$

Note that this assumes access to observations of the full state of $\boldsymbol{\mathcal{M}}^*$.

Based on the data $\mathcal{D}_N$, the goal is to design a controller that ensures the satisfaction of ψ with at least probability $p_{\psi,N}$ — where the additional subscript N indicates the dependence of the bound on the given amount of data. As the true parameters θ^* can generally only be inferred approximately, a key requirement of the developed methodology is that the controller must *not* depend on the unknown θ^* directly. To this end, Problem 2.1 can be decomposed into the following two subproblems.

Subproblem 2.1 (System Identification) Based on data $\mathcal{D}_N$ in (2.2) and the parametrised model structure (2.1) with known f and h, construct a set $\Theta \subset \Omega$ containing the true parameters θ^* with a given probability $1 - \alpha \in (0, 1)$, i.e.,

$$\mathbb{P}(\theta^* \in \Theta) \geq 1 - \alpha.$$

Data-driven parameter identification methods provide a solution to Subproblem 2.1. The resulting parameter set Θ will also be referred to as a *credible set*.

Remark 2.2 (Assumptions) This chapter considers Bayesian linear regression to identify the parameter set Θ, which is limited to systems with linear dependence on the unknown parameters θ and additive noise (see Equation 2.3). As the proposed coupling approach is generally applicable, these assumptions can be relaxed by choosing any suitable alternative parameter identification method providing finite-sample guarantees. For simplicity, the exposition is limited to parametric uncertainty in the deterministic part of the dynamics only (f_θ in Equation 2.1), and the noise is assumed Gaussian. Chapter 3 shows how these assumptions can be lifted.

The parameter set Θ identified in Subproblem 2.1 induces a parametrised set of models $\{\boldsymbol{\mathcal{M}}(\theta),\ \theta \in \Theta\}$. Thus, Problem 2.1 can be solved by designing a controller that is uniformly valid for the entire set of models, formulated as follows.

Subproblem 2.2 (Robust Synthesis) For a given scLTL specification ψ and threshold $p_{\psi,N} \in (0, 1)$, design a controller $\boldsymbol{C}$ that robustly satisfies ψ uniformly for the set of models $\{\boldsymbol{\mathcal{M}}(\theta),\ \theta \in \Theta\}$ with probability at least $p_{\psi,N}$, that is,

$$\mathbb{P}\left(\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}}(\theta) \models \psi\right) \geq p_{\psi,N}, \quad \forall \theta \in \Theta.$$

Haesaert and Soudjani (2021) study controller synthesis for stochastic models via probabilistic simulation relations. Although these simulation relations can relate a nominal model to a set of parametrised models $\boldsymbol{\mathcal{M}}(\theta)$, these relations would lead to a control refinement that is still dependent on the concrete parameters θ^*. Therefore, this approach is unfit to solve Subproblem 2.2, since θ^* is unknown.

Key Idea As one of the main results of this chapter, a parameter-independent control refinement is established by developing a novel simulation relation based on a sub-probability coupling (see Section 2.4), synthesising a single controller uniformly robust for all $\theta \in \Theta$.

Fig. 2.3 summarises the steps of the approach developed in this chapter, indicating their contribution to solving Subproblem 2.1 and Subproblem 2.2. The overview already reveals that, as more data becomes available, the nominal model $\widehat{\boldsymbol{\mathcal{M}}}$ can typically remain unchanged — though it may be refined for improved accuracy or conservatism, if desired. In contrast to IMDP-based approaches, incorporating new

data in the presented framework only requires re-running the synthesis step to update the controller.

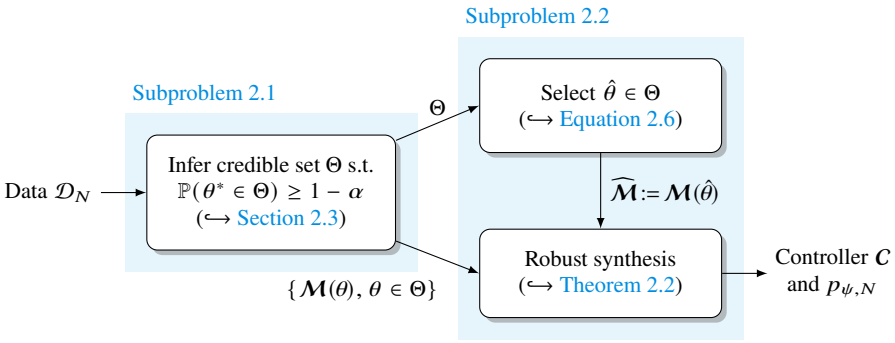


Fig. 2.3: Overview of the proposed Bayesian robust synthesis pipeline: Subproblem 2.1 identifies a credible set of models; Subproblem 2.2 selects a nominal model $\widehat{\mathcal{M}}$ and synthesises a robust controller C (Fig. 2.12 zooms in on this part).

2.3 Data-Driven System Identification via Bayesian Linear Regression

This section presents a solution to Subproblem 2.1 for the subset of nonlinear systems (2.1) linear in the unknown parameters $\theta^\top = [\theta_1, \ldots, \theta_{n_x}]^\top \in \Omega = \mathbb{R}^{n_x \times m}$, i.e.,

$$\mathcal{M}(\theta): \begin{cases} x_{t+1} = \theta^\top f(x_t, u_t) + w_t \\ y_t \quad = h(x_t) \end{cases}, \tag{2.3}$$

with state $x_t \in \mathbb{R}^{n_x}$, i.i.d. zero-mean Gaussian noise w_t with a given covariance matrix $\Sigma \in \mathbb{R}^{n_x \times n_x}$, and known functions $f: \mathbb{X} \times \mathbb{U} \to \mathbb{R}^m$ and $h: \mathbb{X} \to \mathbb{Y}$. Note that $f := [f_1, \ldots, f_m]^\top$ can be a comprehensive library of nonlinear functions, thus, a plethora of systems can be expressed in the form of (2.3).

Bayesian Linear Regression

Based on the state-input data $\mathcal{D}_N$ in (2.2), *Bayesian linear regression* (e.g., Bishop, 2006, Section 3.3) learns a *parameter estimate* $\hat{\theta} \in \Omega$ that approximates the true unknown parameters θ^* and constructs a *credible set* $\Theta \subset \Omega$ that contains θ^* with a given confidence $1 - \alpha \in (0, 1)$.

Credible Set

Definition 2.2 (Credible Set) A set Θ is called a *credible set* of θ^* with confidence level $1 - \alpha$ if $\mathbb{P}(\theta^* \in \Theta) \geq 1 - \alpha$.

Fig. 2.4 visualises how the credible set Θ (in blue) contracts and the estimate $\hat{\theta}$ moves towards the unknown θ^* as more data is introduced (indicated by '$+\mathcal{D}$').

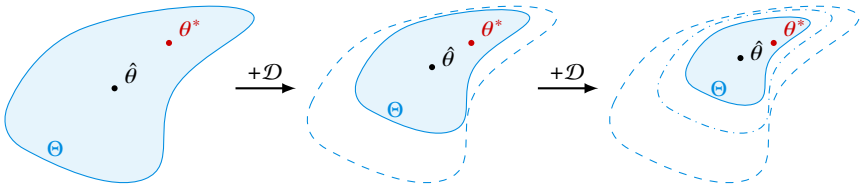


Fig. 2.4: The unknown true parametrisation θ^* — approximated by an estimate $\hat{\theta}$ — is contained within the credible set Θ (in blue) with a confidence of $1-\alpha$. The set Θ contracts as more data is introduced (indicated by '$+\mathcal{D}$').

Remark 2.3 (Credible vs Confidence Sets) Credible sets and *confidence sets* are two related notions stemming from two fundamentally different statistical paradigms. As a product of Bayesian statistics, a credible set of a random variable represents a set of plausible values characterising the uncertainty in the estimation of the random variable up to a desired confidence threshold, i.e., a fixed quantile of the posterior distribution that contains a prescribed fraction of the posterior mass. A confidence set, on the other hand, is based on frequentist theory and is computed based on the frequency of random observations of a fixed true parameter. Bayesian credible sets treat their bounds as fixed and the estimated parameter as a random variable, whereas frequentist confidence sets treat their bounds as random variables and include the true fixed parameter with certain probability. In the following, with an abuse of terminology, "true value" of a random variable is used in the context of credible sets. The reader is referred to Jaynes and Kempthorne (1976) for a thorough exposition.

Remark 2.4 The sub-simulation relations presented in subsequent sections are not restricted to the class of systems in (2.3) and can be used in conjunction with any other estimation method that produces a credible set Θ with $\hat{\theta} \in \Theta$.

2.3.1 Parameter Estimate

To obtain an estimate of the parameters $\theta = [\theta_1, \ldots, \theta_{n_x}] \in \mathbb{R}^{m \times n_x}$, a prior probability distribution is placed on the uncertain weight vector $\bar{\theta} := [\theta_1; \ldots; \theta_{n_x}] \in \mathbb{R}^{mn_x}$:

$$p(\bar{\theta}) = \mathcal{N}(\bar{\theta} | \mu_0, \Sigma_0),$$

where the mean $\mu_0 \in \mathbb{R}^{mn_x}$ and covariance $\Sigma_0 \in \mathbb{R}^{mn_x \times mn_x}$ are either based on prior knowledge of the likelihood of θ, if available, or set to 0 and $\sigma_0^2 I_{n_x m}$, respectively, with σ_0^2 a large number. Based on the prior distributions, Bayesian linear regression computes the posterior distribution of the unknown parameters given data $\mathcal{D}_N$ in (2.2). For this, define the state matrix

$$X_+ := [x_{+,1}^1, \ldots, x_{+,1}^N, \ldots, x_{+,n_x}^1, \ldots, x_{+,n_x}^N]^\top$$

and the *design matrix* Φ of the form

$$\Phi := \begin{bmatrix} f_1(x^1, u^1) & \dots & f_m(x^1, u^1) \\ \vdots & \ddots & \vdots \\ f_1(x^N, u^N) & \dots & f_m(x^N, u^N) \end{bmatrix}.$$

With this, the posterior distribution $p(\bar{\theta}|\mathcal{D}_N)$ is given by

$$p(\bar{\theta}|\mathcal{D}_N) = \mathcal{N}(\bar{\theta}|\mu_N, \Sigma_N), \quad \text{with} \quad \begin{cases} \mu_N & := \Sigma_N(\Sigma_0^{-1}\mu_0 + (\Sigma^{-1} \otimes \Phi^\top)X_+) \\ \Sigma_N^{-1} & := \Sigma_0^{-1} + \Sigma^{-1} \otimes \Phi^\top\Phi \end{cases}, \tag{2.4}$$

where $\otimes$ denotes the *Kronecker product* (Bishop, 2006). Note that the parameter estimate $\hat{\theta}$ is set to the mean, i.e., the matrix $\hat{\theta} = [\mu_{N,1}, \dots, \mu_{N,n_x}] \in \mathbb{R}^{m \times n_x}$ is constructed by concatenating the vectors $\mu_{N,i} \in \mathbb{R}^m$, $i \in \{1, \dots, n_x\}$.

2.3.2 Credible Set

For a desired confidence bound $1 - \alpha \in (0, 1)$, the corresponding credible set

$$\Theta = \left\{\theta \in \mathbb{R}^{m \times n_x} \,\middle|\, (\bar{\theta} - \mu_N)^\top \Sigma_N^{-1} (\bar{\theta} - \mu_N) \leq n_x \cdot \chi^{-1}(1 - \alpha | n_x)\right\} \tag{2.5}$$

is obtained, where $\chi^{-1}(p|\nu) := \inf\{\zeta : p \leq \chi(\zeta|\nu)\}$ denotes the inverse cumulative distribution function or *quantile function* of the *chi-squared* distribution

$$\chi(\zeta|\nu) := \int_0^\zeta \frac{t^{(\nu-2)/2} e^{-t/2}}{2^{\nu/2}\Gamma(\nu/2)} dt$$

with ν degrees of freedom and the *Gamma function* $\Gamma(\nu) := \int_0^\infty t^{\nu-1} e^{-t} dt$ (Chew, 1966).

Solution Subproblem 2.1 is solved by computing the credible set Θ in (2.5).

The next section works towards answering Subproblem 2.2 via an abstraction-based control synthesis scheme. It introduces a new class of simulation relations that enables parameter-independent control refinement from a nominal model to the original concrete model belonging to a parametrised class.

2.4 Control Refinement via Sub-Simulation Relations

The concept of simulation relations is employed to answer Subproblem 2.2. In essence, simulation relations allow to compute a controller for a nominal model $\widehat{\mathcal{M}}$

and specification ψ, and transfer the guarantees of satisfying ψ to the original system $\mathcal{M}$ by quantifying their similarity. Based on the results of Section 2.3, this section constructs a nominal model $\widehat{\mathcal{M}}$, based on which a controller will be designed:

$$\widehat{\mathcal{M}} : \begin{cases} \hat{x}_{t+1} = \hat{f}(\hat{x}_t, \hat{u}_t) + \hat{w}_t, & \hat{w}_t \sim \mathcal{N}(\cdot\,|0, \Sigma) \\ \hat{y}_t \quad = \hat{h}(\hat{x}_t) \end{cases} . \tag{2.6}$$

This nominal model could, for example, be $\mathcal{M}(\theta)$ but for a given valuation of parameters $\theta = \hat{\theta}$, i.e., $\widehat{\mathcal{M}} := \mathcal{M}(\hat{\theta})$ if $\hat{f} \equiv f_{\hat{\theta}}$ and $\hat{h} \equiv h$, or any other model constructed by space reduction or discretisation. Furthermore, a set of models $\{\mathcal{M}(\theta),\ \theta \in \Theta\}$ is constructed using the credible set Θ computed via the approach in Section 2.3 or any alternative approach that solves Subproblem 2.1.

The systems $\widehat{\mathcal{M}}$ in (2.6) and $\mathcal{M}(\theta)$ with a given fixed θ can equivalently be modelled as *general MDPs*, studied previously for formal verification and synthesis of controllers (Haesaert et al., 2017). Given $\hat{x}_t$ and $\hat{u}_t$, the stochastic state transitions of $\widehat{\mathcal{M}}$ can be modelled with a probability kernel $\hat{\mathbf{t}}(\cdot\,|\hat{x}_t, \hat{u}_t)$ that is computed based on $\hat{f}$ and $\mathcal{N}(\cdot\,|0, \Sigma)$ (similarly for $\mathcal{M}(\theta)$). This leads to the representation of the systems as general MDPs, which extend generic MDPs by a metric output space.

Definition 2.3 (Metric Space) A *metric space* $(\mathbb{Y}, \mathbf{d}_{\mathbb{Y}})$ is a set $\mathbb{Y}$ with a *metric* $\mathbf{d}_{\mathbb{Y}} : \mathbb{Y} \times \mathbb{Y} \to \mathbb{R}_{\geq 0}$ that imposes a notion of distance between elements of $\mathbb{Y}$. Formally, a metric $\mathbf{d}_{\mathbb{Y}}$ is a non-negative function satisfying the following conditions for all points $y_1, y_2, y_3 \in \mathbb{Y}$:

(a) $\mathbf{d}_{\mathbb{Y}}(y_1, y_2) = 0$ iff $y_1 = y_2$;
(b) $\mathbf{d}_{\mathbb{Y}}(y_1, y_2) = \mathbf{d}_{\mathbb{Y}}(y_2, y_1)$; and
(c) $\mathbf{d}_{\mathbb{Y}}(y_1, y_3) \leq \mathbf{d}_{\mathbb{Y}}(y_1, y_2) + \mathbf{d}_{\mathbb{Y}}(y_2, y_3)$.

A common example of a metric is the norm function $\mathbf{d}_{\mathbb{Y}}(y_1, y_2) := \|y_1 - y_2\|$. With this, a general MDP is defined as follows.

Definition 2.4 (General Markov Decision Process (gMDP)) A *gMDP* is a tuple $\mathcal{M} = (\mathbb{X}, \mathbb{X}_0, \mathbb{U}, \mathbf{t}, h, \mathbb{Y})$, with General MDP

- $\mathbb{X} \subset \mathbb{R}^{n_x}$, the state space, with states $x \in \mathbb{X}$;
- $\mathbb{X}_0 \subset \mathbb{X}$, the set of initial states $x_0 \in \mathbb{X}_0$;
- $\mathbb{U} \subset \mathbb{R}^{n_u}$, the input space, with control inputs $u \in \mathbb{U}$;
- $\mathbf{t} \colon \mathbb{X} \times \mathbb{U} \times \mathscr{B}(\mathbb{X}) \to [0, 1]$, the transition probability kernel assigning to each state-input pair $(x, u) \in \mathbb{X} \times \mathbb{U}$ a probability measure $\mathbf{t}(\cdot\,|x, u)$ on $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$;
- $\mathbb{Y} \subset \mathbb{R}^{n_y}$, the output space, decorated with a metric $\mathbf{d}_{\mathbb{Y}}$; and
- $h \colon \mathbb{X} \to \mathbb{Y}$, the output map.

The definition of a regular MDP is recovered by eliminating the output space $\mathbb{Y}$ and the output map h, i.e., $\mathcal{M} = (\mathbb{X}, \mathbb{X}_0, \mathbb{U}, \mathbf{t})$ would be an MDP.

In each execution of a gMDP, the consecutive state $x_{t+1} \in \mathbb{X}$ is obtained as a realisation $x_{t+1} \sim \mathbf{t}(\cdot\,|x_t, u_t)$ of the controlled transition (probability) kernel. Note that for a parametrised gMDP $\mathcal{M}(\theta)$ its transition kernel also depends on θ, denoted

as $\mathbf{t}_\theta(\,\cdot\,|x_t, u_t)$. As an example, the transition kernel of the parametrised system in (2.1) is given by

$$\mathbf{t}_\theta(dx_+|x,u) = \int_{\mathbb{W}} \delta_{f_\theta(x,u)+w}(dx_+)\, \mathcal{N}(dw|0,\Sigma).$$

Output traces of the gMDP are obtained as mappings of (finite) state executions, namely $\mathbf{y} := y_0, y_1, y_2, \ldots$ (respectively, $\mathbf{y}_N := y_0, y_1, y_2, \ldots, y_N$), where $y_t = h(x_t)$.

Analogous to representing systems as gMDPs, satisfaction of an scLTL specification ψ can be checked using its alternative representation as a *deterministic finite-state automaton* $\mathcal{A}_\psi$.

Deterministic Finite-State Automaton

Definition 2.5 (Deterministic Finite-State Automaton (DFA)) A DFA is a tuple $\mathcal{A} = (Q, q_0, \mathbf{\Sigma}, F, \tau)$, where Q is the finite set of locations of the DFA, $q_0 \in Q$ is the initial location, $\mathbf{\Sigma}$ the finite alphabet, $F \subset Q$ is the set of accepting locations, and $\tau\colon Q \times \mathbf{\Sigma} \to Q$ is the transition function.

For any $n \in \mathbb{N}$, a word $\boldsymbol{\omega}_n = l_0 l_1 l_2 \ldots l_{n-1} \in \mathbf{\Sigma}^n$ is accepted by a DFA $\mathcal{A}$ if there exists a finite run $q = (q_0, \ldots, q_n) \in Q^{n+1}$ such that q_0 is the initial location, where $q_{i+1} = \tau(q_i, l_i)$ for all $0 \le i \le n$, and $q_n \in F$. See Fig. 2.9 for an example DFA.

2.4.1 Control Refinement

Control Refinement

Having defined the concepts of gMDPs and DFAs, consider a set of models $\{\boldsymbol{\mathcal{M}}(\theta),\ \theta \in \Theta\}$ and suppose that an abstract (nominal) model $\widehat{\boldsymbol{\mathcal{M}}}$ has been chosen. Based on $\widehat{\boldsymbol{\mathcal{M}}}$, the objective is to design a *single* controller alongside the probability of satisfying ψ over all models $\boldsymbol{\mathcal{M}}(\theta)$ in the set of models. In this context, the act of adapting a controller for $\widehat{\boldsymbol{\mathcal{M}}}$ to a controller for $\boldsymbol{\mathcal{M}}(\theta)$ is called *control refinement*.

This section first formalises the notion of a *state mapping* and an *interface function*, that together form the control refinement. Then, conditions are investigated under which a single controller for $\widehat{\boldsymbol{\mathcal{M}}}$ can be refined to a controller for all $\boldsymbol{\mathcal{M}}(\theta)$ *independent* of the concrete valuation of θ. Note that this is the central contribution of this work, since instead of synthesising a *parametrised* controller $\boldsymbol{\mathcal{C}}(\theta)$ for $\boldsymbol{\mathcal{M}}(\theta)$ a single *parameter-independent* controller $\boldsymbol{\mathcal{C}}$ is synthesised that works for all models $\boldsymbol{\mathcal{M}}(\theta)$ in the set of models $\{\boldsymbol{\mathcal{M}}(\theta),\ \theta \in \Theta\}$. This leads to the novel concept of *sub-probability* couplings and simulation relations.

Consider the nominal model $\widehat{\boldsymbol{\mathcal{M}}} = (\hat{\mathbb{X}}, \hat{\mathbb{X}}_0, \hat{\mathbb{U}}, \hat{\mathbf{t}}, \hat{h}, \hat{\mathbb{Y}})$, the concrete model $\boldsymbol{\mathcal{M}}(\theta) = (\mathbb{X}, \mathbb{X}_0, \mathbb{U}, \mathbf{t}_\theta, h, \mathbb{Y})$, and a nominal controller $\widehat{\boldsymbol{\mathcal{C}}}$ for $\widehat{\boldsymbol{\mathcal{M}}}$. To refine the nominal $\widehat{\boldsymbol{\mathcal{C}}}$ to a controller for $\boldsymbol{\mathcal{M}}(\theta)$, a pair of interfacing functions are defined:

- A *state mapping*, translating concrete states $x \in \mathbb{X}$ to nominal states $\hat{x} \in \hat{\mathbb{X}}$; and
- An *interface function*, refining nominal control inputs $\hat{u} \in \hat{\mathbb{U}}$ to concrete control inputs $u \in \mathbb{U}$.

Interface Function

Interface Function An interface function (Girard and Pappas, 2009, 2011) is defined as a Borel measurable probability kernel $\mathbf{i}\colon \hat{\mathbb{X}}\times\mathbb{X}\times\hat{\mathbb{U}}\times\mathscr{B}(\mathbb{U}) \to [0,1]$ refining any nominal input $\hat{u}_t$ of $\widehat{\boldsymbol{\mathcal{M}}}$ to a concrete input u_t for $\boldsymbol{\mathcal{M}}(\theta)$ via

$$u_t \sim \mathbf{i}(\,\cdot\,|\hat{x}_t, x_t, \hat{u}_t). \tag{2.7}$$

State Mapping

State Mapping In its most general form, a state mapping is defined as a probability kernel $\mathbf{s}\colon \hat{\mathbb{X}}\times\mathbb{X}\times\mathbb{X}\times\hat{\mathbb{U}} \to \mathscr{P}(\hat{\mathbb{X}})$ mapping the current state $\hat{x}_t$ and input $\hat{u}_t$ of the nominal model $\widehat{\boldsymbol{\mathcal{M}}}$ to the distribution of the next state $\hat{x}_{t+1}$ via

$$\hat{x}_{t+1} \sim \mathbf{s}(\,\cdot\,|\hat{x}_t, x_t, x_{t+1}, \hat{u}_t).$$

This state mapping is coupled with the concrete model $\boldsymbol{\mathcal{M}}(\theta)$ via its states x_t, x_{t+1} and depends implicitly on u_t through the interface function (2.7).

Fig. 2.5 illustrates how the nominal controller $\widehat{\boldsymbol{\mathcal{C}}}$ produces a control input $\hat{u}_t$ as a function of the nominal state $\hat{x}_t$. Based on the state mapping $\mathbf{s}$ and the interface

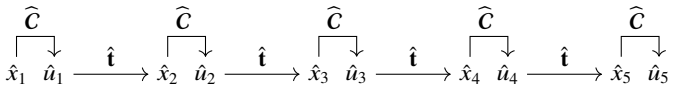


Fig. 2.5: Controller $\widehat{\boldsymbol{\mathcal{C}}}$ of the nominal model $\widehat{\boldsymbol{\mathcal{M}}}$.

function $\mathbf{i}$, the nominal controller $\widehat{\boldsymbol{\mathcal{C}}}$ can be refined to obtain a concrete controller $\boldsymbol{\mathcal{C}}$ for $\boldsymbol{\mathcal{M}}(\theta)$, as depicted in Fig. 2.6. In this figure, the states x_t of the concrete model $\boldsymbol{\mathcal{M}}(\theta)$ are mapped to the nominal states $\hat{x}_t$ via the state mapping $\mathbf{s}$, the nominal control inputs $\hat{u}_t$ are obtained using $\widehat{\boldsymbol{\mathcal{C}}}$ and $\hat{x}_t$, and these inputs are then refined to control inputs u_t for the concrete model $\boldsymbol{\mathcal{M}}(\theta)$ using the interface function $\mathbf{i}$.

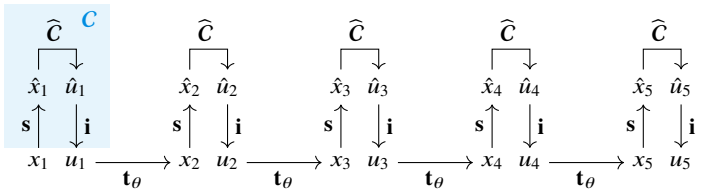


Fig. 2.6: Control refinement using the nominal controller $\widehat{\boldsymbol{\mathcal{C}}}$ to control the concrete model $\boldsymbol{\mathcal{M}}(\theta)$.

Valid Refinement

Subproblem 2.2 begs the question under which conditions on the interface function and the models a parameter-independent refinement is *valid*, that is, preserving the satisfaction probability for all $\boldsymbol{\mathcal{M}}(\theta)$. This is addressed in the next subsection.

2.4.2 Valid Control Refinement and Sub-Simulation Relations

Before diving into the definition of a valid control refinement that can be applied to models with parametric uncertainty, a relaxed version of approximate simulation relations (Haesaert et al., 2017) based on sub-probability couplings is introduced.

To this end, the formal definition of a relation is needed (Himmelberg, 1975), which will be used to associate nominal and concrete states.

Relation

Definition 2.6 (Relation) Given sets A, B, a measurable *relation* $\mathcal{R} \subset \mathscr{B}(A \times B)$ is a set of ordered pairs from $\mathscr{B}(A \times B)$. The relation $\mathcal{R}$ *relates* $x \in \mathscr{B}(A)$ and $y \in \mathscr{B}(B)$ if $(x, y) \in \mathcal{R}$.

Before providing the definition of a sub-simulation relation, the following instrumental definition of a *sub-probability coupling* between two probability measures with respect to a given relation $\mathcal{R} \subset \mathscr{B}(\hat{\mathbb{X}} \times \mathbb{X})$ is given.

Sub-Probability Coupling

Definition 2.7 (Sub-Probability Coupling) Given $p \in \mathscr{P}(\mathbb{X})$, $\hat{p} \in \mathscr{P}(\hat{\mathbb{X}})$, $\mathcal{R} \subset \mathscr{B}(\hat{\mathbb{X}} \times \mathbb{X})$, and a value $\delta \in [0, 1]$, a sub-probability measure v over $(\hat{\mathbb{X}} \times \mathbb{X}, \mathscr{B}(\hat{\mathbb{X}} \times \mathbb{X}))$ with $v(\hat{\mathbb{X}} \times \mathbb{X}) \geq 1 - \delta$ is a *sub-probability coupling* of $\hat{p}$ and p over $\mathcal{R}$ if

(a) the probability mass of v is entirely located on $\mathcal{R}$, i.e., $v(\hat{\mathbb{X}} \times \mathbb{X}) = v(\mathcal{R})$;
(b) for all measurable sets $A \subset \hat{\mathbb{X}}$, it holds that $v(A \times \mathbb{X}) \leq \hat{p}(A)$; and
(c) for all measurable sets $A \subset \mathbb{X}$, it holds that $v(\hat{\mathbb{X}} \times A) \leq p(A)$.

Intuitively, conditions (b)–(c) state that a sub-probability coupling v of $\hat{p}$ and p lives on the product space $\hat{\mathbb{X}} \times \mathbb{X}$, and its marginals minorise the measures $\hat{p}$ and p, respectively. According to condition (a), v is fully supported on the subset $\mathcal{R}$, i.e., it captures only the probability mass that $\hat{p}$ and p mutually place on $\mathcal{R}$. Note that condition (a) furthermore implies that this probability mass is $v(\mathcal{R}) \geq 1 - \delta$.

"Full" Probability Coupling

Proposition 2.1 (Recovering Probability Coupling) *A sub-probability coupling v as in Definition 2.7 can be completed to obtain a probability coupling $\bar{v}\colon \mathscr{B}(\hat{\mathbb{X}} \times \mathbb{X}) \to [0, 1]$ that couples the probability measures $\hat{p}$ and p:*

$$\bar{v}(d\hat{x}_+ \times dx_+) = v(d\hat{x}_+ \times dx_+) + \frac{\big(p(dx_+) - v(\hat{\mathbb{X}} \times dx_+)\big)\big(\hat{p}(d\hat{x}_+) - v(d\hat{x}_+ \times \mathbb{X})\big)}{1 - v(\mathcal{R})}. \tag{2.8}$$

*In other words, $\bar{v}$ in (2.8) satisfies the conditions of a δ-*lifting *(Haesaert et al., 2017, Definition 8), which corresponds to a "full" probability coupling, for which $\bar{v}(\hat{\mathbb{X}} \times \mathbb{X}) = 1$ and $\bar{v}(\mathcal{R}) \geq 1 - \delta$.*

The first term of the coupling (2.8) puts only weight on $\mathcal{R}$, whilst the second term assigns the remaining probability mass in an independent fashion. Note that the factor $1/(1 - v(\mathcal{R}))$ normalises the remaining probability mass such that $\bar{v}$ satisfies the coupling properties $\bar{v}(d\hat{x}_+ \times \mathbb{X}) = \hat{p}(d\hat{x}_+)$ and $\bar{v}(\hat{\mathbb{X}} \times dx_+) = p(dx_+)$ (cf. δ-lifting by Haesaert et al. (2017, Definition 8)).

Proof. First, it is shown that $\bar{v}$ is a probability measure by showing that $\bar{v}$ is fully supported on $\hat{\mathbb{X}} \times \mathbb{X}$, i.e., $\bar{v}(\hat{\mathbb{X}} \times \mathbb{X}) = 1$:

$$\bar{v}(\hat{\mathbb{X}} \times \mathbb{X}) = v(\hat{\mathbb{X}} \times \mathbb{X}) + \frac{\big(1 - v(\hat{\mathbb{X}} \times \mathbb{X})\big)^2}{1 - v(\mathcal{R})} = 1,$$

which utilises property (a) of the sub-probability coupling (Definition 2.7). The coupling properties are easily derived by calculating the marginals of $\bar{v}$, i.e.,

$$\begin{aligned}
\bar{v}(d\hat{x}_+ \times \mathbb{X}) &= v(d\hat{x}_+ \times \mathbb{X}) + \frac{\big(1 - v(\hat{\mathbb{X}} \times \mathbb{X})\big)\big(\hat{p}(d\hat{x}_+) - v(d\hat{x}_+ \times \mathbb{X})\big)}{1 - v(\mathcal{R})}, \\
&= \hat{p}(d\hat{x}_+), \\
\bar{v}(\hat{\mathbb{X}} \times dx_+) &= v(\hat{\mathbb{X}} \times dx_+) + \frac{\big(p(dx_+) - v(\hat{\mathbb{X}} \times dx_+)\big)\big(1 - v(\hat{\mathbb{X}} \times \mathbb{X})\big)}{1 - v(\mathcal{R})}, \\
&= p(dx_+).
\end{aligned}$$

Thus, $\bar{v}$ satisfies all the conditions of Definition 8 by Haesaert et al. (2017). □

Remark 2.5 (Parameter-Dependent Coupling) For the case where p is dependent on a parametrisation θ, the corresponding sub-probability coupling v_θ may likewise depend on θ. Furthermore, although Definition 2.7 defines a sub-probability coupling v as a probability measure over $\mathbb{X} \times \hat{\mathbb{X}}$, in the remainder, it is often used in its kernel form $\boldsymbol{v}\colon \hat{\mathbb{X}} \times \mathbb{X} \times \hat{\mathbb{U}} \times \mathscr{B}(\hat{\mathbb{X}} \times \mathbb{X}) \to [0, 1]$. A concrete probability measure v is obtained from $\boldsymbol{v}$ for a fixed choice of $(\hat{x}, x, \hat{u})$ (see Table 2.1).

Table 2.1: Notation used for the transition kernels and associated measures of the different systems.

Description	Model	Transition Kernel	Associated Measure
Concrete system	$\boldsymbol{\mathcal{M}}(\theta)$	$\mathbf{t}_\theta(dx_+ \mid x, u)$	$t_\theta(dx_+)$
Nominal system	$\widehat{\boldsymbol{\mathcal{M}}}$	$\hat{\mathbf{t}}(d\hat{x}_+ \mid \hat{x}, \hat{u})$	$\hat{t}(d\hat{x}_+)$
Composed system	$\widehat{\boldsymbol{\mathcal{M}}} \bowtie \boldsymbol{\mathcal{M}}(\theta)$	$\mathbf{t}^{\times}_\theta(d\hat{x}_+ \times dx_+ \mid \hat{x}, x, \hat{u})$	$t^{\times}_\theta(d\hat{x}_+ \times dx_+)$
Sub-probability coupling		$\boldsymbol{v}_\theta(d\hat{x}_+ \times dx_+ \mid \hat{x}, x, \hat{u})$	$v_\theta(d\hat{x}_+ \times dx_+)$

The previous definitions are needed to formalise (ε, δ)-*sub-simulation relations*, which are used to quantify the similarity of stochastic systems, where ε indicates the error in the output mapping and δ quantifies the closeness in the probabilistic evolution of the two systems.

Definition 2.8 ((ε, δ)-Sub-Simulation Relation (SSR)) Consider two gMDPs $\boldsymbol{\mathcal{M}} = (\mathbb{X}, \mathbb{X}_0, \mathbb{U}, \mathbf{t}, h, \mathbb{Y})$ and $\widehat{\boldsymbol{\mathcal{M}}} = (\hat{\mathbb{X}}, \hat{\mathbb{X}}_0, \hat{\mathbb{U}}, \hat{\mathbf{t}}, \hat{h}, \mathbb{Y})$, a relation $\mathcal{R} \subset \mathscr{B}(\hat{\mathbb{X}} \times$

Sub-Simulation Relation

$\mathbb{X}$), and an interface function $\mathbf{i}\colon \hat{\mathbb{X}} \times \mathbb{X} \times \hat{\mathbb{U}} \times \mathscr{B}(\mathbb{U}) \to [0,1]$. If there exists a sub-probability kernel $\boldsymbol{v}\colon \hat{\mathbb{X}} \times \mathbb{X} \times \hat{\mathbb{U}} \times \mathscr{B}(\hat{\mathbb{X}} \times \mathbb{X}) \to [0,1]$ such that

(a) $\forall x_0 \in \mathbb{X}_0, \exists \hat{x}_0 \in \hat{\mathbb{X}}_0\colon\ (\hat{x}_0, x_0) \in \mathcal{R}$;
(b) $\forall (\hat{x}, x) \in \mathcal{R}, \forall \hat{u} \in \hat{\mathbb{U}}\colon\ \boldsymbol{v}(d\hat{x}_+ \times dx_+ | \hat{x}, x, \hat{u})$ is a sub-probability coupling of $\hat{\mathbf{t}}(d\hat{x}_+|\hat{x},\hat{u})$ and $\mathbf{t}(dx_+|x, \mathbf{i}(u|\hat{x},x,\hat{u}))$ over $\mathcal{R}$ with respect to δ (see Definition 2.7);
(c) $\forall (\hat{x}, x) \in \mathcal{R}\colon\ \mathbf{d}_{\mathbb{Y}}(\hat{h}(\hat{x}), h(x)) \leq \varepsilon$;

then $\widehat{\boldsymbol{\mathcal{M}}}$ is in an (ε,δ)-SSR with $\boldsymbol{\mathcal{M}}$, denoted by $\widehat{\boldsymbol{\mathcal{M}}} \preceq^{\delta}_{\varepsilon} \boldsymbol{\mathcal{M}}$.

To illustrate the SSR concept, consider two gMDPs $\boldsymbol{\mathcal{M}}$ and $\widehat{\boldsymbol{\mathcal{M}}}$ in an SSR $\widehat{\boldsymbol{\mathcal{M}}} \preceq^{\delta}_{\varepsilon} \boldsymbol{\mathcal{M}}$. Then, Definition 2.8 places three conditions on the composed system $\widehat{\boldsymbol{\mathcal{M}}} \bowtie \boldsymbol{\mathcal{M}}$ evolving on the product space $\hat{\mathbb{X}} \times \mathbb{X}$. Note that $\mathcal{R}$ defines a subset of $\hat{\mathbb{X}} \times \mathbb{X}$. According to condition (a), both systems start in $\mathcal{R}$ upon initialisation. In any subsequent time step, once in $\mathcal{R}$, condition (b) certifies that the systems remain in $\mathcal{R}$ for the next time step with a probability of at least $(1-\delta)$ for all control inputs $\hat{u} \in \hat{\mathbb{U}}$ and $u \sim \mathbf{i}(u|\hat{x},x,\hat{u})$. Finally, given the two systems are in $\mathcal{R}$, the corresponding outputs $\hat{y} := \hat{h}(\hat{x})$ and $y := h(x)$ are ε-close (condition (c)).

Remark 2.6 Both Definition 2.7 and Definition 2.8 are technical relaxations of the definitions of δ-lifting and (ε,δ)-stochastic simulation relations provided by Haesaert et al. (2017). These relations quantify the similarity between two models by bounding their transition probabilities with δ and their output distances with ε.

For a parametrised model class $\boldsymbol{\mathcal{M}}(\theta)$, with nominal $\hat{\mathbf{t}}$ and concrete $\mathbf{t}_\theta$, Definition 2.8 above admits an interface function $\mathbf{i}$ that is independent of θ but a sub-probability kernel $\boldsymbol{v}_\theta$ which may depend on θ. To solve Subproblem 2.2, in addition, the state mapping $\mathbf{s}$ must also be independent of θ. This leads to a general condition under which the state mapping $\mathbf{s}$ gives a *valid control refinement*, formalised next.

Valid Control Refinement

Definition 2.9 (Valid Control Refinement) Consider an interface function $\mathbf{i}\colon \hat{\mathbb{X}} \times \mathbb{X} \times \hat{\mathbb{U}} \times \mathscr{B}(\mathbb{U}) \to [0,1]$ and a sub-probability kernel $\boldsymbol{v}\colon \hat{\mathbb{X}} \times \mathbb{X} \times \hat{\mathbb{U}} \times \mathscr{B}(\hat{\mathbb{X}} \times \mathbb{X}) \to [0,1]$ as in Definition 2.8. A state mapping $\mathbf{s}\colon \hat{\mathbb{X}} \times \mathbb{X} \times \mathbb{X} \times \hat{\mathbb{U}} \to \mathscr{P}(\hat{\mathbb{X}})$ defines a *valid control refinement* if the composed transition kernel

$$\mathbf{t}^{\times}(d\hat{x}_+ \times dx_+ | \hat{x}, x, \hat{u}) := \mathbf{s}(d\hat{x}_+ | \hat{x}, x, x_+, \hat{u})\, \mathbf{t}(dx_+ | x, \mathbf{i}(u|\hat{x},x,\hat{u})) \tag{2.9}$$

majorises the sub-probability coupling $\boldsymbol{v}$, that is

$$\mathbf{t}^{\times}(A|\hat{x},x,\hat{u}) \geq \boldsymbol{v}(A|\hat{x},x,\hat{u}) \tag{2.10}$$

for all sets $A \in \mathscr{B}(\hat{\mathbb{X}} \times \mathbb{X})$, $(\hat{x},x) \in \mathcal{R}$, and $\hat{u} \in \hat{\mathbb{U}}$.

Note that for a parametrised model class $\boldsymbol{\mathcal{M}}(\theta)$ that is in a relation with a model $\widehat{\boldsymbol{\mathcal{M}}}$, the right-hand side of (2.10) may depend on θ and the left-hand side may depend on θ only via the dynamics of $\boldsymbol{\mathcal{M}}(\theta)$ represented by the transition kernel $\mathbf{t}_\theta$, that is,

$$\mathbf{t}^{\times}_{\theta}(A|\hat{x},x,\hat{u}) \equiv \iint_{(\hat{x}_+,x_+)\in A} \mathbf{s}(d\hat{x}_+|\hat{x},x,x_+,\hat{u})\, \mathbf{t}_\theta(dx_+|x,\mathbf{i}(u|\hat{x},x,\hat{u})) \geq \boldsymbol{v}_\theta(A|\hat{x},x,\hat{u}).$$

The next result states that similar to the (ε, δ)-approximate simulation relation defined by Haesaert et al. (2017), there always exists at least one valid control refinement for the newly defined SSR.

Proposition 2.2 (Existence) *For two gMDPs $\widehat{\boldsymbol{\mathcal{M}}}$ and $\boldsymbol{\mathcal{M}}(\theta)$ with $\widehat{\boldsymbol{\mathcal{M}}} \preceq_{\varepsilon}^{\delta} \boldsymbol{\mathcal{M}}(\theta)$, there always exists a valid control refinement.*

Proof. If $\widehat{\boldsymbol{\mathcal{M}}} \preceq_{\varepsilon}^{\delta} \boldsymbol{\mathcal{M}}(\theta)$, then this implies that there also exists a probability kernel $\bar{\boldsymbol{\nu}}_\theta(d\hat{x}_+ \times dx_+|\cdot)$ (see Proposition 2.1) satisfying the conditions for an (ε, δ)-approximate probabilistic simulation relation (cf. Haesaert et al., 2017, Definition 9). Let $\mathbf{t}_\theta^\times = \bar{\boldsymbol{\nu}}_\theta$. Then, the conditional kernel obtained from $\mathbf{t}_\theta^\times$ (which defines a t_θ-almost surely unique measurable map for every $(\hat{x}, x, \hat{u})$ (Borkar, 2012, Corollary 3.1.2)) yields a parameter-independent state mapping $\mathbf{s}$, such that

$$\mathbf{t}_\theta^\times(d\hat{x}_+ \times dx_+|\hat{x}, x, \hat{u}) = \mathbf{s}(d\hat{x}_+|\hat{x}, x, x_+, \hat{u})\, \mathbf{t}_\theta(dx_+|x, \mathbf{i}(u|\hat{x}, x, \hat{u})).$$

Note that the existence of $\bar{\boldsymbol{\nu}}_\theta$ implies the existence of a sub-probability coupling minorising $\mathbf{t}_\theta^\times$, thus satisfying (2.10). □

In general, there exists more than one valid control refinement for given models $\widehat{\boldsymbol{\mathcal{M}}}$ and $\boldsymbol{\mathcal{M}}(\theta)$ in an SSR $\widehat{\boldsymbol{\mathcal{M}}} \preceq_{\varepsilon}^{\delta} \boldsymbol{\mathcal{M}}(\theta)$. This permits the selection of an interface function $\mathbf{i}$ that does not depend on θ. Proposition 2.2 above states that the new framework fully recovers the results of Haesaert et al. (2017) for non-parametrised models.

Remark 2.7 Although Definition 2.9 gives a sufficient condition for a valid refinement, it is not a necessary condition. For specific control specifications, different refinement strategies such as the one presented by Haesaert et al. (2018) can be used, where information on the value function and relation is used to refine the control policy.

The next theorem establishes that the new similarity relation is transitive. This property is extremely useful when the nominal model is constructed from a concrete model in multiple stages of abstraction. This property is exploited in the case studies in Section 2.6.

Theorem 2.1 (Transitivity) *Suppose $\widetilde{\boldsymbol{\mathcal{M}}} \preceq_{\varepsilon_1}^{\delta_1} \widehat{\boldsymbol{\mathcal{M}}}$ and $\widehat{\boldsymbol{\mathcal{M}}} \preceq_{\varepsilon_2}^{\delta_2} \boldsymbol{\mathcal{M}}$. Then, it follows that $\widetilde{\boldsymbol{\mathcal{M}}} \preceq_{\varepsilon}^{\delta} \boldsymbol{\mathcal{M}}$ with $\delta = \delta_1 + \delta_2$ and $\varepsilon = \varepsilon_1 + \varepsilon_2$.* Transitivity

Proof. This proof is similar to the one given by Haesaert et al. (2017) applied to sub-probability couplings. Consider three gMDPs $\boldsymbol{\mathcal{M}} = (\mathbb{X}, \mathbb{X}_0, \mathbb{U}, \mathbf{t}, h, \mathbb{Y})$, $\widehat{\boldsymbol{\mathcal{M}}} = (\hat{\mathbb{X}}, \hat{\mathbb{X}}_0, \hat{\mathbb{U}}, \hat{\mathbf{t}}, \hat{h}, \mathbb{Y})$, and $\widetilde{\boldsymbol{\mathcal{M}}} = (\tilde{\mathbb{X}}, \tilde{\mathbb{X}}_0, \tilde{\mathbb{U}}, \tilde{\mathbf{t}}, \tilde{h}, \mathbb{Y})$. Given $\widetilde{\boldsymbol{\mathcal{M}}} \preceq_{\varepsilon_1}^{\delta_1} \widehat{\boldsymbol{\mathcal{M}}}$ and $\widehat{\boldsymbol{\mathcal{M}}} \preceq_{\varepsilon_2}^{\delta_2} \boldsymbol{\mathcal{M}}$ with $(\mathcal{R}_1, \boldsymbol{\nu}_1)$ and $(\mathcal{R}_2, \boldsymbol{\nu}_2)$, respectively, as formalised in Definition 2.8, define the relation $\mathcal{R} := \{(\tilde{x}, x) \in \mathscr{B}(\tilde{\mathbb{X}} \times \mathbb{X}) \,|\, \exists \hat{x} \in \hat{\mathbb{X}} : (\tilde{x}, \hat{x}) \in \mathcal{R}_1, (\hat{x}, x) \in \mathcal{R}_2\}$.

ε-Deviation: From the definition of $\mathcal{R}$, it follows that $\forall(\tilde{x}, x) \in \mathcal{R}$, $\exists\hat{x} \in \hat{\mathbb{X}}$: $(\tilde{x}, \hat{x}) \in \mathcal{R}_1, (\hat{x}, x) \in \mathcal{R}_2$. Given $(\tilde{x}, x, \hat{x})$ and the mutual output metric $\mathbf{d}_{\mathbb{Y}}$, the output error can be bounded from above as

$$\mathbf{d}_{\mathbb{Y}}(\tilde{h}(\tilde{x}), h(x)) \leq \mathbf{d}_{\mathbb{Y}}(\tilde{h}(\tilde{x}), \hat{h}(\hat{x})) + \mathbf{d}_{\mathbb{Y}}(\hat{h}(\hat{x}), h(x)) \leq \varepsilon_1 + \varepsilon_2.$$

δ-Deviation: The sub-probability couplings $\boldsymbol{\nu}_1$ and $\boldsymbol{\nu}_2$ can be completed to probability kernels $\bar{\boldsymbol{\nu}}_1$ and $\bar{\boldsymbol{\nu}}_2$ that couple $(\tilde{\mathbf{t}}, \hat{\mathbf{t}})$ and $(\hat{\mathbf{t}}, \mathbf{t})$, respectively, (cf. δ-lifting by Haesaert et al. (2017, Definition 8)) via (2.8). Utilising the proof by Haesaert et al. (2017, Appendix D.2), it follows that there exists a probability kernel $\bar{\boldsymbol{\nu}}$ over $(\tilde{\mathbb{X}} \times \mathbb{X}, \mathscr{B}(\tilde{\mathbb{X}} \times \mathbb{X}))$ that couples $\tilde{\mathbf{t}}$ and $\mathbf{t}$ over $\mathcal{R}$. Furthermore, for all $z \in \mathbb{Z} := \{(\tilde{x}, x, \tilde{u}) | (\tilde{x}, x) \in \mathcal{R}, \tilde{u} \in \tilde{\mathbb{U}}\}$ it holds that $\bar{\boldsymbol{\nu}}(\mathcal{R}|z) \leq \delta_1 + \delta_2$. From (2.8), it follows that for all measurable sets $d\tilde{x}_+ \in \mathscr{B}(\tilde{\mathbb{X}})$ and $dx_+ \in \mathscr{B}(\mathbb{X})$ it holds that $\bar{\boldsymbol{\nu}}(d\tilde{x}_+ \times dx_+) \geq \boldsymbol{\nu}(d\tilde{x}_+ \times dx_+)$. Hence, there exists an SSR $\widetilde{\boldsymbol{\mathcal{M}}} \preceq^{\delta}_{\varepsilon} \boldsymbol{\mathcal{M}}$ with $\delta \leq \delta_1 + \delta_2$, concluding the proof. □

2.4.3 Temporal Logic Control via Sub-Simulation Relations

To design a controller satisfying a temporal logic property ψ from the scLTL subset, the specification is considered in its alternative representation as a DFA $\mathcal{A}_\psi = (Q, q_0, \boldsymbol{\Sigma}, F, \tau)$ (see Definition 2.5). Then, a robust version of the dynamic programming framework presented by Haesaert and Soudjani (2021) is used to obtain robust data-driven bounds on the satisfaction probability, by solving a reachability problem on the composition of $\widehat{\boldsymbol{\mathcal{M}}}$ and $\mathcal{A}_\psi$ (Kamgarpour et al., 2013). It follows a brief summary of the framework presented by Haesaert and Soudjani (2021), which encodes the effects of both ε and δ.

Robust Dynamic Programming

Robust Dynamic Programming Define the ε-neighbourhood of an element $y \in \mathbb{Y}$,

$$B_\varepsilon(\hat{y}) := \{y \in \mathbb{Y} | \, \mathbf{d}_{\mathbb{Y}}(y, \hat{y}) \leq \varepsilon\},$$

and a Markov policy $\rho \colon \hat{\mathbb{X}} \times Q \to \mathscr{P}(\hat{\mathbb{U}})$. Similar to dynamic programming with perfect model knowledge (Sutton and Barto, 2018), value functions $V^\rho_l \colon \hat{\mathbb{X}} \times Q \to [0, 1]$ — defined for each iteration $l \in \{0, 1, 2, \ldots\}$ on the product space $\hat{\mathbb{X}} \times Q$ — are used to represent the probability of starting in an initial state pair $(\hat{x}_0, q_0)$ and reaching the accepting set F in l steps. These value functions are connected recursively via operators associated with the dynamics of the system and DFA.

The dynamic programming scheme evolves as follows. The value function is initialised with $V_0 = 0$, and the (ε, δ)-*robust Bellman operator* $\mathbf{T}^\rho_{\varepsilon,\delta}$ acting on value functions is defined as

$$\mathbf{T}^\rho_{\varepsilon,\delta}(V)(\hat{x}, q) := \left[\int_{\hat{\mathbb{X}}} \min_{q_+ \in \tau_\varepsilon(q, \hat{x}_+)} \max\{\mathbb{1}_F(q_+), V(\hat{x}_+, q_+)\} \, \hat{\mathbf{t}}(d\hat{x}_+ | \hat{x}, \rho(\hat{x}, q)) - \delta\right]_0^1, \tag{2.11}$$

with the ε-overapproximating transition function $\tau_\varepsilon(q, \hat{x}_+) := \{\tau(q, \alpha),\ \alpha \in \mathcal{L}(B_\varepsilon(\hat{h}(\hat{x}_+)))\}$ and the truncation function $[\,\cdot\,]_0^1 : \mathbb{R} \to [0, 1]$, given as $[\,\cdot\,]_0^1 := \min(1, \max(0, \cdot))$. Note that δ acts as a discount factor. The value functions $V_1^\rho, V_2^\rho, \ldots$ are computed recursively via iterative application of the operator $\mathbf{T}_{\varepsilon,\delta}^\rho$:

$$V_{l+1}^\rho = \mathbf{T}_{\varepsilon,\delta}^\rho(V_l^\rho), \quad l \in \{0, 1, 2, \ldots\}. \tag{2.12}$$

Consequently, the *optimal* (ε, δ)-robust Bellman operator is given by

$$\mathbf{T}_{\varepsilon,\delta}^*(V)(\hat{x}, q) := \sup_\rho \mathbf{T}_{\varepsilon,\delta}^\rho(V)(\hat{x}, q),$$

where the outer supremum is taken over Markov policies ρ on the product space $\hat{\mathbb{X}} \times Q$. Fig. 2.7 visualises the dynamic programming scheme conceptually.

0	0	1
0	0	0
0	0	0

(a) $\max\{\mathbb{1}_F(q_0), 0\}$

0	0.5	1
0	0	0
0	0	0

(b) $V_1 = \mathbf{T}(V_0)$

0.4	0.6	1
0	0.1	0
0	0	0

(c) $V_2 = \mathbf{T}(V_1)$

0.7045	0.7369	1
0.5194	0.3165	0
0.445	0.1939	0.113

(d) $V_{10} = \mathbf{T}(V_9)$

Fig. 2.7: Conceptual visualisation of the dynamic programming scheme for a finite state space $\hat{\mathbb{X}}$ of 9 states and one goal state with value 1 (simplified for $\delta = 0$).

The following result, adapted from Haesaert and Soudjani (2021, Corollary 4), provides a robust lower bound on the satisfaction probability, by considering the converging value function in the limit case $l \to \infty$.

Robust Satisfaction Probability

Theorem 2.2 ((ε, δ)-Robust Satisfaction Probability) *Suppose $\widehat{\mathcal{M}} \preceq_\varepsilon^\delta \mathcal{M}$ with $\delta > 0$, and the specification ψ expressed as a DFA $\mathcal{A}_\psi$. Then, a controller $\mathcal{C}$ can be constructed such that the specification is satisfied by $\mathcal{C} \bowtie \mathcal{M}$ with probability at least $\mathcal{S}_{\varepsilon,\delta}^*$. This quantity is the* (ε, δ)-robust satisfaction probability*, defined as a function of the initial state of $\mathcal{M}$, namely*

$$\mathcal{S}_{\varepsilon,\delta}^*(x_0) := \min_{q_+ \in \tau_\varepsilon(q_0, \hat{x}_0)} \max\left\{\mathbb{1}_F(q_+), V_\infty^*(\hat{x}_0, q_+)\right\}, \quad \hat{x}_0 \sim \mathbf{s}(\,\cdot\,|x_0, \ldots),$$

where V_∞^* is the unique solution to the fixed-point equation

$$V_\infty^* = \mathbf{T}_{\varepsilon,\delta}^*(V_\infty^*), \tag{2.13}$$

obtained from $V_\infty^* := \lim_{l\to\infty}(\mathbf{T}_{\varepsilon,\delta}^*)^l(V_0)$ with $V_0 = 0$. The nominal controller $\widehat{\mathcal{C}}$ is the stationary Markov policy ρ^* maximising the right-hand side of (2.13),

> *i.e.,* $\rho^* = \arg\sup_\rho \mathbf{T}^\rho_{\varepsilon,\delta}(V^*_\infty)$. *The concrete controller* $\boldsymbol{C}$ *is the refined controller obtained from the nominal controller* $\widehat{\boldsymbol{C}}$*, the interface function* $\mathbf{i}$*, and the state mapping* $\mathbf{s}$.

A corresponding *upper* bound on the satisfaction probability can be derived analogously via the (ε, δ)-*optimistic Bellman operator* (Haesaert and Soudjani, 2021):

$$\mathbf{T}^\rho_{-\varepsilon,-\delta}(V)(\hat{x}, q) := \left[\int_{\hat{\mathbb{X}}} \max_{q_+ \in \tau_\varepsilon(q,\hat{x}_+)} \max\{\mathbb{1}_F(q_+), V(\hat{x}_+, q_+)\}\, \hat{\mathbf{t}}(d\hat{x}_+|\hat{x}, \rho(\hat{x}, q)) + \delta\right]_0^1.$$

Remark 2.8 (Finite-Horizon Specifications) For a finite-horizon specification ψ and $\delta \geq 0$, a finite number of value functions $V^*_{l+1} = \mathbf{T}^*_{\varepsilon,\delta}(V_l)$, $l \in \{0, 1, 2, \ldots, n_\psi\}$, are computed, with $V_0 = 0$ and n_ψ being the horizon of the specification. This also yields a non-stationary control policy.

Remark 2.9 (Bounded State Space) To compute solutions for systems with unbounded support over a bounded state space $\mathbb{X}$, a sink state is added to the composed system $\widehat{\boldsymbol{\mathcal{M}}} \times \mathcal{A}_\psi$ to capture the transitions leaving the bounded set of states.

Remark 2.10 (Computational Complexity) The complexity of the operator $\mathbf{T}^\rho_{\varepsilon,\delta}$ in (2.11) is examined to shed light on the computational complexity of the value update. As indicated in Fig. 2.7, in practise, the value function is computed over the Cartesian product of a *finite* state space $\hat{\mathbb{X}}$ and the DFA locations Q. The complexity with respect to the number of discrete states, $n_{\hat{\mathbb{X}}}$, control inputs, $n_{\hat{\mathbb{U}}}$, and DFA locations, n_Q, is given by $O((n_{\hat{\mathbb{X}}} n_Q) \times (n_{\hat{\mathbb{X}}} n_{\hat{\mathbb{U}}} n_Q))$. Naturally, the complexity is exponential with respect to the dimensionality of the (potentially reduced-order) model, i.e., $O((\eta^{n_x}_{\hat{\mathbb{X}}} n_Q) \times (\eta^{n_x}_{\hat{\mathbb{X}}} \eta^{n_u}_{\hat{\mathbb{U}}} n_Q))$, where $\eta_{\hat{\mathbb{X}}}$ and $\eta_{\hat{\mathbb{U}}}$ denote the numbers of discrete partitions per dimension of the state and input spaces, respectively, and n_x and n_u are the corresponding numbers of dimensions. This highlights the significance of MOR, demonstrated for a 7-dimensional system in Subsection 2.6.2.

2.5 Data-Driven Simulation Relations

In this section, the previously defined concepts are combined to construct data-driven simulation relations and answer Problem 2.1. Furthermore, whilst the scope of the exposition is limited to systems with known Gaussian noise, the approach is generally applicable to other (potentially uncertain) distributions, as detailed in Chapter 3, where it is shown how to extend the approach based on approximations of arbitrary continuous noise distributions via finite Gaussian mixture models.

System Models Consider an unknown nonlinear system $\boldsymbol{\mathcal{M}}(\theta^*)$ of the form (2.3). Let data $\mathcal{D}_N$ in (2.2) be obtained from this unknown true system $\boldsymbol{\mathcal{M}}(\theta^*)$. Based on $\mathcal{D}_N$, the true parametrisation θ^* is approximated with an estimate $\hat{\theta}$ and a credible set

Θ is computed as described in Section 2.3. The fixed parameter estimate $\hat{\theta}$ generates the nominal model $\widehat{\mathcal{M}} := \mathcal{M}(\hat{\theta})$, given by

$$\widehat{\mathcal{M}}\colon \begin{cases} \hat{x}_{t+1} = \hat{\theta}^\top f(\hat{x}_t, \hat{u}_t) + \hat{w}_t \\ \hat{y}_t \quad\;\; = \hat{h}(\hat{x}_t) \end{cases}. \tag{2.14}$$

The noise $w_t, \hat{w}_t \sim \mathcal{N}(\,\cdot\,|0,\Sigma)$ is assumed to have Gaussian distribution with a full-rank covariance matrix $\Sigma \in \mathbb{R}^{n_x \times n_x}$ and output map $\hat{h} \equiv h$.

Transition Kernels To find a valid control refinement for the systems $\mathcal{M}(\theta)$ and $\widehat{\mathcal{M}}$ in (2.3) and (2.14), the corresponding transition kernels are formulated. The probability of $\mathcal{M}(\theta)$ transitioning from a state x with input u to a state $x_+ \in S \subset \mathbb{X}$ is given by

$$\mathbb{P}(x_+ \in S | x, u) = \int_S \mathbf{t}_\theta(dx_+|x,u),$$

where the corresponding transition kernel $\mathbf{t}_\theta$ is given by

$$\mathbf{t}_\theta(dx_+|x,u) = \mathcal{N}(dx_+|\theta^\top f(x,u),\Sigma) = \int_{\mathbb{W}} \delta_{\theta^\top f(x,u)+w}(dx_+)\,\mathcal{N}(dw|0,\Sigma). \tag{2.15}$$

Similarly, the transition kernel of $\widehat{\mathcal{M}}$ is constructed as

$$\hat{\mathbf{t}}(d\hat{x}_+|\hat{x},\hat{u}) = \mathcal{N}(d\hat{x}_+|\hat{\theta}^\top f(\hat{x},\hat{u}),\Sigma) = \int_{\hat{\mathbb{W}}} \delta_{\hat{\theta}^\top f(\hat{x},\hat{u})+\hat{w}}(d\hat{x}_+)\,\mathcal{N}(d\hat{w}|0,\Sigma). \tag{2.16}$$

Error Dynamics The dynamics of the target system $\mathcal{M}(\theta)$ in (2.3) can be rewritten in terms of the nominal dynamics ($\widehat{\mathcal{M}}$) plus an additive disturbance term: Error Dynamics

$$x_{t+1} = \underbrace{\hat{\theta}^\top f(x_t,u_t)}_{\text{nominal dynamics}} + \underbrace{(\theta-\hat{\theta})^\top f(x_t,u_t) + w_t}_{\text{disturbance}}, \quad w_t \sim \mathcal{N}(\,\cdot\,|0,\Sigma).$$

The disturbance term consists of an error in the dynamics — caused by a deviation of the unknown parameters θ from the nominal parametrisation $\hat{\theta}$ — and the stochastic influence of the noise $w_t \sim \mathcal{N}(\,\cdot\,|0,\Sigma)$. Assume that the former can be bounded as

$$\|(\theta-\hat{\theta})^\top f(x,u)\| \le \Delta(x,u), \quad \forall (x,u,\theta) \in \mathbb{X}\times\mathbb{U}\times\Theta,$$

for some function $\Delta\colon \mathbb{X}\times\mathbb{U} \to \mathbb{R}_{\ge 0}$.

Valid Refinement Choosing an interface function $u \sim \mathbf{i}(\,\cdot\,|\hat{x},x,\hat{u}) := \hat{u}$ and the noise coupling defined via $\hat{w} \equiv \gamma_\theta(x,u;\hat{\theta}) + w$, with the offset Noise Coupling

$$\gamma_\theta(x,u;\hat{\theta}) := (\theta-\hat{\theta})^\top f(x,u), \tag{2.17}$$

yields a state mapping

$$\hat{x}_+ = x_+ + \hat{\theta}^\top(f(\hat{x},\hat{u}) - f(x,u)) \tag{2.18}$$

that is both deterministic and not dependent on the unknown parameters θ, i.e.,

$$\mathbf{s}(\hat{x}_+|\hat{x}, x, x_+, \hat{u}) := \delta_{x_+ + \hat{\theta}^\top (f(\hat{x},\hat{u}) - f(x,u))}(\hat{x}_+). \tag{2.19}$$

Establishing the SSR To establish an SSR $\widehat{\boldsymbol{\mathcal{M}}} \preceq^{\delta}_{\varepsilon} \boldsymbol{\mathcal{M}}(\theta)$, the identity relation

$$\mathcal{R} := \left\{ (\hat{x}, x) \in \hat{\mathbb{X}} \times \mathbb{X} \,\middle|\, x = \hat{x} \right\} \tag{2.20}$$

is selected. Condition (a) (Definition 2.8) holds by setting $\hat{\mathbb{X}}_0 = \mathbb{X}_0$. Condition (c) is satisfied with $\varepsilon = 0$ since both systems use the same output mapping. For condition (b), define the following sub-probability coupling $\boldsymbol{\upsilon}_\theta$ over $\mathcal{R}$:

$$\begin{aligned} \boldsymbol{\upsilon}_\theta(d\hat{x}_+ \times dx_+ | \cdot) = \iint_{\hat{\mathbb{W}} \times \mathbb{W}} & \delta_{\hat{\theta}^\top f(\hat{x},\hat{u}) + \hat{w}}(d\hat{x}_+) \delta_{\theta^\top f(x,u) + w}(dx_+) \delta_{\gamma_\theta + w}(d\hat{w}) \\ & \times \min\{\mathcal{N}(dw|0,\Sigma), \mathcal{N}(dw| - \gamma_\theta, \Sigma)\}, \end{aligned} \tag{2.21}$$

that takes the minimum of two Gaussian probability measures. Note that the arguments of $\boldsymbol{\upsilon}_\theta$ and γ_θ were dropped to lighten the notation.

Theorem 2.3 (Establishing an SSR) *Let data $\mathcal{D}_N$ be given. Then, the nonlinear models (2.3) and (2.14) are in an SSR $\widehat{\boldsymbol{\mathcal{M}}} \preceq^{\delta}_{\varepsilon} \boldsymbol{\mathcal{M}}(\theta)$ with the interface function $u_t = \hat{u}_t$, relation (2.20), and sub-probability coupling (2.21). For a given confidence level $1 - \alpha$, the state mapping (2.19) defines a valid control refinement (Definition 2.9) with $\varepsilon = 0$ and*

$$\begin{aligned} \delta(\hat{x}, \hat{u}) &= 1 - 2 \cdot \mathrm{cdf}\left(-\frac{\sqrt{r}}{2} \|f(\hat{x}, \hat{u})\| \right), \\ r &:= \left\| \Sigma^{-1} \right\| \|\Sigma_N\| \cdot n_x \cdot \chi^{-1}(1 - \alpha | n_x), \end{aligned} \tag{2.22}$$

with covariance matrix Σ_N computed via (2.4) and Σ in (2.3).

The proof of Theorem 2.3 has been deferred to Appendix A.1. Note that this theorem yields a δ as a function of $(\hat{x}, \hat{u})$, thus providing less conservative results compared to a constant global δ. Furthermore, δ directly encodes the desired confidence $1-\alpha$ and data amount N. Chapter 3 extends Theorem 2.3 to systems with general continuous noise distributions and parametric uncertainty in both the deterministic and stochastic parts of the dynamics.

Solution Given (ε, δ) from Theorem 2.3, Theorem 2.2 solves Subproblem 2.2 by providing a parameter-independent concrete controller $\boldsymbol{C}$ that is uniformly valid for all $\boldsymbol{\mathcal{M}}(\theta)$, $\theta \in \Theta$ with robust satisfaction probability $\mathcal{S}^*_{\varepsilon,\delta}$, i.e.,

$$\mathbb{P}\left(\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}}(\theta) \models \psi\right) \geq \mathcal{S}^*_{\varepsilon,\delta}, \quad \forall \theta \in \Theta.$$

Visualisation of the Sub-Probability Coupling The sub-probability coupling in (2.21) can be reformulated using the definition of γ_θ in (2.17) to yield

$$\boldsymbol{v}_\theta(d\hat{x}_+ \times dx_+ | \cdot) = \int_{\mathbb{W}} \delta_{\hat{x}_+}(dx_+)\delta_{\hat{\theta}^\top f(x,u)+\hat{w}}(d\hat{x}_+) \min\{\mathcal{N}(d\hat{w}|\gamma_\theta, \Sigma), \mathcal{N}(d\hat{w}|0, \Sigma)\},$$

which is illustrated in Fig. 2.8 on a 2D plane for a concrete valuation of $(\hat{x}, x, \hat{u})$. The coupling v_θ arises from the sub-probability coupling $\bar{w}_\theta$ of the two noise distributions,

$$\bar{w}_\theta(d\hat{w} \times dw) := \delta_{\gamma_\theta + w}(d\hat{w}) \min\{\mathcal{N}(dw|0, \Sigma), \mathcal{N}(dw| - \gamma_\theta, \Sigma)\},$$

where γ_θ quantifies the offset between the two distributions. With the approach

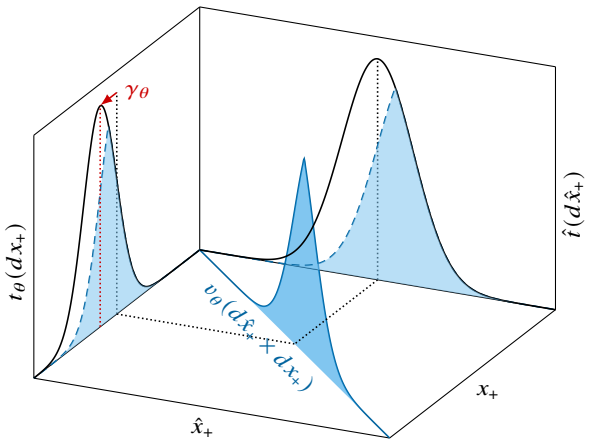


Fig. 2.8: 2D representation of the sub-probability coupling v for the transition kernels $\hat{t}$ and t given some $(\hat{x}, x, \hat{u})$. Displayed is the coupling over the product space $\hat{\mathbb{X}} \times \mathbb{X}$ as well as its marginals (see Definition 2.7(b)-(c)).

presented in this chapter, parametric uncertainty is compensated by shifting the noise distributions relative to each other by the offset γ_θ in the sub-probability coupling. Hence, epistemic uncertainty can only be compensated on state variables that are perturbed with noise.

Based on the previous results and solutions to Subproblem 2.1 and Subproblem 2.2, the following corollary provides a solution to the overarching data-driven Problem 2.1.

Corollary 2.1 (Data-Driven Robust Satisfaction Probability) *Given the dataset $\mathcal{D}_N$ in (2.2), let Θ be a credible set of θ^* with confidence level $1 - \alpha$ (Definition 2.2). Let $\boldsymbol{C}$ be the refined controller with (ε, δ)-robust satisfaction*

probability $\mathcal{S}^*_{\varepsilon,\delta}$ for all $\theta \in \Theta$ (Theorem 2.2). Then, it follows that

$$\mathbb{P}(\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}}(\theta^*) \models \psi \mid \mathcal{D}_N) \geq \mathcal{S}^*_{\varepsilon,\delta}(1-\alpha).$$

Proof. This result follows from the (ε, δ)-robust satisfaction probability $\mathcal{S}^*_{\varepsilon,\delta}$ (Theorem 2.2) and the results from Section 2.3. The law of total probability yields

$$\begin{aligned}\mathbb{P}(\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}}(\theta^*) \models \psi \mid \mathcal{D}_N) &= \mathbb{P}(\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}}(\theta^*) \models \psi \mid \theta^* \in \Theta, \mathcal{D}_N)\, \mathbb{P}(\theta^* \in \Theta) \\ &\quad + \mathbb{P}(\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}}(\theta^*) \models \psi \mid \theta^* \notin \Theta, \mathcal{D}_N)\, \mathbb{P}(\theta^* \notin \Theta), \\ &\geq \mathcal{S}^*_{\varepsilon,\delta}(1-\alpha).\end{aligned}$$

□

Remark 2.11 (Practical Considerations) Corollary 2.1 highlights the linear dependence of the data-driven robust satisfaction probability on the desired confidence $1-\alpha$. Additionally, $1-\alpha$ dictates the size of the credible set Θ, which determines the magnitude of δ (Theorem 2.3). As δ acts as a discount factor in the value update (2.12), large values of δ translate directly into a loss of satisfaction probability.

From a practical standpoint, when testing the proposed approach on a new case study, an initial confidence value $(1-\alpha) \to 0$ is chosen, for which the credible set shrinks to a singleton and the resulting robust satisfaction probability $\mathcal{S}^*_{\varepsilon,\delta}$ from Theorem 2.2 is maximised. Note that this equates to the model-based case where θ is assumed known. However, as θ is unknown, the corresponding data-driven robust satisfaction probability from Corollary 2.1 is zero. To obtain non-trivial results, the desired confidence level $1-\alpha$ is increased progressively, inflating the confidence set Θ to improve the resulting data-driven robust satisfaction probability.

Theorem 2.3 computes δ considering the worst-case parameters $\theta \in \Theta$ in a given credible set agnostic of the posterior distribution $p(\theta|\mathcal{D}_N)$ used to obtain it (see Section 2.3). The following result leverages the posterior distribution — which spans a probability distribution over individual values of θ — to obtain strengthened guarantees once a controller has been computed.

Proposition 2.3 (Strengthened Robust Satisfaction Probability) *Once a controller $\boldsymbol{C}$ has been obtained, a less conservative robust satisfaction probability than $\mathcal{S}^*_{\varepsilon,\delta}$ from Theorem 2.2 can be obtained by integrating with respect to the posterior distribution $p(\theta|\mathcal{D}_N)$ in (2.4), namely*

$$\mathbb{P}(\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}}(\theta^*) \models \psi \mid \mathcal{D}_N) = \int_{\theta \in \Omega} \mathbb{P}(\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}}(\theta) \models \psi \mid \theta, \mathcal{D}_N)\, p(\theta|\mathcal{D}_N).$$

2.6 Case Studies

The effectiveness of the presented approach is demonstrated on three case studies, including a complex specification, a 7-dimensional system, and the nonlinear Van der Pol Oscillator.

2.6.1 Linear System with Complex Specification

Consider the uncertain system in (2.3) with linear dynamics $f(x_k, u_k) = [x_k, u_k]^\top$, $h(x_k) = x_k$, and $\Sigma = \sqrt{0.1} I_2$, modelling an agent moving in a 2D space. Define the state space $\mathbb{X} = [-6, 6]^2 \subset \mathbb{R}^2$, input space $\mathbb{U} = [-1, 1]^2$, and output space $\mathbb{Y} = \mathbb{X}$. The goal is to compute a controller for a package delivery problem (see Fig. 2.13). For this, the controller should navigate the agent to pick up a parcel at region p_1 and deliver it to target region p_3. If the agent passes the region p_2 on its path, it loses the package and must pick up a new one at p_1. This behaviour is captured by the specification $\psi = \Diamond(p_1 \wedge (\neg p_2 \mathbin{\mathsf{U}} p_3))$. Note that the corresponding DFA presented in Fig. 2.9 contains a backward loop from location q_1 and is hence not expressible as a reach-avoid specification over the state space only. Here, the regions $p_1 = [3, 6] \times [-2.5, 1]$, $p_2 = [-1, 1] \times [-4, 3]$, and $p_3 = [-6, -3] \times [-6, -3]$ are used.

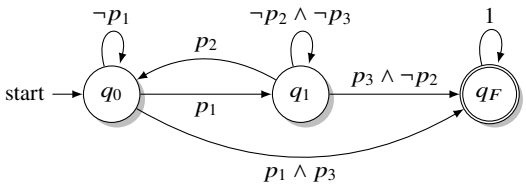


Fig. 2.9: DFA corresponding to the specification of the package delivery problem, where q_F is the accepting location.

System Identification To estimate the unknown parameters, uniform excitations u are sampled from $\mathbb{U}$, prior distributions are set to $\mathcal{N}(\theta_j | 0, 10 I_4)$, $j \in \{1, 2\}$, and N data points are drawn from the target system $\mathcal{M}^*$ with parametrisation $\theta^* = [0.6, 0.3, 1.2, 0; 0.2, 0.7, 0, 1.4]^\top$. Using Bayesian linear regression, a parameter estimate $\hat{\theta}$ and a credible set Θ are obtained (displayed for $(\hat{\theta}_{11}, \hat{\theta}_{31})$ in Fig. 2.10) for a lower confidence bound of $(1 - \alpha) = 0.9$, as outlined in Section 2.3. Note how the credible set in Fig. 2.10 is contracting for increasing amounts of data (see also Fig. 2.4). The nominal model $\widehat{\mathcal{M}} = \mathcal{M}(\hat{\theta})$ of the form (2.14) is selected with $\hat{h} \equiv h$, for which $\varepsilon_1 = 0$. δ_1 is computed via (2.22) as a function of the state and input. The maximum value of δ_1 for $\hat{u} \in \hat{\mathbb{U}}$ is displayed in Fig. 2.11. δ_1 grows rapidly for states

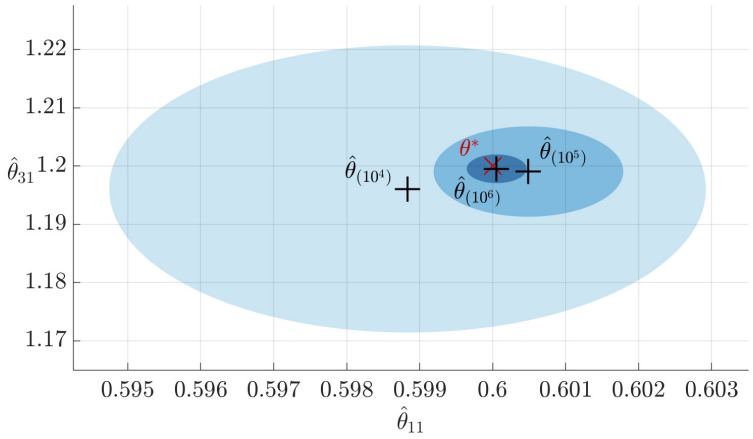


Fig. 2.10: Contracting credible set Θ and estimate $\hat{\theta}_{(N)}$ for data amounts $N = 10^4, 10^5, 10^6$ (light to dark blue) and a confidence level of $(1-\alpha) = 0.9$.

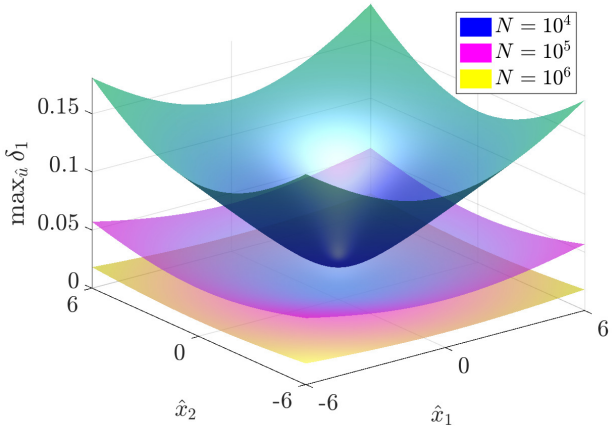


Fig. 2.11: Maximum value of δ_1 with respect to $\hat{u} \in \hat{\mathbb{U}}$ as a function of the initial state $(\hat{x}_1, \hat{x}_2)$ for $N = 10^4, 10^5, 10^6$ (top to bottom) for the package delivery case study with a confidence bound of $(1-\alpha) = 0.9$.

away from the central region, indicating that a global upper bound on δ_1 would result in poor lower bounds on the satisfaction probability.

Abstraction-Based Control Synthesis The following steps are performed using the SySCoRe toolbox (Van Huijgevoort et al., 2023), which was developed in the course of this thesis (see Section 6.2). A second abstract model $\widetilde{\mathcal{M}}$ is computed by discretising the space of $\widehat{\mathcal{M}}$, i.e., $\hat{\mathbb{X}}$ into 2500^2 and $\hat{\mathbb{U}}$ into 5 partitions, yielding $\widetilde{\mathcal{M}} \preceq^{\delta_2}_{\varepsilon_2} \widehat{\mathcal{M}}$ with $\varepsilon_2 = 0.034$ and $\delta_2 = 0.004$. Note that the magnitude of the error bounds $(\varepsilon_2, \delta_2)$ is linked directly to the precision of the chosen discretisation, which is an inherent aspect of abstraction-based methods. The values of $(\varepsilon_2, \delta_2)$ can be reduced by refining the abstraction or adopting an adaptive gridding strategy (Soudjani and Abate, 2013; Kattenbelt et al., 2010). As the development of such techniques is orthogonal to the work presented here, an equidistant gridding is used to simplify the presentation. Using the transitivity property in Theorem 2.1, it follows that $\widetilde{\mathcal{M}} \preceq^{\delta}_{\varepsilon} \mathcal{M}(\theta)$ with $\delta = \delta_1 + \delta_2$ and $\varepsilon = \varepsilon_1 + \varepsilon_2$ (see the layered setup in Fig. 2.12). The robust probability of satisfying the specification with this ε and δ is computed based on Corollary 2.1 and is depicted in Fig. 2.13 as a function of the initial state of $\mathcal{M}(\theta)$. Fig. 2.14 shows the maximum and minimum of the robust satisfaction probability with respect to the initial condition for increasing amounts of data. This indicates that the robust satisfaction probability is improved as more data is introduced.

Validation To visualise the conservatism of the robust satisfaction probability, the controller synthesised for $N = 10^6$ is applied to the system. The resulting robust satisfaction probability is compared to the actual satisfaction probability estimated via Monte Carlo simulation. While any parametrisation $\theta \in \Theta$ could be used to validate robustness, the Monte Carlo estimate is computed for θ^*, which yields the most conservative comparison. To this end, 500 simulations are run per initial state

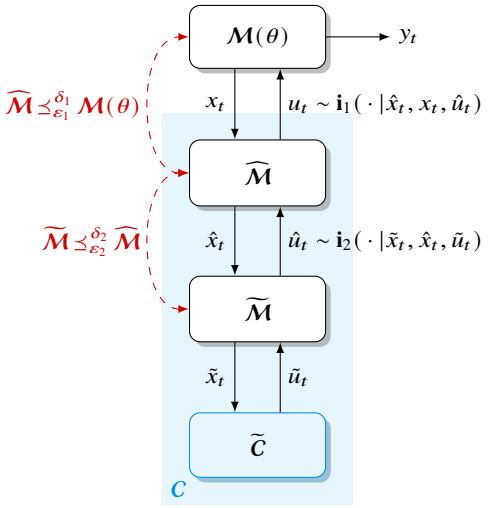


Fig. 2.12 The approach uses two layers of abstraction with two corresponding simulation relations. The first relation is an SSR between the set of feasible models $\{\mathcal{M}(\theta)|\theta \in \Theta\}$ and the continuous-space nominal model $\widehat{\mathcal{M}}$, compensating the parametric uncertainty. The second relation is an approximate simulation relation between the nominal model and its discrete abstraction, as established by Haesaert and Soudjani (2021), compensating the discretisation error.

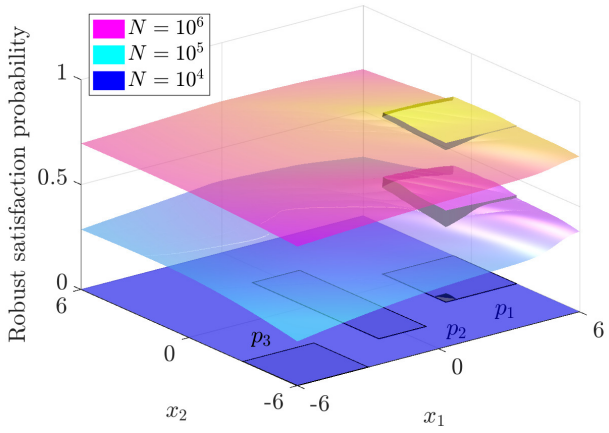


Fig. 2.13: Robust satisfaction probability as a function of the initial state for $N = 10^4, 10^5, 10^6$ (bottom to top) for the package delivery case study with a confidence bound of $(1 - \alpha) = 0.9$.

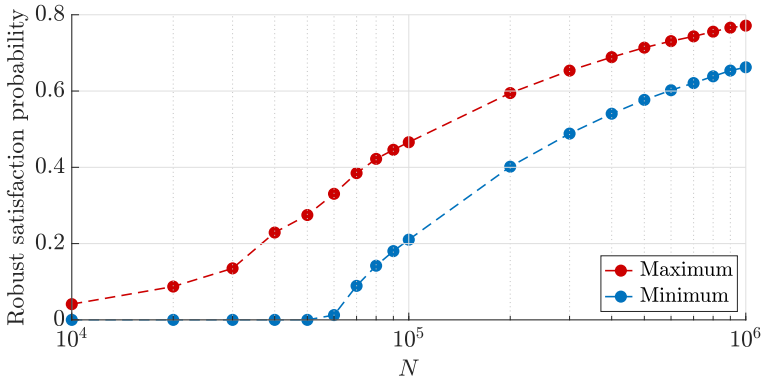


Fig. 2.14: Maximum and minimum of the robust lower bound on the satisfaction probability with respect to the initial state $x_0 \in \mathbb{X}_0 \equiv \mathbb{X}$ as a function of the size of the dataset, N.

with a maximum horizon of 60 time steps. Fig. 2.15 shows the robust satisfaction probability (in red) alongside the actual satisfaction probability (in blue) estimated via Monte Carlo simulation using Chebychev's inequality with a confidence level of 90%. Note that even for the case of $N \to \infty$ the actual satisfaction probability can not generally attain the value of 1 due to the intrinsic stochastic effects of the system. Only the uncertainty associated with the model parameters can be reduced by sampling more data. In Fig. 2.16, 100 example trajectories of the controlled system

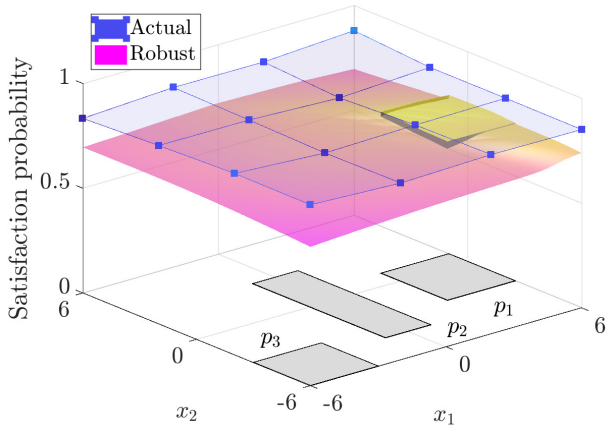


Fig. 2.15: Comparison of the robust satisfaction probability (in red) and the actual satisfaction probability estimated via Monte Carlo simulation as a function of the initial state for $N = 10^6$ for the package delivery case study.

initialised at $x_0 = [-5, -5]^\top$ are shown, 99 of which satisfy the specification. This emphasises that the proposed approach is able to synthesise controllers guaranteeing the satisfaction of complex temporal logic specifications over infinite-horizon runs.

2.6.2 7-Dimensional Building Automation System

The scalability of the presented approach is demonstrated on a 7D affine system of a building automation system, showcasing its compatibility with MOR techniques. The system dynamics can be rewritten similar to (2.3) as

$$\boldsymbol{\mathcal{M}}(\theta): \begin{cases} x_{t+1} = \begin{bmatrix}\theta & A\end{bmatrix}^\top f(x_t, u_t) + w_t \\ y_t \quad = \begin{bmatrix}1 & 0 & 0 & 0 & 0 & 0 & 0\end{bmatrix} x_t \end{cases},$$

where $\theta \in \mathbb{R}^{8\times 2}$ is unknown and $f(x_k, u_k) = [x_k; u_k]^\top$, a matrix $A \in \mathbb{R}^{8\times 5}$, and the noise distribution $w_k \sim \mathcal{N}(\cdot\,|0, \Sigma)$ are known. It is worth mentioning that Σ is non-diagonal. The system performs on a state space $\mathbb{X} \subset \mathbb{R}^7$, input space $\mathbb{U} \subset \mathbb{R}$, and output space $\mathbb{Y} \subset \mathbb{R}$. More details on the system are provided by Cauchi and Abate (2018) and Van Huijgevoort et al. (2023, Section 4.3). Here, the objective is to synthesise a controller maintaining the temperature in the first zone of the building (represented by y_t) at 20 ± 0.5 °C for 6 consecutive time steps.

Following similar steps to the previous case study, the uncertain parameters θ are estimated by drawing $N = 10^6$ data points from the target system. Using the theo-

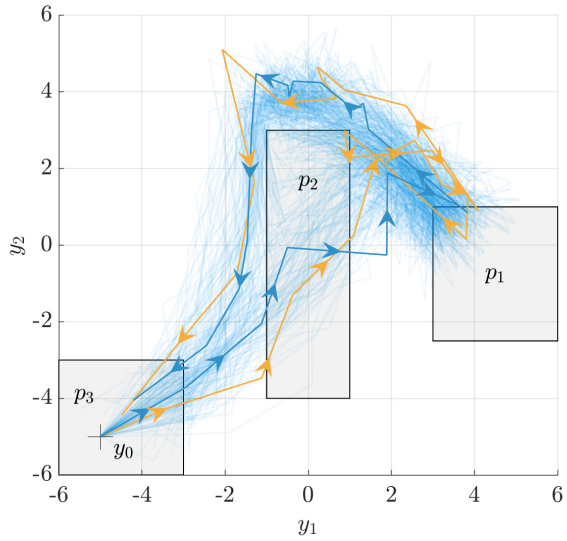


Fig. 2.16: 100 example trajectories of the package delivery case study with two trajectories satisfying the specification highlighted. Note that the yellow agent loses a parcel in p_2 and has to take an additional loop.

retical results from Haesaert et al. (2017); van Huijgevoort and Haesaert (2022), the identified 7D system is reduced to a 2D system alongside the corresponding simulation relation. An abstraction of the reduced-order system is generated by discretising $\hat{\mathbb{X}}$ into 3000^2 and $\hat{\mathbb{U}}$ into 3 partitions and the corresponding error parameters are computed to establish the simulation relation. The robust satisfaction probability is computed via Corollary 2.1, depicted in Fig. 2.17 as a function of the initial state of the 2D reduced system. Note that the results can be projected back onto the state space of the original 7D system.

Comparison All steps required in obtaining the controller and robust satisfaction probability from the dataset $\mathcal{D}_N$ amount to less than 100 s on an AMD Ryzen 5950X (single core used). To put this into perspective with other available approaches, consider an IMDP generated via GP regression as the state of the art (Jackson et al., 2020; Jiang et al., 2022). This approach is limited to systems with dimension-wise independent dynamics and establishes formal guarantees by leveraging concentration inequalities to obtain statistical error bounds for the predictions of the GP models (Fiedler et al., 2021). However, not only is this approach not scalable as it is incompatible with MOR and severely limited by the disadvantages of GPs themselves, but the resulting error bounds are often extremely conservative. To highlight this, consider a setup similar to the one presented here, with $N = 10^6$ data samples, confidence $(1 - \alpha) = 0.9$, and a state and input discretisation of size 9^7 and 3, respectively, to obtain an abstraction of similar size as above. Since exact inference via

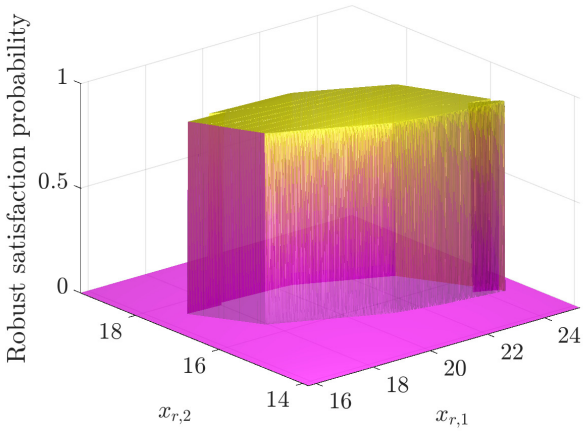


Fig. 2.17: Robust satisfaction probability as a function of the reduced initial state for $N = 10^6$ for the building automation system with a confidence bound of $(1-\alpha) = 0.9$.

GPs has a computational complexity of $O(N^3)$ (Rasmussen and Williams, 2006), this renders obtaining accurate large-scale models computationally excruciating. For sparse options, there are no results available that grant formal guarantees. Provided GP models with sample size N can be obtained, the dimension-wise error bound

$$\mathbb{P}\left(|\mu_{N,i}(x,u) - f_i(x,u;\theta^*)| \leq \mathcal{E}_i(x,u)\right) \geq 1-\alpha, \quad i \in \{1,\ldots,n_x\},$$

by Fiedler et al. (2021, Proposition 2) can be extrapolated[4] to be of magnitude $[\min_{x,u} \mathcal{E}_i(x,u)]_{i=1}^{n_x} := [3.1, 3.0, 3.4, 3.3, 3.2, 3.5, 3.4]$, where $\mu_{N,i}$ denotes the posterior mean function of the i^{th} GP, and $f := [f_1; \ldots; f_n]$ is the true dynamics function as defined in (2.2). Subsequently, constructing the IMDP from the GPs involves nonlinear optimisations with respect to $x \in \mathbb{X}$ and the error bounds, for a total of more than 68 trillion possible transitions for each of two transition matrices. This amounts to a serial complexity of $O(2(\kappa^{n_x} n_Q) \times (\eta_{\hat{\mathbb{X}}}^{n_x} n_Q) \times (\eta_{\hat{\mathbb{X}}}^{n_x} \eta_{\hat{\mathbb{U}}}^{m} n_Q))$, with $(\kappa^{n_x} n_Q)$ being the complexity of a single nonlinear optimisation (Wooding and Lavaei, 2024; Mathiesen et al., 2024). Storing one such dense matrix with double precision would take 500 TB of memory. In contrast, SySCoRe can leverage sparse matrices and tensor multiplications to circumvent this. Using GPs with an active set of 2000 data points, the same machine as before can compute 3.6 transition optimisations per second, rendering the IMDP construction infeasible. The subsequent control synthesis has an added complexity of $O(\eta_{\hat{\mathbb{X}}}^{n_x} \eta_{\hat{\mathbb{U}}}^{m} n_Q \log(\eta_{\hat{\mathbb{X}}}^{n} n_Q))$. As the error bounds $\mathcal{E}_i$ are large in comparison to the size of the partitions and the noise variance, the approach yields the trivial lower bound of zero for the satisfaction probability. In contrast,

[4] The average minimum error is computed over 10 runs for $N = 1000, 1500, \ldots, 5000$, and extrapolated linearly.

the approach proposed in this chapter can obtain comparatively tight statistical error bounds due to the linearity of the model in the uncertain parameters. Note that the robust satisfaction probability can be further strengthened via Proposition 2.3. This emphasises the powerful capabilities of the proposed synthesis approach via probabilistic simulation relations with respect to both scalability and accuracy.

2.6.3 Van der Pol Oscillator

As a renowned example of a nonlinear system, consider the state evolution of the Van der Pol Oscillator in discrete time, given by

$$x_{1,t+1} = x_{1,t} + x_{2,t}\tau_s + w_{1,t},$$
$$x_{2,t+1} = x_{2,t} + (-x_{1,t} + a(1 - x_{1,t}^2)x_{2,t})\tau + u_t + w_{2,t},$$

with parameter a and sampling time $\tau_s = 0.1$. The dynamics can be rewritten in the form of (2.3), where $f(x,u) = [x_1, x_2, x_1^2 x_2, u]^\top$, $h(x) = x$, and $\Sigma = 0.2I_2$. The following spaces are considered: the state space $\mathbb{X} = \mathbb{R}^2$, input space $\mathbb{U} = [-1,1]$, and output space $\mathbb{Y} = \mathbb{X}$. The goal is to design a controller such that the system remains inside region $p_S = [-3,3]^2$ until reaching target region $p_T = [2,3]\times[-1,1]$, written as $\psi = p_S \mathsf{U} p_T$.

For estimating the parameters, uniform inputs u are sampled from $\mathbb{U}$, the prior distributions are set to $\mathcal{N}(\theta_j|0, 10I_4)$, $j \in \{1,2\}$, and N data points are drawn from the unknown target system with parametrisation $\theta^* = [1, \tau_s, 0, 0; -\tau_s, 1+\tau_s a, -\tau_s a, 1]^\top$, where $a = 0.9$ and $\tau_s = 0.1$. Following similar steps as in previous case studies, the models $\widehat{\mathcal{M}}$, $\widetilde{\mathcal{M}}$ (discretising $\hat{\mathbb{X}}$ into 200^2 and $\hat{\mathbb{U}}$ into 5 partitions) and corresponding simulation relations are obtained with $\varepsilon_1 = 0$, $\max_{\hat{u}} \delta_1$ in Fig. 2.18 (top) and $\varepsilon_2 = 0.1$ and δ_2 given in Fig. 2.18 (bottom). The probability of satisfying the specification with $\varepsilon = \varepsilon_1 + \varepsilon_2$ and $\delta = \delta_1 + \delta_2$ is computed based on Corollary 2.1, and is given in Fig. 2.19 as a function of the initial state of $\mathcal{M}(\theta)$.

2.7 Conclusion

This chapter presented a new simulation relation for stochastic systems that can establish a quantitative connection between a parametrised class of models and a simpler nominal model. With this, it extends the applicability of previous work by enabling the synthesis of robust controllers from data and providing quantified guarantees for unknown nonlinear systems with unbounded noise and complex infinite-horizon specifications. Notably, the approach is compatible with MOR techniques and can hence be applied to higher-dimensional systems.

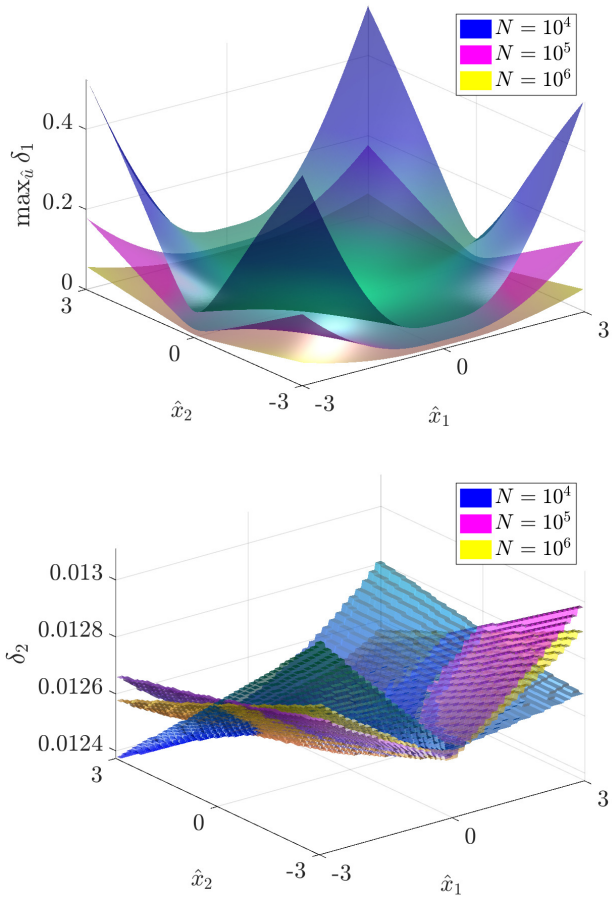


Fig. 2.18: Maximum value of δ_1 (top) with respect to the input as a function of the initial state, and δ_2 as a function of the initial state (bottom) for $N = 10^4, 10^5, 10^6$ for the package delivery case study with a confidence bound of $(1 - \alpha) = 0.9$.

The proposed approach is presented under certain simplifying assumptions, some of which can be directly relaxed — such as extensions to more expressive specification semantics and system classes.

2.7.1 Temporal Logic Extensions

Properties expressed as scLTL specifications can be represented as DFAs (Definition 2.5). Based on this fact, this chapter reformulated the satisfaction of an scLTL

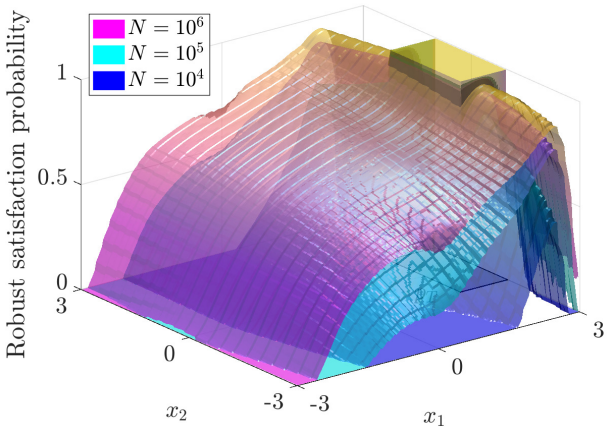


Fig. 2.19: Robust satisfaction probability as a function of the initial state for $N = 10^4, 10^5, 10^6$ (bottom to top) for the Van der Pol Oscillator with a confidence bound of $(1 - \alpha) = 0.9$.

specification as a reachability problem on the product system composed of the DFA and the abstracted system.

Signal Temporal Logic

Signal Temporal Logic Note that scLTL subsumes *signal temporal logic* (STL) (Maler and Nickovic, 2004; Lindemann et al., 2023). With scLTL, time can be explicitly encoded using, e.g., multiple instances of the next operator. In contrast to LTL, which is considered a qualitative semantic, STL specifications provide a quantitative semantic expressing the *degree* of satisfaction or robustness of the result against deviations. Thus, STL can be used to express by how much a signal satisfies the requirements or how far a signal is from satisfying the requirements. To this end, STL is founded on the following syntax (Maler and Nickovic, 2004):

$$\psi ::= \text{true} \mid p \mid \neg\psi \mid \psi_1 \wedge \psi_2 \mid \psi_1 \, \mathsf{U}_{[t_a,t_b]} \, \psi_2,$$

where $p \in \mathbb{Y}$ is an atomic proposition over a well-behaved signal in $\mathbb{Y} = \mathbb{R}^{n_x}$ and $t_a, t_b \in \mathbb{R}_{\geq 0}$ are discrete time instances with $a, b \in \mathbb{N}_0$. Analogously, propositions can be assigned to regions of the output space $\mathbb{Y}$ through the labelling function $\mathcal{L}$. Furthermore, the until operator $\psi_1 \, \mathsf{U}_{[t_a,t_b]} \, \psi_2$ is expressed via scLTL as follows:

$$\psi_1 \, \mathsf{U}_{[t_a,t_b]} \, \psi_2 = \vee_{n=a}^{b} \left(\bigcirc^n \psi_2 \wedge \left(\wedge_{i=0}^{n-1} \bigcirc^i \psi_1 \right) \right).$$

From this, all other STL expressions can be derived recursively using the structure of the formula and passing the negation through the formula using De Morgan's Laws in logic. Therefore, any STL specification can be expressed using a finite combination

of operators $\bigcirc$, $\wedge$, $\vee$, and negation applied to the atomic propositions. This renders STL a subset of scLTL, and thus the results of this chapter are directly applicable.

Finite-Trace Linear Temporal Logic

Finite-Trace Linear Temporal Logic The method described in this chapter is naturally extendable to *linear temporal logic* specifications over *finite-traces* (LTL_F) using the native approach outlined by Wells et al. (2020). LTL_F semantics are useful for formalising finite-horizon planning problems. With the approach by Wells et al. (2020), the results for SSRs can be extended to LTL_F by first translating the LTL_F formula to an equivalent first-order logic formula (Giacomo and Vardi, 2013) and subsequently generating a minimal DFA using, e.g., the tool MONA (Henriksen et al., 1995). Thereafter, the procedure remains the same, by constructing a product MDP and solving the reachability problem (see Haesaert and Soudjani, 2021).

2.7.2 System Extensions

While this chapter focused on systems with Gaussian noise, the approach is more broadly applicable to other (uncertain) noise distributions, as demonstrated in Chapter 3. Furthermore, to establish formal guarantees based on noisy data (i.e., when the observations are extended to be probabilistically conditioned on the latent state), relation (2.20) must be relaxed to accommodate the inflicted uncertainty. This requires adapting the presented coupling accordingly.

Currently, no definition of probabilistic simulation relations exists for *non*-Markovian systems. There is some work using CMEs addressing this challenge by reasoning over entire trajectories (Thorpe et al., 2022). Extending the simulation relations developed here to such non-Markovian or partially observed settings — beyond those linear POMDPs studied by (Haesaert et al., 2018) — remains an open and highly relevant direction for future research.

Chapter 3
Correct-by-Design Synthesis for Networks of Uncertain Systems

Abstract In this chapter, the synthesis approach from Chapter 2 is extended to CPSs composed of multiple dynamically interconnected uncertain subsystems. Networked discrete-time nonlinear systems with additive stochastic noise and parametrised model uncertainty are considered, as arise when systems interact in uncertain environments and only observational data is available. Two main limitations of the previous approach are addressed. First, Gaussian noise is often assumed for tractability, but in networked settings, the heterogeneity of subsystems and their interactions typically leads to more complex, non-Gaussian behaviour. Second, whilst model uncertainty was handled monolithically in the previous chapter, here, a compositional view is taken by reasoning about uncertainty at the subsystem level and synthesising guarantees for the composed system. Abstraction-based synthesis using sub-probability couplings is extended accordingly, with controllers constructed from data-driven parameter sets and arbitrary noise approximated via Gaussian mixtures.



3.1 Introduction

In the previous chapter, it was shown that efficient formal controller synthesis for nonlinear systems subject to both aleatoric and epistemic uncertainty can be enabled through abstraction-based techniques when a parametric model structure is available and uncertainty is confined to unknown parameters. The approach proved effective in producing guarantees, particularly when combined with model order reduction to mitigate computational complexity. However, it was also evident that scalability remains a core challenge: as system dimensionality increases, the number of discrete states grows exponentially, rendering synthesis for a single monolithic abstraction increasingly intractable.

This limitation becomes only more pronounced when considering realistic environments, where autonomous systems operate not in isolation but in coordination with other agents. For instance, in traffic, manoeuvres must be performed amidst surrounding vehicles and in response to previously unseen situations. Such interactions give rise to *networked* systems — interconnected collections of dynamical agents. While the individual dynamics may again be given in parametric form, the relevant parameters are often uncertain and must be estimated from data. Treating the overall network as a single system is typically infeasible due to the resulting state explosion.

This class of problems arises in numerous domains, including transportation, smart grids, industrial automation, and robotic swarms. To enable tractable design, a decomposition-based strategy is typically desirable, where synthesis is performed at subsystem level and network-level correctness is inferred compositionally. This concept — referred to as *compositionality* — has been extensively studied within the formal methods community (Arcak et al., 2016; Alur et al., 2001; Hahn et al., 2013). However, the majority of abstraction-based approaches for networked stochastic systems remain limited to the model-based setting. One exception is Lavaei et al. (2023), who use model-free reinforcement learning, relying on known Lipschitz constants and limited to finite-horizon specifications. In fact, most work not only assumes that exact models of all subsystems are available (Mallik et al., 2018; Lavaei et al., 2021), but may additionally restrict stochastic behaviour to be either bounded (Majumdar et al., 2023a) or Gaussian (Lavaei et al., 2020b). While these assumptions can simplify analysis, they are often misaligned with the complex and heterogeneous stochastic behaviour exhibited by real-world networks (Blackmore et al., 2010; Sun et al., 2006), a characteristic largely neglected in prior work.

In contrast, the approach presented here extends formal synthesis to networked systems subject to both epistemic and *complex* aleatoric uncertainty. To this end, the notion of sub-simulation relations (SSRs) introduced in Chapter 2 — originally applied to monolithic parametrised systems — is now considered in a compositional setting. It is shown that local SSRs between uncertain subsystems can be composed to form a global relation at network level, enabling decentralised synthesis for global specifications formulated as conjunctions of local properties. Global guarantees are established for both cascaded and cyclic network structures. To accommodate realistic stochastic behaviour, arbitrary additive noise distributions are approximated using *Gaussian mixture models* (GMMs), which serve as surrogate models of the unknown true noise. Note that this contribution extends beyond the compositional framework presented here. While parametric uncertainty was previously considered confined to the deterministic component of the dynamics, with the stochastic law known a priori, this chapter relaxes that assumption. Importantly, it is shown that the approximation error introduced by the GMM representation can be bounded even when the true noise distribution is unknown. This enables the construction of abstractions that are robust to both model uncertainty and complex stochastic disturbances — advancing compositional synthesis under uncertainty toward real-world applicability.

Organisation The chapter is organised as follows. In Section 3.2, the class of networked systems, local controllers, and specifications is defined, and a two-stage problem statement is formulated. Solutions to these subproblems are presented in Section 3.3 and Section 3.4, respectively. In particular, Section 3.3 extends the SSR definition to networked systems and establishes compositional results. Section 3.4 quantifies the closeness between two systems with GMM noise distributions. Finally, the capabilities of the proposed compositional approach are demonstrated on a nonlinear package delivery case study and a vehicle platoon in Section 3.5.

3.2 Preliminaries and Problem Statement

The setup considered in this chapter is formalised by first introducing the networked system structure, and the associated local control policies and specifications. Based on this, a two-stage control synthesis problem is derived.

3.2.1 Networks of Uncertain Stochastic Systems

This chapter considers networked discrete-time uncertain nonlinear systems with two sources of uncertainty: additive stochastic noise and parametric uncertainty. Fig. 3.1 shows a vehicle platoon as an illustrative example of such a networked system. Systems of this class can be represented by a model $\boldsymbol{\mathcal{M}}(\theta)$ parametrised by $\theta = [\theta^1; \ldots; \theta^N] \in \Omega^1 \times \ldots \times \Omega^N$ and partitioned into N subsystems

$$\boldsymbol{\mathcal{M}}^i(\theta^i): \begin{cases} x_{t+1}^i = f_{\theta^i}^i(x_t^i, u_t^i) + w_t^i, & w_t^i \sim p_w^i(\cdot\,|\theta^i) \\ y_t^i = h^i(x_t^i) \end{cases}, \quad i \in \{1, \ldots, N\}, \tag{3.1}$$

where $x_t^i \in \mathbb{X}^i \subset \mathbb{R}^{n_x^i}$, $u_t^i \in \mathbb{U}^i \subset \mathbb{R}^{n_u^i}$, and $y_t^i \in \mathbb{Y}^i \subset \mathbb{R}^{n_y^i}$ denote the state, input, and observation of subsystem $\boldsymbol{\mathcal{M}}^i(\theta^i)$ at time-step $t \in \mathbb{N}_0$, respectively. The state evolution and observation mapping are described by the functions $f^i : \mathbb{X}^i \times \mathbb{U}^i \times \Omega^i \to \mathbb{X}^i$ and $h^i : \mathbb{X}^i \to \mathbb{Y}^i$. The additive noise $w_t^i \in \mathbb{W}$ is an i.i.d. sequence with *arbitrary* continuous distribution $w_t^i \sim p_w^i(\cdot\,|\theta^i)$. Note that in contrast to Chapter 2, here, both the state evolution $f_{\theta^i}^i$ and the additive noise distribution $p_w^i(\cdot\,|\theta^i)$ are conditioned on an uncertain parametrisation θ^i — which is assumed to lie within a given uncertainty set Θ^i such that $\Theta := \prod_i \Theta^i$ — and the noise distribution p_w^i may be non-Gaussian. This uncertainty set can be inferred from observed input-output data, e.g., via system identification techniques as showcased in Chapter 2.

Input-Output Configuration

The input and output vectors of each subsystem $\boldsymbol{\mathcal{M}}^i(\theta^i)$, $i \in \{i, \ldots, N\}$ are assumed to be partitioned as $u_t^i = [u_t^{i1}; \ldots; u_t^{iN}]$ and $y_t^i = [y_t^{i1}; \ldots; y_t^{iN}]$, respectively. Internal coupling between subsystems is modelled by setting their *internal* $u_t^{ij} = y_t^{ji}$ for all $i \neq j$ (see Fig. 3.3). The components $u_t^{ii} \in \mathbb{U}^{ii}$ and $y_t^{ii} \in \mathbb{Y}^{ii}$ are referred to as the *external* input and output, respectively. For the overall network $\boldsymbol{\mathcal{M}}(\theta)$, the state,

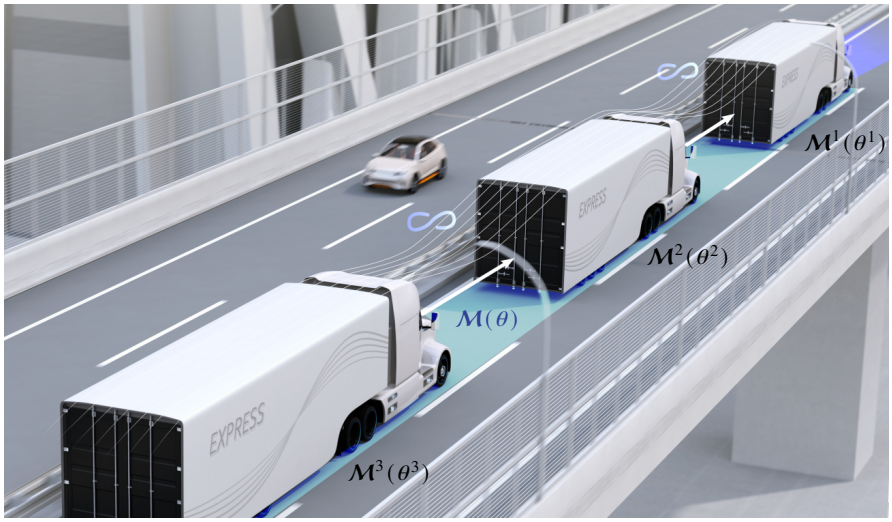


Fig. 3.1: A vehicle platoon as an example of a networked system: The platoon $\mathcal{M}(\theta)$ arises as the interconnection of the individual vehicles $\mathcal{M}^i(\theta^i)$, $i \in \{1, 2, 3\}$. The coordinated motion enables improved safety, fuel efficiency, and extended situational awareness through shared sensing.

input, and output vectors are recovered by concatenation:

$$x_t = [x_t^1; \ldots; x_t^N], \quad u_t = [u_t^1; \ldots; u_t^N], \quad y_t = [y_t^1; \ldots; y_t^N].$$

Notation For convenience, a column vector $x = [x^1; \ldots; x^n] \in \mathbb{R}^n$ deprecated by its i^{th} component is written as $x^{-i} := [x^1; \ldots; x^{i-1}; x^{i+1}; \ldots; x^n]$. Similarly, the product of sets $\{A^i\}_{i=1}^n$ with the i^{th} set removed is written as $A^{-i} := \prod_{j \neq i} A^j$.

3.2.2 Gaussian Mixture Models

The noise distribution $p_w^i(w_t^i|\theta^i)$ is modelled as a *Gaussian mixture model* (GMM). As a subclass of finite mixture models, GMMs are a widely used modelling framework for approximating probability distributions. Apart from being particularly useful for capturing multiple sources of randomness, GMMs can approximate any continuous distribution with arbitrary precision (McLachlan and Peel, 2000).

Gaussian Mixture Model

Definition 3.1 (Gaussian Mixture Model (GMM)) A *GMM* is a probability measure $\mathcal{NW}_K : \mathscr{B}(\mathbb{X}) \to [0, 1]$, which is a weighted sum of finitely many ($K \in \mathbb{N}$) Gaussian *component densities* $\mathcal{N}(x|\mu_k, \Sigma_k)$, $k \in \{1, \ldots, K\}$, i.e.,

$$\mathcal{MN}_K(dx|\pi,\mu,\Sigma) := \sum_{k=1}^{K} \pi_k \, \mathcal{N}(dx|\mu_k,\Sigma_k),$$

with *mixing weights* $\pi := (\pi_1,\ldots,\pi_K)$, $0 \leq \pi_k \leq 1$, $\sum_{k=1}^{K} \pi_k = 1$, mean values $\mu := (\mu_1,\ldots,\mu_K)$, and covariance matrices $\Sigma := (\Sigma_1,\ldots,\Sigma_K)$.

A GMM is called *homoscedastic*, if all K components share the same covariance matrix, and *heteroscedastic* otherwise. Fig. 3.2 depicts an example of a heteroscedastic GMM. When the number of components K is uncertain, the index indicating the number of components is omitted, i.e., $\mathcal{MN}(\,\cdot\,|\pi,\mu,\Sigma)$.

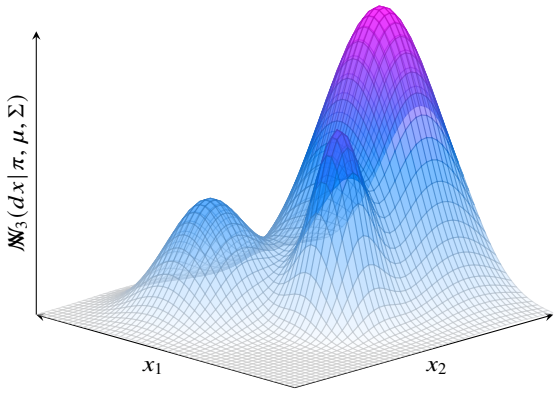


Fig. 3.2 Example of a 2D Gaussian mixture with $K = 3$ components: $(\pi_1 = 0.2, \mu_1 = [1;-1], \Sigma_1 = 0.5I)$, $(\pi_2 = 0.3, \mu_2 = [-1;0], \Sigma_2 = 0.2I)$, and $(\pi_3 = 0.5, \mu_3 = [0;2], \Sigma_3 = I)$.

For the networked system in (3.1), the noise distributions can hence be written as $w_t^i \sim \mathcal{MN}(\,\cdot\,|\pi^i,\mu^i,\Sigma^i) \equiv \mathcal{MN}(\,\cdot\,|\theta^i)$, where the parameters (π^i,μ^i,Σ^i) are contained in the unknown parametrisation θ^i.

Remark 3.1 (Learning a GMM from Data) Any continuous probability distribution can be approximated arbitrarily precise using a GMM (McLachlan and Peel, 2000). One method for determining the parameters (π,μ,Σ) based on data is using maximum likelihood fitting via the *expectation maximisation* (EM) algorithm. The EM algorithm is an iterative procedure, where every iteration $l \in \mathbb{N}$ comprises two steps. In the E-step, the *expected* likelihood of some parametrisation $(\pi,\mu,\Sigma)^{(l)}$ is computed. The consecutive M-step involves the *maximisation* of mentioned likelihood to find the next parametrisation $(\pi,\mu,\Sigma)^{(l+1)}$. This procedure is repeated until convergence, yielding a maximum likelihood estimate $(\hat{\pi},\hat{\mu},\hat{\Sigma})$.

3.2.3 Local Controllers

For each subsystem $\boldsymbol{\mathcal{M}}^i(\theta^i)$ in (3.1) initialised at an initial state x_0^i at $t = 0$ and an input sequence $u_0^i, u_1^i, u_2^i, \ldots$, consecutive states $x_{t+1}^i \in \mathbb{X}^i$ are obtained as realisations $x_{t+1}^i \sim f_{\theta^i}^i(x_t^i, u_t^i; \theta^i) + \mathcal{W}(w_t^i|\theta^i)$. Recall that the execution history $(x_0^i, u_0^i, x_1^i, \ldots, u_{N-1}^i, x_N^i)$ grows with the number of observations (N), and takes values in a history space $\mathbb{H}_N^i := (\mathbb{X}^i \times \mathbb{U}^i)^N \times \mathbb{X}^i$.

Local Control Policies

In contrast to the monolithic treatment in Chapter 2, here, each subsystem $\boldsymbol{\mathcal{M}}^i(\theta^i)$ is equipped with a local controller, which determines which inputs to apply (see Fig. 3.3). Formally, a *local control policy* $\boldsymbol{\rho}^i = (\rho_0^i, \rho_1^i, \rho_2^i, \ldots)$ is a sequence of universally measurable maps $\rho_t^i : \mathbb{H}_t^i \to \mathcal{P}(\mathbb{U}^{ii})$, $t \in \mathbb{N}_0$, from the execution histories to distributions over external inputs $u^{ii} \in \mathbb{U}^{ii}$. As in Chapter 2, it is sufficient to consider finite memory policies for the satisfaction of temporal logic specifications on the system executions. As the deterministic implementation of such a policy, the local controller for a given subsystem $\boldsymbol{\mathcal{M}}^i(\theta^i)$ acts as a state-feedback law

$$u_t^{ii} = \boldsymbol{C}^i(x_t^i),$$

with the closed-loop composition denoted as $\boldsymbol{C}^i \bowtie \boldsymbol{\mathcal{M}}^i(\theta^i)$.

3.2.4 Local Specifications

It is assumed that each local subsystem $\boldsymbol{\mathcal{M}}^i$ associates a corresponding labelling function $\mathcal{L}^i : \mathbb{Y}^i \to \Sigma^i$ assigning a letter to each output. Using this labelling map, temporal logic specifications can be defined over the output of the subsystem as described in Subsection 2.2.3. Analogous to the monolithic case, the probability of the controlled subsystem $\boldsymbol{C}^i \bowtie \boldsymbol{\mathcal{M}}^i$ satisfying its specification ψ^i is given by $\mathbb{P}(\boldsymbol{C}^i \bowtie \boldsymbol{\mathcal{M}}^i \models \psi^i)$.

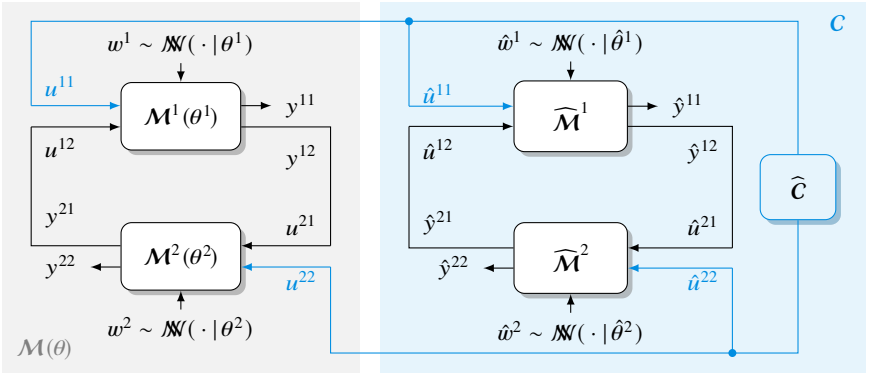
The approach presented in this chapter is restricted by the common assumption that global specifications can be decomposed into local subspecifications.

Assumption 3.1 (Local Specifications) Let the system be decomposable into N subsystems $\boldsymbol{\mathcal{M}}^1, \ldots, \boldsymbol{\mathcal{M}}^N$ and the global specification ψ into N local scLTL specifications $\psi^1, \ldots, \psi^N$ (see Definition 2.1 for the scLTL syntax) such that $\psi = \wedge_{i=1}^N \psi^i$.

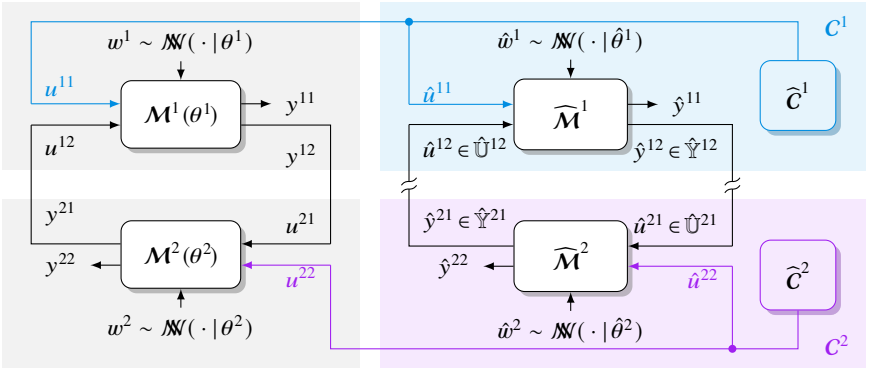
3.2.5 Problem Statement

When synthesising a controller for the entire network, the problem size grows exponentially with the number of subsystems involved — for instance, when designing a global safety controller (see Fig. 3.3(a)). To mitigate this complexity, compositional reasoning is employed: the global synthesis problem is decomposed into local control tasks (see Fig. 3.3(b)) by decoupling subsystems under the assumption of

bounded influence (as commonly done in related work (Lavaei et al., 2022a)). The local control problems operate over the much lower-dimensional dynamics of the individual subsystems, substantially reducing the computational burden imposed by the exponential complexity of formal synthesis.



(a) Global control loop (simplified with identity interface functions $\mathbf{i}^1, \mathbf{i}^2$).



(b) Local control loops (simplified with identity interface functions $\mathbf{i}^1, \mathbf{i}^2$).

Fig. 3.3: Global controller (a) vs local controllers (b).

Given the structure of the network and the form of the local specifications defined earlier, the problem is approached by solving the following two subproblems.

Subproblem 3.1 (Composition) Consider a networked uncertain system $\boldsymbol{\mathcal{M}}(\theta)$ with N subsystems $\boldsymbol{\mathcal{M}}^i(\theta^i)$ of the form (3.1) and a global specification ψ satisfying Assumption 3.1. Let thresholds $p_{\psi^i} \in (0,1)$ be given for each subsystem $i \in \{1,\ldots,N\}$. Design a global controller $\boldsymbol{C}$ with lower bound $p_\psi \in (0,1)$ such that

$$\mathbb{P}\left(\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}}(\theta) \models \psi\right) \geq p_{\psi}, \quad \forall \theta \in \Theta = \textstyle\prod_i \Theta^i,$$

by designing a local controller $\boldsymbol{C}^i$ for each $\boldsymbol{\mathcal{M}}^i(\theta^i)$, $i \in \{1, \ldots, N\}$, such that

$$\mathbb{P}\left(\boldsymbol{C}^i \bowtie \boldsymbol{\mathcal{M}}^i(\theta^i) \models \psi^i \,|\, \boldsymbol{\omega}^j \models \psi^j, j \neq i\right) \geq p_{\psi^i}, \quad \forall \theta^i \in \Theta^i.$$

Given a solution to Subproblem 3.1, the resulting N local control problems can be addressed individually as follows.

Subproblem 3.2 (Robust Synthesis (Extended)) For a given scLTL specification ψ^i and threshold $p_{\psi^i} \in (0, 1)$, design a controller $\boldsymbol{C}^i$ that robustly satisfies ψ^i uniformly for the set of models $\{\boldsymbol{\mathcal{M}}^i(\theta^i),\ \theta^i \in \Theta^i\}$ with probability at least p_{ψ^i}:

$$\mathbb{P}\left(\boldsymbol{C}^i \bowtie \boldsymbol{\mathcal{M}}^i(\theta^i) \models \psi^i\right) \geq p_{\psi^i}, \quad \forall \theta^i \in \Theta^i.$$

Analogous to Subproblem 2.2 in Chapter 2, Subproblem 3.2 is solved for each subsystem $\boldsymbol{\mathcal{M}}^i(\theta^i)$, $i \in \{1, \ldots, N\}$, by computing a robust controller for a selected nominal system $\widehat{\boldsymbol{\mathcal{M}}}^i$ in the set of feasible models $\{\boldsymbol{\mathcal{M}}^i(\theta^i) | \theta^i \in \Theta^i\}$, and constructing an SSR between the nominal system and the set of models. Recall that a valid refinement requires neither the interface function nor the state refinement to be dependent on the concrete parametrisation θ^i. Note that here, in contrast to Chapter 2, both the deterministic and the stochastic part of the system dynamics are considered uncertain and the noise distribution is non-Gaussian.

To solve Subproblem 3.1, local controllers are synthesised by bounding the internal inputs and outputs of the network and including the satisfaction of these bounds in the local specifications. A lower bound on the satisfaction probability of the networked system $\boldsymbol{\mathcal{M}}(\theta)$ is derived by showing that the local SSRs induce a global SSR, similar to the compositional approach by Lavaei et al. (2021). To solve Subproblem 3.2, the results for establishing SSRs presented in Chapter 2 are extended to systems with parametrised GMM noise distributions.

3.3 Sub-Simulation Relations for Networks of Uncertain Systems

Following the approach by Lavaei et al. (2021), the solution to Subproblem 3.1 is based on constructing individual nominal models $\widehat{\boldsymbol{\mathcal{M}}}^i$ for each subsystem $\boldsymbol{\mathcal{M}}^i$. Local abstraction-based controllers $\widehat{\boldsymbol{C}}^i$ are then synthesised for the nominal abstractions. Using local SSRs between each pair $(\boldsymbol{\mathcal{M}}^i, \widehat{\boldsymbol{\mathcal{M}}}^i)$, the controllers $\widehat{\boldsymbol{C}}^i$ can subsequently be refined to concrete controllers $\boldsymbol{C}^i$ for the original subsystems $\boldsymbol{\mathcal{M}}^i$ (see Fig. 3.3(b)). Fig. 3.4 illustrates the overall setup.

In this section, it is shown how the SSR in Definition 2.8 can be extended to networks of uncertain systems, and the associated compositional results are derived.

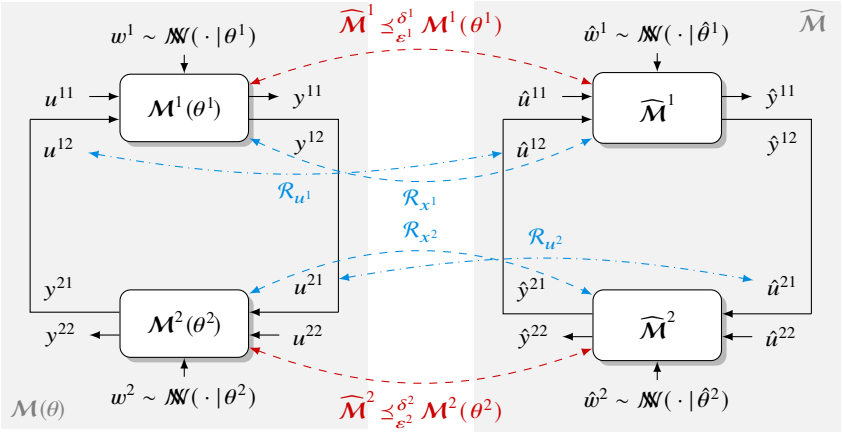


Fig. 3.4: The network with parametric uncertainty, $\mathcal{M}(\theta)$, (left) and its nominal abstraction $\widehat{\mathcal{M}}$ (right) are related by establishing simulation relations for each subsystem pair. These simulation relations quantify the similarity of the systems based on relations on the states and internal inputs.

3.3.1 Models and Interconnections

The individual subsystems $\mathcal{M}^i$ in (3.1) are represented as gMDPs, as defined in Definition 2.4. To model the network of subsystems, the *interconnection* of N gMDPs is formally defined as follows (see Fig. 3.4).

Interconnection

Definition 3.2 (Interconnection of Subsystems) Consider $N \in \mathbb{N}$ gMDPs $\mathcal{M}^i = (\mathbb{X}^i, \mathbb{X}_0^i, \mathbb{U}^i, \mathbf{t}^i, h^i, \mathbb{Y}^i)$, $i \in \{1, \dots, N\}$, with an input-output configuration that can be partitioned as outlined in Subsection 3.2.1. Then, the *interconnection* of the subsystems is itself a gMDP $\mathcal{M} = (\mathbb{X}, \mathbb{X}_0, \mathbb{U}, \mathbf{t}, h, \mathbb{Y})$, also denoted as $\mathcal{I}(\mathcal{M}^1, \dots, \mathcal{M}^N)$, with $\mathbb{X} = \prod_i \mathbb{X}^i$, $\mathbb{X}_0 = \prod_i \mathbb{X}_0^i$, $\mathbb{U} = \prod_i \mathbb{U}^i$, $\mathbf{t} = \prod_i \mathbf{t}^i$, $h = [h^{11}; \dots; h^{NN}]$, $\mathbb{Y} = \prod_i \mathbb{Y}^i$, and *interconnection constraints*

$$\forall i \neq j: \quad u_t^{ij} = y_t^{ji}, \quad \mathbb{Y}^{ji} \subseteq \mathbb{U}^{ij}. \tag{3.2}$$

The SSR definition for networked systems is presented in the next subsection.

3.3.2 Relation on Subsystem Level

For each subsystem $\mathcal{M}^i$ in (3.1), an abstract nominal model is constructed:

$$\widehat{\mathcal{M}}^i: \begin{cases} \hat{x}_{t+1}^i = f_{\hat{\theta}^i}^i(\hat{x}_t^i, \hat{u}_t^i) + \hat{w}_t^i, \quad \hat{w}_t^i \sim \mathcal{W}(\cdot|\hat{\theta}^i) \\ \hat{y}_t^i \quad = \hat{h}^i(x_t^i) \end{cases}, \tag{3.3}$$

with nominal parameters $\hat{\theta}^i$. Note that $\widehat{\mathcal{M}}^i \equiv \mathcal{M}^i(\hat{\theta}^i)$ if $\hat{h}^i \equiv h^i$. Individual SSRs are established for each subsystem pair $(\mathcal{M}^i(\theta^i), \widehat{\mathcal{M}}^i)$ using the sub-probability coupling introduced in Definition 2.7, which are independent of the concrete uncertain parametrisations θ^i.

The following definition is central to the approach developed in this chapter, extending the original SSR from Definition 2.8 to systems with both internal and external inputs.

Local SSR

Definition 3.3 ($(\varepsilon^i, \delta^i)$-Sub-Simulation Relation (SSR)) Consider two gMDPs $\mathcal{M}^i = (\mathbb{X}^i, \mathbb{X}_0^i, \mathbb{U}^i, \mathbf{t}^i, h^i, \mathbb{Y}^i)$ and $\widehat{\mathcal{M}}^i = (\hat{\mathbb{X}}^i, \hat{\mathbb{X}}_0^i, \hat{\mathbb{U}}^i, \hat{\mathbf{t}}^i, \hat{h}^i, \mathbb{Y}^i)$, relations $\mathcal{R}_{x^i} \subset \mathscr{B}(\hat{\mathbb{X}}^i \times \mathbb{X}^i)$ and $\mathcal{R}_{u^{-i}} \subset \mathscr{B}(\hat{\mathbb{U}}^{-i} \times \mathbb{U}^{-i})$, and an interface function $\mathbf{i}^i : \hat{\mathbb{X}}^i \times \mathbb{X}^i \times \hat{\mathbb{U}}^i \times \mathscr{B}(\mathbb{U}^{ii}) \to [0,1]$. If there exists a sub-probability kernel $\boldsymbol{v}^i : \hat{\mathbb{X}}^i \times \mathbb{X}^i \times \hat{\mathbb{U}}^i \times \mathbb{U}^{-i} \times \mathscr{B}(\hat{\mathbb{X}}^i \times \mathbb{X}^i) \to [0,1]$, such that

(a) $\forall x_0^i \in \mathbb{X}_0^i, \exists \hat{x}_0^i \in \hat{\mathbb{X}}_0^i \colon (\hat{x}_0^i, x_0^i) \in \mathcal{R}_{x^i}$;
(b) $\forall (\hat{x}^i, x^i) \in \mathcal{R}_{x^i}, \forall \hat{u}^i \in \hat{\mathbb{U}}^i$ there is $u^{ii} \sim \mathbf{i}^i(\,\cdot\,|\hat{x}^i, x^i, \hat{u}^i)$ s.t. $\forall u^{-i} \in \mathcal{R}_{u^{-i}}^{-1} \hat{u}^{-i}$, $\boldsymbol{v}^i(d\hat{x}_+^i \times dx_+^i | \hat{x}^i, x^i, \hat{u}^i, u^{-i})$ is a sub-probability coupling of $\hat{\mathbf{t}}^i(d\hat{x}_+^i | \hat{x}^i, \hat{u}^i)$ and $\mathbf{t}^i(dx_+^i | x^i, u^i)$ over $\mathcal{R}_{x^i}$ with respect to δ^i (see Definition 2.7);
(c) $\forall (\hat{x}^i, x^i) \in \mathcal{R}_{x^i} \colon \mathbf{d}_{\mathbb{Y}}(\hat{h}^i(\hat{x}^i), h^i(x^i)) \leq \varepsilon^i$;

then $\widehat{\mathcal{M}}^i$ is in an $(\varepsilon^i, \delta^i)$-*SSR* with $\mathcal{M}^i$, denoted by $\widehat{\mathcal{M}}^i \preceq_{\varepsilon^i}^{\delta^i} \mathcal{M}^i$.

The intuition is similar to the original monolithic SSR: Definition 3.3 imposes three conditions on the composed system $\widehat{\mathcal{M}}^i \times \mathcal{M}^i$ evolving on the product space $\hat{\mathbb{X}}^i \times \mathbb{X}^i$. They roughly correspond to its initial state, state transition, and output mapping (see Fig. 3.4). Condition (a) requires the states of both systems to start in the relation $\mathcal{R}_{x^i}$ upon initialisation. Once in $\mathcal{R}_{x^i}$, condition (b) certifies that for any external input $\hat{u}^{ii}$ to $\widehat{\mathcal{M}}^i$ there exists a corresponding external input u^{ii} to $\mathcal{M}^i$ such that the systems stay in relation $\mathcal{R}_{x^i}$ with probability at least $(1 - \delta^i)$ provided that the internal inputs of the two systems remain in relation $\mathcal{R}_{u^{-i}}$; In effect, $\mathcal{R}_{u^{-i}}$ acts as a contract on the allowed subsystem interactions. Note that the external input $\hat{u}^{ii}$ is mapped onto u^{ii} by an interface function. Finally, according to condition (c), given that the two systems are in relation $\mathcal{R}_{x^i}$, the corresponding outputs shall be ε^i-close.

Definition 3.3 reduces to the original definition of (ε, δ)-sub-simulation relations for monolithic systems (Definition 2.8) if $N = 1$ (thus equivalently $i \in \{1\}$).

In the next section, it is shown that — as Lavaei et al. (2021) demonstrated for approximate simulation relations — local SSRs between the interconnected subsystems induce a global SSR between the resulting networks.

3.3.3 Relation on Network Level

The introduced framework is now applied to networks subject to parametric uncertainty. In contrast to Chapter 2, where an SSR was directly established on

the entire system and its nominal counterpart, here, individual SSRs $\widehat{\mathcal{M}}^i \preceq^{\delta^i}_{\varepsilon^i} \mathcal{M}^i$ are established between each pair of subsystems $(\mathcal{M}^i(\theta^i), \widehat{\mathcal{M}}^i)$ as defined in (3.1) and (3.3). To establish the relation, the interface function is chosen as $u^{ii} \sim \mathbf{i}^i(\cdot|\hat{x}^i, x^i, \hat{u}^i) := \hat{u}^{ii}$, and the noise is coupled via

Noise Coupling

$$\hat{w}^i \equiv \gamma^i_{\theta^i}(x^i, u^i) + w^i, \qquad \text{with an offset} \tag{3.4}$$

$$\gamma^i_{\theta^i}(x^i, u^i) := f^i_{\theta^i}(x^i, u^i) - f^i_{\hat{\theta}^i}(x^i, u^i), \tag{3.5}$$

so that the resulting state mapping is independent of the uncertain parameters θ:

$$\hat{x}^i_+ = x^i_+ - f^i_{\hat{\theta}^i}(x^i, u^i) + f^i_{\hat{\theta}^i}(\hat{x}^i, \hat{u}^i). \tag{3.6}$$

Notably, the additive noise coupling in (3.4) implies that the distributions $\mathcal{W}_{K^i}(w^i|\theta^i)$ and $\mathcal{W}_{\hat{K}^i}(\hat{w}^i|\hat{\theta}^i)$ must be of pairwise identical covariance structure, i.e., with a matching number of components $\hat{K}^i = K^i$ and shared covariances $\hat{\Sigma}^i_k = \Sigma^i_k$ for all $k \in \{1, \dots, K\}$. Importantly, this is not overly restrictive, as the resulting formulation does not limit the expressiveness of addressable noise distributions and is more general than assuming homoscedasticity. The following relations are used to find a convenient, closed-form coupling expression:

$$\mathcal{R}_{x^i} := \left\{(\hat{x}^i, x^i) \in \hat{\mathbb{X}}^i \times \mathbb{X}^i | \hat{x}^i = x^i\right\}, \tag{3.7}$$

$$\mathcal{R}_{u^{-i}} := \left\{(\hat{u}^{-i}, u^{-i}) \in \hat{\mathbb{U}}^{-i} \times \mathbb{U}^{-i} \,|\, \hat{u}^{-i} = u^{-i}\right\}, \tag{3.8}$$

$$\begin{aligned}\mathcal{R}_x &:= \left\{(\hat{x}, x) \in \hat{\mathbb{X}} \times \mathbb{X} \,|\, \hat{x} = x\right\}, \\ &\equiv \left\{(\hat{x}, x) \in \hat{\mathbb{X}} \times \mathbb{X} \,|\, \forall i \in \{1, 2\} : \ (\hat{x}^i, x^i) \in \mathcal{R}_{x^i}\right\}.\end{aligned} \tag{3.9}$$

Condition (a) of Definition 3.3 is satisfied by setting the initial states $\hat{x}^i_0 = x^i_0$, for all $i \in \{1, \dots, N\}$. For condition (b), a sub-probability coupling $\boldsymbol{v}^i_{\theta^i}$ (Definition 2.7) of the transition kernels $\hat{\mathbf{t}}^i$ and $\mathbf{t}^i_{\theta^i}$ over $\mathcal{R}_{x^i}$ is defined as

$$\begin{aligned}\boldsymbol{v}^i_{\theta^i}(d\hat{x}^i_+ \times dx^i_+ | \cdot) = \int_{w^i} &\delta_{f^i_{\hat{\theta}^i}(x^i,u^i)+w^i+\gamma^i_{\theta^i}}(d\hat{x}^i_+)\delta_{f^i_{\theta^i}(x^i,u^i)+w^i}(dx^i_+) \\ &\min\left\{\mathcal{W}_{K^i}(dw^i|\theta^i), \mathcal{W}_{K^i}(dw^i|\hat{\pi}^i, \hat{\mu}^i - \gamma^i_{\theta^i}, \Sigma^i)\right\},\end{aligned} \tag{3.10}$$

where the dependence of $\boldsymbol{v}^i_{\theta^i}$ and $\gamma^i_{\theta^i}$ on (x^i, u^i) is omitted for brevity. This coupling $\boldsymbol{v}^i$ guarantees that the marginal distributions form lower bounds on the transition probabilities of $\mathcal{M}^i(\theta^i)$ and $\widehat{\mathcal{M}}^i$, thus ensuring that condition (b) of the SSR definition is met.

Based on these components, the following result shows that the local SSRs induce a global SSR between the corresponding interconnected systems.

Induced Compositional SSR

Theorem 3.1 (Induced Compositional SSR) *Consider gMDPs* $\{\mathcal{M}^i(\theta^i)\}_{i=1}^N$ *and* $\{\widehat{\mathcal{M}}^i\}_{i=1}^N$ *as in* (3.1) *and* (3.3) *with a common metric* $\mathbf{d}_{\mathbb{Y}} = \|\cdot\|$, *and let*

$$\boldsymbol{\mathcal{M}}(\theta) = \mathcal{J}(\boldsymbol{\mathcal{M}}^1(\theta^1), \ldots, \boldsymbol{\mathcal{M}}^N(\theta^N)) \quad \textit{and} \quad \widehat{\boldsymbol{\mathcal{M}}} = \mathcal{J}(\widehat{\boldsymbol{\mathcal{M}}}^1, \ldots, \widehat{\boldsymbol{\mathcal{M}}}^N)$$

be the corresponding interconnections (Definition 3.2). For all $i \in \{1, \ldots, N\}$, *let SSRs* $\widehat{\boldsymbol{\mathcal{M}}}^i \preceq^{\delta^i}_{\varepsilon^i} \boldsymbol{\mathcal{M}}^i(\theta^i)$ *with relations* $\mathcal{R}_{x^i}$ *and* $\mathcal{R}_{u^{-i}}$ *in* (3.7)–(3.8), *sub-probability couplings* $\boldsymbol{v}^i_{\theta^i}$ *in* (3.10), *and interface functions* $u^{ii} = \hat{u}^{ii}$ *be given. If for all* $i \in \{1, \ldots, N\}$, $j \neq i$, *and* $\hat{u}^{ij} \equiv h^{ji}(x^j)$, *the condition*

$$\forall (\hat{x}^i, x^i) \in \mathcal{R}_{x^i} : \ (\hat{u}^{-i}, u^{-i}) \in \mathcal{R}_{u^{-i}}$$

holds, then $\widehat{\boldsymbol{\mathcal{M}}}$ *is in an* induced compositional (ε, δ)-SSR *with* $\boldsymbol{\mathcal{M}}(\theta)$ *(Definition 2.8) — denoted as* $\widehat{\boldsymbol{\mathcal{M}}} \preceq^{\delta}_{\varepsilon} \boldsymbol{\mathcal{M}}(\theta)$ *— with relation* $\mathcal{R}_x$ *in* (3.9) *and error parameters* $\varepsilon = \sum_i \varepsilon^i$ *and* $\delta = 1 - \prod_i (1 - \delta^i)$. *The corresponding sub-probability coupling* $\boldsymbol{v}_\theta = \prod_i \boldsymbol{v}^i_{\theta^i}$ *and the interface function* $\mathbf{i} = \prod_i \mathbf{i}^i$ *follow by applying the interconnection constraints* (3.2).

The proof has been deferred to Appendix A.2. Based on Theorem 3.1, global guarantees for the networked system can be provided.

3.3.4 Global Guarantees

The probability of a local controlled subsystem $\boldsymbol{C}^i \bowtie \boldsymbol{\mathcal{M}}^i$ satisfying its specification ψ^i is conditioned on the behaviour of the other subsystems to which it is connected. The probability that the interconnected system $\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}}$ satisfies the global specification ψ is obtained as the probability of the intersection of events $(\boldsymbol{C}^i \bowtie \boldsymbol{\mathcal{M}}^i \models \psi^i)$.

Results are established for two distinct network structures. In a *cascaded* network, the interconnection between subsystems is unidirectional, with each subsystem influenced only by its predecessors. In a *cyclic* network, at least one cycle is present in the interconnection graph of the subsystems. In the following, the *predecessors* of the i^{th} subsystem in the network are indexed by $\text{Pre}(i) \subset \mathbb{N}$.

Theorem 3.2 (Global Guarantees (Cascaded)) *Consider a cascaded network* $\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}}$ *of locally controlled subsystems* $\boldsymbol{C}^i \bowtie \boldsymbol{\mathcal{M}}^i$, $i \in \{1, \ldots, N\}$. *The global satisfaction probability with respect to* ψ *is lower bounded by*

$$\mathbb{P}(\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}} \models \psi) \geq \prod_{i=1}^{N} \min_{\mathbf{y}^j \in \mathbf{Y}^j_{\psi^j}} \mathbb{P}\Big(\boldsymbol{C}^i \bowtie \boldsymbol{\mathcal{M}}^i \models \psi^i \,\Big|\, \mathbf{y}^j, \forall j \in \text{Pre}(i)\Big), \quad (3.11)$$

where $\mathbf{Y}^j_{\psi^j} := \{\mathbf{y}^j : \mathcal{L}^j(\mathbf{y}^j) \models \psi^j\}$ *are sets of satisfying output traces.*

Proof. Due to the acyclic network structure, the satisfaction probability can be expressed in terms of conditional probabilities, analogous to Bayesian networks, namely

$$\mathbb{P}\left(\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}} \models \psi\right) = \mathbb{P}\Big(\wedge_{i=1}^{N} \left(\boldsymbol{C}^i \bowtie \boldsymbol{\mathcal{M}}^i \models \psi^i\right) \Big),$$
$$= \prod_{i=1}^{N} \mathbb{P}\left(\boldsymbol{C}^i \bowtie \boldsymbol{\mathcal{M}}^i \models \psi^i \,\middle|\, \boldsymbol{C}^j \bowtie \boldsymbol{\mathcal{M}}^j \models \psi^j, \forall j \in \mathrm{Pre}(i)\right). \quad (3.12)$$

Recall that a subsystem $\boldsymbol{C}^j \bowtie \boldsymbol{\mathcal{M}}^j$ is said to (deterministically) satisfy its specification ψ^j if all of its output traces $\mathbf{y}^j$ satisfy $\mathcal{L}^j(\mathbf{y}^j) \models \psi^j$. The expression (3.11) is obtained by taking the minimum of the terms in the right-hand side of (3.12) with respect to the output traces $\mathbf{y}^j$ for which $\boldsymbol{C}^j \bowtie \boldsymbol{\mathcal{M}}^j \models \psi^j$. This concludes the proof. □

For a general *circular* network configuration — i.e., when feedback loops or cycles are present in the interconnection graph of the subsystems — a more conservative lower bound is obtained. To this end, the internal outputs of the subsystems are bounded to sets $\mathcal{B}^i \subset \mathbb{Y}^i$ for all $i \in \{1, \ldots, N\}$, and are assigned output labels $p_{\mathcal{B}^i}$. The event of a subsystem $\boldsymbol{C}^i \bowtie \boldsymbol{\mathcal{M}}^i$ satisfying its bound $\mathcal{B}^i$ is defined as the safety specification $\mathcal{E}^i := (\boldsymbol{C}^i \bowtie \boldsymbol{\mathcal{M}}^i \models \Box p_{\mathcal{B}^i})$, where "□" denotes the *always* LTL operator (Baier and Katoen, 2008). The probability that the interconnected system $\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}}$ satisfies the global specification ψ is then expressed as the probability of the intersection of events $(\boldsymbol{C}^i \bowtie \boldsymbol{\mathcal{M}}^i \models \psi^i \wedge \mathcal{E}^i)$.

Theorem 3.3 (Global Guarantees (Cyclic)) *Consider a cyclic network $\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}}$ of locally controlled subsystems $\boldsymbol{C}^i \bowtie \boldsymbol{\mathcal{M}}^i$. The global probability of satisfaction with respect to ψ is lower bounded by*

$$\mathbb{P}\left(\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}} \models \psi\right) \geq \prod_{i=1}^{N} \min_{\mathbf{y}^j \in \mathbf{Y}^j_{\mathcal{E}^j}} \mathbb{P}\left(\left(\boldsymbol{C}^i \bowtie \boldsymbol{\mathcal{M}}^i \models \psi^i\right) \wedge \mathcal{E}^i \,\middle|\, \mathbf{y}^j, \forall j \in \mathrm{Pre}(i)\right), \quad (3.13)$$

where $\mathbf{Y}^j_{\mathcal{E}^j} := \left\{\mathbf{y}^j : \mathcal{L}^j(\mathbf{y}^j) \models \mathcal{E}^j\right\}$ are sets of satisfying output traces.

Proof. The safety specifications $\mathcal{E}^i$ are added to get

$$\mathbb{P}\left(\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}} \models \psi\right) \geq \mathbb{P}\left(\left(\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}} \models \psi\right) \wedge \left(\wedge_{i=1}^{N} \mathcal{E}^i\right)\right),$$
$$= \mathbb{P}\Big(\wedge_{i=1}^{N} \left(\boldsymbol{C}^i \bowtie \boldsymbol{\mathcal{M}}^i \models \psi^i\right) \wedge \mathcal{E}^i \Big).$$

Next, a similar technique to the one used by Lavaei et al. (2023, Theorem 3.3) is used to derive a lower bound based on the local satisfaction probabilities. This is achieved by considering the worst-case output traces of the individual predecessors $j \in \mathrm{Pre}(i)$ that satisfy their local safety specification $\mathcal{E}^j$, i.e., $\mathbf{Y}^j_{\mathcal{E}^j} := \left\{\mathbf{y}^j : \mathcal{L}^j(\mathbf{y}^j) \models \mathcal{E}^j\right\}$. Based on this, the global satisfaction probability is lower-bounded as shown in (3.13), concluding the proof. □

Solution Depending on the network structure, either Theorem 3.2 or Theorem 3.3 can be used to obtain a solution to Subproblem 3.1.

Comparison and Feasibility The main distinction between Theorem 3.2 and Theorem 3.3 lies in how the sets of admissible output traces — $\mathbf{Y}^j_{\psi^j}$ and $\mathbf{Y}^j_{\mathcal{E}^j}$, respectively — are determined. In the cascaded case, synthesis proceeds from the roots of the system graph towards its leaves. The set $\mathbf{Y}^j_{\psi^j}$ of the j-th subsystem emerges naturally after its local controller has been synthesised and subsequently constrains the synthesis of its successors. To ensure feasibility of these downstream problems, the outputs of preceding subsystems must typically remain within bounded domains. Such boundedness often follows directly from the support of a system or from the local specification ψ^i, which commonly includes safety-type constraints that restrict the dynamics to a compact region — this is also a practical necessity for numerical computation on bounded domains (see Remark 2.9). In contrast, cyclic interconnections preclude this sequential synthesis: each subsystem depends, directly or indirectly, on its own successors. To resolve this circular dependence, explicit safety contracts $\mathcal{E}^i$ are introduced to bound the internal outputs before synthesis. These bounds serve as fixed-point constraints that can, if required, be iteratively refined to obtain tighter guarantees.

3.4 Sub-Simulation Relations for Gaussian Mixture Models

Now that global guarantees for the interconnected system can be computed from local guarantees on the individual subsystems $\boldsymbol{\mathcal{M}}^i$, the remaining challenge in addressing Subproblem 3.2 is to show how an SSR (Definition 3.3) can be established for systems subject to non-Gaussian noise — specifically, systems with GMM noise as in (3.1). As such, beyond their role within the presented compositional framework, the following formulation establishes GMMs as a general mechanism for extending SSRs to systems exhibiting non-Gaussian stochasticity. Note that in contrast to Theorem 2.3 the noise distribution is additionally considered to be parametrised by the uncertain parameters θ as well. For notational simplicity, since each subsystem is treated individually in the following, the superscript i is omitted.

Theorem 3.4 (Establishing an SSR (Extended)) *Let subsystems $\boldsymbol{\mathcal{M}}(\theta)$ and $\widehat{\boldsymbol{\mathcal{M}}}$ in (3.1) and (3.3) be in an SSR $\widehat{\boldsymbol{\mathcal{M}}} \preceq^{\delta}_{\varepsilon} \boldsymbol{\mathcal{M}}(\theta)$ with interface function $u_t = \hat{u}_t$, relation (3.7), and sub-probability coupling (3.10). Then, the state mapping (3.6) defines a valid control refinement (Definition 2.9) with error parameters $\varepsilon = 0$ and*

$$\delta(x,u) = \sup_{\theta\in\Theta} \left\{ 1 - \sum_{k=1}^{K} \left[A_{k,\theta} + B_{k,\theta} \right] \right\}, \tag{3.14}$$

where the coefficients are given by

$$
\begin{aligned}
A_{k,\theta} &:= \pi_k \mathrm{cdf}\left(-\left(\frac{1}{2} - \eta_{k,\theta}\right)\left\|\beta_{k,\theta}\right\|\right), \\
B_{k,\theta} &:= \hat{\pi}_k \mathrm{cdf}\left(-\left(\frac{1}{2} + \eta_{k,\theta}\right)\left\|\beta_{k,\theta}\right\|\right), \\
\eta_{k,\theta}(x,u) &:= \frac{1}{\beta_{k,\theta}(x,u)^\top \beta_{k,\theta}(x,u)} \log \frac{\pi_k}{\hat{\pi}_k}, \\
\beta_{k,\theta}(x,u) &:= \mu_k - \hat{\mu}_k + \gamma_\theta(x,u),
\end{aligned}
$$

and the offset γ_θ defined in (3.5).

Theorem 3.4 extends the results from Chapter 2 to more general, uncertain noise distributions. Theorem 2.3 is recovered for $\mu_k = \hat{\mu}_k$ and $K = 1$ (implying $\pi_k = \hat{\pi}_k = 1$). The proof has been deferred to Appendix A.3.

Solution Given $(\varepsilon^i, \delta^i)$ from Theorem 3.4, Theorem 2.2 solves Subproblem 3.2 by providing a parameter-independent concrete local controller $\boldsymbol{C}^i$ that is uniformly valid for all $\boldsymbol{\mathcal{M}}^i(\theta^i)$, $\theta^i \in \Theta^i$, with robust sat. prob. $\mathcal{S}^*_{\varepsilon^i,\delta^i}$, i.e.,

$$
\mathbb{P}\left(\boldsymbol{C}^i \bowtie \boldsymbol{\mathcal{M}}^i(\theta^i) \models \psi^i\right) \geq \mathcal{S}^*_{\varepsilon^i,\delta^i}, \quad \forall \theta^i \in \Theta^i.
$$

3.5 Case Studies

This section demonstrates the capabilities of the proposed extensions on a monolithic package delivery case study and a vehicle platoon with two coupled subsystems.

3.5.1 Package Delivery

This case study, adapted from that in Subsection 2.6.1, is extended to incorporate nonlinear dynamics and uncertain GMM noise into the system

$$
\boldsymbol{\mathcal{M}}(\theta): \begin{cases} x_{t+1} = \begin{bmatrix} 0.9x_{t,1} + 0.6\sin(0.1x_{t,2}) + 1.7\bar{\theta}u_{t,1} \\ 0.9x_{t,2} + 1.7u_{t,2} \end{bmatrix} + w_t \\ y_t \quad = x_t \end{cases},
$$

with state $x_t = [x_{t,1}; x_{t,2}]$ and input $u_t = [u_{t,1}; u_{t,2}]$, describing an agent translating in a 2D space. Again, the goal is to compute a controller to navigate the agent to collect a parcel in region p_1 and deliver it to the target region p_3. If the agent visits

the avoid region p_2 on its path, it loses the package and must collect a new one at p_1. This behaviour is captured by the specification $\psi = \Diamond(p_1 \wedge (\neg p_2 \,\mathsf{U}\, p_3))$ (with the corresponding DFA in Fig. 2.9). Recall that this is a complex specification that cannot be expressed as a reach-avoid specification. The homoscedastic noise distribution $w_t \sim \mathcal{MN}_2(\cdot|\pi, \mu, \Sigma)$ is defined with a shared covariance matrix $\Sigma := (\bar{\Sigma}, \bar{\Sigma})$, $\bar{\Sigma} = [\sqrt{0.2}, 0; 0, \sqrt{0.2}]$. The uncertain parametrisation is given by $\theta := (\bar{\theta}, \pi, \mu)$. The spaces $\mathbb{X}, \mathbb{U}, \mathbb{Y}$ and regions p_1, p_2, p_3 are kept identical to the ones in Subsection 2.6.1.

The uncertainty set Θ — as can be identified from data via system identification methods as shown in Chapter 2 — is defined by the elements $\bar{\theta} \in [0.950, 1.050]$, $\pi = (0.8, 0.2)$, $\mu_1 \in [-0.010, 0.010] \times [0.790, 0.810]$, and $\mu_2 \in [-0.810, -0.790]^2$. A nominal model $\widehat{\mathcal{M}} = \mathcal{M}(\hat{\theta})$ is selected by choosing a nominal parametrisation $\hat{\theta}$ with the elements $\bar{\theta} = 0.99$, $\pi = (0.8, 0.2)$, $\mu_1 = [0; 0.8]$, and $\mu_2 = [-0.8; -0.8]$, yielding $\varepsilon_1 = 0$. An input-dependent δ_1 is computed using (3.14).

Abstraction-Based Control Synthesis The following steps are again performed using the SySCoRe toolbox (Van Huijgevoort et al., 2023). A second abstract model $\widetilde{\mathcal{M}}$ is computed by discretising the space of $\widehat{\mathcal{M}}$ and an SSR $\widehat{\mathcal{M}} \preceq^{\delta_2}_{\varepsilon_2} \widetilde{\mathcal{M}}$ is established with $\varepsilon_2 = 0.060$ and a state-dependent δ_2. Via the transitivity property in Theorem 2.1, it follows that $\widetilde{\mathcal{M}} \preceq^{\delta}_{\varepsilon} \mathcal{M}$ with $\delta = \delta_1 + \delta_2$ and $\varepsilon = \varepsilon_1 + \varepsilon_2$.

Sim-to-Real Gap The resulting robust satisfaction probability $\mathcal{S}^*_{\varepsilon,\delta}$ — computed using this (ε, δ) from Theorem 2.2 — is compared against the actual satisfaction probability for a concrete parametrisation θ^* with components $\bar{\theta} = 1$, $\pi = (0.8, 0.2)$, $\mu_1 = [0; 0.8]$, and $\mu_2 = [-0.8; -0.8]$. The actual probability is estimated via Monte Carlo simulation for several representative initial states, with 10^5 simulations performed per initial state and a maximum horizon of 30 time steps. In Fig. 3.5, the robust satisfaction probability $\mathcal{S}^*_{\varepsilon,\delta}$ (in blue) is plotted as a function of the initial state of $\mathcal{M}(\theta)$ and shown alongside the estimated actual satisfaction probability (blue mesh). Moreover, Fig. 3.5 shows (in red) the robust satisfaction probability of a controller synthesised for a tighter uncertainty set $\bar{\Theta} \subset \Theta$ — as can be obtained using a bigger data set — with components $\bar{\theta} \in [0.999, 1.001]$, $\pi = (0.8, 0.2)$, $\mu_1 \in [-0.001, 0.001] \times [0.799, 0.801]$, and $\mu_2 \in [-0.801, -0.799]^2$.

3.5.2 Car Platoon

Now, a networked system consisting of two subsystems is considered, modelling the dynamics of a platoon of two cars driving behind each other (cf. Fig. 3.1). The dynamics of the *leading vehicle* $\mathcal{M}^1(\theta^1)$ are given by

$$\mathcal{M}^1(\theta^1): \begin{cases} \begin{bmatrix} s^1_{t+1} \\ v^1_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & \tau \\ 0 & 0.9 \end{bmatrix} \begin{bmatrix} s^1_t \\ v^1_t \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{\tau}{m^1} \end{bmatrix} u^1_t + w^1_t \\ y^1_t \quad = x^1_t \end{cases},$$

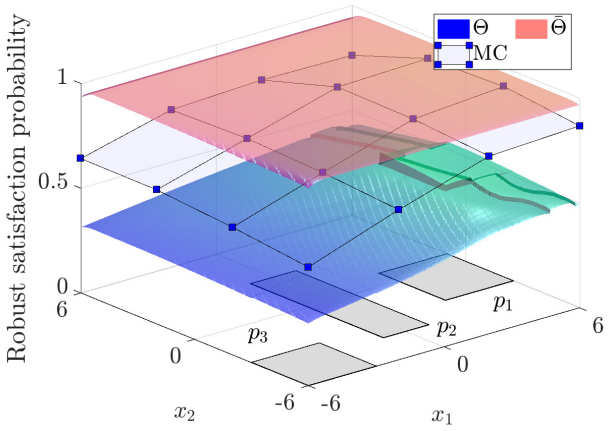


Fig. 3.5: Robust satisfaction probability as a function of the initial state and the actual satisfaction probability estimated via Monte Carlo simulation for the package delivery case study for Θ in blue. The robust lower bound is boosted for a contracting uncertainty set ($\bar{\Theta}$ in red).

with a constant time discretisation $\tau = 0.5$, and a homoscedastic noise distribution $w_t^1 \sim \mathcal{MW}_2(\cdot\,|\pi^1, \mu^1, \Sigma^1)$ with a shared covariance matrix $\bar{\Sigma}^1 := [0.050, 0; 0, 0.025]$, i.e., $\Sigma^1 := (\bar{\Sigma}^1, \bar{\Sigma}^1)$. The uncertain parameters are given by $\theta^1 := (m^1, \pi^1, \mu^1)$. The state, input, and output spaces are defined as $\mathbb{X}^1 = [0, 5] \times [0.95, 1.05]$, $\mathbb{U}^1 = [0.3, 1]$, and $\mathbb{Y}^1 = \mathbb{X}^1$, respectively.

The dynamics of the *following vehicle* $\mathcal{M}^2(\theta^2)$ are given by

$$\mathcal{M}^2(\theta^2)\colon \begin{cases} \begin{bmatrix} s_{t+1}^2 \\ v_{t+1}^2 \\ d_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 1.8\tau & 0 \\ 0 & 0.9 & 0 \\ 0 & -\tau & 1 \end{bmatrix} \begin{bmatrix} s_t^2 \\ v_t^2 \\ d_t \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{\tau}{m^2} \\ 0 \end{bmatrix} u_t^2 + \begin{bmatrix} 0 \\ 0 \\ \tau \end{bmatrix} v_t^1 + w_t^2 \\ y_t^2 \quad = x_t^2 \end{cases},$$

where the third state variable — capturing the distance between the vehicles — couples the two subsystems via the velocity v_t^1. Similarly as before, the system has a noise distribution $w_t^2 \sim \mathcal{MW}_2(\cdot\,|\pi^2, \mu^2, \Sigma^2)$ with common covariance matrix $\bar{\Sigma}^2 := [0.075, 0, 0; 0, 0.025, 0; 0, 0, 0.050]$, $\Sigma^2 := (\bar{\Sigma}^2, \bar{\Sigma}^2)$. Note how the GMM allows to capture sensor noise on the velocity measurements of $\mathcal{M}^1(\theta^1)$. The uncertain parameters are $\theta^2 := (m^2, \pi^2, \mu^2)$. Define the state space $\mathbb{X}^2 = [0, 5] \times [0.9, 1.1] \times [0.2, 2.2]$, input space $\mathbb{U}^2 = [0.5, 1.5]$, and output space $\mathbb{Y}^2 = \mathbb{X}^2$.

Specification The global specification can be decomposed into local subspecifications as follows (see Assumption 3.1):

- The goal of $\boldsymbol{\mathcal{M}}^1$ is to start in the initial region $p^1_{\text{init}} := [1.2, 2.2] \times [0.95, 1.05]$ and reach target region $p^1_{\text{targ}} := [4, 5] \times [0.95, 1.05]$ whilst remaining in a safe region $p^1_{\text{safe}} := [1.2, 5] \times [0.95, 1.05]$, written as $\psi^1 := p^1_{\text{init}} \wedge \bigcirc(p^1_{\text{safe}} \mathbin{\mathsf{U}} p^1_{\text{targ}})$.
- Analogously, the specification of $\boldsymbol{\mathcal{M}}^2$ is defined as $\psi^2 := p^2_{\text{init}} \wedge \bigcirc(p^2_{\text{safe}} \mathbin{\mathsf{U}} p^2_{\text{targ}})$, with regions $p^2_{\text{init}} := [0, 1] \times [0.9, 1.1] \times [0.2, 2.2]$, $p^2_{\text{targ}} := [2.8, 3.8] \times [0.9, 1.1] \times [0.2, 2.2]$, and $p^2_{\text{safe}} := [1, 3.8] \times [0.9, 1.1] \times [0.2, 2.2]$.

Note that enforcing p^1_{safe} and p^2_{safe} bounds the velocities of the vehicles to contract-type sets $v^1_t \in \mathcal{B}^1 := [0.95, 1.05]$ and $v^2_t \in \mathcal{B}^2 := [0.9, 1.1]$ for all $t \in \mathbb{N}_0$ — thus ensuring feasibility despite unbounded noise (see Theorem 3.2).

Parametrisations The uncertainty sets $\hat{\Theta}^i$ are given as follows:

- $\hat{\Theta}^1$, with elements $m^1 \in [3.9, 4.1]$, $\pi^1 = (0.3, 0.7)$, and $\mu^1 \in ([0.09, 0.11] \times [-0.01, 0.01], [-0.01, 0.01] \times [-0.01, 0.01])$; and
- $\hat{\Theta}^2$, with elements $m^2 \in [3.8, 4.0]$, $\pi^2 = (0.3, 0.7)$, and $\mu^2 \in ([-0.010, 0.010]^2 \times [0.009, 0.011], [-0.010, 0.010]^2 \times [-0.011, -0.009])$.

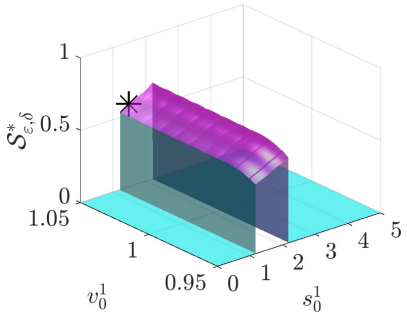
Nominal models $\widehat{\boldsymbol{\mathcal{M}}}^i = \boldsymbol{\mathcal{M}}^i(\hat{\theta}^i)$ are selected based on nominal parametrisations:

- $\hat{\theta}^1$, with elements $m^1 = 4$, $\pi^1 = (0.3, 0.7)$, $\mu^1 = ([0.1; 0], [0; 0])$; and
- $\hat{\theta}^2$, with elements $m^2 = 3.9$, $\pi^2 = (0.3, 0.7)$, $\mu^2 = ([0; 0; 0.01], [0; 0; -0.01])$.

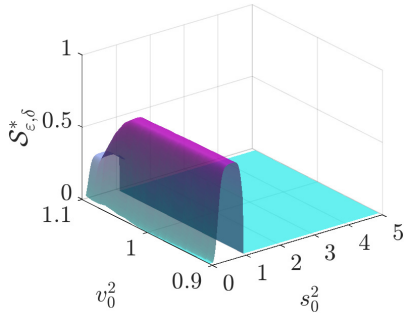
This choice yields $\varepsilon^i_1 = 0$. Input-dependent δ^i_1 are computed using (3.14).

Abstraction-Based Control Synthesis SySCoRe constructs a second batch of abstract models $\widetilde{\boldsymbol{\mathcal{M}}}^i$, $i \in \{1, 2\}$, by discretising the space of the models $\widehat{\boldsymbol{\mathcal{M}}}^i$, alongside corresponding SSRs $\widetilde{\boldsymbol{\mathcal{M}}}^i \preceq^{\delta^i}_{\varepsilon^i} \widehat{\boldsymbol{\mathcal{M}}}^i$ with $(\varepsilon^1_2 = 0.006, \delta^1_2 = 0.002)$ and $(\varepsilon^2_2 = 0.030, \delta^2_2 = 0.007)$. Notably, the error inflicted by the internal input v^1_t is captured by augmenting the discretisation error in the invariance constraint (see van Huijgevoort and Haesaert, 2022, Equation (20d)). Using the transitivity property Theorem 2.1, it follows that $\widetilde{\boldsymbol{\mathcal{M}}}^i \preceq^{\delta^i}_{\varepsilon^i} \boldsymbol{\mathcal{M}}^i$ with $\delta^i = \delta^i_1 + \delta^i_2$ and $\varepsilon^i = \varepsilon^i_1 + \varepsilon^i_2$. To synthesise local controllers, the velocity difference $(v^1_t - v^2_t) \in [-0.15, 0.15]$ is bounded, with the worst-case difference considered in each time step. The robust satisfaction probability $\mathcal{S}^*_{\varepsilon,\delta}$ of each subsystem with respect to their local specification is computed based on Theorem 2.3 and depicted in Fig. 3.6(a)–(b) as functions of the initial state of $\boldsymbol{\mathcal{M}}^1(\theta^1)$ and $\boldsymbol{\mathcal{M}}^2(\theta^2)$, respectively. The resulting global robust satisfaction probability $\mathcal{S}^*_{\varepsilon,\delta}$ of the cascaded network $\boldsymbol{\mathcal{C}} \bowtie \boldsymbol{\mathcal{M}}$ is computed based on the individual probabilities using Theorem 3.2, visualised in Fig. 3.6(c) as a function of the initial state of $\boldsymbol{\mathcal{M}}^2(\theta^2)$, and $\boldsymbol{\mathcal{M}}^1(\theta^1)$ initialised in $[s^1_0; v^1_0] = [1.5; 1.05]$.

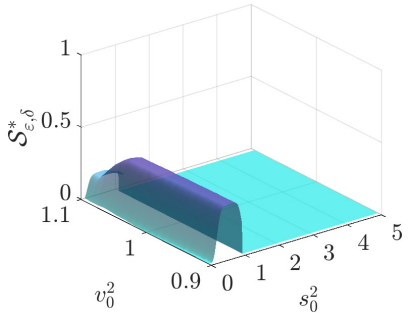
Scalability The case study demonstrates compositional reasoning for two interconnected subsystems. For larger networks or cyclic interconnections, the guarantees from Theorem 3.2 and Theorem 3.3 tend to become more conservative due to the multiplicative structure of the bounds. More details on this scalability aspect and possible remedies are provided in the chapter conclusion.



(a) Robust satisfaction prob. of $\mathcal{C}^1 \bowtie \mathcal{M}^1(\theta^1)$



(b) Robust satisfaction prob. of $\mathcal{C}^2 \bowtie \mathcal{M}^2(\theta^2)$



(c) Robust satisfaction prob. of $\mathcal{C} \bowtie \mathcal{M}(\theta)$

Fig. 3.6: Local robust satisfaction probability $\mathcal{S}^*_{\varepsilon,\delta}$ as a function of the initial state of the (a) leading vehicle and (b) following vehicle, as well as (c) the global robust satisfaction probability $\mathcal{S}^*_{\varepsilon,\delta}$ as a function of the initial state of the following vehicle, with the leading car initialised at $[s_0^1; v_0^1] = [1.5; 1.05]$ (black asterisk in (a)).

3.6 Conclusion

In this chapter, the original SSR definition introduced in Chapter 2 was extended to interconnected systems, enabling quantitative relations between networks of uncertain systems and their abstract counterparts. The proposed approach supports complex, non-Gaussian stochasticity by approximating arbitrary noise distributions using Gaussian mixture models and allows the synthesis of robust controllers even when the true noise is unknown (Section 3.4). Note that this contribution stands independently of the compositional framework. These extensions were demonstrated on two intricate case studies.

The main limitation of the presented compositional approach lies in the accumulating error bounds (ε, δ) in Theorem 3.1 and the resulting multiplicative nature of the global guarantees obtained via Theorem 3.2 and Theorem 3.3. This generally leads to more conservative satisfaction probabilities compared to the monolithic ap-

proach from Chapter 2. Compositionality thus trades conservatism for computational tractability. For large-scale systems, alternative techniques such as max small-gain reasoning (Lavaei et al., 2020b) may reduce conservatism, since the associated global ε is not directly tied to the number of subsystems in the network.

While the methods developed here are suitable when interactions between subsystems are mild and fully observable, many real-world systems involve tightly coupled agents with only partial and noisy observations. Extending compositional synthesis to such partially observed networks remains an important open challenge. First steps toward this direction are outlined in Chapter 7, building on ongoing research into situational awareness under uncertainty.

PART II
LEARNING UNDER UNCERTAINTY

The world we inhabit is infinitely complex. The art of learning is knowing what's relevant — and knowing the limits of what remains unknown.

Chapter 4
Learning Formal Abstractions via Binary-Tree Gaussian Processes

Abstract Previous chapters have demonstrated that abstract system representations are essential for scalable formal reasoning, and that their structure critically affects compatibility with verification and synthesis algorithms. This chapter investigates how such models can be learned directly from data whilst preserving formal guarantees, moving beyond parametrised models to more expressive classes of uncertain systems. Binary-tree Gaussian processes (BTGPs) are introduced as a data-driven method for modelling stochastic systems with additive Gaussian noise, producing piecewise-constant posterior representations that naturally induce a discretisation of the system domain. The result: a discrete abstraction learned directly from data. This structure makes BTGPs ideally suited for verification based on interval Markov chains (IMCs), eliminating the need for cumbersome discretisation and error quantification procedures. As a result, generating an IMC of the latent system — typically a computational bottleneck — is accelerated by orders of magnitude. The approach is applied to reach-avoid verification, providing formal guarantees in the absence of an explicit model architecture. Instead, the discrete structure emerges directly from the kernel, demonstrating how learning can yield models that are both data-driven and verification-ready by design.



4.1 Introduction

The previous chapters explored robust control synthesis under aleatoric and epistemic uncertainty, assuming the availability of a known parametrised model of the target system. This enabled the use of stochastic simulation relations to formally reason about system behaviour by linking continuous-space models — derived as surrogates of the physical systems — with discrete symbolic abstractions suitable for formal synthesis. In this setting, model uncertainty was confined to a set of learned

parameters. However, such approaches rely fundamentally on the existence of a suitable parametrised model — an assumption that may not hold in practice due to limited modelling resources or insufficient system insight. Furthermore, even when expert-derived simulation models are available, they are often too complex to support formal verification directly — instead serving merely as another intermediary within a data-driven verification pipeline. Worse still, if the chosen model structure does not accurately describe the true system behaviour, it is unclear whether any derived guarantees remain valid. In all such settings, data-driven learning becomes essential. Moreover, rather than learning a detailed continuous-space model only to abstract it subsequently for formal verification, it may be preferable to learn a formally amenable representation directly, whilst quantifying the abstraction error relative to the data-generating system.

This chapter focuses on learning such formal abstractions directly from data without requiring a pre-defined parametrised model. To simplify the exposition, the focus is restricted to formal *verification*, rather than synthesis. Within this setting, *interval Markov chains*[1] (IMCs) serve as an effective framework: by representing transitions via upper and lower bounds on transition probabilities, they provide an expressive formalism capable of capturing both stochasticity and epistemic uncertainty (Suilen et al., 2025). Once an IMC is obtained, it can be efficiently verified against various temporal logic specifications using model-checking algorithms such as the *interval iteration algorithm* (IIA) (Haddad and Monmege, 2018), which in comparison to value iteration provides a principled stopping criterion and convergence rate.

Most existing data-driven IMC approaches rely on non-parametric techniques — such as GP regression (Jackson et al., 2020; Jiang et al., 2022) or non-parametric estimation (Zhang et al., 2024) — with others limited to finite horizons and either linear systems and reach-avoid specifications (Badings et al., 2023), or assuming known Lipschitz constants (Lavaei et al., 2022b; Kazemi et al., 2024a; Nazeri et al., 2025). Related kernel-based work by Thorpe et al. (2022); Romao et al. (2023) is restricted to distributionally robust guarantees. While GPs can offer statistical error bounds (Chowdhury and Gopalan, 2017; Fiedler et al., 2021), NN-based methods for learning IMDPs (Adams et al., 2022) provide no such guarantees and would require additional NN verification, which is often intractable.

Consequently, when formal robustness is required, IMDP-based frameworks built on GPs or non-parametric techniques remain among the most promising approaches (Jackson et al., 2020; Jiang et al., 2022; Zhang et al., 2024). However, a key bottleneck remains: generating discrete-space IMCs from continuous-space models such as vanilla GPs is computationally intensive, as the conversion requires solving (non-convex) optimisation problems for every possible transition to determine worst-case probability bounds (see Subsection 4.5.1). To overcome this, this chapter employs *binary-tree Gaussian processes* (BTGPs) (Cohen et al., 2022b) — a recently introduced GP variant with a discontinuous kernel designed to reduce the computational burden of GP regression significantly — increasing the scalability of GPs to higher

[1] An interval Markov decision process (IMDP) — discussed previously — reduces to an interval Markov chain (IMC) when its input set is empty. Since this chapter focuses on formal verification, the term IMC is used broadly to refer to both IMCs and IMDPs where appropriate.

dimensional systems and larger data sets. BTGPs produce piecewise-constant posterior mean and covariance functions, thereby inducing a natural discrete partitioning of the input domain. This structure is exploited here to generate IMCs rapidly, accelerating verification while preserving formal correctness.

Whereas conventional GP-based IMC construction requires optimising both upper and lower probability bounds for every possible transition, BTGPs reduce this to two integrations per transition — one for each bound — yielding a polynomial speed-up. Based on this, a data-driven verification approach is presented, using BTGP-induced IMCs to verify continuous-state stochastic systems against infinite-horizon reach-avoid specifications. Statistical guarantees on the accuracy of the learned BTGP model are derived by combining measure-theoretic results with a BTGP-specific abstraction error. Additionally, theoretical results are derived that show how discontinuous kernels naturally yield finite abstractions of the latent system; closed-form characterisations of the resulting function space are provided.

Organisation The remainder of this chapter is structured as follows. After a brief discussion of additional related work, Section 4.3 introduces kernel methods and reproducing kernel Hilbert spaces, along with the formalised problem statement. Section 4.4 presents system identification using BTGP regression, including results on statistical error bounds. Section 4.5 shows how BTGPs enable efficient construction of IMC models and provides a suitable verification algorithm. Finally, Section 4.6 illustrates the approach on a nonlinear case study.

4.2 Related Work

This chapter explores the use of BTGPs for the abstraction of nonlinear, continuous-space stochastic systems. The general idea of approximating a nonlinear vector field via a hybrid automaton is a well-established concept known as *hybridisation* (Alur et al., 2000; Asarin et al., 2003) — though it has largely been restricted to deterministic settings. Related learning-based variants have recently gained traction. For example, *neural abstractions* construct approximate dynamical systems using neural networks (Abate et al., 2022b; Majumdar et al., 2023b). In this line of work, the approach developed by Edwards et al. (2023) is the closest to the one presented here, learning piecewise-constant abstractions of continuous-time deterministic systems with bounded non-deterministic disturbances. To ensure model correctness, Edwards et al. (2023) use a counterexample-guided refinement loop relying on SMT solvers to iteratively bound the error of the learned abstraction. However, such schemes generally lack convergence guarantees and often fail in practice for more complex systems due to scalability issues.

Further related work includes a contraction-guided adaptive partitioning approach proposed by Harapanahalli et al. (2023a) — again, within a deterministic framework. Similarly, Banse et al. (2023) construct dynamics-guided abstractions via adaptive partitioning based on Cantor and Kantorovich distances between system models.

4.3 Preliminaries and Problem Statement

Notation Let $\mathbb{B} := \{0, 1\}$ be the set of binary numbers and $s \in \mathbb{B}^q$ a binary string of length $|s| := q$, where its substring prefix of length $l \in \mathbb{N}$ is denoted by $[s]_{\leq l}$. To save space, a column vector $X_N := [x_1; \ldots; x_N] \in \mathbb{X}^N$ may subsequently be written compactly as $X_N := [x_i]_{i=1}^N$. The element-wise evaluation of a function $f : \mathbb{X} \to \mathbb{Y}$ on X_N is denoted as $f(X_N) := [f(x_i)]_{i=1}^N$. Similarly, $A = [a_{ij}]_{i,j=1}^N$ indicates a matrix with elements a_{ij}. The trace of such a matrix is given by $\mathrm{Tr}(A) := \sum_i a_{ii}$. For $b \in \{\text{true}, \text{false}\}$, the indicator function is defined as $\mathbb{1}(b) := \{1, \text{ if } b = \text{ true, and 0 otherwise}\}$. The uniform probability distribution over $\mathbb{X}$ is denoted as $\mathcal{U}_{\mathbb{X}}$, with realisations $x \sim \mathcal{U}_{\mathbb{X}}$.

4.3.1 Reproducing Kernel Hilbert Spaces

Kernel Methods

Kernel methods — often seen in direct opposition to neural networks — are a foundational class of algorithms in machine learning, underpinning models such as support vector machines (SVMs), Gaussian processes (GPs), and principal component analysis (PCA) (Schölkopf and Smola, 2002). Their core idea revolves around the concept of embedding data into a possibly infinite-dimensional feature space via a (nonlinear) feature map (see Fig. 4.1), and predicting expressive behaviours by interpolating linearly between the embedded data points. Thus, this transformation converts nonlinear problems into linear problems in feature space, offering a significant computational advantage.

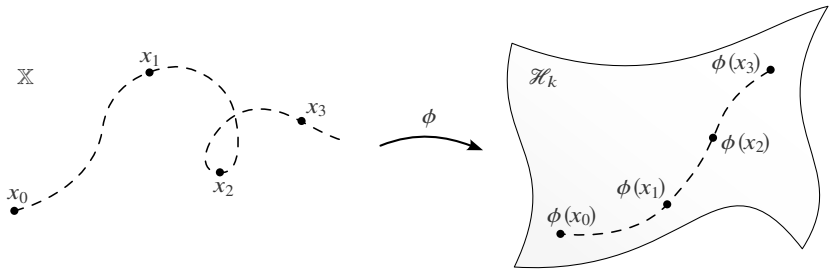


Fig. 4.1: A kernel k associates a (nonlinear) feature map ϕ embedding input data (e.g., state sequences) into a (high-/infinite-dimensional) RKHS.

Reproducing Kernel Hilbert Spaces

The feature map is defined implicitly by a positive-definite *kernel* function, which induces a *reproducing kernel Hilbert space* (RKHS) of learnable functions, inheriting structural properties from the chosen kernel (e.g., a linear kernel yields linear functions, a polynomial kernel yields polynomials, etc.). Thanks to *representer the-*

orems, solutions can be expressed entirely in terms of kernel evaluations on data points, bypassing any explicit density estimation or computations in the (possibly infinite-dimensional) feature space — a property known as the *kernel trick*.[2] Thus, kernel methods are non-parametric approaches, where the performance depends on the task-specific choice and tuning of a suitable kernel.

This chapter leverages the kernel architecture of a specific type of GP model to obtain an explicit mathematical characterisation of the underlying function space. This property is essential for deriving formal error bounds on the learned system dynamics. To this end, a brief summary of RKHS theory is provided, whilst the interested reader is referred to the monograph by Berlinet and Thomas-Agnan (2004) for a comprehensive study of RKHSs.

RKHS Theory A symmetric function $k : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ is called a (positive-definite) *kernel* (note the distinction from *probability kernels*) if $\sum_{i=1}^{N} \sum_{j=1}^{N} a_i a_j k(x_i, x_j) \geq 0$ for all $a_1, \ldots, a_N \in \mathbb{R}$, $N \in \mathbb{N}$, $x_1, \ldots, x_N \in \mathbb{X} \subset \mathbb{R}^n$. A prominent example is the *squared exponential* kernel (Rasmussen and Williams, 2006; Kanagawa et al., 2018)

Squared Exponential Kernel

$$k_{\mathbb{X}}(x, x') := \sigma_f^2 \exp\left(-\tfrac{1}{2}(x - x')^\top \Sigma\,(x - x')\right), \quad \Sigma := \operatorname{diag}(\sigma_l)^{-2}, \tag{4.1}$$

with amplitude $\sigma_f^2 \geq 0$ and lengthscale coefficients $\sigma_l \in \mathbb{R}^n$. Note that a positive-definite kernel gives rise to a positive-definite *Gram matrix* $K := [k(x_i, x_j)]_{i,j=1}^{N}$. Given a kernel k on a non-empty set $\mathbb{X}$, there exists a corresponding unique *reproducing kernel Hilbert space* (RKHS) $\mathscr{H}_k$ of functions $f : \mathbb{X} \to \mathbb{R}$, equipped with an inner product $\langle \cdot, \cdot \rangle_{\mathscr{H}_k}$ with the celebrated *reproducing property*, such that for any function $f \in \mathscr{H}_k$ evaluated at any $x \in \mathbb{X}$ it holds that $f(x) = \langle f, k(\cdot, x) \rangle_{\mathscr{H}_k}$. Note that $k(\cdot, x) : \mathbb{X} \to \mathscr{H}_k$ is a real-valued function, which is also called an implicit *canonical embedding* or *feature map* ϕ such that $k(x, x) = \langle \phi(x), \phi(x') \rangle_{\mathscr{H}_k}$ for all $x, x' \in \mathbb{X}$. For an RKHS $\mathscr{H}_k$, the associated feature map ϕ and kernel k are used interchangeably for ease of notation and comprehensibility. The inner product induces the norm $\|f\|_{\mathscr{H}_k} := \sqrt{\langle f, f \rangle_{\mathscr{H}_k}}$ of the RKHS. A kernel k is called *translation-invariant* if it can be expressed as a function of distance between points $x, x' \in \mathbb{X}$, i.e., there exists a function $\Delta : \mathbb{X} \to \mathbb{R}$ such that $k(x, x') = \Delta(x - x')$.

Gram Matrix

Translation Invariance

Given N i.i.d. samples $X_N := [x^i]_{i=1}^{N}$ with $x^i \in \mathbb{X}$, the *Gram matrix* of k is given by $K_X^N := [k(x^i, x^j)]_{i,j=1}^{N}$. Furthermore, the vector-valued function $k_X^N(x) := [k(x, x^i)]_{i=1}^{N}$ is defined.

[2] Unlike Koopman-operator-based methods, that rely on manually chosen observables to approximate nonlinear dynamics (Mauroy et al., 2020), kernel methods provide a systematic nonlinear feature map with closed-form expressions via the reproducing property, avoiding computations in high-dimensional feature spaces.

4.3.2 Problem Statement

Before formalising the problem statement, the class of target systems and specifications are declared. As the focus of this chapter is on learning, without loss of generality, the exposition is limited to the verification of autonomous systems against reachability properties for simplicity.

System Dynamics An *autonomous*[3] stochastic system $\boldsymbol{\mathcal{M}}$ is considered, given by

$$\boldsymbol{\mathcal{M}}\colon \begin{cases} x_{t+1} = f(x_t) + w_t, & w_t \sim \mathcal{N}(\,\cdot\,|0, \sigma_w^2 I_n) \\ y_t \quad = x_t \end{cases}, \tag{4.2}$$

where the state $x_t \in \mathbb{X} \subset \mathbb{R}^{n_x}$ is fully observed via the output $y_t := [y_{t,1}; \ldots; y_{t,n_x}] \in \mathbb{Y} \equiv \mathbb{X}$. The transition function $f\colon \mathbb{X} \to \mathbb{X}$ is unknown and assumed to lie in the RKHS of some translation-invariant kernel k, i.e., $f \in \mathscr{H}_k$. This assumption is not restrictive; for example, the squared exponential kernel (4.1) encompasses all smooth functions. The process noise $w_t \sim \mathcal{N}(\,\cdot\,|0, \sigma_w^2 I_{n_x})$ is assumed to be i.i.d. zero-mean Gaussian with known covariance $\sigma_w^2 I_{n_x}$, implying noise independence across dimensions. The covariance can often be inferred (approximately) from data. Analogous to the representation of systems as gMDPs used previously (see Definition 2.4), the autonomous system $\boldsymbol{\mathcal{M}}$ in (4.2) can be modelled as a *Markov chain* (MC).

Markov Chain

Definition 4.1 (Markov Chain (MC)) An *MC* is a tuple $\boldsymbol{\mathcal{M}} = (\mathbb{X}, \mathbb{X}_0, \mathbf{t})$, with

- $\mathbb{X} \subset \mathbb{R}^{n_x}$, the state space, with states $x \in \mathbb{X}$;
- $\mathbb{X}_0 \subset \mathbb{X}$, the set of initial states $x_0 \in \mathbb{X}_0$; and
- $\mathbf{t}\colon \mathbb{X} \times \mathscr{B}(\mathbb{X}) \to [0, 1]$, the transition probability kernel assigning to each state $x \in \mathbb{X}$ a probability measure $\mathbf{t}(\,\cdot\,|x)$ on $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$.

Reachability Specifications

Reachability Specifications The previous chapters addressed scLTL specifications by solving a reachability problem on the composition of the automaton representation of the specification and the symbolic system abstraction (see Subsection 2.4.3 and Kamgarpour et al. (2013)). The same setup applies here; however, for simplicity, the scope is reduced to reachability on the system only (see Fig. 4.2): given a target set $\boldsymbol{\mathcal{T}} \subset \mathbb{X}$, the system $\boldsymbol{\mathcal{M}}$ satisfies the (infinite-horizon) reachability specification $\psi_{\boldsymbol{\mathcal{T}}}$ with probability at least $p_{\psi_{\boldsymbol{\mathcal{T}}}} \in (0, 1)$ if the probability that the trajectories remain in $\mathbb{X}$ until eventually reaching $\boldsymbol{\mathcal{T}}$ is no less than $p_{\psi_{\boldsymbol{\mathcal{T}}}}$, written as $\mathbb{P}(\boldsymbol{\mathcal{M}} \models \psi_{\boldsymbol{\mathcal{T}}}) \geq p_{\psi_{\boldsymbol{\mathcal{T}}}}$.

Let $\mathcal{D}_N := \{x^i, y^i\}_{i=1}^N \subset \mathbb{X} \times \mathbb{X}$ be a set of $N \in \mathbb{N}$ data samples from the latent true system in (4.2), such that $y^i := f(x^i) + w^i$ are output samples for uniformly drawn $x^i \in \mathcal{U}_{\mathbb{X}}$ and i.i.d. noise $w^i \sim \mathcal{N}(\,\cdot\,|0, \sigma_w^2 I_{n_x})$. The data may also be written as $\mathcal{D}_N = (X_N, Y_N)$, where $X_N := [x_1, \ldots, x_N]$ and $Y_N := [y_1, \ldots, y_N]^\top$.

[3] In control theory, an *autonomous* system is characterised by dynamics that do not explicitly depend on external variables such as time t or control inputs u_t. Here, the term emphasises that $\boldsymbol{\mathcal{M}}$ evolves without an external control input.

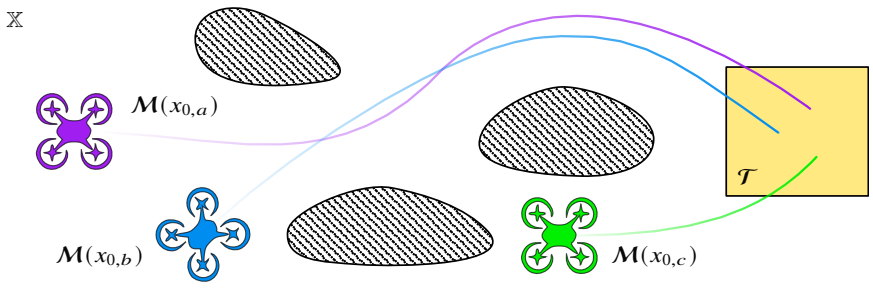


Fig. 4.2: Example reachability problem: Initialised at different positions $x_{0,a}, x_{0,b}, x_{0,c} \in \mathbb{X}$ ($\mathbb{X}$ being the complement of the obstacles and target), a drone $\mathcal{M}$ shall eventually reach the target $\mathcal{T}$.

Problem 4.1 (Data-Driven Reachability) Let data $\mathcal{D}_N$ be given from the system $\mathcal{M}$ in (4.2) with unknown transition function $f \in \mathcal{H}_k$. For a reachability specification $\psi_{\mathcal{T}}$, verify that $\mathcal{M}$ satisfies $\psi_{\mathcal{T}}$ with probability at least $p_{\psi_{\mathcal{T}}} \in (0, 1)$, i.e., $\mathbb{P}(\mathcal{M} \models \psi_{\mathcal{T}}) \geq p_{\psi_{\mathcal{T}}}$.

To address this problem, the idea is to construct a function estimator $\hat{f}\colon \mathbb{X} \to \mathbb{X}$ based on the data $\mathcal{D}_N$ such that $\hat{f}$ approximates the unknown function $f \in \mathcal{H}_k$ in (4.2). To this end, a kernel $\hat{k}$ is specified that characterises the class of learnable functions $\hat{f} \in \mathcal{H}_{\hat{k}}$. To be able to address a wealth of functions f and perform the function approximation in a function space $\mathcal{H}_{\hat{k}} \not\supset \mathcal{H}_k$, the following assumption is raised.

Assumption 4.1 Let the unknown function $f \in \mathcal{H}_k$ and the function estimator $\hat{f} \in \mathcal{H}_{\hat{k}}$, with known kernels $k, \hat{k}\colon \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ and complexity bound $B \geq \|f\|_{\mathcal{H}_k} \geq 0$.

Assumption 4.1 is raised to allow for greater flexibility in choosing the kernel $\hat{k}$ used for estimation. In particular, in this chapter, a discontinuous $\hat{k}$ is used. To still be able to address more arbitrary functions — such as continuous/smooth functions living in the RKHS of the more general squared exponential kernel — the error introduced by constructing an estimator in a different RKHS must be quantified. The RKHS norm $\|f\|_{\mathcal{H}_k}$ of a function acts as a measure of its complexity in $\mathcal{H}_k$. Thus, B puts a bound on the maximum permitted complexity of the latent dynamics f.

Note that the choice of k can be arbitrarily general, but also allows the incorporation of prior knowledge about the latent system dynamics — conceptually similar to providing a parametrised model as in Chapter 2, but in a more abstract and therefore less restrictive manner. Moving forward, the accent "^" is used to denote symbols referring to the estimated model.

4.4 System Identification via Binary-Tree Gaussian Processes

This section gives a brief overview of the classical GP regression framework, followed by an examination of GPs based on the so-called binary-tree kernel. For the latter, the underlying RKHS is characterised and a statistical error bound on its predictions is derived.

4.4.1 Gaussian Process Regression

GP Regression

A *Gaussian process* (GP) is a parameter-free regression model that allows the user to learn dimension-wise independent input-output mappings of the form $\hat{f}_d : \mathbb{X} \to \mathbb{R}$, $d \in \{1, \dots, n_x\}$, from empirical data $\mathcal{D}_N = \{x^i, y^i\}_{i=1}^N$, where $y^i := [y_1^i, \dots, y_{n_x}^i]^\top$, such that the prediction error

$$\mathrm{err}_d(\hat{f}_d) := \underbrace{\tfrac{1}{N} \sum_{i=1}^N \left\| y_d^i - \hat{f}_d(x^i) \right\|^2}_{\text{empirical error}} + \underbrace{\sigma_w^2 \, \|f\|_{\mathscr{H}_k}}_{\text{regularisation}}$$

is minimised for all dimensions $d \in \{1, \dots, n_x\}$ (Rasmussen and Williams, 2006; Berlinet and Thomas-Agnan, 2004; Kanagawa et al., 2018). Recall that the process noise affects the outputs of the system (4.2) dimension-wise independently, allowing to train individual GPs for each output dimension independently. This is without loss of generality, and the results of this chapter are applicable to non-diagonal covariance matrices by applying a linear transformation on the data gathered from the system.

In the following, the subscript d indicating a specific dimension may be dropped for brevity. As an infinite-dimensional generalisation of the multivariate Gaussian distribution, GPs are particularly popular for their ability to capture statistical information about the latent function. The following definition is analogous to Definition 2.2 by Kanagawa et al. (2018).

Definition 4.2 (Gaussian Process (GP)) Let $\hat{k} : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ be a (symmetric, positive-definite) *kernel* or *covariance function*, and let $\hat{\mu}_d : \mathbb{X} \to \mathbb{R}$ be a function. A random function $\hat{f}_d : \mathbb{X} \to \mathbb{R}$ is said to be a *GP*, denoted by $\hat{f}_d \sim \mathcal{GP}(\hat{\mu}_d, \hat{k})$, if for any finite set $X_N := [x^1, \dots, x^N]$, with $x^i \in \mathbb{X}$, $N \in \mathbb{N}$, the vector $\hat{f}_{d,N} := [\hat{f}_d(x^i)]_{i=1}^N$ follows a multivariate distribution

$$\hat{f}_{d,N} \sim \mathcal{N}(\cdot \,|\hat{\mu}_{d,N}, \hat{K}_N),$$

where $\hat{\mu}_{d,N} := [\hat{\mu}_d(x^i)]_{i=1}^N$ and $\hat{K}_N := [\hat{k}(x^i, x^j)]_{i,j=1}^N$ denote the mean vector and Gram matrix, respectively.

The aim of GP regression is to produce predictions of the output at test points $x \in \mathbb{X}$ not included in the data $\mathcal{D}_N$. Although the unknown true data-generating function f is not assumed to be a sample from a GP with kernel $\hat{k}$ (see Assumption 4.1), a *posterior GP* can nevertheless be constructed to serve as a data-driven surrogate

model of f. Moreover, a probabilistic bound on its prediction error can be established (e.g., via Theorem 2 by Chowdhury and Gopalan (2017) or the results of Fiedler et al. (2021)).

Definition 4.3 (Posterior GP) Let $\mathcal{D}_N = (X_N, Y_N)$ be data from the system (4.2). Then, the *posterior* distribution of $\hat{f}_d(x)$ given $\mathcal{D}_N$ is Posterior GP

$$\hat{f}_d(x) \sim \mathcal{N}(\cdot\,|\hat{\mu}_{N,d}(x), \hat{\sigma}^2_{N,d}(x)), \tag{4.3}$$

characterised by the posterior mean and covariance functions

$$\hat{\mu}_{N,d}(x) := \hat{k}_N(x)^\top [\hat{K}_N + \sigma_w^2 I_N]^{-1} Y_{N,d} \quad \text{and} \tag{4.4}$$

$$\hat{\sigma}^2_{N,d}(x) := \hat{k}(x,x) - \hat{k}_N(x)^\top [\hat{K}_N + \sigma_w^2 I_N]^{-1} \hat{k}_N(x), \tag{4.5}$$

where $\hat{k}_N(x) := [\hat{k}(x^i, x)]_{i=1}^N$, and $Y_{N,d}$ denotes the d^{th} column of Y_N.

Based on Definition 4.3, predictions of the next state of the system (4.2) for any current state $x \in \mathbb{X}$ can be obtained via

$$\hat{y}_d(x) \sim \mathcal{N}(\cdot\,|\hat{\mu}_{N,d}(x), \hat{\sigma}^2_{N,d}(x) + \sigma_w^2).^4$$

An illustration is provided in Fig. 4.3, following the standard practice of visualising a GP based on its posterior mean and a double standard deviation band, corresponding to a 95% confidence interval. A probabilistic bound on the prediction error is derived

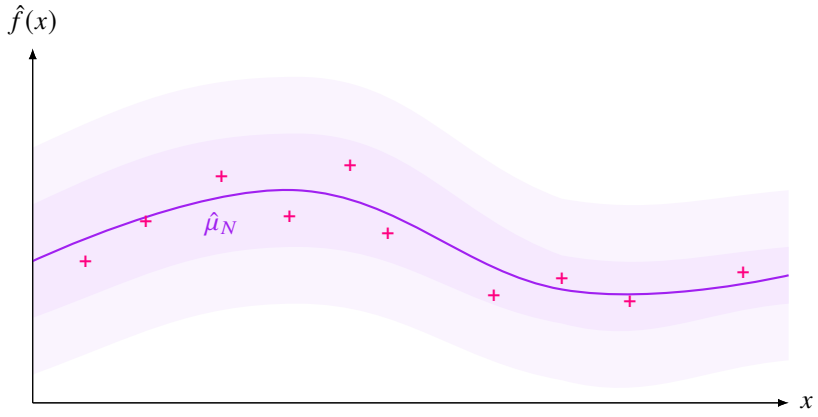


Fig. 4.3: Illustrative posterior GP ($d = 1$): based on the observed data points (in red), the GP is characterised by the posterior mean function $\hat{\mu}_N$ (purple curve) and single/double standard deviation bands $\hat{\mu}_N \pm j\hat{\sigma}_N$, $j \in \{1, 2\}$ (shaded regions).

[4] By default, the posterior distribution (4.3) models the predictive mean and variance for a noise-*free* process. To compute predictions of the original, *noisy* system (4.2), the noise variance σ_w^2 is added (Rasmussen and Williams, 2006, Section 2.2).

in Subsection 4.4.3, following the introduction of the BTGP in the next subsection.

4.4.2 Binary-Tree Gaussian Processes

This chapter explores GPs based on the so-called *binary-tree* (BT) kernel (Cohen et al., 2022b), which is defined on a finite space of binary strings and offers several practical advantages which render it particularly well-suited for formal approaches. In particular, the resulting *binary-tree Gaussian process* (BTGP) will be used to bridge between the continuous state space and a discrete space of symbolic binary strings, thereby linking the physical system to its formal abstraction (see Fig. 4.4). To this end, a mapping $\Psi_q : \mathbb{X} \to \mathbb{B}^q$ is introduced, which encodes every state $x \in \mathbb{X}$ from the continuous space $\mathbb{X}$ as a binary string $s \in \mathbb{B}^q$ in the finite space of binary strings of length q. This encoding induces a partitioning of the state space $\mathbb{X}$ into 2^q disjoint subsets; accordingly, $q \in \mathbb{N}$ is referred to as the *bit-depth* or *precision* of the resulting discrete abstraction.

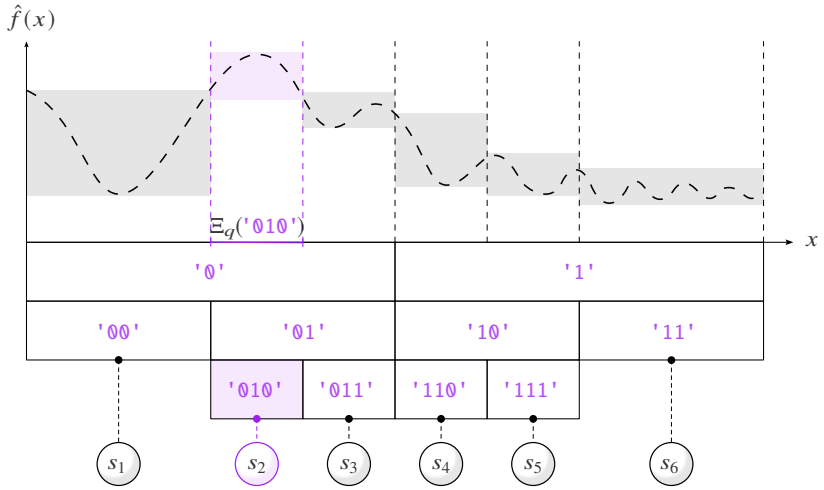


Fig. 4.4: Continuous world vs discrete world: The BTGP links the physical system to its formal abstraction. Every lowest-level string associates a symbolic state s_i.

Binary Abstraction In the following, the subset of states from the continuous space $\mathbb{X}$ that is represented by a binary string $s \in \mathbb{B}^q$ is denoted by $\Xi_q(s) \subset \mathbb{X}$, that is, $\Psi_q(x) = s$ for all $x \in \Xi_q(s)$. As an example, Fig. 4.4 illustrates how the subset $\Xi_q(\texttt{'010'}) \subset \mathbb{R}$ of the x-axis is encoded by the binary string $s = \texttt{'010'}$. It follows that the state space is partitioned as $\mathbb{X} = \bigcup_{s \in \mathbb{B}^q} \Xi_q(s)$, for any precision

$q \in \mathbb{N}$. This yields a binary abstraction of the original domain into finitely many equivalence classes of states, with the equivalence resolution dictated by the precision q. Conversely, given a set $A \subset \mathbb{X}$, its projection onto $\mathbb{B}^q$ is written as $\Psi_q(A) \subset \mathbb{B}^q$. For notational convenience, the truncation of the binary string $\Psi_q(x)$ to its first i bits is written as $\Psi_i(x) := [\Psi_q(x)]_{q \leq i}$, for all $x \in \mathbb{X}$ and $i \in \mathbb{N}$.

Symbolic States The binary strings $s \in \mathbb{B}^q$ serve as abstract symbolic states, previously denoted by $\hat{x} \in \hat{\mathbb{X}}$. The notation s is retained to emphasise the distinction between continuous *numerical* states $x \in \mathbb{X} \subset \mathbb{R}^{n_x}$ of the physical system and finite binary *strings* $s \in \mathbb{B}^q$ acting as symbolic abstract states. This representation forms the basis for a more abstract system representation, and is further explored in Chapter 5 through spectral symbolic abstractions.

With this, the BT kernel is formally defined as follows.

Binary-Tree Kernel

Definition 4.4 (Binary-Tree Kernel) Given a map $\Psi_q : \mathbb{X} \to \mathbb{B}^q$, the *BT kernel* $\hat{k}_q : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ of *bit-depth* or *precision* $q \in \mathbb{N}$ is defined as

$$\hat{k}_q(x, x') := \sum_{i=1}^{q} \mathrm{w}_i \mathbb{1}\big(\Psi_i(x) = \Psi_i(x')\big), \tag{4.6}$$

with weight coefficients $\mathrm{w}_i \in \mathbb{R}$ such that $\sum_{i=1}^{q} \mathrm{w}_i = 1$.

Intuitively, the BT kernel $\hat{k}_q$ assigns its covariance mass according to whether points $x, x' \in \mathbb{X}$ fall into the same partition. The final covariance value is computed as a weighted sum of these coincidences across all discretisation levels up to precision q.

Binary-Tree Gaussian Processes

Binary-Tree Gaussian Processes Notably — unlike well-known alternatives such as the *squared-exponential* or *Matérn* kernels — the BT kernel is discontinuous, since it is translation-*variant*. As a consequence, it induces discontinuous functional mappings when used in the context of GP regression. Fig. 4.5 depicts the resulting piecewise-constant posterior mean $\hat{\mu}_N$ and double standard deviation band $\hat{\mu}_N \pm 2\hat{\sigma}_N$ as functions of state for a GP utilising a BT kernel with precision $q = 2$. In the following, such a GP is referred to as a *binary-tree Gaussian process* (BTGP), which naturally yields discrete abstractions of the modelled system. To simplify notation, the associated posterior functions $\hat{\mu}_N(s)$ and $\hat{\sigma}^2_N(s)$ may be written explicitly as functions of the discrete strings $s \in \mathbb{B}^q$. Moving forward, it is worth noting that although function samples from a BTGP (or any GP) do not lie in the RKHS of the associated kernel almost surely, the resulting posterior mean function does (Kanagawa et al., 2018, Section 4).

Remark 4.1 (Kernel Validity) The BT kernel violates the conditions required for a continuous RKHS (Berlinet and Thomas-Agnan, 2004, Theorem 17, condition a)). Despite its discontinuity, it remains a valid positive-definite kernel, as shown by Cohen et al. (2022b, Proposition 1). Hence, by the Moore-Aronszajn theorem, it

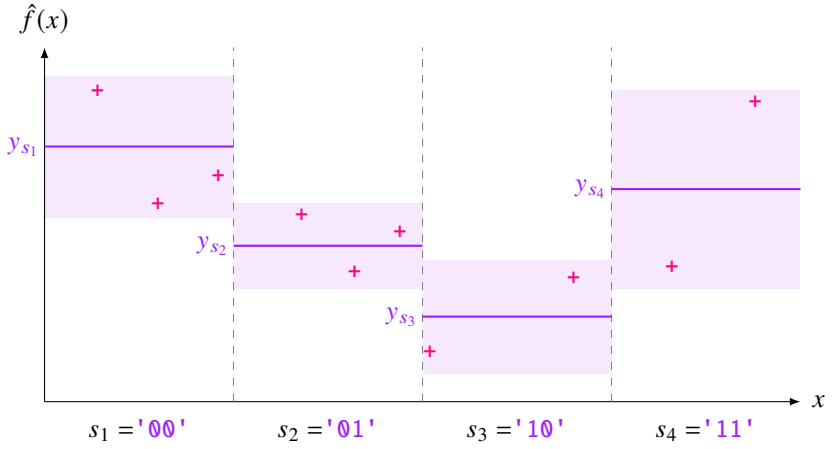


Fig. 4.5: The BTGP produces functions that are piecewise constant over the partitions of $\mathbb{X}$. The posterior mean $\hat{\mu}_N$ and double standard deviation $\hat{\mu}_N \pm 2\hat{\sigma}_N$ are shown as a piecewise-constant graph and an enclosing shaded band, respectively. For this example with chosen precision $q = 2$, the posterior mean takes values $\hat{\mu}_N \in \{y_{s_1}, \ldots, y_{s_4}\}$.

serves as a reproducing kernel of a unique RKHS (Berlinet and Thomas-Agnan, 2004, Theorem 3), which contains only discontinuous functions.

Remark 4.2 (Gram Matrix Size) An implication of employing the BT kernel of precision q in GP regression is that the resulting Gram matrix $\hat{K}_N$ is bounded in size by $2^q \times 2^q$. This is because any dataset $\{x^i, y^i\}_{i=1}^N$ can be replaced with a modified dataset $\{x_s, y_s\}_{s \in \mathbb{B}^q}$, where each $x_s \in \Xi_q(s)$ is a representative state, and y_s is defined as the average of all outputs y^i associated with the same partition s, i.e.,

$$y_s := \frac{1}{|A(s)|} \sum_{i \in A(s)} y^i, \quad \text{with } A(s) := \left\{ i \,\middle|\, x^i \in \Xi_q(s) \right\}.$$

In this sense, BTGPs can be interpreted as a type of sparse GP with naturally arising inducing points (Titsias, 2009).

Remark 4.3 (Hierarchical Efficiency of BTGPs) The performance benefits of the BTGP arise from its construction as a sequence of successive partitionings of the input space, in which the precision of each partition is either refined or retained at every discretisation level. At each level, functional mass — as represented by the corresponding weight coefficient — is applied whenever two points fall into the same partition. As demonstrated by Cohen et al. (2022b, Theorem 1), this structure enables an efficient computation of the Gram matrix and its inverse: the matrix inversion may be carried out with a complexity of $O(N)$ for each precision level $i \in \{1, \ldots, q\}$, rather than $O(iN)$, resulting in an overall complexity of $O(qN)$.

Analysis of the Binary-Tree RKHS The next two results provide further insights into the functions that can be learned by the BTGP by characterising the RKHS of the BT kernel. The first result establishes a closed-form representation of the functions in the RKHS of the BT kernel.

Proposition 4.1 (BT RKHS) *Let $\hat{k}_q : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ denote the BT kernel of finite precision $q \in \mathbb{N}$ as defined in (4.6). Then, the associated RKHS $\mathcal{H}_{\hat{k}_q}$ is finite-dimensional and contains only functions $f \in \mathcal{H}_{\hat{k}_q}$ that admit the representation*

$$f(x) = \sum_{i=1}^{q} \sqrt{\mathrm{w}_i} \sum_{s \in \mathbb{B}^i} y_s \mathbb{1}\big(x \in \Xi_i(s)\big), \tag{4.7}$$

for some coefficients $y_s \in \mathbb{R}$. The resulting RKHS $\mathcal{H}_{\hat{k}_q}$ has dimension $\sum_{i=1}^{q} 2^i$ and is characterised by the piecewise-constant feature map

$$\hat{\phi}_q(x) := \left[\sqrt{\mathrm{w}_i}\, \mathbb{1}(x \in \Xi_i(s))\right]_{s \in \mathbb{B}^i, i \in \{1,2,\ldots,q\}}. \tag{4.8}$$

The proof of Proposition 4.1 has been deferred to Appendix A.4.

The RKHS norm provides a measure for the complexity of the functions in a given RKHS. For the BT kernel, an upper bound on the RKHS norm can be computed as follows.

Proposition 4.2 (Norm Bound) *Let $\hat{k}_q : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ be the BT kernel in (4.6), and let $\mathcal{H}_{\hat{k}_q}$ denote its associated RKHS. Then, for any function $f \in \mathcal{H}_{\hat{k}_q}$ of the form (4.7), the following bound on its RKHS norm holds:*

$$\|f\|_{\mathcal{H}_{\hat{k}_q}} \leq \sqrt{\sum_{i=1}^{q} \mathrm{w}_i \sum_{s \in \mathbb{B}^i} y_s^2}.$$

The proof of Proposition 4.2 has been deferred to Appendix A.5.

4.4.3 Statistical Error Bound

To integrate GP regression into a formal framework, it is necessary to quantify the error between its predicted posterior mean $\hat{\mu}_N \in \mathcal{H}_{\hat{k}_q}$ (4.4) and the actual data-generating function $f \in \mathcal{H}_k$ (4.2). Note that the accuracy of the predictions depends on the available amount of data. Furthermore, to accommodate a broad class of functions f — beyond those discontinuous functions residing in the BT-RKHS $\mathcal{H}_{\hat{k}_q}$ — Assumption 4.1 is leveraged to quantify the additional error introduced by the mismatch in function spaces. The following theorem establishes a statistical

error bound on the BTGP predictions, drawing upon the results by Hsu et al. (2012, Theorem 2.1) and Fiedler et al. (2021, Proposition 2) — the latter being an adaptation of Theorem 2 by Chowdhury and Gopalan (2017). As previously mentioned, each output dimension (indexed by d) is modelled by an independent GP, thus the error bound is established dimension-wise.

Statistical Error Bound

Theorem 4.1 (Piecewise-Constant Error Bound) *Consider the BT kernel $\hat{k}_q$ in (4.6) with precision $q \in \mathbb{N}$, and let $\mathcal{H}_{\hat{k}_q}$ denote its associated RKHS. Suppose that the unknown function $f \in \mathcal{H}_k$, where k is a known translation-invariant kernel satisfying $c^2 \geq k(x,x) \geq 0$ for all $x \in \mathbb{X}$ and some constant $c > 0$, and assume $\|f\|_{\mathcal{H}_k} \leq B$ for a known complexity bound $B \geq 0$ as per Assumption 4.1. Given data $\mathcal{D}_N = (X_N, Y_N)$ and a desired confidence level $1-\alpha \in (0,1)$, it holds for each output dimension $d \in \{1,\dots,n\}$ that the corresponding BTGP posterior mean $\hat{\mu}_{N,d}(s) \in \mathcal{H}_{\hat{k}_q}$ in (4.4) satisfies*

$$\mathbb{P}\Big(\forall s \in \mathbb{B}^q : \sup_{x \in \Xi(s)} \left|\hat{\mu}_{N,d}(s) - f_d(x)\right| \leq \varepsilon_d(s)\Big) \geq 1-\alpha, \tag{4.9}$$

where the total error $\varepsilon_d(s) := \varepsilon_{d,1}(s) + \varepsilon_{d,2}(s) + \varepsilon_{d,3}(s)$ consists of the following components:

$$\varepsilon_{d,1}(s) := \min\Bigg\{\left\|\hat{C}_N^{-1}\hat{k}_N(s)\right\| \sqrt{N + 2\sqrt{N\log\left(\alpha^{-1}\right)} + 2\log\left(\alpha^{-1}\right)},$$
$$\frac{\hat{\sigma}_{N,d}(s)}{\sigma_w}\sqrt{\mathrm{Tr}(\Sigma) + 2\sqrt{\mathrm{Tr}(\Sigma^2)\log\left(\alpha^{-1}\right)} + 2\|\Sigma\|\log\left(\alpha^{-1}\right)}\Bigg\},$$
$$\varepsilon_{d,2}(s) := B\sqrt{2\left|c^2 - k^+_{x_s}(s)\right|}, \quad \text{and}$$
$$\varepsilon_{d,3}(s) := B\sqrt{\left|\hat{k}_N(s)^\top \hat{C}_N^{-1}\left(K_N\hat{C}_N^{-1}\hat{k}_N(s) - 2k_N(x_s)\right) + c^2\right|},$$

with $\hat{C}_N := [\hat{K}_N + \sigma_w^2 I_N]$, $\hat{k}_N(s) := [k(x^i, \Psi_q(s))]_{i=1}^N$, posterior covariance $\hat{\sigma}^2_{N,d}(s)$ as in (4.5), $\Sigma := \hat{K}_N\hat{C}_N^{-1}$, $k^+_{x_s}(s) := \sup_{x\in\Xi(s)} k(x_s, x)$ for a fixed set of representative points $\{x_s\}_{s\in\mathbb{B}^q} \subset \mathbb{X}$, $K_N := [k(x^i,x^j)]_{i,j=1}^N$, and $k_N(x) := [k(x^i,x)]_{i=1}^N$.

The proof of Theorem 4.1 is deferred to Appendix A.6.

Theorem 4.1 establishes a non-asymptotic bound on the error between the BTGP posterior mean $\hat{\mu}_{N,d}$ and the true, unknown function f. By imposing a complexity bound B on $\|f\|_{\mathcal{H}_k}$, error guarantees can be provided even when $f \notin \mathcal{H}_{\hat{k}}$, i.e., in the so-called "misspecified case", such as when f is continuous.

Unlike Fiedler et al. (2021), the error ε_d is quantified without requiring a global bound on the kernel approximation error. This is advantageous, as such a bound would be maximised in the present setting due to the general mismatch between $\hat{k}_q$

and k, even for $q \to \infty$. In fact, whilst the error terms $\varepsilon_{d,2}$ and $\varepsilon_{d,3}$ can be reduced by increasing the BTGP precision level q, this is not the case for $\varepsilon_{d,1}$.

The first term in the minimisation of $\varepsilon_{d,1}$ corresponds to the well-specified case and reflects the statistical error, which is generally very conservative. To mitigate this, a second term is introduced.

Computation of $\varepsilon_{d,3}$ necessitates the evaluation of Gram matrices involving both k and $\hat{k}_q$, which diminishes the scalability benefits of the BTGP. This motivates future work on quantifying ε_d without requiring access to K_N.

Remark 4.4 (Limitations of Approximating Continuous Kernels) Although it may appear tempting to use the BT kernel $\hat{k}_q$ (or a similar hierarchical kernel) to approximate a continuous kernel (such as the squared-exponential) in the limit $q \to \infty$ to simplify Theorem 4.1, no such convergent kernel exists. For any fixed $x \in \mathbb{X}$ and variable $x' \in \mathbb{X}$, the kernels $\hat{k}_q$ and $\hat{k}_{q'}$ with $q' > q$ define two discrete distance metrics that concentrate their increasing precision solely within the neighbourhood of x. As such, raising the precision enhances smoothness locally but leaves the covariance structure elsewhere unaffected.

Comparison to Simulation Relations In contrast to the probabilistic simulation relations developed in earlier chapters — where error bounds between the concrete system and its discrete abstraction were established via the computation of local error parameters (ε, δ) — the error bound $\varepsilon_d(s)$ in Theorem 4.1 is expressed as a global formula. When evaluated for individual partitions $s \in \mathbb{B}^q$, this formula yields analogous local error estimates, removing the need for computing error parameters for each partition individually. However, this advantage comes at the potential cost of conservatism, as the resulting bound may not be uniformly tight across all regions of the state space.

4.5 Verification Approach

The verification approach adopted here is based on the construction of a finite-state abstraction of the dynamical system $\boldsymbol{\mathcal{M}}$ in (4.2) via the BTGP introduced in Subsection 4.4.2. For the learned BTGP, let $\mathbb{B}^q$ denote a finite set of binary strings representing the abstract states. Recall that the BTGP induces a mapping $\Psi_q : \mathbb{X} \to \mathbb{B}^q$ associating each continuous state $x \in \mathbb{X}$ of $\boldsymbol{\mathcal{M}}$ with a binary string $s \in \mathbb{B}^q$. Thus, each string corresponds to a partition of $\mathbb{X}$ and is associated with a chosen representative point $x_s \in \mathbb{X}$, here chosen as the centre point of the respective partition.

While this yields a natural discretisation of the continuous-state system $\boldsymbol{\mathcal{M}}$, the statistical error characterised in Theorem 4.1 introduces ambiguity in the transition probability distributions, precluding the construction of a discrete Markov chain. To model this uncertainty, interval Markov chains (IMCs) are employed, in which transition probabilities between abstract states are represented by intervals. Prior work has demonstrated that IMCs can be learned from data via GP regression

(Jackson et al., 2020; Jiang et al., 2022). However, for standard continuous GPs, the construction of such IMCs constitutes a significant computational bottleneck.

Here, the same strategy is adopted, but regression is carried out using the BTGP (Subsection 4.4.2), leading to a substantial reduction in the computational cost of IMC construction. A formal definition is given below.

Interval Markov Chain

Definition 4.5 (Interval Markov Chain (IMC)) A (finite-state) *IMC* is a tuple $\widehat{\mathcal{M}} := (\mathbb{S}, \mathbb{S}_0, \underline{\mathbf{t}}, \bar{\mathbf{t}})$, with

- $\mathbb{S}$, a finite state space, with states $s \in \mathbb{S}$;
- $\mathbb{S}_0 \subset \mathbb{S}$, the set of initial states $s_0 \in \mathbb{S}_0$;
- $\underline{\mathbf{t}}, \bar{\mathbf{t}} \colon \mathbb{S} \times \mathbb{S} \to [0,1]$, lower and upper bounds on the transition probabilities, specifying an interval $0 \leq \underline{\mathbf{t}}(s_+|s) \leq \bar{\mathbf{t}}(s_+|s) \leq 0$ for each state pair $(s, s_+) \in \mathbb{S}$.

4.5.1 BTGP-Based IMC Construction

The continuous-state MC $\mathcal{M} = (\mathbb{X}, \mathbb{X}_0, \mathbf{t})$ in (4.2) is abstracted into a finite-state IMC $\widehat{\mathcal{M}} := (\mathbb{S}, \mathbb{S}_0, \underline{\mathbf{t}}, \bar{\mathbf{t}})$ by employing the BT-partitioned state space $\mathbb{S} := \mathbb{B}^q$, with the initial states given by $\mathbb{S}_0 := \Psi_q(\mathbb{X}_0)$. Recall that dynamics of $\mathcal{M}$ are modelled using one BTGP for each output dimension. Consequently, n_x BTGPs $\hat{f}_d \sim \mathcal{GP}(\hat{\mu}_{N,d}, \hat{\sigma}^2_{N,d})$, $d \in \{1, \ldots, n_x\}$, are obtained, where the same precision $q \in \mathbb{N}$ is adopted to generate a common partitioning $\mathbb{S} \equiv \Psi_q(\mathbb{X})$.

The discrete transition bounds $\underline{\mathbf{t}}, \bar{\mathbf{t}}$ are computed from the learned BTGPs s.t.

$$\underline{\mathbf{t}}(s_+|s) \leq \mathbf{t}(\Psi_q(s_+)|\Psi_q(s)) \leq \bar{\mathbf{t}}(s_+|s), \quad \forall s, s_+ \in \mathbb{S}.$$

In contrast to continuous GPs, the piecewise-constant nature of the posterior mean and covariance functions of the BTGP renders this computation comparatively straightforward (see Fig. 4.6(b)). More specifically, for every partition $s \in \mathbb{S}$, the consecutive state $s_+ := [s_{+,d}]_{d=1}^{n_x} \in \mathbb{S}$ has Gaussian distribution

$$x_{+,d} \sim \mathcal{N}(\cdot\,|\hat{\mu}_{N,d}(s) + \epsilon_d(s),\, \hat{\sigma}^2_{N,d}(s)), \quad d \in \{1, \ldots, n_x\},$$

where the error $\epsilon_d(s) \in E_d(s) := [-\varepsilon_d(s), \varepsilon_d(s)]$ is obtained via the piecewise-constant statistical error bound ε_d derived in Theorem 4.1.

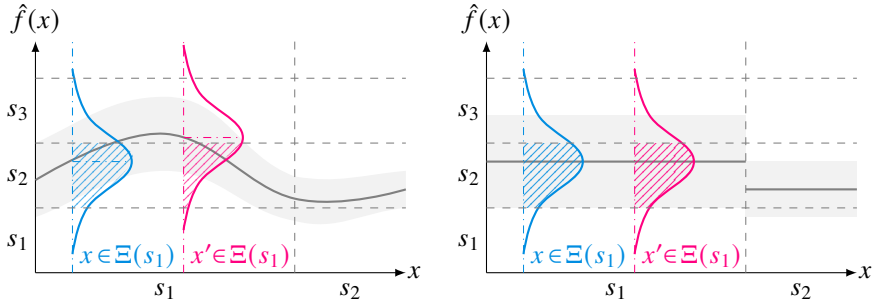
Transition Probability Bounds Computation

Lower and upper bounds on the transition probability from s to any $s_+ \in \mathbb{S}$ are thus obtained as

$$\underline{\mathbf{t}}(s_+|s) = \min_{\epsilon \in E(s)} g(s, \epsilon) \quad \text{and} \quad \bar{\mathbf{t}}(s_+|s) = \max_{\epsilon \in E(s)} g(s, \epsilon), \tag{4.10a}$$

$$g(s, \epsilon) := \int_{x_+ \in \Xi(s_+)} \prod_{d=1}^{n_x} \mathcal{N}(dx_{+,d}|\hat{\mu}_{N,d}(s) + \epsilon_d, \hat{\sigma}^2_{N,d}(s)), \tag{4.10b}$$

where $E(s) := \prod_{d=1}^{n_x} E_d(s)$, $\epsilon = [\epsilon_d]_{d=1}^{n_x}$, and $x_+ = [x_{+,d}]_{d=1}^{n_x}$. In the case where the partitions of the BTGP are axis-aligned hyperrectangles, the integration in (4.10b)



(a) Continuous GP (b) BTGP

Fig. 4.6: Computing transition probabilities from a continuous GP (a) vs from a BTGP (b).

and the subsequent optimisations in (4.10a) can be performed independently for each dimension.

Comparison to Continuous GPs By contrast, for a GP with a kernel that is not piecewise-constant — such as illustrated in Fig. 4.6(a) for the squared-exponential kernel —, the bounds must be computed by performing an additional optimisation optimise $\in \{\max, \min\}$ in each function call

$$g(s, \epsilon) = \underset{x \in \Xi(s)}{\text{optimise}} \int_{x_+ \in \Xi(s_+)} \prod_{d=1}^{n_x} \mathcal{N}(dx_{+,d} | \hat{\mu}_{N,d}(x) + \epsilon_d(x), \hat{\sigma}^2_{N,d}(x)) \tag{4.11}$$

for each transition pair (s, s_+). Thus, the BTGP removes the necessity for this inner optimisation, resulting in a significant reduction of the computational burden involved in the IMC construction. Consequently, the computation is limited to a single integral per bound per transition, which can be performed efficiently using the CDF as shown in Algorithm 1. Here, $\text{sgn}(\cdot)$ and "$\odot$" denote the sign function and element-wise product, respectively. The functions ComputeRep and ComputeBounds are defined as follows:

- ComputeRep determines the centre point of a given partition $s \in \mathbb{S}$ as the corresponding representative state $x_s \in \mathbb{X}$.
- ComputeBounds computes dimension-wise bounds $\text{lb}_{s,d}, \text{ub}_{s,d} \in \mathbb{X}_d$ (where $\mathbb{X} = \mathbb{X}_1 \times \cdots \times \mathbb{X}_d$) for each axis-aligned hyperrectangular partition $s \in \mathbb{S}$.

4.5.2 Verification via Interval Iteration Algorithm

Interval Iteration Algorithm

Once the finite-state IMC has been constructed via Algorithm 1, upper and lower bounds on the probability of satisfying a given reachability specification $\psi_{\mathcal{T}}$ may be

Algorithm 1 BTGP-Based IMC Construction

Input: MC $\mathcal{M} = (\mathbb{X}, \mathbb{X}_0, \mathbf{t})$, BTGP posteriors $(\hat{\mu}_{N,d}, \hat{\sigma}^2_{N,d})_{d=1}^{n_x}$, error bounds $(\varepsilon_d)_{d=1}^{n_x}$

$\mathbb{S} \leftarrow \mathbb{B}^q$
$X_s \leftarrow [x_s]_{s\in\mathbb{S}} \leftarrow \textsc{ComputeRep}(\mathbb{S})$
$(\mathrm{lb}_{s,d}, \mathrm{ub}_{s,d})_{s\in\mathbb{S},d\in\{1,\dots,n_x\}} \leftarrow \textsc{ComputeBounds}(\mathbb{X}, \mathbb{S})$
$\mathbb{S}_0 \leftarrow \Psi_q(\mathbb{X}_0)$
$\underline{\mathbf{t}}, \bar{\mathbf{t}} \leftarrow 1, 1$
for $d \in \{1, \dots, n_x\}$ **do**
  for $s \in \mathbb{S}$ **do**
    $\Delta \leftarrow \hat{\mu}_{N,d}(s) - X_{s,d}$ ▷ Distances from mean to representative states
    $\Delta_{\max} \leftarrow |\Delta| + \varepsilon_d(s)$ ▷ "Worst-case" error
    $\hat{\mu}_{\text{worst}} \leftarrow X_{s,d} + \mathrm{sgn}(\Delta) \odot \Delta_{\max}$
    $\underline{p} \leftarrow \mathrm{cdf}\left(\mathrm{ub}_{s,d} \,\middle|\, \hat{\mu}_{\text{worst}}, \hat{\sigma}^2_{N,d}(s)\right) - \mathrm{cdf}\left(\mathrm{lb}_{s,d} \,\middle|\, \hat{\mu}_{\text{worst}}, \hat{\sigma}^2_{N,d}(s)\right)$ ▷ Lower bound
    $\Delta_{\min} \leftarrow \max\{0, |\Delta| - \varepsilon_d(s)\}$ ▷ "Best-case" error
    $\hat{\mu}_{\text{best}} \leftarrow X_{s,d} + \mathrm{sgn}(\Delta) \odot \Delta_{\min}$
    $\overline{p} \leftarrow \mathrm{cdf}\left(\mathrm{ub}_{s,d} \,\middle|\, \hat{\mu}_{\text{best}}, \hat{\sigma}^2_{N,d}(s)\right) - \mathrm{cdf}\left(\mathrm{lb}_{s,d} \,\middle|\, \hat{\mu}_{\text{best}}, \hat{\sigma}^2_{N,d}(s)\right)$ ▷ Upper bound
    $\bar{\mathbf{t}}(\cdot\,|s) \leftarrow \bar{\mathbf{t}}(\cdot\,|s) \odot \overline{p}$
    $\underline{\mathbf{t}}(\cdot\,|s) \leftarrow \underline{\mathbf{t}}(\cdot\,|s) \odot \underline{p}$
  end for
end for
Output: IMC $\widehat{\mathcal{M}} = (\mathbb{S}, \mathbb{S}_0, \underline{\mathbf{t}}, \bar{\mathbf{t}})$

obtained using, for instance, the *interval iteration algorithm* (IIA) proposed by Haddad and Monmege (2018). Similar to the dynamic programming framework outlined in Subsection 2.4.3, the IIA is based on the iterative update of (two) value functions, $\underline{V}_i, \overline{V}_i : \mathbb{S} \to [0, 1]$ defined on the discrete space $\mathbb{S}$. These functions form an envelope or *interval* around the true probability of satisfying the specification starting from some initial partition $s_0 \in \mathbb{S}_0$. The corresponding algorithmic procedure is provided in Algorithm 2.

Inter alia, Haddad and Monmege (2018) establish the following result.

Proposition 4.3 (Value Convergence) *For a given IMC $\widehat{\mathcal{M}} := (\mathbb{S}, \mathbb{S}_0, \underline{\mathbf{t}}, \bar{\mathbf{t}})$ and an (infinite-horizon) reachability specification $\psi_{\mathcal{T}}$, let $V^* : \mathbb{S} \to [0, 1]$ be the* optimal value function *such that $\mathbb{P}(\widehat{\mathcal{M}} \models \psi_{\mathcal{T}}) = V^*(s_0)$, for $\widehat{\mathcal{M}}$ initialised in $s_0 \in \mathbb{S}_0$. Then, the IIA (Algorithm 2) converges to V^* as $i \to \infty$, for any $\nu > 0$. Furthermore, for every iteration $i \in \mathbb{N}_0$ it holds that $\underline{V}_i(s_0) \leq V^*(s_0) \leq \overline{V}_i(s_0)$.*

Before giving a summary of Algorithm 2, the following *linear program* (LP) is defined, where optimise $\in \{\max, \min\}$:

Algorithm 2 Interval Iteration Algorithm (IIA)

Input: IMC $\widehat{\mathcal{M}} = (\mathbb{S}, \mathbb{S}_0, \underline{\mathbf{t}}, \overline{\mathbf{t}})$, reachability specification $\psi_{\mathcal{T}}$, stopping threshold $\nu > 0$

1: **for** $s \in \mathbb{S}$ **do**
2: $\quad \overline{r}(s) \leftarrow \min\left\{\sum_{s_+ \in \Psi_q(\mathcal{T})} \overline{\mathbf{t}}(s_+|s),\ 1\right\}$
3: $\quad \underline{r}(s) \leftarrow \sum_{s_+ \in \Psi_q(\mathcal{T})} \underline{\mathbf{t}}(s_+|s)$
4: $\quad \overline{l}(s) \leftarrow 1 - \sum_{s_+ \in \mathbb{S}} \underline{\mathbf{t}}(s_+|s)$
5: $\quad \underline{l}(s) \leftarrow \max\left\{0,\ 1 - \sum_{s_+ \in \mathbb{S}} \overline{\mathbf{t}}(s_+|s)\right\}$
6: **end for**
7: $i \leftarrow 0, \overline{V}_0 \leftarrow 1, \underline{V}_0 \leftarrow 0$
8: **while** $\|\overline{V}_i - \underline{V}_i\|_\infty > \nu$ **do**
9: $\quad r, \mathbf{t} \leftarrow$ solve min LP in (4.12)
10: $\quad$ **for** $s \in \mathbb{S}$ **do**
11: $\quad\quad \overline{V}_{i+1}(s) \leftarrow r(s) + \sum_{s_+ \in \mathbb{S} \setminus \Psi_q(\mathcal{T})} \mathbf{t}(s_+|s) \overline{V}_i(s_+)$
12: $\quad\quad \underline{V}_{i+1}(s) \leftarrow r(s) + \sum_{s_+ \in \mathbb{S} \setminus \Psi_q(\mathcal{T})} \mathbf{t}(s_+|s) \underline{V}_i(s_+)$
13: $\quad$ **end for**
14: $\quad i \leftarrow i + 1, \overline{V}_i \leftarrow \overline{V}_{i+1}, \underline{V}_i \leftarrow \underline{V}_{i+1}$
15: **end while**
16: $V_{\min} \leftarrow \underline{V}_i$ ▷ Take underapproximation
17: $i \leftarrow 0, \overline{V}_0 \leftarrow 1, \underline{V}_0 \leftarrow 0$
18: **while** $\|\overline{V}_i - \underline{V}_i\|_\infty > \nu$ **do**
19: $\quad r, \mathbf{t} \leftarrow$ solve max LP in (4.12)
20: $\quad$ **for** $s \in \mathbb{S}$ **do**
21: $\quad\quad \overline{V}_{i+1}(s) \leftarrow r(s) + \sum_{s_+ \in \mathbb{S} \setminus \Psi_q(\mathcal{T})} \mathbf{t}(s_+|s) \overline{V}_i(s_+)$
22: $\quad\quad \underline{V}_{i+1}(s) \leftarrow r(s) + \sum_{s_+ \in \mathbb{S} \setminus \Psi_q(\mathcal{T})} \mathbf{t}(s_+|s) \underline{V}_i(s_+)$
23: $\quad$ **end for**
24: $\quad i \leftarrow i + 1, \overline{V}_i \leftarrow \overline{V}_{i+1}, \underline{V}_i \leftarrow \underline{V}_{i+1}$
25: **end while**
26: $V_{\max} \leftarrow \overline{V}_i$ ▷ Take overapproximation

Output: Bounds $(V_{\min}, V_{\max})$

$$
\begin{aligned}
\underset{r,\mathbf{t}}{\text{optimise}} \quad & r(s) + \textstyle\sum_{s_+ \in \mathbb{S} \setminus \Psi_q(\mathcal{T})} \mathbf{t}(s_+|s) V_i(s_+), \\
\text{subject to} \quad & \underline{\mathbf{t}}(s_+|s) \le \mathbf{t}(s_+|s) \le \overline{\mathbf{t}}(s_+|s), \\
& \underline{r}(s) \le r(s) \le \overline{r}(s), \\
& \underline{l}(s) \le l(s) \le \overline{l}(s), \\
& r(s) + l(s) + \textstyle\sum_{s_+ \in \mathbb{S} \setminus \Psi_q(\mathcal{T})} \mathbf{t}(s_+|s) = 1.
\end{aligned}
\tag{4.12}
$$

Lines 2 to 5 of Algorithm 2 compute functional lower and upper bounds $(\underline{r}, \overline{r})$ on the *reward* — the probability of reaching the projected target region $\Psi_q(\mathcal{T}) \subset \mathbb{S}$ in one time step starting from the state $s \in \mathbb{S}$. Analogously, $(\overline{l}, \underline{l})$ represent functional bounds on the *loss* — the probability of leaving the state space $\mathbb{S}$. In line 7, the overapproximation $\overline{V}_0$ and underapproximation $\underline{V}_0$ of the *value function* — the probability of *eventually* reaching the target $\mathcal{T}$ starting from partition s — are initialised. Until convergence, the following two steps are repeated. First, for every state pair $s, s_+ \in \mathbb{S}$, the LP (4.12) is minimised in line 9 to find feasible rewards r and transition probabilities $\mathbf{t}(s_+|s)$ that capture the worst-case behaviour of the

system. Then, $\overline{V}_0$ and $\underline{V}_0$ are updated via the *Bellman equation* in lines 11 and 12. The while-loop converges when the maximal difference between the over- and under-approximation is less than the predefined closeness bound $\nu > 0$. After convergence, the under-approximation $\underline{V}_i$ is taken as the lower bound $V_{\min}$ on the value function in line 16. The same procedure is repeated to obtain an overapproximation $V_{\max}$ in lines 17 to 26, where the LP is maximised in line 19. From Proposition 4.3, the true latent value V^* lies between the obtained upper and lower bounds, i.e., $V_{\min} \leq V^* \leq V_{\max}$.

Based on Theorem 4.1 and Proposition 4.3, the following result establishes bounds on the robust satisfaction probability of the latent data-generating system $\boldsymbol{\mathcal{M}}$ in (4.2) with respect to a reachability specification $\psi_{\mathcal{T}}$.

Theorem 4.2 (Robust Satisfaction Probability) *Consider the setup of Theorem 4.1. Let $\hat{f}_d \sim \mathcal{GP}(\hat{\mu}_{N,d}, \hat{\sigma}^2_{N,d})$, $d \in \{1, \ldots, n\}$, be the posterior BTGPs obtained based on data $\mathcal{D}_N$ from the system $\boldsymbol{\mathcal{M}}$ in (4.2). Let $\widehat{\boldsymbol{\mathcal{M}}}$ be the corresponding IMC constructed via Algorithm 1. Then, for an (infinite-horizon) reachability specification $\psi_{\mathcal{T}}$ and $\boldsymbol{\mathcal{M}}$ initialised at some $x_0 \in \mathbb{X}_0$, it holds that*

$$V_{\min}(s_0) \leq \mathbb{P}(\boldsymbol{\mathcal{M}}(x_0) \models \psi_{\mathcal{T}}) \leq V_{\max}(s_0),$$

where $s_0 := \Psi_q(x_0)$, and $(V_{\min}, V_{\max})$ are the bounds obtained via the IIA (Algorithm 2) for any $\nu > 0$.

The result follows directly from Theorem 4.1 and Proposition 4.3.

Solution Theorem 4.2 solves Problem 4.1, where $\mathbb{P}(\boldsymbol{\mathcal{M}} \models \psi_{\mathcal{T}}) \geq V_{\min}$.

4.6 Case Study

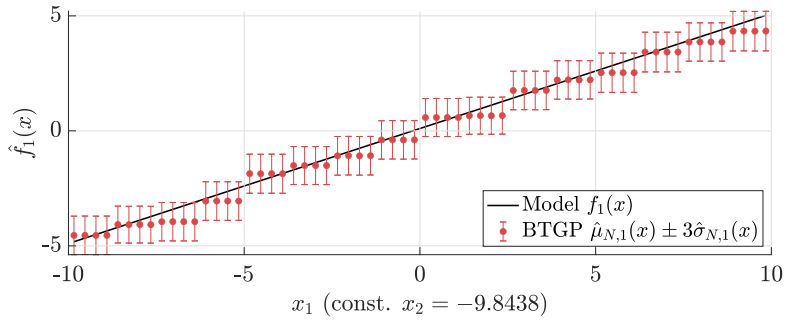
Consider the autonomous discrete-time nonlinear system

$$\boldsymbol{\mathcal{M}}\colon \begin{cases} \begin{bmatrix} x_{1,t+1} \\ x_{2,t+1} \end{bmatrix} = \begin{bmatrix} x_{1,t} - \tau_s x_{1,t} + 0.5\tau_s \sin(x_{2,t}) \\ x_{2,t} - \tau_s x_{2,t} + 0.5\tau_s \sin(x_{1,t}) \end{bmatrix} + w_t, \quad w_t \sim \mathcal{N}(\cdot\,|0, \sigma_w^2 I_2), \end{cases}$$

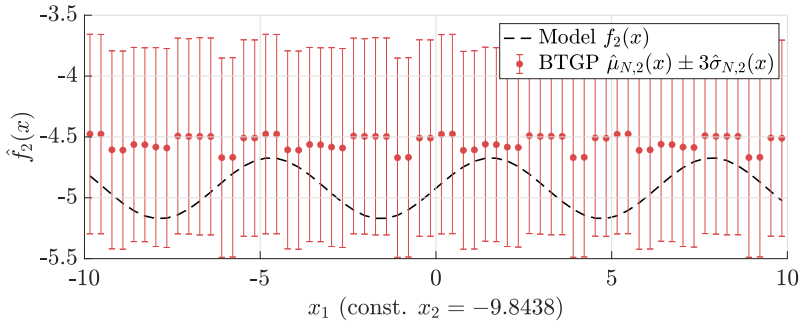
evolving on the continuous state space $\mathbb{X} = [-10, 10]^2 \subset \mathbb{R}^2$, with a time discretisation of $\tau_s := 0.5$, and additive Gaussian noise of standard deviation $\sigma_w = 3.16$. The objective is to compute a lower bound on the probability of the system reaching the target set $\mathcal{T} = [-3, 3]^2 \subset \mathbb{X}$ from a given initial position $x_0 \in \mathbb{X}_0 := \mathbb{X}$. To this end, $N = 5000$ transitions are generated from the system as outlined in Subsection 4.3.2. Based on the data, two BTGPs with precision $q = 6$ are trained — one for each output dimension of the system. For the computation of the statistical error bounds, a squared-exponential kernel (4.1) with hyperparameters $\sigma_{f,1} = 12$,

$\sigma_{f,2} = 7$, $\sigma_{l,1} = [4000, 2500]$, and $\sigma_{l,2} = [500, 2000]$ is employed. The complexity bounds for the two output dimensions are set to $B_1 = 0.015$ and $B_2 = 0.006$.

Training and evaluation of the BTGPs at the representative states takes 42 s on an Apple M1 Pro CPU. Fig. 4.7 illustrates the model fit of the BTGPs. A statistical



(a) Dimension 1 ($d = 1$)



(b) Dimension 2 ($d = 2$)

Fig. 4.7: BTGP model fit $\hat{\mu}_{N,d}(x) \pm 3\hat{\sigma}_{N,d}(x)$ versus true $f_d(x)$.

error bound is computed via Theorem 4.1 for a confidence level of $1 - \alpha := 0.8$. The representative states $\{x_s\}_{s \in \mathbb{B}^q}$ correspond to the centre points of the generated partitions. Owing to the piecewise-constant posterior of the BTGP, the IMC can be constructed via (4.10a) in under 3 s (serial implementation). In contrast, abstracting a comparable continuous GP using (4.11) requires approximately 12 h. The IMC is verified using the IIA, which converges after 3 s with a termination threshold of $\nu = 10^{-8}$. The resulting lower bound on the probability of reaching $\mathcal{T}$ is shown in Fig. 4.8 as a function of the initial state.

This example illustrates why existing IMC generation tools rely crucially on massive parallelisation (Wooding and Lavaei, 2024; Mathiesen et al., 2024). However, such parallelism does not address the exponential growth in complexity with system

dimension and can be applied equally within the BTGP–IMC framework, making it an orthogonal consideration rather than a distinguishing feature.

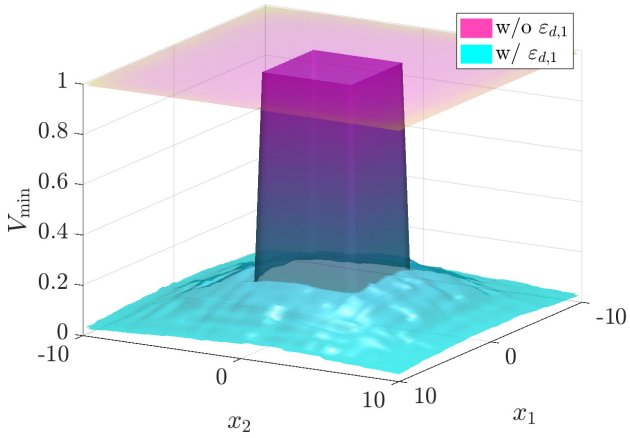


Fig. 4.8: Lower bound on the infinite-horizon reachability probability as a function of the initial state.

Comparison The verification procedure is repeated with the first statistical error component $\varepsilon_{d,1}$ in Theorem 4.1 set to zero, yielding a reachability probability lower bound close to 1 (see Fig. 4.8). This demonstrates that the conservativeness of the proposed method is predominantly driven by this term, which arises directly from standard statistical bounds used widely in GP-based formal frameworks (Fiedler et al., 2021; Chowdhury and Gopalan, 2017). Neglecting this term places the approach on par with existing methods that either assume model correctness or offer only asymptotic guarantees — such as NN-based dynamics models (Adams et al., 2022) or the reachability method explored by Lew and Pavone (2021). Crucially, while such approaches remain oblivious to statistical error under finite data, the BTGP framework can explicitly account for it and reveals its substantial influence on the final verification result. Interestingly, in the limit where the total error $\varepsilon_d = \varepsilon_{d,1} + \varepsilon_{d,2} + \varepsilon_{d,3}$ vanishes — e.g., with sufficient data and negligible approximation error — the BTGP-induced IMC collapses into a Markov chain (see Equation 4.10b). This reduction enables direct application of the extensive model-based MDP literature.

This comparison yields two key insights. First, that accounting for statistical error is not only feasible but has a decisive impact. Second, that under high data availability and high precision q, the BTGP-based method can yield substantially stronger guarantees than existing black-box approaches. At the same time, this highlights the pressing need for tighter, less conservative statistical error bounds for GPs — a challenge orthogonal to this thesis.

While the previous chapters benefited from known parametric models, the final chapter addresses a setting where formally relating statistical error to data availability becomes significantly more intricate. Although theoretical bounds exist (Li et al., 2022) — similar to the case of GPs, explored in this chapter — they are often overly conservative or impractical to compute. Accordingly, moving forward, the error bound is treated as a given input, and its influence on the verification outcome is explored across a range of values.

4.7 Conclusion

This chapter investigated binary-tree Gaussian processes (BTGPs) as a data-driven framework for learning formal abstractions of uncertain stochastic systems directly from data. By employing a binary-tree kernel that induces piecewise-constant posterior GPs, BTGPs abstract continuous-space dynamics into discrete-state IMCs over binary strings. This sidesteps the computational bottlenecks of traditional GP-based abstraction pipelines, which are dominated by explicit discretisation and error quantification procedures. The resulting abstractions support formal verification and synthesis *by design*; here, demonstrated for verification against reachability specifications, with IMCs generated orders of magnitude faster than with conventional GPs. Numerical results underscore the value of encoding structurally aligned priors in learning-enabled models for effective integration with formal tools. Crucially, it was shown that statistical error bounds can be derived to provide robust end-to-end guarantees — without relying on predefined parametric models, as assumed in earlier chapters. Together, these results position BTGPs as a promising interface between latent physical systems and formal frameworks.

While this chapter focused on verification, the same machinery extends naturally to synthesis and scLTL specifications — by composing the IMC with a DFA encoding the temporal property (Kamgarpour et al., 2013), as done in previous chapters. Moreover, IMC/IMDP models broaden applicability to full LTL semantics via an additional preprocessing step on the graph structure of the abstraction (Baier and Katoen, 2008; Haddad and Monmege, 2018).

Despite these strengths, limitations remain. The statistical error bound in Theorem 4.1 reflects two sources of conservatism: first, classical GP error bounds (Fiedler et al., 2021; Chowdhury and Gopalan, 2017) are extremely conservative and exhibit poor scaling with increasing data — highlighting the need to explore sharper bounds; second, even as the BTGP resolution increases (i.e., tree depth grows), the binary-tree kernel does not converge to a smooth kernel, such as the squared exponential kernel. This lack of asymptotic smoothness complicates error analysis and limits approximation quality during learning.

These limitations point toward several interesting directions for future work: designing alternative hierarchical kernels that preserve the efficiency of BTGPs whilst expanding expressiveness; enabling adaptive refinement of abstractions at varying resolutions; and exploring analogues of *deep kernel learning* (DKL) for

tree-structured kernels, akin to recent work on GP–IMDP approaches enhanced with neural feature maps (Reed et al., 2023). In this context, a key challenge is to preserve statistical guarantees: by modifying the feature map, DKL implicitly redefines the underlying space of learnable functions — raising open questions about how to ensure the true dynamics remain contained within it.

The next chapter builds on these findings and ideas, relaxing many of the assumptions made here to address the formal verification of fully black-box stochastic systems.

Chapter 5
Kernel-Based Learning of Black-Box Safety Certificates

Abstract The previous chapter demonstrated that kernel methods offer a powerful framework for learning structured abstractions of latent stochastic systems without relying on strong modelling assumptions. To reach a level of abstraction required for certifying autonomous systems with complex, often opaque AI components in the loop, this chapter adopts a fully black-box perspective. A data-driven safety verification framework is developed, that learns control barrier certificates directly from observed system transitions, without requiring access to the underlying dynamics. Conditional mean embeddings are used to obtain a latent approximation of the unknown stochastic dynamics in a reproducing kernel Hilbert space, capturing uncertainty and enabling the construction of ambiguity sets that facilitate distributionally robust guarantees. A spectral abstraction based on a finite Fourier expansion of the kernel is introduced to cast the data-driven barrier problem into a tractable linear program. Leveraging the fast Fourier transform, the spectral barrier can be computed efficiently, yielding a scalable yet distributionally robust approach to verifying safety. Theoretical results show how the method naturally extends beyond safety to general ω-regular temporal logic specifications. The approach moves beyond restrictive assumptions on system dynamics and uncertainty, as demonstrated on several case studies, including a black-box system with a neural network controller.



5.1 Introduction

The preceding chapters established rigorous results for abstraction-based verification and synthesis under both aleatoric and epistemic uncertainty. However, these guarantees were contingent on specific structural assumptions about the latent system. In fact, formal reasoning under unknown dynamics is only feasible when data is complemented by sufficient regularity assumptions. Previously, this

was achieved either through a parametrised model or — more recently — by imposing a GP structure, modelling uncertainty as additive Gaussian noise[1] superposed on a deterministic regressor. Such assumptions, however, restrict applicability to gray-box systems. Extending these methods to AI-driven systems — e.g., those governed by neural network (NN) controllers — remains elusive, as their complexity and opacity typically necessitate a black-box treatment. In such settings, the burden of constructing an uncertainty-aware model becomes prohibitive, rendering direct reasoning about desired behavioural properties essential.

While the previous chapter employed GPs to regress latent dynamics within an RKHS, recent advances in machine learning have extended RKHS-based techniques toward more expressive characterisations of stochastic systems, allowing to reason about their expected behaviour directly from data without explicit model construction. A particularly notable development is the theory of *conditional mean embeddings* (CMEs) (Song et al., 2009; Klebanov et al., 2020), which embed *conditional* distributions into RKHSs. This representation allows conditional expectations to be expressed as inner products, facilitating probabilistic reasoning while sidestepping explicit density estimation or numerical integration. Compared to existing approaches that rely on explicit parametrisations or structural assumptions, CMEs offer greater flexibility, accommodating systems where linearity, affine dynamics, or fixed stochasticity cannot be presumed. They thus form a natural and more expressive extension of GP-based methods, underpinned by a rigorous theoretical foundation.

However, reasoning solely with respect to the *expected* behaviour of a system can mask important aspects of the underlying stochastic transition process — motivating the development of *distributionally robust* frameworks (Elsayed-Aly et al., 2024). CMEs have demonstrated promise in data-driven chance-constrained control (Thorpe et al., 2022) and distributionally robust abstraction-based synthesis (Romao et al., 2023). While the former yields only approximate solutions without formal guarantees, the latter — as the abstraction-based approaches studied in preceding chapters — suffers from the curse of dimensionality. Notably, computing conditional expectations from data is also central to learning control barrier certificates (CBCs) for uncertain stochastic systems (Prajna et al., 2007). This motivates the application of CMEs to CBC synthesis in a model-free setting.

In contrast to related works relying on control-affine dynamics (Lopez and Slotine, 2023; Cohen et al., 2022a), known Lipschitz constants and exponential sample complexity (Salamati et al., 2024; Wang et al., 2023), or intermediate models based on less expressive GPs or RMDPs under partial model knowledge (Jagtap et al., 2020a; Wajid et al., 2022; Wang et al., 2018; Gracia et al., 2023; Chekan and Langbort, 2023), the approach proposed here learns CBCs *directly* from data. It leverages ambiguity sets over candidate transition kernels without requiring full stochastic model construction. This reformulation admits a distributionally robust interpretation, enabling probabilistic guarantees for the data-generating system via

[1] Gaussianity is a structural assumption crucial to many frameworks, as it often enables closed-form solutions or ensures computational tractability (Park et al., 2012). Developing methods that move beyond this assumption is significantly more challenging, as it exposes the inherent complexity of stochastic systems — no longer merely deterministic regressors masked by static Gaussian noise.

concentration inequalities over finite data sets (Li et al., 2022), and generalises beyond safety to ω-regular temporal logic specifications through automata-based representations and Streett supermartingales (Abate et al., 2024b).

Computationally, the resulting problem constitutes a non-convex semi-infinite program (SIP), a formulation also encountered in many *model-based* CBC approaches. Although barrier methods are typically less susceptible to the curse of dimensionality than abstraction-based techniques, SIPs are inherently challenging to solve monolithically. Here, various solution strategies are analysed and compared, both theoretically and empirically, revealing that many ostensibly abstraction-free approaches still rely on some form of discrete spatial approximation (see Section 5.7).

For the squared-exponential kernel and safety verification, a key innovation is introduced: a Fourier series expansion of the kernel yields an expressive yet tractable spectral abstraction of the dynamics alongside a novel *Fourier control barrier certificate*. Truncating the expansion reduces the CME-based constraint to a form efficiently evaluable via the fast Fourier transform (FFT) on equidistant sample lattices. Inspired by random feature approximations (Rahimi and Recht, 2007), this spectral discretisation generalises the spatial abstraction via binary-tree kernels introduced in Chapter 4, eliminating the need for dense spatial support vectors. This significantly improves tractability while retaining expressiveness. The resulting spectral formulation reduces the SIP to a finitely constrained linear program (LP), leveraging recent bounding results for trigonometric polynomials (Pfister and Bresler, 2018). The method relaxes restrictive assumptions of existing techniques and scales favourably across benchmarks with complex safety specifications and NN-based controllers.

Organisation The remainder of the chapter is structured as follows. After a brief review of additional related work, the definition of CBCs and the formal problem statement are provided in Section 5.3. The theory of kernel mean embeddings for (conditional) probability measures is introduced in Section 5.4. This foundation is used in Section 5.5 to derive data-driven CBC constraints without assuming an explicit system model. Generalisations to temporal logic objectives beyond safety are discussed in Section 5.6. With a focus on safety specifications, the resulting SIP is characterised and possible solutions are compared in Section 5.7. Section 5.8 introduces a tractable spectral solution approach based on Fourier expansions, which yields an LP relaxation of the data-driven verification problem. Numerical results on two benchmarks are presented in Section 5.9, followed by concluding remarks in Section 5.10.

5.2 Related Work

Most black-box approaches to formal verification focus on deterministic systems with known or fully queryable dynamics (Corso et al., 2021), often relying on reachability analysis or abstraction-based methods. In contrast, the approach presented here

enables direct safety verification of stochastic systems from data, without requiring the construction of an explicit model.

Adjacent work on abstraction-free temporal logic control often employs model-free reinforcement learning (Kazemi and Soudjani, 2020; Lavaei et al., 2020a), but typically lacks formal guarantees in the absence of known Lipschitz constants. Lew and Pavone (2021) compute reachable sets for stochastic systems using a sampling-based scheme, which yields only asymptotic guarantees. Others employ GPs to construct reachable sets via mixed monotonicity theory (Cao et al., 2022), but rely on prior knowledge of bounds for unmodelled dynamics. There is also a growing body of work exploring polynomial chaos expansions to propagate stochastic distributions through dynamics (Pan and Faulwasser, 2023; Kim et al., 2013; Seshadri et al., 2020), which is based on representing random variables as a function of a finite-dimensional random vector using an orthogonal polynomial basis.

NN-based methods for synthesising barrier certificates have gained traction due to their functional flexibility (So et al., 2024; Abate et al., 2021, 2024a). These approaches generally involve training neural CBCs on finite samples and verifying them using NN verification tools, such as bound propagation (Wang et al., 2024). However, convergence guarantees remain elusive. The broader field of NN verification is highly active, with numerous tools being developed to ensure reliability and safety in AI systems (Liu et al., 2021), including extensions to closed-loop settings (Harapanahalli et al., 2023b; Jafarpour et al., 2023). Nonetheless, correctness with respect to the underlying data-generating process remains unaddressed.

Closest in spirit to the present work are approaches based on interval probability metrics such as the maximum mean discrepancy (MMD) or the Wasserstein distance. A recent method by Chen et al. (2025) leverages Wasserstein distances for distributionally robust reach–avoid control. Both approaches aim to provide probabilistic guarantees under distributional uncertainty by bounding deviations between the empirical and true transition distributions. However, whereas our method employs MMD uncertainty sets and yields a tractable finite linear program, their formulation embeds the Wasserstein distance within a dynamic programming framework solved using sum-of-squares (SOS) optimisation, demonstrated on comparatively low-dimensional examples.

5.3 Preliminaries and Problem Statement

This chapter requires additional probabilistic concepts extending those introduced in Subsection 2.2.1.

Random Variables and Expectations Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space. For a random variable $X : \Omega \to \mathbb{X}$, define p_X as the pushforward measure of $\mathbb{P}$ under X, i.e., $p_X(A) := \mathbb{P}(X^{-1}(A))$ for all $A \in \mathscr{B}(\mathbb{X})$. The expected value of a measurable function $f : \mathbb{X} \to \mathbb{R}$ is denoted by $\mathbb{E}_{p_X}[f(X)]$, and abbreviated as $\mathbb{E}[f(X)]$ when clear from context. Given measurable spaces $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$ and $(\mathbb{Y}, \mathscr{B}(\mathbb{Y}))$, let X and Y be

random variables taking values in $\mathbb{X}$ and $\mathbb{Y}$, respectively. The conditional expectation of a measurable function $f\colon \mathbb{Y} \to \mathbb{R}$ given X is denoted by $\mathbb{E}_{\mathbf{p}}[f(Y)|X = \cdot\,]$, which is a function of $x \in \mathbb{X}$, where $\mathbf{p}\colon \mathbb{X} \times \mathscr{B}(\mathbb{Y}) \to [0,1]$ is a probability kernel. For a realisation $x \in \mathbb{X}$, $\mathbf{p}(\cdot\,|x)$ is a probability measure on $(\mathbb{Y}, \mathscr{B}(\mathbb{Y}))$, and $\mathbb{E}_{\mathbf{p}}[f(Y)|X = x]$ denotes the expected value of f under that measure.

5.3.1 System and Specification

System Definition This chapter considers black-box systems that can be represented as a Markov decision process (MDP) $\boldsymbol{\mathcal{M}} = (\mathbb{X}, \mathbb{X}_0, \mathbb{U}, \mathbf{t})$, derived directly from the generalised MDP definition (see Definition 2.4) by omitting the explicit output space. In every execution, given a current state $x \in \mathbb{X}$ and control input $u \in \mathbb{U}$, the system $\boldsymbol{\mathcal{M}}$ transitions to a successor state $x_+ \in \mathbb{X}$, drawn as a realisation $x_+ \sim \mathbf{t}(\cdot\,|x,u)$. A wide class of systems can be expressed in this form, encompassing all discrete-time stochastic dynamics with Markovian structure, namely all systems of the form

$$\boldsymbol{\mathcal{M}}\colon \begin{cases} x_{t+1} = f(x_t, u_t, w_t), & w_t \sim p_w \\ y_t \quad\;\; = x_t \end{cases}, \tag{5.1}$$

where the state evolution is described by a function $f\colon \mathbb{X} \times \mathbb{U} \times \mathbb{W} \to \mathbb{X}$, subject to *arbitrary* i.i.d. noise $w_t \sim p_w$ supported on a (possibly unbounded) set $\mathbb{W}$. This represents the most general system class considered in this thesis, extending beyond earlier assumptions of Gaussian and/or additive noise.

Safety Specifications

Safety Specifications The objective in this chapter is to synthesise a controller $\boldsymbol{\mathcal{C}}$ such that the closed-loop system $\boldsymbol{\mathcal{C}} \bowtie \boldsymbol{\mathcal{M}}$ satisfies a given safety specification. To this end, the goal is to design a controller $\boldsymbol{\mathcal{C}}\colon \mathbb{X} \to \mathbb{U}$ (see Subsection 2.2.2 for more details), as such policies are sufficient to achieve optimality in the context of MDPs. The closed-loop system $\boldsymbol{\mathcal{C}} \bowtie \boldsymbol{\mathcal{M}}$ is said to be safe if, for any initial state $x_0 \in \mathbb{X}_0$, the system does not enter an unsafe set of states $\mathbb{X}_u \subset \mathbb{X}$ in a given finite horizon $T \in \mathbb{N}$ (see Fig. 5.1). A (finite-horizon) *safety specification* $\psi_{\text{safe}} := (\mathbb{X}_u, T)$ is thus fully characterised by an unsafe set $\mathbb{X}_u$ and a time horizon T. The associated safety probability is abbreviated as $P_{\text{safe}} := \mathbb{P}(\boldsymbol{\mathcal{C}} \bowtie \boldsymbol{\mathcal{M}} \models \psi_{\text{safe}})$ whenever the reference system is clear from context.

Although the main focus of this chapter is on finite-horizon safety specifications, a natural extension to more general temporal properties is provided in Section 5.6.

5.3.2 Control Barrier Certificates

Control Barrier Certificates

Whilst certifying the safety of a continuous-space stochastic system is generally challenging, *control barrier certificates* (CBCs) and *control barrier functions* (CBFs) leverage the concept of set invariance to arrive at an abstraction-free formulation. This

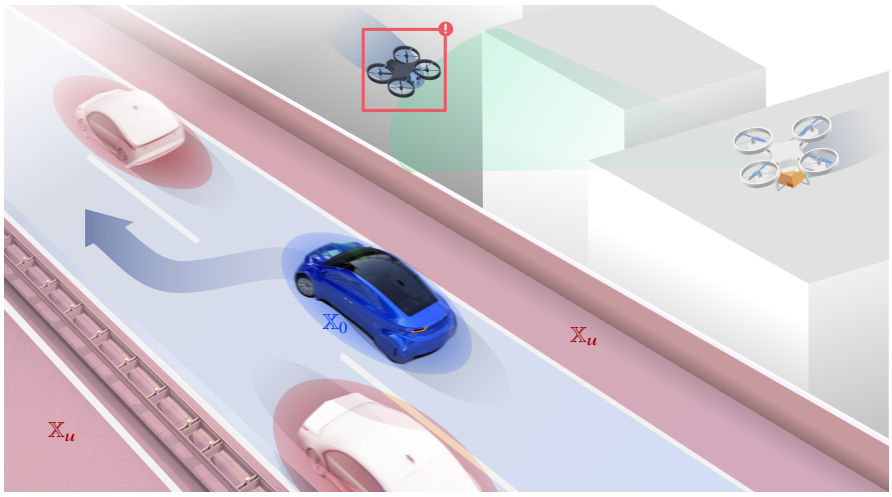


Fig. 5.1: Examples of safety-critical embodied AI systems in the transport sector. The initial set $\mathbb{X}_0$ (in blue) and the unsafe regions $\mathbb{X}_u$ (in red) for the ego vehicle are indicated schematically.

has made CBCs/CBFs popular tools for safety verification and synthesis (Prajna et al., 2007; Cosner et al., 2024). The CBC definition is recalled briefly as follows.

Definition 5.1 (Control Barrier Certificate (CBC)) A non-negative function $\mathcal{B}\colon \mathbb{X} \to \mathbb{R}_{\geq 0}$ is called a *CBC* of an MDP $\boldsymbol{\mathcal{M}} = (\mathbb{X}, \mathbb{X}_0, \mathbb{U}, \mathbf{t})$ with reference to an unsafe set $\mathbb{X}_u$, if the following conditions hold (cf. Fig. 5.2):

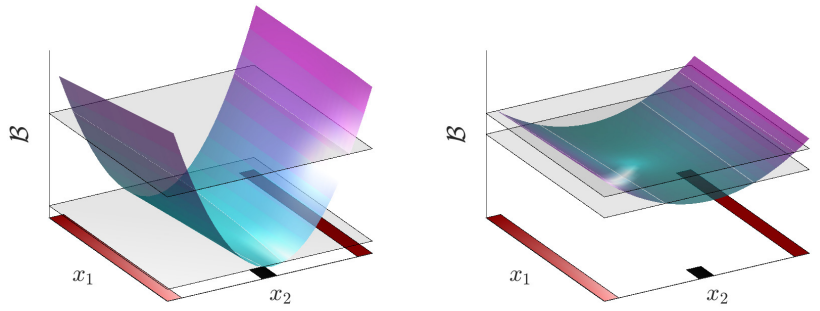
(a) $\forall x_0 \in \mathbb{X}_0\colon\ \mathcal{B}(x_0) \leq \eta$;
(b) $\forall x_u \in \mathbb{X}_u\colon\ \mathcal{B}(x_u) \geq \gamma$; and
(c) $\forall x \in \mathbb{X}, \exists u \in \mathbb{U}\colon\ \mathbb{E}_{\mathbf{t}}[\,\mathcal{B}(X_+)|X = x,\, U = u\,] - \mathcal{B}(x) \leq c$;

for some constants $\gamma > \eta \geq 0$ and $c \geq 0$.

Intuitively, condition (c) of Definition 5.1 restricts $\mathcal{B}$ to elicit a relaxed *supermartingale* property for any non-zero c. Note that this implies the existence of a stationary controller $\boldsymbol{\mathcal{C}}\colon \mathbb{X} \to \mathbb{U}$ generating the corresponding control inputs $u \in \mathbb{U}$.

If a CBC can be found for a system $\boldsymbol{\mathcal{M}}$, then, a lower bound on the probability of $\boldsymbol{\mathcal{M}}$ being safe is given by the following proposition due to Kushner (1967, Theorem 3).

Proposition 5.1 (Finite-Horizon Safety) *Consider an MDP $\boldsymbol{\mathcal{M}} = (\mathbb{X}, \mathbb{X}_0, \mathbb{U}, \mathbf{t})$ and a safety specification $\psi_{\text{safe}} = (\mathbb{X}_u, T)$. Suppose there exists a CBC $\mathcal{B}$ (Definition 5.1) with reference to $\mathbb{X}_u$ with constants η, γ, and c. Then, there exists a stationary controller $\boldsymbol{\mathcal{C}}$ such that*



(a) High level set separation $\gamma >> \eta$ (b) Low level set separation $\gamma > \eta$

Fig. 5.2: Conceptual illustration of CBCs $\mathcal{B}$ for an initial region $\mathbb{X}_0$ (in black), unsafe regions $\mathbb{X}_u$ (in red), and two different pairs of level sets γ, η (in gray).

$$\mathbb{P}(\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}} \models \psi_{\text{safe}}) \geq 1 - \frac{\eta + cT}{\gamma}.$$

See Fig. 5.2 for a conceptual visualisation of this result.

Remark 5.1 (Infinite-Horizon Safety) Finding a CBC (Definition 5.1) for a non-negative constant $c > 0$ is generally easier than for $c = 0$. If, however, a CBC can be obtained for $c = 0$, Proposition 5.1 provides safety guarantees for an unbounded time horizon $T \to \infty$.

5.3.3 Problem Statement

The aim is to synthesise a safety controller $\boldsymbol{C}$ for a black-box system $\boldsymbol{\mathcal{M}}$, where the transition kernel $\mathbf{t}$ is unknown and accessible only through empirical data. Specifically, a finite set of observed transitions of the form $\{x^i, u^i, x_+^i\}_{i=1}^N$, $N \in \mathbb{N}$, is available, with $x_+^i \sim \mathbf{t}(\cdot\,|x^i, u^i)$ and uniformly sampled states $x^i \sim \mathcal{U}_{\mathbb{X}}$ and inputs $u^i \sim \mathcal{U}_{\mathbb{U}}$. This corresponds to drawing i.i.d. samples from the joint distribution

$$(x, u, x_+) \sim \mathbf{t}(x_+|x, u)\, \mathcal{U}_{\mathbb{X}}(dx)\, \mathcal{U}_{\mathbb{U}}(du). \tag{5.2}$$

Note that this setting admits a plethora of target systems. The problem statement is formalised as follows.

Problem 5.1 (Black-Box Safety) Given a safety specification ψ_{safe} and a confidence level $1-\alpha \in (0, 1)$, and based only on i.i.d. observations $\{x^i, u^i, x_+^i\}_{i=1}^N$,

> $N \in \mathbb{N}$, determine a controller $\boldsymbol{C}$ and threshold $p_N \in (0,1)$ such that $\mathbb{P}(\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}} \models \psi_{\text{safe}}) \geq p_N$ with probability at least $1-\alpha$.

In this chapter, Problem 5.1 is addressed under the assumption of minimal system information (see Assumption 5.1), using CBCs to provide probabilistic safety guarantees. The main challenge arises from the chance constraint inherent to the CBC condition (c) (see Definition 5.1). It is shown how CBCs can be constructed from data by forming an empirical mean embedding of the latent transition kernel in a reproducing kernel Hilbert space (RKHS) via state-of-the-art kernel methods, thereby expressing the chance constraint as a linear inner product. An RKHS ambiguity set centred at the empirical mean embedding is constructed, that can be inflated to account for out-of-distribution dynamics, i.e., transitions not captured by the data. The resulting controller synthesis problem is cast as a semi-infinite program (SIP). For the special case in which a controller $\boldsymbol{C}$ is fixed — i.e., for system *verification* — an efficient algorithmic solution is presented. In particular, by employing a finite Fourier expansion of the kernel, the SIP is relaxed to a tractable linear program (LP).

In the following section, techniques for embedding (conditional) probability measures into RKHSs are provided.

5.4 Kernel Mean Embeddings

In the previous chapter, a single RKHS associated with a kernel k was sufficient to encode uncertain dynamics in the form of a GP, which implicitly assumes additive Gaussian noise. In contrast, the present chapter considers *conditional* probability distributions, without making parametric assumptions about the stochastic law. This shift introduces a key structural change: two separate kernels are required — one for the conditioning variable X and another for the target variable Y. These are denoted by $k_{\mathbb{X}} \colon \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ and $k_{\mathbb{Y}} \colon \mathbb{Y} \times \mathbb{Y} \to \mathbb{R}$, respectively. The use of conditional distributions naturally gives rise to operator-valued representations, which are formally constructed within the tensor product of the respective RKHSs.

Recall the brief summary on reproducing kernel Hilbert spaces (RKHSs) in Section 4.3.1. Throughout this chapter, all kernels are assumed to be bounded on their domains, e.g., $\mathbb{E}[k_{\mathbb{X}}(x,x)] < \infty$ for $x \in \mathbb{X}$, and all RKHSs are assumed to be *separable* (Schölkopf and Smola, 2002).

Tensor Product Hilbert Spaces For two kernels $k_{\mathbb{X}} \colon \mathbb{X}\times\mathbb{X} \to \mathbb{R}$ and $k_{\mathbb{Y}} \colon \mathbb{Y}\times\mathbb{Y} \to \mathbb{R}$ with associated RKHSs $\mathscr{H}_{\mathbb{X}}$ and $\mathscr{H}_{\mathbb{Y}}$, $\mathscr{H}_{\mathbb{X}\mathbb{Y}} := \mathscr{H}_{\mathbb{X}} \otimes \mathscr{H}_{\mathbb{Y}}$ is the *tensor product of the Hilbert spaces* $\mathscr{H}_{\mathbb{X}}$ *and* $\mathscr{H}_{\mathbb{Y}}$ with reproducing kernel $k_{\mathbb{X}\mathbb{Y}}((x,y),(x',y')) = k_{\mathbb{X}}(x,x')\, k_{\mathbb{Y}}(y,y')$, for $x,x' \in \mathbb{X}$ and $y,y' \in \mathbb{Y}$. $\mathscr{H}_{\mathbb{X}} \otimes \mathscr{H}_{\mathbb{Y}}$ is equipped with the inner product $\langle \varphi_{\mathbb{X}} \otimes \varphi_{\mathbb{Y}}, \varphi'_{\mathbb{X}} \otimes \varphi'_{\mathbb{Y}} \rangle_{\mathscr{H}_{\mathbb{X}} \otimes \mathscr{H}_{\mathbb{Y}}} = \langle \varphi_{\mathbb{X}}, \varphi'_{\mathbb{X}} \rangle_{\mathscr{H}_{\mathbb{X}}} \langle \varphi_{\mathbb{Y}}, \varphi'_{\mathbb{Y}} \rangle_{\mathscr{H}_{\mathbb{Y}}}$, for $\varphi_{\mathbb{X}}, \varphi'_{\mathbb{X}} \in \mathscr{H}_{\mathbb{X}}$ and $\varphi_{\mathbb{Y}}, \varphi'_{\mathbb{Y}} \in \mathscr{H}_{\mathbb{Y}}$, where its elements $(\varphi_{\mathbb{X}} \otimes \varphi_{\mathbb{Y}})$ are called *tensors*. The associated norm is given by $\|\phi_{\mathbb{X}} \otimes \phi_{\mathbb{Y}}\|_{\mathscr{H}_{\mathbb{X}} \otimes \mathscr{H}_{\mathbb{Y}}} = \|\phi_{\mathbb{X}}\|_{\mathscr{H}_{\mathbb{X}}} \|\phi_{\mathbb{Y}}\|_{\mathscr{H}_{\mathbb{Y}}}$ (Arai, 2018, Chapter 2). For feature vectors $\phi_{\mathbb{X}}(x)$ and $\phi_{\mathbb{Y}}(y)$, $\phi_{\mathbb{X}}(x) \otimes \phi_{\mathbb{Y}}(y)$ indicates their outer product.

5.4.1 Embedding Probability Measures

To reason about the expected value of a random variable, embedding the variable into a (higher dimensional) space is a well-established concept in machine learning (Schölkopf and Smola, 2002; Steinwart and Christmann, 2008). The *(kernel) mean embedding* (ME) follows the same reasoning and represents the projection of a probability measure into an RKHS (Smola et al., 2007; Muandet et al., 2017).

Mean Embedding

Definition 5.2 (Mean Embedding (ME)) Given an RKHS $\mathscr{H}_{k_\mathbb{X}}$ spanned by a kernel $k_\mathbb{X} \colon \mathbb{X} \times \mathbb{X} \to \mathbb{R}$, the *mean embedding* of a probability measure $p \colon \mathscr{B}(\mathbb{X}) \to [0,1]$ is computed via the *mean map* $\Psi_{k_\mathbb{X}} \colon \mathscr{P}(\mathbb{X}) \to \mathscr{H}_{k_\mathbb{X}}$ as

$$\Psi_{k_\mathbb{X}}(p) := \mathbb{E}_p[\phi_\mathbb{X}(X)] = \int_\mathbb{X} \phi_\mathbb{X}(x)\, dp(x).$$

Note that the reproducing property of the kernel carries on to the ME, facilitating the computation of the expected value of a function $f \in \mathscr{H}_{k_\mathbb{X}}$ via the inner product (Smola et al., 2007), that is

$$\mathbb{E}_p[f(X)] = \langle f, \Psi_{k_\mathbb{X}}(p) \rangle_{\mathscr{H}_{k_\mathbb{X}}}. \tag{5.3}$$

A kernel $k_\mathbb{X}$ is called *characteristic*, if the associated ME $\Psi_{k_\mathbb{X}}$ is injective, i.e., any probability measure $p \in \mathscr{P}(\mathbb{X})$ is injectively mapped to a unique ME $\Psi_{k_\mathbb{X}}(p) \in \mathscr{H}_{k_\mathbb{X}}$ (Gretton et al., 2012). This gives rise to the definition of a distance metric between embeddings in the RKHS — the *maximum mean discrepancy* (MMD). For a characteristic kernel $k_\mathbb{X}$, the MMD between two probability measures $p, p' \in \mathscr{P}(\mathbb{X})$ in $\mathscr{H}_{k_\mathbb{X}}$ is defined as $\left\| \Psi_{k_\mathbb{X}}(p) - \Psi_{k_\mathbb{X}}(p') \right\|_{\mathscr{H}_{k_\mathbb{X}}}$ (Gretton et al., 2012).

5.4.2 Embedding Probability Kernels

Analogous to generic probability measures, there exists a similar notion for the conditional case, that is, for probability kernels of the form $\mathbf{p} \colon \mathbb{X} \times \mathscr{B}(\mathbb{Y}) \to [0,1]$. More specifically, the measure-theoretic definition of a *conditional mean embedding* (CME), introduced by Park and Muandet (2020), is employed.

Conditional Mean Embedding

Definition 5.3 (Conditional Mean Embedding (CME)) Given two RKHSs $\mathscr{H}_{k_\mathbb{X}}$ and $\mathscr{H}_{k_\mathbb{Y}}$ with the associated kernels $k_\mathbb{X} \colon \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ and $k_\mathbb{Y} \colon \mathbb{X} \times \mathbb{X} \to \mathbb{R}$, the *CME* of a probability kernel $\mathbf{p} \colon \mathbb{X} \times \mathscr{B}(\mathbb{Y}) \to [0,1]$ is an X-measurable random variable taking values in $\mathscr{H}_{k_\mathbb{Y}}$, given by

$$\Psi_{k_\mathbb{Y}|k_\mathbb{X}}(\mathbf{p})(\,\cdot\,) := \mathbb{E}_\mathbf{p}[\phi_\mathbb{Y}(Y)|X = \,\cdot\,].$$

Refer to Park and Muandet (2020) for a comprehensive mathematical dissemination of the CME. Analogous to the non-conditional case, the CME can be used

to compute the conditional expectation of a function $f \in \mathcal{H}_{k_\mathbb{Y}}$ via its inner product, i.e., almost surely it holds that

$$\mathbb{E}_{\mathbf{p}}[f(Y)|X = x] = \langle f, \Psi_{k_\mathbb{Y}|k_\mathbb{X}}(\mathbf{p})(x)\rangle_{\mathcal{H}_{k_\mathbb{Y}}}. \tag{5.4}$$

Since the CME of a probability kernel $\mathbf{p}$ is often unknown (and potentially infinite dimensional), an empirical estimate may be obtained from a finite set of training data.

Empirical CME

Proposition 5.2 (Empirical CME) *Let a finite dataset* $(\hat{X}_N, \hat{Y}_N) := ([\hat{x}^i]_{i=1}^N, [\hat{y}^i]_{i=1}^N)$ *of samples* $(\hat{x}^i, \hat{y}^i) \in \mathbb{X}\times\mathbb{Y}$ *from a probability kernel* $\mathbf{p}\colon \mathbb{X}\times\mathcal{B}(\mathbb{Y}) \to [0,1]$ *be given such that* $\hat{y}^i \sim \mathbf{p}(\,\cdot\,|\hat{x}^i)$. *For kernels* $k_\mathbb{X}$, $k_\mathbb{Y}$, *the* empirical CME *of* $\mathbf{p}$ *given* $(\hat{X}_N, \hat{Y}_N)$ *is defined as*

$$\hat{\Psi}^N_{k_\mathbb{Y}|k_\mathbb{X}}(\,\cdot\,) := k_{\hat{X}}(\,\cdot\,)^\top \left[K^N_{\hat{X}} + N\lambda I_N\right]^{-1} \phi_\mathbb{Y}(\hat{Y}_N),$$

with a regularisation factor $\lambda \geq 0$. *For* $N \to \infty$, *the empirical CME converges in expectation, that is*

$$\mathbb{E}_{\mathcal{U}_\mathbb{X}\mathcal{U}_\mathbb{U}}\left[\left\|\hat{\Psi}^N_{k_\mathbb{Y}|k_\mathbb{X}}(X,U) - \Psi_{k_\mathbb{Y}|k_\mathbb{X}}(\mathbf{p})(X,U)\right\|_{\mathcal{H}_{k_\mathbb{Y}}}\right] \to 0. \tag{5.5}$$

Here and in the following, it is assumed that $[K^N_{\hat{X}} + N\lambda I_N]$ (and equivalent terms) are invertible. Note that this is always true if the regularisation factor λ is strictly positive. Given sufficient data, i.e., N sufficiently large, it holds by the virtue of the reproducing property for any function $f \in \mathcal{H}_{k_\mathbb{Y}}$ almost surely that

$$\mathbb{E}_{\mathbf{p}}[f(Y)|X = x] \approx \langle f, \hat{\Psi}^N_{k_\mathbb{Y}|k_\mathbb{X}}(x)\rangle_{\mathcal{H}_{k_\mathbb{Y}}} = k^N_{\hat{X}}(x)^\top \left[K^N_{\hat{X}} + N\lambda I_N\right]^{-1} f(\hat{Y}_N). \tag{5.6}$$

From (5.5) it follows that the empirical estimate improves in expectation as more data becomes available.

Remark 5.2 (Comparison) The empirical conditional expectation in (5.6) closely resembles the GP posterior mean in Equation 4.4; they coincide when $\sigma_w^2 = N\lambda$. An in-depth comparison between GPs and kernel methods is provided in Kanagawa et al. (2018).

Unlike GPs, which assume additive Gaussian noise, the CME enables conditional expectation estimates for arbitrary probability kernels. This flexibility is afforded by the use of characteristic kernels, which — analogous to characteristic functions in probability theory — encode all distributional information (in the form of their generalised moments) in the rich RKHS. As a result, the embedding becomes injective. In contrast, the BT kernel used in Chapter 4 intentionally introduced equivalence classes through piecewise-constant representations to yield a discrete-state model. This sacrificed expressiveness for tractability. A similar trade-off will be revisited

in Section 5.8, where a different strategy is introduced to derive representations that are both expressive and formally verifiable by design.

While CMEs are effective for estimating conditional means, recovering full conditional distributions remains challenging. Parametric methods based on *kernel moment matching* (Song et al., 2008) offer one route, but rely on a predefined family of distributions. Non-parametric alternatives (Schuster et al., 2020) exist, but were found to be numerically ill-conditioned, often failing to produce valid probability distributions.

5.5 Data-Driven Barrier Formulation

In this section, the supermartingale-like property (c) in Definition 5.1 is established for a system with unknown transition kernel $\mathbf{t}$, using its CME based on i.i.d. data

$$\mathcal{D}_N := (\hat{X}_N, \hat{U}_N, \hat{X}_N^+) := \left([\hat{x}^i]_{i=1}^N, [\hat{u}^i]_{i=1}^N, [\hat{x}_+^i]_{i=1}^N\right), \quad \hat{x}_+^i \sim \mathbf{t}(\cdot\,|\hat{x}^i, \hat{u}^i), \tag{5.7}$$

generated from the unknown black-box system as shown in (5.2). In the following, three kernels are used to represent each part of the data $\hat{X}_N$, $\hat{U}_N$, and $\hat{X}_N^+$:

$$k_x \colon \mathbb{X} \times \mathbb{X} \to \mathbb{R}, \quad k_u \colon \mathbb{U} \times \mathbb{U} \to \mathbb{R}, \quad k_+ \colon \mathbb{X} \times \mathbb{X} \to \mathbb{R}, \tag{5.8}$$

with associated RKHSs $\mathcal{H}_x$, $\mathcal{H}_u$, and $\mathcal{H}_+$, respectively. It is assumed that the output kernel k_+ is characteristic. Furthermore, the tensor product space $\mathcal{H}_{xu} := \mathcal{H}_x \otimes \mathcal{H}_u$ is defined, with the associated kernel $k_{xu} := k_x \cdot k_u$. Table 5.1 gives a brief summary of the notation used in the remainder of the chapter.

RKHS	Kernel	Feature Map
$\mathcal{H}_x$	$k_x \colon \mathbb{X} \times \mathbb{X} \to \mathbb{R}$	$\phi_x \colon \mathbb{X} \to \mathcal{H}_x$
$\mathcal{H}_u$	$k_u \colon \mathbb{U} \times \mathbb{U} \to \mathbb{R}$	$\phi_u \colon \mathbb{U} \to \mathcal{H}_u$
$\mathcal{H}_+$	$k_+ \colon \mathbb{X} \times \mathbb{X} \to \mathbb{R}$	$\phi_+ \colon \mathbb{X} \to \mathcal{H}_+$
$\mathcal{H}_{xu}$	$k_{xu} \colon (\mathbb{X} \times \mathbb{U}) \times (\mathbb{X} \times \mathbb{U}) \to \mathbb{R}$	not used
$\mathcal{G}$	$\Gamma \colon (\mathbb{X} \times \mathbb{U}) \times (\mathbb{X} \times \mathbb{U}) \to \mathcal{L}(\mathcal{H}_+)$	not used

Table 5.1: Overview of the kernels and RKHSs.

Hypothesis Space

To reason about the unknown dynamics in a mathematically rigorous way, the transition kernel $\mathbf{t}$ is assumed to admit a CME that lies in a known vector-valued RKHS $\mathcal{G}$ of mappings $\mathbb{X} \times \mathbb{U} \to \mathcal{H}_+$ (Park and Muandet, 2020, Section 2.3)[2]. In the following, the space $\mathcal{G}$ is referred to as the latent *hypothesis (function) space* of the system. Under standard assumptions, $\mathcal{G}$ is dense in $L_2(\mathbb{X} \times \mathbb{U}, \mathcal{B}(\mathbb{X} \times \mathbb{U}), \mathbf{t}; \mathcal{H}_+)$,

[2] For each fixed $(x, u) \in \mathbb{X} \times \mathbb{U}$, the vector-valued RKHS $\mathcal{G}$ reduces to the regular RKHS $\mathcal{H}_+$.

where $L_2(\mathbb{X}\times\mathbb{U}, \mathscr{B}(\mathbb{X}\times\mathbb{U}), \mathbf{t}; \mathscr{H}_+)$ is the space of strongly $(\mathscr{B}(\mathbb{X}\times\mathbb{U}) - \mathscr{B}(\mathscr{H}_+))$-measurable and Bochner square-integrable functions $\mathbb{X}\times\mathbb{U} \to \mathscr{H}_+$ with respect to $\mathbf{t}$, with $\mathscr{B}(\mathbb{X}\times\mathbb{U})$ and $\mathscr{B}(\mathscr{H}_+)$ denoting the Borel σ-fields of $\mathbb{X}\times\mathbb{U}$ and $\mathscr{H}_+$, respectively (Li et al., 2022).

This assumption imposes minimal structural constraints on the underlying dynamics. In fact, in contrast to many existing approaches based on finite Taylor approximations of the unknown dynamics — which are limited to approximating analytic functions of bounded degree — the RKHS framework adopted here allows approximation of a far broader class, e.g., of all smooth functions via the squared exponential kernel. This property affords significantly greater modelling flexibility and expressiveness. Moreover, the choice of kernel provides a natural mechanism for incorporating prior knowledge, where available. With this, the following standard assumption is raised.

Assumption 5.1 Let the CME $\Psi_{k_+|k_{xu}}(\mathbf{t})$ be *well-specified*, i.e., $\Psi_{k_+|k_{xu}}(\mathbf{t}) \in \mathscr{G}$. For a given confidence $1-\alpha \in (0,1)$, let a radius $\mathfrak{R} \geq 0$ be known s.t.

$$\mathbb{P}\left(\left\|\Psi_{k_+|k_{xu}}(\mathbf{t}) - \hat{\Psi}^N_{k_+|k_{xu}}\right\|_{\mathscr{G}} \leq \mathfrak{R}\right) \geq 1-\alpha,$$

where $\hat{\Psi}^N_{k_+|k_{xu}}$ is the empirical CME for the data $\mathcal{D}_N$ in (5.7), given by

$$\hat{\Psi}^N_{k_+|k_{xu}}(\,\cdot\,) := k^N_{\hat{X}\hat{U}}(\,\cdot\,)^\top \left[K^N_{\hat{X}\hat{U}} + N\lambda I_N\right]^{-1} \phi_+(\hat{X}^+_N). \tag{5.9}$$

It is common practice to introduce a robustness radius $\mathfrak{R}$ (as in Assumption 5.1) to account for uncertainty in empirical estimates such as the estimated conditional expectation in (5.6), by constructing an RKHS ambiguity set centred at the empirical CME. Although theoretical results exist that relate this radius to the probability of the true CME of the black-box dynamics $\mathbf{t}$ lying within the ambiguity set — achieving minimax rates of order $O(\log(N)/N)$ (Li et al., 2022; Mollenhauer et al., 2022) — such concentration bounds tend to be excessively conservative in practice. Consequently, the radius is often chosen manually or calibrated using data-driven techniques such as MMD-based bootstrapping (see, e.g., Nemmour et al., 2022, for an approach using non-conditional mean embeddings).

To learn CBCs from data, the conditional expectation in the left-hand side of condition (c) in Definition 5.1 is first reformulated as an inner product involving the CME $\Psi_{k_+|k_{xu}}(\mathbf{t})$, as shown in (5.4). That is, almost surely

$$\mathbb{E}_{\mathbf{t}}[\mathcal{B}(X_+)|X=x,\, U=u] = \langle \mathcal{B}, \Psi_{k_+|k_{xu}}(\mathbf{t})(x,u)\rangle_{\mathscr{H}_+}.$$

Ambiguity Set

Since the CME $\Psi_{k_+|k_{xu}}(\mathbf{t})$ is unknown, an ambiguity set $C^N_{\mathfrak{R}} \subset \mathscr{G}$ is constructed around the empirical CME $\hat{\Psi}^N_{k_+|k_{xu}}$, which is computed from data (5.7) using Proposition 5.2. The radius $\mathfrak{R} \geq 0$ is chosen such that the true CME lies within this set with confidence at least $1-\alpha \in (0,1)$, i.e.,

$$\mathbb{P}\left(\Psi_{k_+|k_{xu}}(\mathbf{t}) \in C_{\mathfrak{R}}^N\right) \geq 1-\alpha,$$

where the ambiguity set $C_{\mathfrak{R}}^N$ is defined as

$$C_{\mathfrak{R}}^N := \left\{\Psi \in \mathscr{G} \,\middle|\, \left\|\Psi - \hat{\Psi}_{k_+|k_{xu}}^N\right\|_{\mathscr{G}} \leq \mathfrak{R}\right\}. \tag{5.10}$$

This leads to the following result:

Theorem 5.1 *Let data $\mathcal{D}_N$ as in (5.7) from an unknown MDP $\boldsymbol{\mathcal{M}}$ and kernels k_{xu} and k_+ (characteristic) be given. Let $\hat{\Psi}_{k_+|k_{xu}}^N$ denote the empirical CME as defined in (5.9), and consider the ambiguity set $C_{\mathfrak{R}}^N$ in (5.10) constructed with confidence $1-\alpha$. If there exists a function $\mathcal{B}\colon \mathbb{X} \to \mathbb{R}_{\geq 0}$ with $\mathcal{B} \in \mathscr{H}_+$ s.t.*

$$\forall x \in \mathbb{X},\ \exists u \in \mathbb{U},\ \forall \Psi \in C_{\mathfrak{R}}^N\colon\ \langle \mathcal{B}, \Psi(x,u)\rangle_{\mathscr{H}_+} - \mathcal{B}(x) \leq c \tag{5.11}$$

for some constant $c \geq 0$, then $\mathcal{B}$ satisfies CBC condition (c) of Definition 5.1 with respect to $\boldsymbol{\mathcal{M}}$ with probability at least $1-\alpha$.

The following result provides a tractable way of establishing Theorem 5.1 by reformulating the left-hand side of (5.11) for the RKHS norm-ball ambiguity set in (5.10). This yields a data-driven barrier constraint involving the Gram matrices

$$K_{\hat{X}_+}^N := [k_+(\hat{x}_+^i, \hat{x}_+^j)]_{i,j=1}^N \quad \text{and} \quad K_{\hat{X}\hat{U}}^N := [k_{xu}((\hat{x}^i, \hat{u}^i), (\hat{x}^j, \hat{u}^j))]_{i,j=1}^N.$$

Theorem 5.2 *Consider the setup of Theorem 5.1. If there exists a function $\mathcal{B}\colon \mathbb{X} \to \mathbb{R}_{\geq 0}$, $\mathcal{B} \in \mathscr{H}_+$, with a bound $\overline{\mathcal{B}} \geq \|\mathcal{B}\|_{\mathscr{H}_+}$ such that*

$$\forall x \in \mathbb{X},\ \exists u \in \mathbb{U}:\ w(x,u)^\top \mathcal{B}(\hat{X}_N^+) - \mathcal{B}(x) \leq c - \mathfrak{R}\overline{\mathcal{B}}\kappa(x,u) \tag{5.12}$$

for some constant $c \geq 0$, with $\kappa(x,u) := \sqrt{k_x(x,x)}\sqrt{k_u(u,u)}$, and

$$w(x,u)^\top := k_{\hat{X}\hat{U}}^N(x,u)^\top \left[K_{\hat{X}\hat{U}}^N + N\lambda I_N\right]^{-1} \tag{5.13}$$

for some regularisation factor $\lambda \geq 0$, then $\mathcal{B}$ satisfies CBC condition (c) of Definition 5.1 with respect to $\boldsymbol{\mathcal{M}}$ with probability at least $1-\alpha$. In particular, this implies the existence of a controller $\boldsymbol{C}\colon \mathbb{X} \to \mathbb{U}$ satisfying (5.12).

Proof. The inner product in (5.11) can be rewritten as

$$\langle \mathcal{B}, \Psi(x,u)\rangle_{\mathscr{H}_+} = \langle \mathcal{B}, \Psi(x,u) - \hat{\Psi}_{k_+|k_{xu}}^N(x,u)\rangle_{\mathscr{H}_+} + \langle \mathcal{B}, \hat{\Psi}_{k_+|k_{xu}}^N(x,u)\rangle_{\mathscr{H}_+}.$$

Via (5.6), the latter term has the closed form

$$\langle \mathcal{B}, \hat{\Psi}^N_{k_+|k_{xu}}(x,u)\rangle_{\mathcal{H}_+} = w(x,u)^\top \mathcal{B}(\hat{X}^+_N).$$

To bound the first term, the operator-valued positive definite kernel $\Gamma\colon (\mathbb{X}\times\mathbb{U})\times(\mathbb{X}\times\mathbb{U}) \to \mathscr{L}(\mathcal{H}_+)$ associated with $\mathcal{G}$ (see, e.g., Li et al., 2022, Definition 1) is introduced:

$$\Gamma((x,u),(x',u')) := k_{xu}((x,u),(x',u'))\,\mathrm{Id}_{\mathcal{H}_+},$$

where $\mathrm{Id}_{\mathcal{H}_+}$ is the identity operator on $\mathcal{H}_+$ and $\mathscr{L}(\mathcal{H}_+)$ denotes the Banach space of bounded linear operators from $\mathcal{H}_+$ to $\mathcal{H}_+$.

By the reproducing property of Γ (see, e.g., Li et al., 2022, Equation (2)ff), the following bound can be established:

$$\begin{aligned}
\langle \mathcal{B}, \Psi(x,u) - \hat{\Psi}^N_{k_+|k_{xu}}(x,u)\rangle_{\mathcal{H}_+} &= \langle \Gamma(\,\cdot\,,(x,u))\mathcal{B}, \Psi - \hat{\Psi}^N_{k_+|k_{xu}}\rangle_{\mathcal{G}},\\
&\le \mathfrak{R}\,\|\Gamma(\,\cdot\,,(x,u))\mathcal{B}\|_{\mathcal{G}}\,,\\
&= \mathfrak{R}\sqrt{\langle \Gamma(\,\cdot\,,(x,u))\mathcal{B}, \Gamma(\,\cdot\,,(x,u))\mathcal{B}\rangle_{\mathcal{G}}},\\
&= \mathfrak{R}\sqrt{\langle \mathcal{B}, \Gamma((x,u),(x,u))\mathcal{B}\rangle_{\mathcal{H}_+}},\\
&\le \mathfrak{R}\overline{\mathcal{B}}\sqrt{\|\Gamma((x,u),(x,u))\|},\\
&= \mathfrak{R}\overline{\mathcal{B}}\sqrt{k_{xu}((x,u),(x,u))},
\end{aligned}$$

where the inequality in the penultimate line follows from the assumption $\overline{\mathcal{B}} \ge \|\mathcal{B}\|_{\mathcal{H}_+}$ and the submultiplicativity of the operator norm. Finally, noting that $k_{xu}((x,u),(x,u)) = k_x(x,x)k_u(u,u)$ and substituting both terms back into (5.11), the result follows. □

Intuitively, the conservatism introduced through the worst-case treatment with respect to the unknown transition kernel $\mathbf{t}$ is captured in the additional term $\mathfrak{R}\overline{\mathcal{B}}\kappa$, which is proportional to the radius $\mathfrak{R}$. Ensuring that the inequality in (5.12) holds for an adequate robustness margin $\mathfrak{R}\overline{\mathcal{B}}\kappa$ provides the necessary headroom to account for possible deviations in the true dynamics from those captured by the empirical CME. Note that the required margin is strongly dependent on the amount and representativeness of the available data, as captured through the radius $\mathfrak{R}$.

When designing a candidate CBC $\mathcal{B}$, an upper bound $\overline{\mathcal{B}}$ on the RKHS norm of $\mathcal{B}$ can be estimated either analytically or numerically; see, for example, Berlinet and Thomas-Agnan (2004); Kanagawa et al. (2018).

Based on Theorem 5.2, the following result establishes a quantified safety probability for a black-box system $\boldsymbol{\mathcal{M}}$.

Theorem 5.3 (Data-Driven Finite-Horizon Safety) *Consider the data-driven setup in Theorem 5.2 and a safety specification $\psi_{\text{safe}} = (\mathbb{X}_u, T)$. Suppose there exists a function $\mathcal{B}$ satisfying the conditions in Theorem 5.2 for some constant $c \ge 0$ and a controller $\boldsymbol{C}$. If there exist constants $\gamma > \eta \ge 0$ such that*

(a) $\forall x_0 \in \mathbb{X}_0\colon\ \mathcal{B}(x_0) \le \eta$; and

(b) $\forall x_u \in \mathbb{X}_u \colon \mathcal{B}(x_u) \geq \gamma$;

then, with probability at least $1 - \alpha$*, the following lower bound holds:*

$$\mathbb{P}(\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}} \models \psi_{\text{safe}}) \geq 1 - \frac{\eta + cT}{\gamma}.$$

Theorem 5.3 follows directly from Theorem 5.2 and Proposition 5.1.

5.6 Extension to General Classes of Temporal Logic Specifications

The results presented in the previous section can be extended to develop a data-driven approach for computing a lower bound on the satisfaction of more general specifications beyond safety. Examples include developing AI systems satisfying liveness, fairness, and complex temporal properties in scenarios involving cooperation or human interaction.

5.6.1 Linear Temporal Logic over Finite Traces

The work by Jagtap et al. (2020b) introduces a model-based procedure for using CBCs to compute a lower bound on the probability of satisfying *LTL specifications over finite traces* (LTL_F). The procedure comprises the following steps:

(i) The specification is negated;
(ii) The negated specification is translated into a DFA;
(iii) The resulting DFA is decomposed into a sequence of reachability tasks;
(iv) For each individual reachability task, an upper bound on its satisfaction probability is computed using CBCs; and
(v) The individual bounds are composed, based on the structure of the DFA, to obtain a lower bound on the probability of satisfying the original LTL_F specification.

This procedure relies primarily on knowledge of the specification and uses CBCs as a subroutine to establish bounds on the individual reachability probabilities. In scenarios where a model of the system is *not* available but an LTL_F specification is given, the identical procedure outlined above can be applied in conjunction with data-driven CBCs constructed using the results of Section 5.5. Further technical details are omitted here, as the overall approach remains unchanged from that of Jagtap et al. (2020b), with the only difference being the substitution of model-based CBCs by their data-driven counterparts developed in Section 5.5.

5.6.2 Specifications Modelled via Co-Büchi Automata

In an effort to address a broader class of specifications beyond LTL_F, the work by Anand et al. (2024) proposes a method for computing a lower bound on the probability of satisfying ω-regular specifications in networked stochastic systems. However, the approach is based on a fundamental simplification: the original specification is reduced to an automaton with a co-Büchi acceptance condition — that is, an acceptance criterion requiring certain states to be visited only finitely often. Consequently, the method is effectively restricted to a *strict subset* of ω-regular specifications.

The approach of Anand et al. (2024) comprises the following steps:

(i) The ω-regular specification is modelled as a DFA with a Rabin acceptance condition;
(ii) The specification is strengthened by replacing the Rabin acceptance condition with a co-Büchi acceptance condition;
(iii) A subset of satisfying traces with repeating cycles is extracted from the resulting automaton;
(iv) These traces are decomposed into sequences of safety specifications;
(v) CBCs are employed to compute individual lower bounds on the probability of satisfying these safety specifications; and
(vi) The individual bounds are combined to obtain a lower bound on the probability of satisfying the original specification.

As for the approach proposed by Jagtap et al. (2020a) reviewed previously, the above procedure relies primarily on knowledge of the specification and uses CBCs as a subroutine to obtain bounds on the individual safety probabilities. Thus, the data-driven results of Section 5.5 can be used to extend the procedure to a model-free setting. The underlying steps remain unchanged, and so the mathematical details are omitted.

Given the replacement of the original ω-regular specification with a strengthened version based on a co-Büchi condition, it should be noted that the resulting lower bound is likely to be overly conservative, if not trivial.

5.6.3 ω-Regular Specifications

To address the general class of ω-regular specifications, this section builds on the concept of *Streett supermartingales*, proposed by Abate et al. (2024b), and provides a data-driven version that can be established without model knowledge and based only on data from the latent system. For this, let a *Streett pair* (F, I), i.e., a pair of sets $F, I \in \mathscr{B}(\mathbb{X})$, be given. The following result is an extension of Theorem 3 by Abate et al. (2024b) to non-autonomous stochastic systems.

Street
Supermartingales

Proposition 5.3 (Streett Supermartingale) *For an MDP $\mathcal{M} = (\mathbb{X}, \mathbb{X}_0, \mathbb{U}, \mathbf{t})$ and a Streett pair (F, I), suppose there exist two functions $V\colon \mathbb{X} \to \mathbb{R}$ and $\mathcal{B}\colon \mathbb{X} \to \mathbb{R}_{\geq 0}$, and positive constants $\chi, \nu > 0$, such that for all $x \in \mathbb{X}$ there exists a control input $u \in \mathbb{U}$ satisfying the following conditions:*

(a) $V(x_+) \leq V(x)$ *almost surely, where* $x_+ \sim \mathbf{t}(\cdot\,|x, u)$ *if* $x \in W$;
(b) $\mathbb{E}_{\mathbf{t}}[\,\mathcal{B}(X_+)|X = x,\, U = u\,] - \mathcal{B}(x) \leq -\chi$ *if* $x \in (F \backslash I) \cap W$;
(c) $\mathbb{E}_{\mathbf{t}}[\,\mathcal{B}(X_+)|X = x,\, U = u\,] - \mathcal{B}(x) \leq \nu$ *if* $x \in I \cap W$; *and*
(d) $\mathbb{E}_{\mathbf{t}}[\,\mathcal{B}(X_+)|X = x,\, U = u\,] - \mathcal{B}(x) \leq 0$ *if* $x \in W \backslash (F \cup I)$;

where $W = \{x \in \mathbb{X} | V(x) \leq 1\}$ *is a sublevel set of* V. *Then,* $\mathcal{B}$ *is called a* Streett supermartingale *and there exists a controller* $\boldsymbol{C}\colon \mathbb{X} \to \mathbb{U}$ *under which the MDP, starting from any initial state* $x_0 \in W$, *almost surely either visits* F *a* finite *number of times or visits* I infinitely *often.*

From the four conditions in Proposition 5.3, the last three involve the computation of conditional expectations. These conditions can be replaced with their data-driven version using CMEs. The first condition, however, requires the inequality to hold almost surely. To satisfy the first condition, the following assumption is raised.

Assumption 5.2 (Known Support) For the MDP $\mathcal{M} = (\mathbb{X}, \mathbb{X}_0, \mathbb{U}, \mathbf{t})$, it is assumed that the support of the transition kernel $\mathbf{t}$ is known. That is, a set-valued function $S(x, u)$ is given such that $\mathbf{t}(S(x, u)|x, u) = 1$ for all $x \in \mathbb{X}$ and $u \in \mathbb{U}$.

Remark 5.3 When the system is directly represented in the form of (5.1), Assumption 5.2 can be satisfied by requiring the deterministic component f of the dynamics and the support $\mathbb{W}$ of the noise to be known, while allowing the measure $p_w \in \mathcal{P}(\mathbb{W})$ to remain unknown. In this case, the support can be expressed as $S(x, u) = \{f(x, u, w)|w \in \mathbb{W}\}$. Although this constitutes a restrictive assumption, it is noted that in many practical scenarios a deterministic approximation of the model and a reasonable estimate of the support set for stationary noise can be constructed.

The next result presents the conditions for the CME-based computation of Streett supermartingales under Assumption 5.2.

Theorem 5.4 (Data-Driven Street Supermartingales) *Let data $\mathcal{D}_N$ as defined in (5.7), collected from an unknown MDP $\mathcal{M}$ under Assumption 5.2, and kernels k_{xu} and k_+ (characteristic) be given. Consider the resulting empirical CME $\hat{\Psi}^N_{k_+|k_{xu}}$ in (5.9) and the ambiguity set $C^N_{\mathfrak{R}}$ defined in (5.10) with confidence bound $1 - \alpha$. For a given Streett pair (F, I), suppose that there exist two functions $V\colon \mathbb{X} \to \mathbb{R}$ and $\mathcal{B}\colon \mathbb{X} \to \mathbb{R}_{\geq 0}$, where $\mathcal{B} \in \mathcal{H}_+$, as well as constants $\chi, \nu > 0$, such that for all $x \in \mathbb{X}$, a control input $u \in \mathbb{U}$ exists, satisfying*

(a) $V(x_+) \leq V(x)$, *for all* $x_+ \in S(x, u)$ *if* $x \in I$;
(b) $w(x, u)^\top \mathcal{B}(\hat{X}^+_N) - \mathcal{B}(x) \leq -\chi - \mathfrak{R}\overline{\mathcal{B}}\kappa(x, u)$ *if* $x \in (F \backslash I) \cap W$;
(c) $w(x, u)^\top \mathcal{B}(\hat{X}^+_N) - \mathcal{B}(x) \leq \nu - \mathfrak{R}\overline{\mathcal{B}}\kappa(x, u)$ *if* $x \in I \cap W$; *and*

(d) $w(x,u)^\top \mathcal{B}(\hat{X}_N^+) - \mathcal{B}(x) \leq -\Re\overline{\mathcal{B}}\kappa(x,u)$ *if* $x \in W \backslash (F \cup I)$*;*

where $W = \{x \in \mathbb{X} | V(x) \leq 1\}$, $\overline{\mathcal{B}} \geq \|\mathcal{B}\|_{\mathscr{H}_+}$, $\kappa(x,u) := \sqrt{k_x(x,x)}\sqrt{k_u(u,u)}$, *and* $w(x,u)^\top$ *is defined in* (5.13)*. This implies the existence of a controller* $\boldsymbol{C}: \mathbb{X} \to \mathbb{U}$ *that satisfies the above conditions. Then, with probability at least* $1-\alpha$*, the function* $\mathcal{B}$ *constitutes a* Streett supermartingale*, and the system* $\boldsymbol{\mathcal{M}}$*, starting from any initial state* $x_0 \in W$ *and operating under the controller* $\boldsymbol{C}$*, almost surely either visits* F *only* finitely *many times or visits* I infinitely *often.*

Any ω-regular specification can be represented by a DFA equipped with a *Streett acceptance condition* (Streett, 1981; Baier and Katoen, 2008). A Streett acceptance condition is defined by a finite collection of pairs $\{(F_i, I_i)\}_{i=1}^p$, where a trajectory is accepted if for every $i \in \{1, \ldots, p\}$ either the set F_i is visited only *finitely* often, or the set I_i is visited *infinitely* often. Thus, to verify almost-sure satisfaction of an ω-regular specification, Streett supermartingales may be constructed for each pair (F_i, I_i), with each supermartingale yielding an associated *winning domain* W_i. A lower bound on the satisfaction probability can then be obtained by solving a reachability problem to the intersection of the winning domains $\bigcap_{i=1}^p W_i$.

Theorem 5.5 (Data-Driven ω-Regular Satisfaction) *Consider the data-driven setup of Theorem 5.4 and a Streett acceptance condition given by* $\{(F_i, I_i)\}_{i=1}^p$*. Suppose that for a common controller* $\boldsymbol{C}: \mathbb{X} \to \mathbb{U}$ *each pair* (F_i, I_i) *admits a Street supermartingale* $\mathcal{B}_i$ *according to Theorem 5.4 with confidence at least* $1-\alpha$ *and with associated winning domain* W_i*. Then, the trajectories of* $\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}}$ *satisfy the Streett acceptance condition almost surely from any initial state* $x_0 \in \bigcap_{i=1}^p W_i$*.*

5.7 Problem Characterisation and Comparison of Possible Solutions

To enable a focused investigation of possible computational solutions, the remainder of this chapter is focused on *safety verification*. Thus, the controller $\boldsymbol{C}$ is assumed to be given, and the controlled system is denoted using the same notation after eliminating the input from the mathematical expressions.

First, the optimisation problem identified for computing data-driven CBCs is summarised and characterised:

Data-Driven Barrier SIP

$$\min_{\mathcal{B},c,\eta,\gamma} \quad \frac{\eta + cT}{\gamma}, \tag{5.14a}$$

$$\text{subject to} \quad \forall x_0 \in \mathbb{X}_0 \colon \ \mathcal{B}(x_0) \leq \eta, \tag{5.14b}$$

$$\forall x_u \in \mathbb{X}_u \colon \ \mathcal{B}(x_u) \geq \gamma, \tag{5.14c}$$

$$\forall x \in \mathbb{X} \colon \ w(x)^\top \mathcal{B}(\hat{X}_N^+) - \mathcal{B}(x) \leq c - \mathfrak{R}\overline{\mathcal{B}}\overline{\kappa}, \tag{5.14d}$$

$$\forall x \in \mathbb{X} \colon \ \mathcal{B}(x) \geq 0, \tag{5.14e}$$

$$c \geq 0,\ \gamma > \eta \geq 0,\ \mathcal{B} \in \mathscr{H}_+,$$

where $\overline{\kappa} \geq \sup_{x \in \mathbb{X}} \sqrt{k_x(x,x)}$, complexity bound $\overline{\mathcal{B}} \geq \|\mathcal{B}\|_{\mathscr{H}_+}$, and

$$w(x)^\top := k_{\hat{X}}^N(x)^\top \left[K_{\hat{X}}^N + N\lambda I_N \right]^{-1}.$$

For translation invariant kernels, the bound $\overline{\kappa}$ is exact. The constraint set of (5.14) is infinite, rendering it a semi-infinite program (SIP).

Every candidate barrier function $\mathcal{B} \in \mathscr{H}_+$ admits a representation as a weighted combination of kernel evaluations, that is

$$\mathcal{B}(x) := \sum_i b_i k_+(\bar{x}_i, x), \tag{5.15}$$

where the spatial support vectors $\bar{x}_i \in \mathbb{X}$ and coefficients $b_i \in \mathbb{R}$ are to be optimised. For commonly used kernels such as the Gaussian, squared exponential, or Matérn kernel, the SIP (5.14) is generally non-convex. Possible (partial) dual reformulations of this problem are discussed in Subsection 5.7.1. Broadly, (5.14) falls under the class of *non-convex semi-infinite optimisation problems*. Several relevant techniques have been proposed in the literature on robust and semi-infinite programming; see, e.g., (Djelassi et al., 2021). The following subsections present an overview of the frameworks that were tested as part of this thesis, comparing them both in terms of theoretical complexity and empirical performance. All methods discussed here take the SIP (5.14) as their starting point and rely on Assumption 5.1. Importantly, this assumption can be set arbitrarily general and does not impose any strong structural assumptions such as linear, affine, or polynomial dynamics that could be exploited by special solvers.

5.7.1 Monolithic Solutions

CBCs have been developed to reduce the computational burden associated with formal methods. In particular, they have been developed as a supposedly “abstraction-free" alternative to address the so-called *curse of dimensionality*, that arises from approaches based on the spatial partitioning of the system domain — aptly dubbed “abstraction-based". Over the years, a list of customary assumptions has been es-

tablished in the literature to obtain feasible formulations of the barrier problem that can be efficiently solved monolithically, i.e., without spatial or spectral subdivision. In the model-based case, the most typical assumptions include control-affine polynomial dynamics and polynomial barriers — both of known maximum degree. Recall that the RKHS framework adopted here allows the approximation of non-analytic smooth functions far beyond polynomials (see Assumption 5.1). In the model-free/data-driven setting, additional Lipschitz bounds on the dynamics and intermediate functions (Salamati et al., 2024) are inevitable, knowledge assumed given even though the dynamics are deemed unknown. Evidently, it is unclear how these assumptions can be rigorously justified.

5.7.1.1 Sum of Squares

The *sum-of-squares* (SOS) approach has gained notable popularity for establishing barriers in polynomial system models (Prajna et al., 2005), with Schön et al. (2024) being the first to cast *data-driven* barrier problems into SOS form. This method relies on several assumptions: polynomial barrier functions, sets $\mathbb{X}, \mathbb{X}_0, \mathbb{X}_u \subset \mathbb{R}^{n_x}$ given in semi-algebraic form, and — in the data-driven case — dynamics that are polynomially dependent on the state x (i.e., $\mathscr{H}_x$ limited to the RKHS of the polynomial kernel). Furthermore, the computational complexity hinges directly on the maximum polynomial degree appearing in the resulting program. There exist extensions to bases beyond polynomials, known as *kernel SOS* (KSOS) (Marteau-Ferey et al., 2020). For instance, Bagnell and Farahmand (2015) restrict a kernel machine to functions of SOS form from the product RKHS $\mathscr{H}' := \mathscr{H} \otimes \mathscr{H}$, associating the kernel

$$k'(x, x') := \langle \phi(x) \otimes \phi(x), \phi(x') \otimes \phi(x') \rangle_{\mathscr{H}'} = \langle \phi(x), \phi(x') \rangle_{\mathscr{H}}^2.$$

Further examples of similar methods used for *hard-constraining* kernel machines are provided in Subsection 5.7.2. Note, that in the multivariate case $n_x > 1$, SOS functions only constitute a subset of all non-negative functions, thus introducing conservatism (Parrilo, 2003). Nevertheless, the SOS approach is very popular for obtaining well-behaved/convex relaxations of complex programs such as the SIP in (5.14), as it avoids domain discretisation and nonlinear optimisation.

Although nomenclature-wise abstraction-free, the SOS approach is not immune to the curse of dimensionality, as it covertly solves a semi-definite program (SDP), for which the number of variables scales exponentially with the system dimension n_x and the maximum polynomial degree d. The resulting complexity of the SOS is polynomial in the number of variables. Additionally, solving barrier problems such as the SIP (5.14) usually requires the sets $\mathbb{X}$, $\mathbb{X}_0$, $\mathbb{X}_u$, constants η, γ, $c - \Re\overline{\mathcal{B}}\overline{\kappa}$, and additional Lagrange multipliers to be expressed in the chosen kernel basis. This can be extremely cumbersome — especially in an infinite basis — and increase the size of the SDP substantially. Sampling-based extensions exist that can bypass this additional overhead and even exploit sparsity under the assumption of poisedness (Lofberg and Parrilo, 2004; Cifuentes and Parrilo, 2017). The interested reader is

referred to the paper by Parrilo (2003) for a more detailed dissemination of the SOS approach, as well as Jagtap et al. (2020b, Section V.C.1) and Schön et al. (2024) for its application to barrier certificates.

5.7.1.2 Dual Reformulation

In an effort to reformulate the SIP in (5.14) under minimal structural assumptions into a form that can be efficiently solved monolithically, *dual reformulation* of the constraints may be considered. For example, given a polyhedral unsafe set $\mathbb{X}_u := \{x_u \in \mathbb{X} \mid H_u x_u \leq h_u\}$ with matrix H_u and vector h_u of appropriate dimensions, the dual to the constraint (5.14c) is given by

$$\min_{\lambda_u \leq 0} \mathcal{B}^*(H_u^\top \lambda_u) - \lambda_u^\top h_u \leq -\gamma, \tag{5.16}$$

where $\mathcal{B}^*$ is the Fenchel conjugate[3] of $\mathcal{B}$. Closed-form solutions of (5.16) can be obtained for specific choices of the kernel, as shown in the following example.

Example 5.1 (Exponential Kernel) *Consider $\mathcal{B} \in \mathscr{H}$, with $\mathscr{H}$ the RKHS of the exponential kernel $k(x, x') := \alpha e^{\beta x^\top x'}$, where $\alpha, \beta > 0$. Then, $\mathcal{B}$ can be written in the form $\mathcal{B}(x) = \alpha \sum_{i=1}^{n_x} a_i e^{\beta e_i^\top x}$, where $\{e_i\}_{i=1}^{n_x} \subset \mathbb{X}$ is an orthogonal basis of $\mathbb{X}$. Furthermore, the Fenchel conjugate $\mathcal{B}^*$ is given by*

$$\mathcal{B}^*(\lambda) = \frac{1}{\beta} \sum_{i=1}^{n_x} \frac{1}{\|e_i\|} e_i^\top H_u^\top \lambda.$$

*The expression can be simplified for an ortho*normal *basis $\{e_i\}_{i=1}^{n_x}$, where $\|e_i\| = 1$ for all $i \in \{1, \ldots, n_x\}$. The derivation is omitted for brevity.*

Even though this shows that closed-form convex reformulations of individual constraints can be obtained for specific kernel choices, it is *not* possible to convexify the *entire* SIP (5.14) in this manner. In fact, non-linear convexity of both constraints (5.14b) *and* (5.14c) is contradictory. Furthermore, the Kushner constraint (5.14d) is generally neither convex nor concave. Dual reformulation will be revisited briefly in the context of local optimisations used in branch-and-bound optimisers, described in the next subsection.

5.7.2 Divide-and-Conquer Solutions

The main take-away from the previous subsection is that solving the non-convex SIP (5.14) in a monolithic fashion is only feasible under carefully chosen assumptions.

[3] The Fenchel conjugate of a function $f : \mathbb{X} \to \mathbb{R}$ is defined as $f^*(y) = \sup_{x \in \mathbb{X}} \left(y^T x - f(x)\right)$ (Boyd and Vandenberghe, 2004).

But even then, the solution approach suffers — similarly, albeit less strongly than abstraction-based approaches — from the curse of dimensionality. For those that strive for maximal flexibility and minimal assumptions, this subsection discusses approaches that (at least partially) leverage a spatial subdivision of the domain of the system/barrier. Although bearing similarities to traditional abstraction-based approaches, these approaches leverage the barrier as a mathematical vehicle for storing information and guiding the spatial abstraction process. In comparison, abstraction-based approaches often suffer from significant memory requirements and adaptive abstraction techniques rely on ad-hoc principles such as the approximation of local Lipschitz constants (Zhang et al., 2024; Kazemi et al., 2024a).

One-Stage vs Two-Stage Solutions Approaches to solving the SIP (5.14) can be applied in both one-stage and two-stage frameworks. One-stage frameworks find valid barriers directly, while two-stage methods involve (i) *candidate barrier identification* and (ii) *candidate barrier verification*. For instance, the *counterexample-guided inductive synthesis* (CEGIS) approach is a typical two-stage approach, that iterates between learning candidate barriers and verifying them using a counterexample approach (Edwards et al., 2024). Typically, verifying a candidate barrier is less complex than generating it, rendering two-stage methods generally more computationally efficient. Section 5.8 presents an efficient sampling-based algorithm for generating barriers that are valid *by design*, thus avoiding the computational challenges associated with CEGIS. For completeness, various verification approaches are discussed and compared in the following subsubsections.

5.7.2.1 SAT/SMT Solvers

Satisfiability (SAT) solvers and *satisfiability modulo theories* (SMT) solvers are useful algorithmic tools for determining if a given set of equations can be satisfied. Thus, SMT solvers such as Z3 (De Moura and Bjørner, 2008) and dReal (Gao et al., 2013) have found widespread adoption in applications requiring formal correctness guarantees. For example, Huang et al. (2017) use Z3 to verify the safety of neural networks. Similarly, the recent barrier tool by Edwards et al. (2024) relies on SMT solvers to verify the soundness of generated candidate barriers. Although the aforementioned solvers can handle (some) nonlinear expressions, their numerical complexity is prohibitive. Whilst the exact complexity of solving an SIP such as (5.14) via SMT solvers is problem dependent, the general complexity of δ-SMT programs such as the ones solved by dReal is NP-complete (Gao et al., 2012). The involved δ-decision procedures construct a symbolic representation of the problem by replacing nonlinear expressions with symbolic variables. The constraints are subsequently evaluated over the continuous domain by applying interval arithmetic. The resulting number of paths to be checked scales exponentially in the number of symbolic variables; a phenomenon known as *path explosion*.

5.7.2.2 Global Optimisers

Another tool for solving the SIP (5.14) are global optimisers with deterministic guarantees for finding the global optimum (within tolerance) in finite time (Horst and Tuy, 2013). Global optimisers themselves can be subcategorised depending on their operating principle. For instance, *branch-and-bound* (B&B) optimisers such as BARON (Sahinidis, 1996), SCIP (Bestuzheva et al., 2021), and MAiNGO (Bongartz et al., 2018), solve the problem by iteratively subdividing the domain and performing local convex relaxations to obtain upper and lower bounds on possible function values. Similar to SMT solvers, they decompose the nonlinear program by assigning symbolic variables to nonlinear expressions. For kernel methods such as GPs, the number of nonlinear terms scales quadratically in the number of training samples. See Schweidtmann et al. (2021) for a dissemination of global optimisation of GPs based on branch-and-bound algorithms. For the specific case of neural barrier verification, interval analysis tools such as immrax (Harapanahalli et al., 2024) can be used. In contrast to SMT solvers, global optimisers can provide information beyond logical feasibility and determine the actual (approximate) value of the global optimum. However, SMT solvers typically leverage combinatorial techniques such as conflict-driven clause learning (CDCL) and are generally not less well equipped to solving SIPs over continuous domains than global optimisers leveraging convex relaxations. Notably, the tool MAiNGO provides dedicated relaxations for common kernel functions. The worst-case complexity of any B&B algorithm is exponential. However, the average complexity of the algorithm can be significantly lower on a case-by-case basis. In our experiments using the squared exponential kernel, neither the SMT solver dReal nor any of the previously mentioned B&B optimisers was able to produce results in a reasonable time frame for a minimal example of the SIP (5.14) involving all constraints.

5.7.2.3 Compact Coverings

Most methods enforcing *hard shape constraints* — such as positivity — on kernel-based function approximations rely on compact coverings of the domain. For example, the covering-based strategy proposed by Aubin-Frankowski and Szabó (2020) introduces inequality constraints on function derivatives by sampling a finite number of equidistant points. These methods exploit regularity properties of the function space, similar to global optimisation techniques. However, unlike iterative B&B schemes, a compact covering represents the limit case where the entire domain is partitioned uniformly without discarding regions where constraints are already satisfied (e.g., through convex relaxations). As a result, their computational complexity becomes an inferior $O(r^{n_x})$, where r denotes the resolution of the discretisation and n_x the dimensionality of the state space (Lasserre, 2001). To provide some intuition, in the case of the squared exponential kernel (4.1), the required resolution r for a given approximation accuracy is highly dependent on the kernel amplitude σ_f^2.

5.7.2.4 Other Approaches

Several additional methods fall outside the previously discussed categories or are currently impractical. For instance, the non-convex scenario optimisation approach introduced by Garatti and Campi (2024) reformulates the solution of a SIP as the identification of the least complex finite subproblem — i.e., one involving the fewest active constraints. However, the identification of such a subproblem is generally obviously difficult. By contrast, B&B methods identify a relaxed problem via a systematic iterative procedure.

Next, the proposed solution to the SIP (5.14) is introduced. In contrast to previous methods that operate directly on the spatial domain, a kernel inducing a Fourier basis is employed to derive a spectral representation of the problem. Specifically, a finite Fourier feature expansion of the squared exponential kernel is used, serving as a spectral abstraction of the system. Building on the mathematical foundation of kernel methods, correct-by-design barriers are constructed through an explicit mathematical characterisation of both the candidate barrier space and the hypothesis space $\mathscr{G}$ of the system.

5.8 Fourier Barrier Solution

To address the SIP in (5.14), a central observation is made: although modelling nonlinear state evolutions, at its core, a CME corresponds to a *linear* operator in RKHS. However, as characteristic RKHSs are typically infinite-dimensional, performing explicit computations within RKHS is generally not possible. To obtain a tractable approximation, a Fourier series expansion of the characteristic squared exponential kernel is employed. This yields a finite-dimensional spectral approximation of the associated RKHS, which remains expressive enough to capture both the system dynamics and candidate barrier functions.

In contrast to compact covering methods (see Subsubsection 5.7.2.3), where the number of spatial support vectors scales exponentially with the dimensionality of the system, the spectral density of the squared exponential kernel is concentrated near the origin, allowing for a compact and efficient basis requiring comparatively few spectral support vectors. The resulting formulation is a semi-infinite linear program (SILP), which can be solved efficiently as a standard linear program (LP) using trigonometric bounding techniques (Pfister and Bresler, 2018).

Truncating the Fourier expansion introduces a spectral abstraction, yielding equivalence classes of functions analogous to the spatial abstraction introduced by the BTGP in Subsection 4.4.2. Notably, unlike before, the ambiguity arises here in the frequency domain, introducing uncertainty over the frequency response of the system. While this chapter focuses on establishing the overall framework, quantifying the induced abstraction error and identifying principled strategies for selecting an optimal spectral basis remain open challenges for future work.

5.8.1 Finite Fourier Feature Expansion

It is well known that for translation-invariant kernels, Bochner's theorem provides an infinite Fourier series expansion characterising the kernel in terms of a kernel-specific *spectral measure* (Rahimi and Recht, 2007). In this chapter, focus is placed on the versatile squared exponential kernel (4.1), which admits a Fourier expansion via the Gaussian measure as follows:

$$k(x, x') \equiv \sigma_f^2 \int_{\mathbb{R}^{n_x}} \mathcal{N}(d\omega|0, \Sigma)\, e^{\mathrm{i}\omega^\top (P(x) - P(x'))}, \quad \Sigma := \operatorname{diag}(\sigma_l)^{-2}, \tag{5.17}$$

with the imaginary unit $\mathrm{i} := \sqrt{-1}$, and where the projection $x \mapsto P(x)$ maps the domain $\mathbb{X}$ into the unit hypercube $[0, 1]^{n_x}$. In what follows, the kernel k in (5.17) — with its associated infinite-dimensional RKHS $\mathscr{H}$ — is used to replace both the conditioning kernel k_x and the output kernel k_+. Thus, the corresponding subscripts are dropped for simplicity.

Recall that for the CME $\Psi(\mathbf{t})(x) \in \mathscr{H}$ for all $x \in \mathbb{X}$. To obtain a tractable formulation, the barrier function is chosen deliberately from a smaller, finite-dimensional function space $\mathscr{H}_M$ approximating[4] $\mathscr{H}$, spanned by a truncated Fourier basis derived from (5.17), namely

$$\phi_M(x) = \sigma_f \begin{bmatrix} w_0 \\ \sqrt{2} w_1 \cos(\omega_1^\top P(x)) \\ \sqrt{2} w_1 \sin(\omega_1^\top P(x)) \\ \vdots \\ \sqrt{2} w_M \cos(\omega_M^\top P(x)) \\ \sqrt{2} w_M \sin(\omega_M^\top P(x)) \end{bmatrix},$$

where $M \in \mathbb{N}$ denotes the number of selected *wavenumbers* (i.e., spatial frequencies) $\omega_j := 2\pi\zeta_j \in \mathbb{R}^{n_x}$, $j \in \{0, \ldots, M\}$, indexed by multi-indices $\zeta_j \in (\mathbb{N}_0)^{n_x}$. The weights $w_0, \ldots, w_M \in \mathbb{R}_{\geq 0}$ are determined by the Gaussian spectral density via the multivariate CDF:

$$w_j^2 := \int_{(2\zeta_j - 1)\pi}^{(2\zeta_j + 1)\pi} \mathcal{N}(d\xi|0, \Sigma), \quad j \in \{0, \ldots, M\}.$$

Note that $\zeta_0 := (0, \ldots, 0)$ corresponds to the zero frequency component.

Fourier Barrier Certificate

Using this basis, any barrier $\mathcal{B} \in \mathscr{H}_M$ can be expressed as a finite Fourier series with $2M + 1$ coefficients:

$$\mathcal{B}(x) = \phi_M(x)^\top b = \alpha_0 + \sum_{i=1}^{M} \alpha_i \cos\left(\omega_i^\top P(x)\right) + \beta_i \sin\left(\omega_i^\top P(x)\right), \tag{5.18}$$

[4] Strictly speaking, $\mathscr{H}_M$ is not a subspace of the original RKHS $\mathscr{H}$ since, for instance, the constant function (included in $\mathscr{H}_M$) is not contained in $\mathscr{H}$; see, e.g., Kanagawa et al. (2018).

which is linear in the corresponding spectral amplitudes

$$b = \left[\frac{\alpha_0}{\sigma_f^2 w_0^2} \ \frac{\alpha_1}{2\sigma_f^2 w_1^2} \ \frac{\beta_1}{2\sigma_f^2 w_1^2} \ \cdots \ \frac{\alpha_M}{2\sigma_f^2 w_M^2} \ \frac{\beta_M}{2\sigma_f^2 w_M^2} \right]^\top \in \mathbb{R}^{2M+1}$$

associated with the basis functions. Naturally, barriers of the form (5.18) will be referred to as *Fourier control barrier certificates*. For illustrative purposes, the first basis functions of an example Fourier barrier are visualised in Fig. 5.3. A candidate barrier $\mathcal{B} \in \mathscr{H}_M$ arises as a weighted sum of these basis functions.

Compared to the typical representer form (5.15) based on spatial support vectors $\bar{x}_1, \bar{x}_2, \ldots \in \mathbb{X}$, the spectral form (5.18) is a much more efficient representation for higher-dimensional systems. The number of spatial support vectors needed to construct adequately expressive barriers grows exponentially in the dimensionality of the system. This is avoided here. In fact, the same principle is leveraged by memory-efficient file formats such as JPEG.

Apart from the memory efficiency of the spectral barrier form, the same idea can be applied to collapse the complex CME term in (5.14d), accelerated by the fast Fourier transform (FFT). To that end, a spectral coefficient matrix $H \in \mathbb{R}^{(2M+1)\times(2M+1)}$ is computed such that

$$k_{\hat{X}}^N(x)^\top \left[K_{\hat{X}}^N + N\lambda I_N \right]^{-1} \left[\phi_M(\hat{x}_+^i)^\top \right]_{i=1}^N b \approx \phi_M(x)^\top H b. \tag{5.19}$$

Whilst projection onto a truncated Fourier basis introduces an approximation error, the error decreases exponentially with the cardinality of wavenumbers in the basis (Rahimi and Recht, 2007).

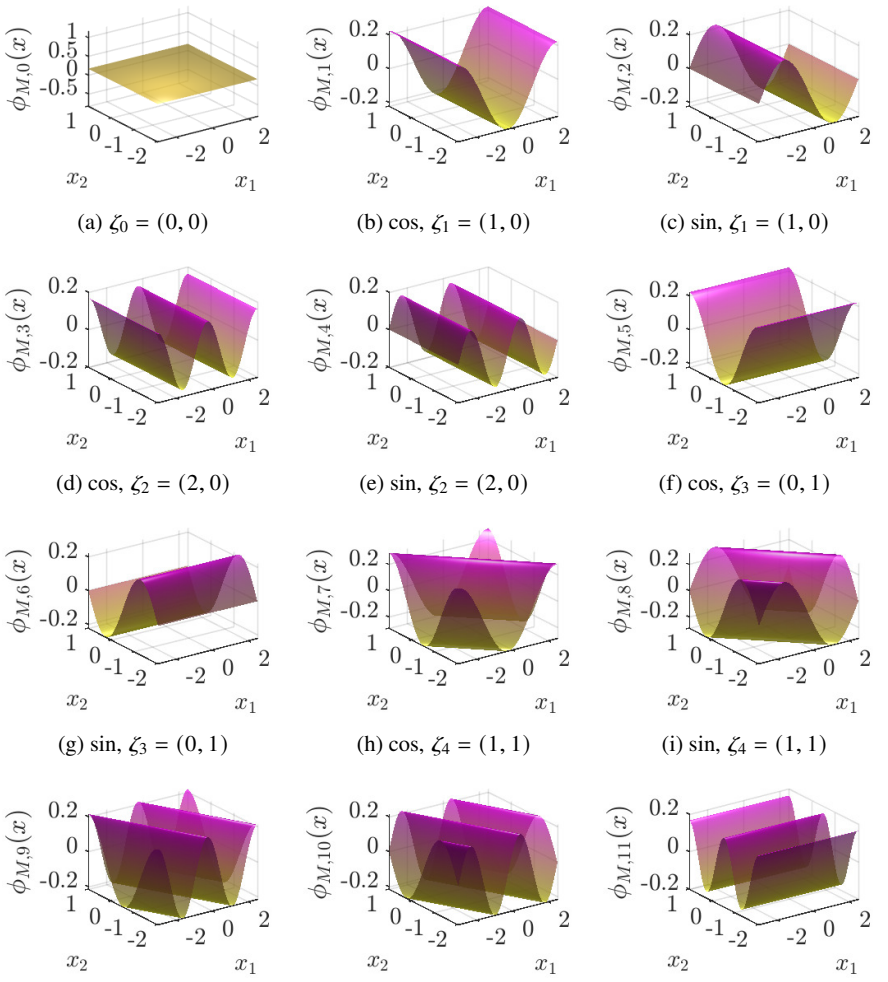
Substituting the barrier $\mathcal{B} \in \mathscr{H}_M$ of the form (5.18) and the approximated CME term in (5.19) into the SIP (5.14), the following *semi-infinite linear program* (SILP) is obtained:

Data-Driven Barrier SILP

$$\begin{aligned}
\min_{b,c,\eta,\gamma} \quad & \frac{\eta + cT}{\gamma}, && \text{(5.20a)}\\
\text{subject to} \quad & \forall x_0 \in \mathbb{X}_0 : \ \phi_M(x_0)^\top b \le \eta, && \text{(5.20b)}\\
& \forall x_u \in \mathbb{X}_u : \ \phi_M(x_u)^\top b \ge \gamma, && \text{(5.20c)}\\
& \forall x \in \mathbb{X} : \ \phi_M(x)^\top (Hb - b) \le c - \Re\overline{\mathcal{B}}\overline{\kappa}, && \text{(5.20d)}\\
& \forall x \in \mathbb{X} : \ \phi_M(x)^\top b \ge 0, && \text{(5.20e)}\\
& c \ge 0, \ \gamma > \eta \ge 0, \ b \in \mathbb{R}^{2M+1},
\end{aligned}$$

with coefficients $\overline{\kappa} \ge \sigma_f$ and $\overline{\mathcal{B}} \ge \|b\|_2$.

Spectral Symbolic Representation In the previous chapters, discrete representative states (Chapter 2 & Chapter 3) and binary strings (Chapter 4) served as symbolic abstractions of the continuous system state. Here, the truncated



(a) $\zeta_0 = (0, 0)$ (b) cos, $\zeta_1 = (1, 0)$ (c) sin, $\zeta_1 = (1, 0)$

(d) cos, $\zeta_2 = (2, 0)$ (e) sin, $\zeta_2 = (2, 0)$ (f) cos, $\zeta_3 = (0, 1)$

(g) sin, $\zeta_3 = (0, 1)$ (h) cos, $\zeta_4 = (1, 1)$ (i) sin, $\zeta_4 = (1, 1)$

(j) cos, $\zeta_5 = (2, 1)$ (k) sin, $\zeta_5 = (2, 1)$ (l) cos, $\zeta_6 = (0, 2)$

Fig. 5.3: Spectral basis functions used for the barrier in Subsection 5.9.1. A Fourier barrier is thus composed of a superposition of standing waves yielding a Fourier series expansion.

Fourier expansion of the kernel yields a *spectral* abstraction of both the barrier and the embedded dynamics, as expressed in (5.18) and (5.19). In this setting, the wavenumbers $\omega_j := 2\pi\zeta_j \in \mathbb{R}^{n_x}$, $j \in \{0, \dots, M\}$, act as symbolic coordinates over frequency space. Unlike the previous cases, these symbols do *not* encode equivalence classes of physical states $x \in \mathbb{X}$, but rather abstract the frequency response of the system. Designing more adaptive or efficient

wavenumber selection procedures for spectral abstraction remains an interesting direction for future work.

5.8.2 Finite-Constraint Relaxation

Deterministic sampling-based bounds for multivariate trigonometric polynomials have been established in the literature (Cifuentes and Parrilo, 2017; Pfister and Bresler, 2018), which can be directly applied to the bandlimited spectral barrier in (5.18). Using these results, the SILP in (5.20) can be relaxed to a finite-constraint linear program (LP) by sampling a finite subset $X_{\hat{N}} \subset \mathbb{X}$ of cardinality $\hat{N}$. To this end, $\hat{N} = Q^{n_x}$ equidistant points are selected to form a discrete lattice of points covering the state space. The corresponding constraints can then be evaluated efficiently in $O(Q \log(Q))$ operations via the inverse FFT. In addition, the same procedure also considerably reduces the number of evaluations of the computationally expensive CME term in (5.14d), which requires $O(N^2)$ operations per evaluation. For a further details, the reader is referred to the Petersen-Middleton theorem (Petersen and Middleton, 1962), which generalises the Nyquist-Shannon sampling theorem to higher-dimensional Euclidean domains. The following result by Pfister and Bresler (2018) provides the technical foundation for this relaxation:

Trigonometric Bounds

Lemma 5.1 (Trigonometric Bounds) *Let $f\colon [0,1]^{n_x} \to \mathbb{R}$ be a real-valued trigonometric polynomial (e.g., (5.18)) of maximum degree $d \in \mathbb{N}_0$, and define a sampling lattice $\Theta_Q := \{\frac{2\pi l}{Q} | l = 0, \dots, Q-1\}$ with $Q \geq 2d+1$. Let $\hat{f} := \max_{z \in \Theta_Q^n} f(z)$ and $\check{f} := \min_{z \in \Theta_Q^n} f(z)$. Then, the following bounds hold for all $z \in [0,1]^{n_x}$:*

$$\frac{1}{2}\left(\hat{f} + \check{f} - C_{Q^{n_x}}\left(\hat{f} - \check{f}\right)\right) \leq f(z) \leq \frac{1}{2}\left(\hat{f} + \check{f} + C_{Q^{n_x}}\left(\hat{f} - \check{f}\right)\right),$$

where the coefficient $C_{Q^{n_x}} \geq 1$ is defined as

$$C_{Q^{n_x}} := \left(1 - \frac{2d}{Q}\right)^{-\frac{n_x}{2}}.$$

Finite-Constraint Relaxation

From Lemma 5.1, a finite-constraint relaxation of the SIP (5.14) can be derived via a constraint-tightening argument:

$$\frac{1}{2}\left(\hat{\mathcal{B}}_{\hat{N}}^{\mathbb{X}} + \check{\mathcal{B}}_{\hat{N}}^{\mathbb{X}} - C_{\hat{N}}\left(\hat{\mathcal{B}}_{\hat{N}}^{\mathbb{X}} - \check{\mathcal{B}}_{\hat{N}}^{\mathbb{X}}\right)\right) \geq \mathrm{lb} \quad \Longrightarrow \quad \forall x \in \mathbb{X}\colon\ f(x) \geq \mathrm{lb},$$

for a desired lower bound $\mathrm{lb} \in \mathbb{R}$, where the barrier function $\mathcal{B}(x) = \phi_M(x)^\top b$ attains the following maximum and minimum values on the finite lattice:

$$\hat{\mathcal{B}}^{\mathbb{X}}_{\hat{N}} := \max_{x \in X_{\hat{N}}} \left\{ \phi_M(x)^\top b \right\} \quad \text{and} \quad \check{\mathcal{B}}^{\mathbb{X}}_{\hat{N}} := \min_{x \in X_{\hat{N}}} \left\{ \phi_M(x)^\top b \right\}.$$

Consequently, a sufficient condition can be expressed in the form

$$\frac{1}{2}\left(\hat{\mathcal{B}}^{\mathbb{X}}_{\hat{N}} + \check{\mathcal{B}}^{\mathbb{X}}_{\hat{N}} - C_{\hat{N}}\left(\hat{\mathcal{B}}^{\mathbb{X}}_{\hat{N}} - \check{\mathcal{B}}^{\mathbb{X}}_{\hat{N}}\right)\right) \geq \text{lb} \quad \Leftrightarrow \quad \check{\mathcal{B}}^{\mathbb{X}}_{\hat{N}} \geq \frac{2}{C_{\hat{N}}+1}\text{lb} + \frac{C_{\hat{N}}-1}{C_{\hat{N}}+1}\hat{\mathcal{B}}^{\mathbb{X}}_{\hat{N}}.$$

An analogous relaxation can be derived for upper bounds $f(x) \leq \text{ub}$ for all $x \in \mathbb{X}$, which yields the constraint

$$\hat{\mathcal{B}}^{\mathbb{X}}_{\hat{N}} \geq \frac{2}{C_{\hat{N}}+1}\text{ub} + \frac{C_{\hat{N}}-1}{C_{\hat{N}}+1}\check{\mathcal{B}}^{\mathbb{X}}_{\hat{N}}.$$

These inequalities thus enable the reformulation of the original semi-infinite program as a tractable finite-dimensional LP. Notably, this approach yields *deterministic* coverage guarantees, in contrast to many alternative sampling-based methods that can only provide statistical guarantees, which are generally insufficient to ensure the validity of the barrier (see Definition 5.1).

5.8.3 Linear Program

The finite-constraint relaxation of the SILP in (5.20) is now presented. In addition to the discretised set $X_{\hat{N}} \subset \mathbb{X}$, discrete subsets $\{x_0^{(1)}, \ldots, x_0^{(\hat{N}_0)}\} \subset \mathbb{X}_0$ and $\{x_u^{(1)}, \ldots, x_u^{(\hat{N}_u)}\} \subset \mathbb{X}_u$ of cardinality $\hat{N}_0 \in \mathbb{N}$ and $\hat{N}_u \in \mathbb{N}$, respectively, are formed. For given values of $\overline{\mathcal{B}}$ and γ, the following LP is obtained:

Relaxed Data-Driven Barrier LP

$$\begin{aligned}
\min_{\substack{b,c,\eta \\ \check{\mathcal{B}}^{\mathbb{X}_0}_{\hat{N}}, \hat{\mathcal{B}}^{\mathbb{X}_u}_{\hat{N}}, \hat{\mathcal{B}}^{\mathbb{X}}_{\hat{N}}, \check{\mathcal{B}}_\Delta}} \quad & \eta + cT, & & \\
\text{subject to} \quad & \phi_M(x_0^{(i)})^\top b \leq \hat{\eta}, & & i = 1, \ldots, \hat{N}_0, \\
& \phi_M(x_u^{(i)})^\top b \geq \hat{\gamma}, & & i = 1, \ldots, \hat{N}_u, \\
& \phi_M(x^{(i)})^\top (Hb - b) \leq \hat{\Delta}, & & i = 1, \ldots, \hat{N}, \\
& \phi_M(x^{(i)})^\top b \geq \hat{\xi}, & & i = 1, \ldots, \hat{N}, \\
& \check{\mathcal{B}}^{\mathbb{X}_0}_{\hat{N}} \leq \phi_M(x_0^{(i)})^\top b, & & i = 1, \ldots, \hat{N}_0, \\
& \hat{\mathcal{B}}^{\mathbb{X}_u}_{\hat{N}} \geq \phi_M(x_u^{(i)})^\top b, & & i = 1, \ldots, \hat{N}_u, \\
& \hat{\mathcal{B}}^{\mathbb{X}}_{\hat{N}} \geq \phi_M(x^{(i)})^\top b, & & i = 1, \ldots, \hat{N}, \\
& \check{\mathcal{B}}_\Delta \leq \phi_M(x^{(i)})^\top (Hb - b), & & i = 1, \ldots, \hat{N}, \\
& c \geq 0,\ \gamma > \eta \geq 0,\ b \in \mathbb{R}^{2M+1}, & &
\end{aligned} \tag{5.21}$$

with $\overline{\kappa} \geq \sigma_f$, $\overline{\mathcal{B}} \geq \|b\|_2$, and constraint-tightening coefficients

$$\hat{\eta} := \frac{2}{C_{\hat{N}}+1}\eta + \frac{C_{\hat{N}}-1}{C_{\hat{N}}+1}\check{\mathcal{B}}_{\hat{N}}^{\mathbb{X}_0}, \qquad \hat{\gamma} := \frac{2}{C_{\hat{N}}+1}\gamma + \frac{C_{\hat{N}}-1}{C_{\hat{N}}+1}\hat{\mathcal{B}}_{\hat{N}}^{\mathbb{X}_u},$$
$$\hat{\Delta} := \frac{2}{C_{\hat{N}}+1}\left(c - \Re\overline{\mathcal{B}}\overline{\kappa}\right) + \frac{C_{\hat{N}}-1}{C_{\hat{N}}+1}\check{\mathcal{B}}_{\Delta}, \qquad \hat{\xi} := \frac{C_{\hat{N}}-1}{C_{\hat{N}}+1}\hat{\mathcal{B}}_{\hat{N}}^{\mathbb{X}}.$$

Any solution to the LP (5.21) satisfying the norm constraint $\|b\|_2 \leq \overline{\mathcal{B}}$ constitutes a feasible solution to the original SIP (5.14).

Solution For a given controller $\boldsymbol{C}$, any feasible solution to the LP (5.21) satisfying $\|b\|_2 \leq \overline{\mathcal{B}}$ solves Problem 5.1 via Proposition 5.1, namely

$$\mathbb{P}(\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}} \models \psi_{\text{safe}}) \geq 1 - \frac{\eta + cT}{\gamma}.$$

Further details regarding the derivation of the LP (5.21) from the SILP (5.20) via the constraint-tightening procedure introduced in Subsection 5.8.2 are provided in Appendix B.

5.9 Case Studies

The proposed Fourier barrier approach is evaluated on several benchmark problems. In all experiments, $N = 1000$ data points are sampled from the unknown system. The inverse of the Gram matrix is computed with a regularisation parameter $\lambda = 10^{-5}$. To construct the finite-constraint LP (5.21), a discrete lattice of $\hat{N}$ points is generated using the FFT, with the value of $\hat{N}$ specified per benchmark. For simplicity, a rectangular sampling lattice is used, leaving the exploration of more complex (e.g., optimal) sampling lattices for future work.

All computations are performed on a machine equipped with an AMD Ryzen 9 5950X 16-core CPU and 64 GB of RAM. The LP is generated using a serial implementation. For solving the LP, Gurobi (Gurobi Optimization, LLC, 2024) is used, which is allowed to execute solution methods concurrently using its default settings. The feasibility tolerance is set to the minimum possible value of 10^{-9}, keeping the remaining settings at default.

5.9.1 Complex Safety Specification

The first benchmark is adapted from the deterministic example $\texttt{Barr}_3$ by Abate et al. (2021), extended here to include stochastic process noise. The resulting two-

dimensional nonlinear stochastic dynamics are given by

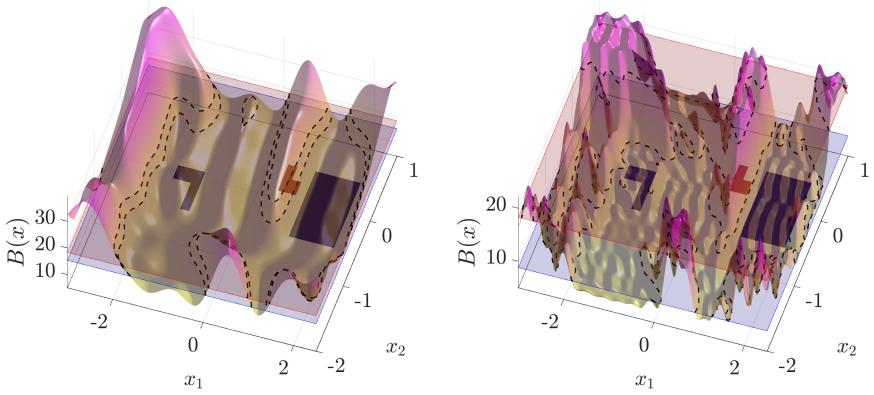
$$\mathcal{M}\colon \left\{ \begin{bmatrix} x_{1,t+1} \\ x_{2,t+1} \end{bmatrix} = \begin{bmatrix} x_{2,t} \\ \frac{1}{3}x_{1,t}^3 - x_{1,t} - x_{2,t} \end{bmatrix} + w_t, \quad w_t \sim \mathcal{N}(\cdot\,|0, 0.01 I_2). \right.$$

The objective is to compute a certified lower bound on the probability that the system initialised in $[x_{1,0}, x_{2,0}]^\top \in \mathbb{X}_0$ (blue regions in Fig. 5.4) avoids the unsafe regions $\mathbb{X}_u$ (in red) for a time horizon of $T = 10$. The close proximity of $\mathbb{X}_0$ and $\mathbb{X}_u$ renders the barrier synthesis particularly challenging.

The LP is generated based on a fixed lattice of $\hat{N} = 11^2$ equidistant points. Two configurations of spectral barrier bases are considered:

(a) $M = 6^2 - 1 = 35$ wavenumbers, yielding $2M + 1 = 71$ barrier coefficients; and
(b) $M = 16^2 - 1 = 255$ wavenumbers, resulting in 511 coefficients.

In both cases, the kernel hyperparameters are fixed to $\sigma_f = 19.456$ and $\sigma_l = [30, 23.568]$, with a complexity bound $\bar{\mathcal{B}} = 25$, robustness radius $\mathfrak{R} = 0.001$, and threshold level $\gamma = 18.312$. The resulting barriers are displayed in Fig. 5.4. For (a),



(a) $M = 35$, $p_N = 16\%$ (b) $M = 255$, $p_N = 52\%$

Fig. 5.4: Robustified CBCs and level sets γ (in red) and η (in blue), computed for the benchmark in Subsection 5.9.1, robustified with margin $\mathfrak{R}\overline{\mathcal{B}}\sigma_f = 0.4864$ and using (a) $M = 35$ wavenumbers ($\|b\| = 10.394$), and (b) $M = 255$ wavenumbers ($\|b\| = 21.397$), respectively.

the optimisation yields $\eta = 15.337$ and $c = 0$, certifying a lower bound of 16% on the safety probability via Theorem 5.3. As $c = 0$, this certificate remains valid as $T \to \infty$ (see Remark 5.1). The barrier basis (b) is much richer, leading to $\eta = 8.826$ and $c = 0$, certifying a safety probability of at least 52% for infinite time horizons.

The actual safety probability P_{safe} is estimated via Monte Carlo simulation using 10 000 sample trajectories, each with a maximum length of 500 time steps. By

applying Chebyshev's inequality at a 90% confidence level, the safety probability is estimated to lie within the interval 75–78%. It is noted that the certified lower bound may not converge to the actual safety probability even in the limit $M \to \infty, N \to \infty$, due to the general conservatism inherent to the barrier method. This is in contrast to abstraction-based approaches explored in earlier chapters, for which the certified probability converges to the true one as the abstraction is refined and more data is accumulated (Laurenti and Lahijanian, 2023).

5.9.2 Overtaking Scenario Featuring a Neural Network Controller

Now, a system controlled by a NN controller is considered. The scenario involves an AV overtaking another vehicle, as depicted in Fig. 5.1. The dynamics of the ego vehicle are given by Dubins car model with additive noise:

$$\boldsymbol{\mathcal{M}}\colon \begin{cases} \begin{bmatrix} x_{t+1} \\ y_{t+1} \\ \phi_{t+1} \end{bmatrix} = \begin{bmatrix} x_t \\ y_t \\ \phi_t \end{bmatrix} + \begin{bmatrix} \tau v \cos(\phi_t) \\ \tau v \sin(\phi_t) \\ \tau u_t \end{bmatrix} + \begin{bmatrix} w_{1,t} \\ w_{2,t} \\ w_{3,t} \end{bmatrix}, \quad w_t \sim \mathcal{N}(\cdot\,|0,\Sigma). \end{cases} \tag{5.22}$$

where (x, y) denotes the relative position in longitudinal and lateral direction, and ϕ is the heading angle. The noise $w_{i,t}$, $i \in \{1, 2, 3\}$, is zero-mean Gaussian with covariance matrix $\Sigma := \mathrm{diag}([0.01^2, 0.01^2, 0.001^2])$.

The steering angle $u_t = \boldsymbol{C}(x_t)$ is supplied by a NN controller $\boldsymbol{C}\colon \mathbb{X} \to \mathbb{U}$, trained using the *twin-delayed deep deterministic* (TD3) policy gradient algorithm. The controller is trained to avoid the unsafe regions $\mathbb{X}_u$ and reach a target set, as illustrated in Fig. 5.5. The NN is equipped with two layers of 200 neurons each and produces actions $u_t \in [-\pi, \pi]$. A fixed $v = 1$ and time discretisation $\tau = 0.5$ are used. The objective is to certify that the system initialised in $[x_0, y_0, \phi_0]^\top \in \mathbb{X}_0$ remains safe, i.e., avoids $\mathbb{X}_u$, over $T = 50$ time steps.

An LP is formulated on a lattice of $\hat{N} = 30^3$ equidistant points. A spectral basis with $M = 4^3 - 1$ wavenumbers is used, yielding a barrier parametrised by $2M + 1 = 127$ coefficients. The kernel hyperparameters are set as $\sigma_f = 7.764$ and $\sigma_l = [8.1212, 6.8174, 6.3307]$. For $\mathfrak{R} = 0$, $c = 0$, and level sets $\gamma = 1.775$ and $\eta = 0.325$, the resulting barrier (shown in Fig. 5.6a) certifies a safety probability of at least 82%. As $c = 0$, this safety guarantee holds even for the infinite-horizon case $T \to \infty$ (see Remark 5.1). Note that this is only certifying safety with respect to the empirical distribution, i.e., the CME constructed from the observed data. To robustify the result to out-of-distribution behaviour, the robustness coefficient $\mathfrak{R}$ is increased. Its effect on the certified robust safety probability p_N is shown in Fig. 5.7. Intuitively, increasing $\mathfrak{R}$ tightens the Kushner constraint (5.14d) by enforcing the barrier to decay by at least $\mathfrak{R}\overline{\mathcal{B}}\sigma_f$ in every time step. This requires higher barrier complexity, reflected in an exponentially increasing complexity of the barrier — measured through the norm $\|b\|$. Simultaneously, the separation between the level sets γ and η shrinks, prompting the robust safety probability p_N to decline. For the

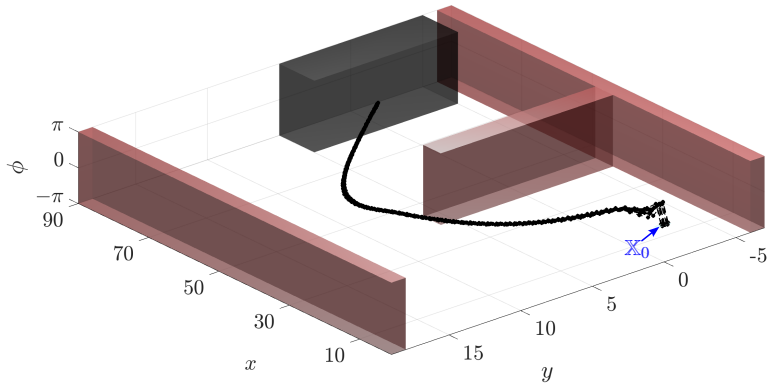
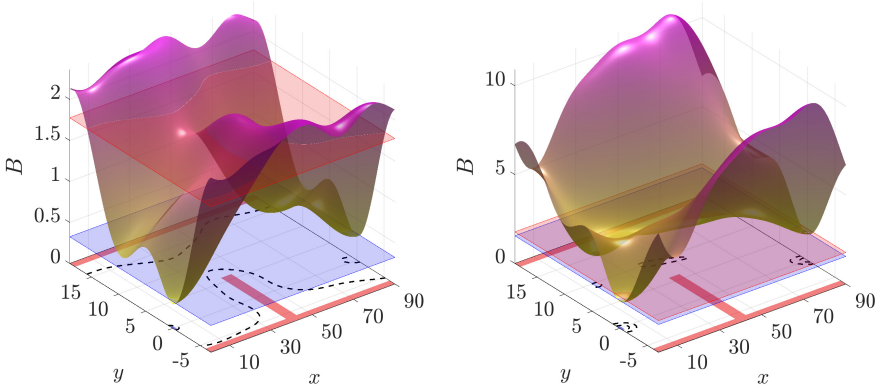


Fig. 5.5: Control performance of the NN controller for the overtaking scenario (Subsection 5.9.2), demonstrated on 10 random trajectories initialised in $\mathbb{X}_0$ (in blue), trained to avoid the unsafe regions $\mathbb{X}_u$ (in red) and reach a target set (in black).



(a) Empirical CBC ($\mathfrak{R}\overline{\mathcal{B}}\sigma_f = 0$), $p_N = 82\%$ (b) Robust CBC ($\mathfrak{R}\overline{\mathcal{B}}\sigma_f = 0.2502$), $p_N = 12\%$

Fig. 5.6: CBCs and level sets γ (in red) and η (in blue) computed for the overtaking scenario (Subsection 5.9.2) shown for a fixed $\phi = 0$, (a) based only on the empirical data, and (b) robustified with $\mathfrak{R} = 0.006$ and $\overline{\mathcal{B}} = 5.37029$. The intersections of the barrier and the level sets are indicated as dashed lines.

largest considered value $\mathfrak{R} = 0.006$, the required decay reaches $\mathfrak{R}\overline{\mathcal{B}}\sigma_f = 0.2502$, and the resulting barrier (shown in Fig. 5.6b) certifies a robust safety probability of 12%.

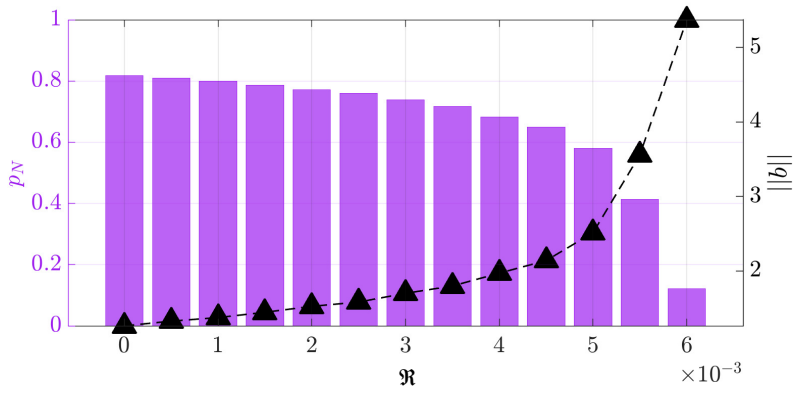


Fig. 5.7: Robust safety probability p_N (in purple) and required barrier complexity $\|b\|$ (in black) for the overtaking scenario (Subsection 5.9.2) generated for different robustness radii $\mathfrak{R}$.

Table 5.2: Robust safety probability p_N and runtime (for Fourier projection and LP solving) obtained for the benchmarks for $\mathfrak{R} = 0$ and variations of the following quantities: number of system dimensions n_x, basis size $2M + 1$, number of lattice samples Q^{n_x}.

Benchmark	**#Dimensions**	$2M+1$	**#Samples**	p_N	**Runtime**
Complex specification (Subsection 5.9.1)	2	35	11^2	16%	0.6 s
		255	11^2	52%	36 s
Overtaking scenario (Subsection 5.9.2)	3	127	30^3	81.6%	111 s
		127	40^3	82.4%	219 s
		127	50^3	83.8%	606 s
		249	40^3	83.4%	579 s
		249	50^3	84.7%	1154 s
		431	40^3	84.2%	3596 s

Fig. 5.7 visualises this trade-off between robustness and barrier complexity. Table 5.2 further quantifies the scalability of the approach across the two benchmarks, basis size, and lattice resolution.

5.10 Conclusion

This work contributes to the verification and synthesis of stochastic black-box systems without explicit model knowledge, by leveraging the theory of conditional mean embedding (CMEs) to learn control barrier certificates (CBCs) directly from data. Probabilistic CBC constraints are reformulated into distributionally robust optimisation problems, enabling their characterisation from data with probabilistic correctness guarantees. The approach relaxes many assumptions inherent in existing approaches, allowing it to tackle black-box systems with neural network (NN) controllers.

It is shown how the approach can be generalised to encompass temporal logic specifications beyond safety via automata-based representations and the introduction of Streett supermartingales For the data-driven computation of safety barriers, a finite Fourier expansion has been proposed, allowing the CBC constraints to be encoded as a linear program.

Despite these advances, several limitations remain. The conservatism of the method warrants further investigation, particularly regarding the derivation of tighter ambiguity set radii $\mathfrak{R}$. This includes the potential application of concentration bounds tailored to the barrier function class, e.g., based on Rademacher complexity, and the exploration of alternative algorithmic strategies for solving the Fourier-based CBC optimisation problem (5.20).

While the spectral representation circumvents the need for exponentially many spatial support vectors, the number of samples required by the trigonometric sampling bounds remains exponential in the system dimensionality. Although FFT-based implementations alleviate this burden to some extent, scalability challenges persist, especially in solving the resulting LP. Future work may involve coupling the Fourier-based formulation with global nonlinear optimisers possessing deterministic convergence guarantees. Achieving this is expected to necessitate the development of a dedicated semi-infinite solver that fully exploits the problem structure. Further efforts are also needed to address hyperparameter tuning and the principled selection of the spectral basis. Finally, a complete software implementation supporting temporal logic specifications remains an important direction for future work.

Chapter 6
Conclusion

This chapter summarises the main contributions of the thesis and reflects on their significance and impact. These contributions address core challenges in the formal synthesis and verification of autonomous systems operating under uncertainty, and are organised thematically to highlight the progression from model-rich formulations to learning-based, black-box approaches:

I Addressing Uncertainty via Robust Stochastic Simulation Relations
II Scalable Synthesis via Compositional Reasoning
III Direct Learning of Efficiently-Amenable Formal Abstractions from Data
IV A Framework for Formal Verification of Black-Box Systems

The theoretical contributions presented in this thesis are supported by the development of corresponding software tools, briefly summarised in Section 6.2.

6.1 Summary

I Addressing Uncertainty via Robust Stochastic Simulation Relations Chapter 2 extends stochastic simulation relations to systems with ambiguous dynamics. By relaxing classical probabilistic couplings to sub-probability couplings, this work synthesises robust controllers with formal guarantees despite uncertain system parameters. Paired with Bayesian inference, the resulting data-driven approach provides refined guarantees as data accumulates, bridging the gap between theoretical robustness and practical adaptability. This contribution is foundational for the control design of autonomous CPSs, offering robust guarantees in settings where precise models are unavailable and parameters vary across operational modes or due to manufacturing tolerances.

II Scalable Synthesis via Compositional Reasoning Chapter 3 addresses the scalability limitations of abstraction-based formal methods that hinder their application to large-scale, interconnected CPSs. By developing a compositional framework and

dealing with the challenges of heterogeneous stochasticity, this work shows how local simulation relations and controllers can be systematically combined to guarantee the desired global system behaviour. This enables formal synthesis methods to scale linearly with the number of subsystems, a key step toward practical deployment in complex networks such as smart grids, automated transportation, or industrial automation. The compositional approach thus significantly expands the applicability of robust formal methods to real-world high-risk domains.

III Direct Learning of Efficiently-Amenable Formal Abstractions from Data Chapter 4 advances the thesis focus to *learning* by addressing the challenge of formal verification for systems with completely unknown dynamics. By leveraging structured Gaussian process models, it enables the direct construction of formal abstractions from data that effectively capture learning uncertainty without incurring excessive computational cost. This contribution opens a promising path toward scalable *learning methods designed for verification*, which are essential for extending formal guarantees to realistic autonomous systems with learning-based components.

IV A Framework for Formal Verification of Black-Box Systems Chapter 5 completes the progression from robustness- to learning-centred approaches, addressing a central safety challenge posed by autonomous systems: the combination of randomness and AI components in the loop. Due to their inherent complexity and unknown dynamics, such systems can only be effectively treated as black boxes. They lie beyond the scope of earlier formal frameworks, which rely on assumptions about model structure or noise distributions that do not hold in this setting. Breaking free from previous abstraction paradigms, this work concentrates on learning safety certificates directly from observed system behaviour, eliminating the need for intermediate models and operating directly on observed data points through conditional mean embeddings. This opens a promising direction for establishing formal guarantees in opaque, AI-driven systems. Such capabilities are essential for realising trustworthy autonomy in realistic settings, where safety must ultimately be inferred directly from behaviour learned from multimodal sensor streams.

To bridge theory and practice, the theoretical contributions of this thesis are accompanied by the development of tools that render the results practically accessible.

6.2 Tool Development

SySCoRe – Synthesis via Stochastic Simulation Relations Developed alongside the work in Chapter 2 and Chapter 3, the SySCoRe toolbox[1] is a MATLAB package implementing the key algorithms for correct-by-design synthesis of stochastic systems in an accessible and performant format (Van Huijgevoort et al., 2023). Given a system description and a specification, SySCoRe automates all steps required to

[1] Source code: https://github.com/oliverschoen/SySCoRe; online executable capsule: https://doi.org/10.24433/co.3825821.v2.

construct a symbolic abstraction, translate the formula into a DFA, perform stochastic MOR, and synthesise a robust controller with quantified satisfaction probability. SySCoRe distinguishes itself from existing tools by supporting MOR, nonlinear dynamics, and rich scLTL specifications over infinite horizons. This is achieved by establishing probabilistic simulation relations between an abstract nominal model and the original system, as described in earlier chapters. A key strength is that the computed error does not grow linearly with the specification horizon, enabling meaningful lower bounds even under unbounded disturbances and infinite-horizon properties. Efficient tensor-based representations further accelerate the computation of transition probabilities.

In contrast to widely used tools for probabilistic model checking over finite-state systems — such as PRISM (Kwiatkowska et al., 2002) and MRMC (Katoen et al., 2011) — SySCoRe targets continuous-space systems. It has been benchmarked against comparable tools in the annual ARCH competition (Abate et al., 2022a, 2023, 2025), with most alternative tools in this space relying on IMDPs (Wooding and Lavaei, 2024; Mathiesen et al., 2024), which offer only limited support for MOR and often incur higher computational cost.

LUCID – Learning Uncertain Control Invariant Dynamics Building on the framework introduced in Chapter 5, LUCID is a verification engine under development for certifying the safety of black-box stochastic systems using only data (Casablanca et al., 2026). LUCID provides both a graphical user interface and command-line access. Users can load datasets, specify operating and unsafe regions (e.g., obstacles), and receive a computed barrier certificate that certifies safety with a quantified probability. Internally, LUCID constructs a CME-based expectation oracle and automatically learns a Fourier barrier function that satisfies the probabilistic barrier conditions. The toolbox is being developed in Python, with core components implemented in C++ for performance, in collaboration with Ernesto Casablanca and Prof. Sadegh Soudjani.

Few existing tools offer comparable capabilities. The most closely related is TRUST (Gardner et al., 2025), which can synthesise barrier certificates from data but is limited to deterministic systems with polynomial dynamics. FOSSIL (Edwards et al., 2024) supports arbitrary nonlinear dynamics but is not data-driven and is also limited to deterministic systems. In contrast, LUCID is designed specifically for stochastic black-box systems — essential for certifying the safety of AI-driven autonomous systems.

Future extensions of LUCID will include control synthesis, less conservative barrier variants, and support for more expressive specifications.

BTGP Synthesis and Verification Toolbox The BTGP–IMDP framework introduced in Chapter 4 is currently being implemented as a professional Python toolbox. It enables users to train BTGPs from data through a scikit-learn-compatible interface, abstracting away the internal complexity of the underlying architecture. The sparse structure of the BT kernel accelerates training, and the toolbox provides streamlined functionality for verifying reachability properties and synthesising controllers — both accessible via a single command.

In contrast to related *model-based* tools such as IMPaCT (Wooding and Lavaei, 2024) and IntervalMDP.jl (Mathiesen et al., 2024) — which require a known model and exhibit significantly higher computational cost for IMDP construction — this toolbox directly supports learning from data with orders-of-magnitude speed-up, as demonstrated in Chapter 4.

Chapter 7
Future Directions

Several challenges of advancing formal synthesis and verification to real-world deployment have been addressed in this thesis. At the same time, the works uncovered many promising directions for future work, that shall be discussed here. The following future directions are examined:

I Data-Driven Black-Box Control
II IMDP-Based Learning
III Hybrid Correctness Frameworks
IV Efficient Reasoning in Agent-Environment Interactions
V Control-Theoretical Connections
VI Dynamic Specifications
VII Situational Awareness
VIII Identifying High-Probability Critical Scenarios
IX Inferring Causal Structure
X Partial Observability and Non-Markovian Dynamics
XI Discrete- vs Continuous-Time
XII Multimodal Data Streams
XIII Formal Contracts for AI-Enabled Components

Albeit tightly linked, these extensions can be loosely categorised into (1) methodological extensions, (2) abstraction and sparsity techniques, and (3) extensions addressing expressiveness, integration, and deployment.

7.1 Methodological Extensions

I Data-Driven Black-Box Control A natural extension of the results in Chapter 5 on verifying black-box systems is a CME-based framework for control synthesis. Since the control input enters the problem formulations through the (nonlinear) kernel (see Theorem 5.2 for the details), the resulting optimisation problem is generally non-convex and difficult to solve. A promising direction is to explore whether —

analogous to the Fourier-based formulation considered in this thesis — structural properties of the kernel can be exploited to yield more tractable formulations. Further investigations into tree-like kernels or spectral properties of the CME may offer a fruitful starting point.

II IMDP-Based Learning To obtain stronger guarantees beyond distributional robustness, CME-based approaches could be explored under additional strengthening assumptions, such as Markovian dynamics. In this case, CMEs stand in direct competition with IMDP-based approaches, which offer strong formal guarantees under model uncertainty. When applied in learning-based contexts — particularly relevant for embodied AI systems, where policies must adapt continuously and in real-time to newly acquired data — however, standard GP–IMDP pipelines are often prohibitively slow (see Chapter 4). Future work should therefore investigate how IMDP updates can be performed sparsely in response to newly acquired data that only affects the learned dynamics locally, rather than requiring full regeneration. In particular, it would be promising to exploit the structure of the BTGP to enable such incremental updates, potentially at the cost of slightly more conservative guarantees. Advancing this capability may prove essential for scaling formal methods to continuous learning scenarios, where maintaining safety guarantees during the exploration of the environment remains a key challenge.

III Hybrid Correctness Frameworks There has been only very little effort devoted to analysing the dichotomy between abstraction-*based* and abstraction-*free* approaches, which seems to persist throughout the literature (Laurenti and Lahijanian, 2023). As shown in Chapter 5, however, this distinction becomes increasingly blurred in practice. Supposedly abstraction-free frameworks frequently give rise to optimisation problems that are difficult to solve monolithically, especially for complex systems and/or specifications. As a result, many approaches resort to solution procedures that are abstraction-based in all but name. This suggests that a strict separation between the two methodological frameworks may not be tenable, and future research should instead explore their symbiotic fusion.

In fact, barrier certificates already act as inert local correctness witnesses, whilst abstraction can be seen more as a momentum mechanism that structures the search or verification process. One could therefore try to establish new frameworks that start by identifying local "seed" barrier certificates within coarse partitions of the system domain, and then expand or merge them through successive refinement. This is methodologically similar to an approach explored by Bobiti and Lazar (2018), although their work is limited to deterministic dynamics and relies on estimating Lipschitz bounds. By contrast, kernel methods — extensively used throughout this thesis — may offer a more flexible and scalable foundation for such hybrid strategies.

7.2 Abstraction and Sparsity

IV Efficient Reasoning in Agent-Environment Interactions The importance of identifying and leveraging problem structure — both for scalability and for obtaining strong guarantees — has been emphasised throughout the thesis (Miner and Parker, 2004). Whilst compositional methods (see Chapter 3 and references therein) are effective for networked systems with bounded interaction between their subsystems, this assumption often fails in realistic CPSs, where dynamics are tightly coupled — so-called *reactive* systems — and global specifications cannot be decomposed into local ones. In such settings, efficient reasoning must instead rely on identifying and abstracting the dominant modes of interaction, i.e., in the form of a compressed model. This perspective aligns with the notion of *situational awareness*, which draws from human intuition (Kahneman, 2011): in the presence of many moving parts, attention is placed on the key dynamics that govern the situation.

Accordingly, navigating tightly coupled environments efficiently often requires reasoning under environmental ambiguity. This direction is partially explored by Schön et al. (2025), who embed probabilistic simulation relations into an assume-guarantee framework. Ambiguity is introduced deliberately through model order reduction and partial observability to control complexity. A promising direction for future work is to explore how, for a given situation, such compressed dynamics can be learned from data, as well as how situations themselves can be defined and switched between effectively. Central challenges in this context are quantifying the inflicted statistical abstraction error and learning suitable probabilistic assumptions (Feng et al., 2011).

V Control-Theoretical Connections The idea of imposing sparsity to obtain compressed representations extends beyond the previously discussed intuition and suggests a deeper connection to existing concepts from classical control theory, which remains largely unexplored. For instance, the Fourier-based CME explored in Chapter 5 can be interpreted as a *frequency response* operator, raising the question of whether it admits a principled interpretation within the framework of control theory. Understanding the extent and nature of this connection could potentially untap a new line of research at the intersection of machine learning and control.

To this end, one may investigate alternative operator constructions, such as wavelet bases or tree-structured kernels, and analyse their control-theoretic interpretations. If successful, such work could lay the foundation for a renaissance of kernel-based techniques in formal modelling and reasoning, aligned with established notions from control theory.

7.3 Expressiveness, Integration, and Deployment

VI Dynamic Specifications This thesis assumed time-invariant specifications with a static labelling function. An important direction for future work is to study how

dynamic specifications (with dynamic labelling functions) affect the synthesis process — for example, by introducing symbolic parameters that encode aspects of the environment, such as moving obstacles. Recall that the labelling function defines an implicit abstraction of the output domain of the target system. Designing a dynamic output abstraction could enable efficient planning in dynamic environments without requiring full online controller synthesis (see, e.g., Lacerda et al., 2014).

VII Situational Awareness More broadly, supporting rich temporal specifications could pave the way for autonomous systems that are not only provably safe, but also context-aware. Extending beyond safety properties to expressive (signal) temporal logic objectives — such as those required for systems with situational awareness capabilities — allows for the automatic generation of robust, holistic controllers directly from requirements catalogues. Recall that many of the results from this thesis can already be readily extended to STL (see Section 2.7). For comprehensive requirements catalogues, a sparse integration of the structural constraints imposed by temporal logic specifications with the (e.g., CME-based) representation of the learned dynamics is required. This could be achieved by leveraging the structure of the associated labelling or robustness function throughout the learning and verification process (see, e.g., Gao et al., 2024). To manage and prioritise stacks of potentially conflicting objectives, multi-objective strategies have to be developed (see, e.g., Monir et al., 2024; Forejt et al., 2011b).

VIII Identifying High-Probability Critical Scenarios In a similar vein, identifying critical scenarios — particularly useful for analysing the behaviour of agents in complex environments throughout the design phase of an autonomous system — requires a systematic examination of the interaction between the specification and the system dynamics. Existing work in this direction is limited, with only a few approaches leveraging the structure of temporal logic specifications to isolate scenarios that are both realistic (as quantified by their likelihood) and exhibit low robustness margins (Corso and Kochenderfer, 2020; Tuncali et al., 2019). However, these methods do not provide any guarantees on the coverage of the space of possible scenarios, limiting their usefulness to mere recommendation tools in the system development process rather than supporting rigorous formal verification. In the context of the correct-by-design frameworks considered in this thesis, this raises the question how to extract critical scenarios that constrain the achievable guarantees. Note that this is not only helpful in the development process, but the identification of hazardous scenarios is a requirement of several industry standards such as ISO 34502:2022 (2022).

IX Inferring Causal Structure Taking this one step further, a natural extension is to move toward inferring the underlying causal structure governing the interaction between the system, its environment, and the specification. To this end, one could explore graph-based learning and reasoning methods capable of systematically uncovering interpretable dependencies from data. Such an approach would align well with existing industrial practices, where expert-defined discrete graph models are commonly used in commercial tools for design-space exploration, optimisation, and

counterexample generation. A data-driven approach to learning such structures could offer a robust and scalable alternative to manual specification.

Unlike the correct-by-design frameworks studied in this thesis, inferring causal structure from data may support stronger robustness guarantees under intervention-like perturbations — especially relevant for medical applications. Moreover, access to an explicit causal structure could facilitate more efficient verification, synthesis, and even counterfactual reasoning by constraining the analysis to graph-based optimisations aligned with observed system behaviour.

X Partial Observability and Non-Markovian Dynamics Building on the previous discussion of causal structure, the hierarchical relationships between symbolic variables — such as latent state variables and observable output variables — invite further investigation into relaxing the thesis-wide assumption of full-state access. In practice, systems are typically only partially observed through noisy sensor measurements. There exists work extending kernel-based learning models such as GPs and CMEs to sequential data (Toth and Oberhauser, 2020; Thorpe et al., 2022), however, they have not been studied in a formal context yet. Formal approaches remain limited to model-based settings, suffering mainly from the computational blow-up associated with partially observed systems — see, e.g., Haesaert et al. (2018) for a treatment of Gaussian linear POMDPs.

Partial observability was previously addressed by Schön et al. (2025) as a by-product of abstraction introduced to improve scalability. The resulting information asymmetry was captured via sub-probability couplings, enabling tractable reasoning through abstraction but prone to becoming increasingly conservative as observability degrades. A natural direction for future work is to explore whether more suitable coupling strategies exist and how they can be computed. A promising starting point is the data-driven, kernel-based framework developed in this thesis, which may also clarify how stochastic simulation relations relate to CMEs. Since current simulation relations consider one-step transitions of the system, they become intractable under high degrees of uncertainty. This underscores the need to incorporate sequential information — potentially using sequential kernels (Király and Oberhauser, 2019) — to capture temporal dependencies in partially observed settings.

XI Discrete- vs Continuous-Time When considering temporal aspects, formal frameworks are typically associated with either a discrete- or continuous-time paradigm. Navigating between these two remains conceptually unclear, particularly in data-driven settings. While physical systems evolve in continuous time, discrete-time models are often more natural for learning and verification due to the sampled nature of data and the inherently discrete operation of computational systems. Nonetheless, many methods embed discrete data into continuous-time frameworks by approximating time derivatives from samples — a practice that lacks formal rigour and may introduce artefacts if the time discretisation is not chosen appropriately. The challenge is compounded by asynchronous transitions across system components and sensor streams, rendering unified modelling increasingly difficult. An interesting direction for future work is to explore whether kernel methods, especially those designed for sequential data (Király and Oberhauser, 2019), could offer

a principled medium for fusing asynchronous, multi-rate signals. Such methods may provide a theoretical bridge between discrete- and continuous-time reasoning, although their practical utility in formal verification remains largely unexplored.

XII Multimodal Data Streams A natural extension of the previous discussion arises in the context of sensor fusion, where systems may no longer rely on structured state trajectories but instead operate directly on raw, multimodal sensor data. In such cases, partial observability is compounded by the heterogeneous nature of inputs — ranging from image sequences and point clouds to audio and natural language. Note that while most available formal approaches are tailored to Euclidean spaces, kernel-based approaches can be defined to relate any type of object by defining a suitable feature map (Berlinet and Thomas-Agnan, 2004; Schölkopf and Smola, 2002). The signature method underlying the sequential kernels previously discussed presents a mathematically grounded approach to encoding arbitrary types of sequential data in a unified manner (Lee and Oberhauser, 2023). Whether direct operation on these high-dimensional, unstructured inputs can be made computationally tractable in practice, however, remains an open question. This also involves the exploration of suitable hardware to deploy such algorithms efficiently (Ali et al., 2023; Garcia et al., 2019).

XIII Formal Contracts for AI-Enabled Components Shifting from a system- to a component-level perspective, the kernel-based black-box verification framework introduced in Chapter 5 could be employed to establish formal contracts on the input-output behaviour of AI-enabled components, such as perception modules or learning-based controllers. As these components typically operate on sequential, potentially multimodal data, similar sequential kernels discussed before could be investigated.

References

Abate A, Ahmed D, Edwards A, Giacobbe M, Peruffo A (2021) FOSSIL: A software tool for the formal synthesis of Lyapunov functions and barrier certificates using neural networks. In: Proceedings of the 24th International Conference on Hybrid Systems: Computation and Control (HSCC '21), ACM, pp 1–11, DOI 10.1145/3447928.3456646

Abate A, Blom H, Delicaris J, Haesaert S, Hartmanns A, van Huijgevoort BC, Lavaei A, Ma H, Niehage M, Remke A, Schön O, Schupp S, Soudjani S, Willemsen L (2022a) ARCH-COMP22 category report: Stochastic models. EPiC Series in Computing 90:113–141

Abate A, Edwards A, Giacobbe M (2022b) Neural abstractions. In: Proceedings of the 36th International Conference on Neural Information Processing Systems (NeurIPS '22), Curran Associates Inc., pp 26432 – 26447

Abate A, Blom H, Cauchi N, Delicaris J, Haesaert S, van Huijgevoort B, Lavaei A, Remke A, Schön O, Schupp S, Shmarov F, Soudjani S, Willemsen L, Zuliani P (2023) Arch-comp23 category report: Stochastic models. In: Proceedings of 10th International Workshop on Applied, vol 96, pp 126–150

Abate A, Bogomolov S, Edwards A, Potomkin K, Soudjani S, Zuliani P (2024a) Safe reach set computation via neural barrier certificates. IFAC-PapersOnLine 58(11):107–114, DOI 10.1016/j.ifacol.2024.07.433, 8th IFAC Conference on Analysis and Design of Hybrid Systems (ADHS '24)

Abate A, Giacobbe M, Roy D (2024b) Stochastic omega-regular verification and control with supermartingales. In: Gurfinkel A, Ganesh V (eds) Computer Aided Verification (CAV '24), Springer, pp 395–419, DOI 10.1007/978-3-031-65633-0_18

Abate A, Akbarzadeh O, Blom HA, Haesaert S, Hassani S, Lavaei A, Mathiesen FB, Misra R, Nejati A, Niehage M, et al. (2025) ARCH-COMP25 category report: Stochastic models. arXiv preprint DOI 10.48550/arXiv.2506.17602

Adams S, Lahijanian M, Laurenti L (2022) Formal control synthesis for stochastic neural network dynamic models. IEEE Control Systems Letters 6:2858–2863, DOI 10.1109/LCSYS.2022.3178143

Ajeleye D, Lavaei A, Zamani M (2023) Data-driven controller synthesis via finite abstractions with formal guarantees. IEEE Control Systems Letters 7:3453–3458, DOI 10.1109/LCSYS.2023.3331385

Ali T, Bhowmik D, Nicol R (2023) Domain-specific optimisations for image processing on fpgas. Journal of Signal Processing Systems 95(10):1167–1179, DOI 10.1007/s11265-023-01888-2

Althoff M, Frehse G, Girard A (2021) Set propagation techniques for reachability analysis. Annual Review of Control, Robotics, and Autonomous Systems 4(1):369–395

Alur R, Henzinger TA, Lafferriere G, Pappas GJ (2000) Discrete abstractions of hybrid systems. Proceedings of the IEEE 88(7):971–984, DOI 10.1109/5.871304

Alur R, Grosu R, Lee I, Sokolsky O (2001) Compositional refinement for hierarchical hybrid systems. In: Di Benedetto MD, Sangiovanni-Vincentelli A (eds) Hybrid Systems: Computation and Control (HSCC '01), Springer, pp 33–48, DOI 10.1007/3-540-45351-2_7

Amodei D, Olah C, Steinhardt J, Christiano P, Schulman J, Mané D (2016) Concrete problems in AI safety. arXiv preprint DOI 10.48550/arXiv.1606.06565

Anand M, Lavaei A, Zamani M (2024) Compositional synthesis of control barrier certificates for networks of stochastic systems against ω-regular specifications. Nonlinear Analysis: Hybrid Systems 51(101427), DOI 10.1016/j.nahs.2023.101427

Annaswamy AM, Johansson KH, Pappas G (2024) Control for societal-scale challenges: Road map 2030. IEEE Control Systems Magazine 44(3):30–32, DOI 10.1109/MCS.2024.3382376

Arai A (2018) Analysis on Fock Spaces and Mathematical Theory of Quantum Fields: An Introduction to Mathematical Analysis of Quantum Fields. World Scientific, DOI 10.1142/10367

Araujo H, Mousavi MR, Varshosaz M (2023) Testing, validation, and verification of robotic and autonomous systems: A systematic review. ACM Transactions on Software Engineering and Methodology 32(2):1–61, DOI 10.1145/3542945

Arcak M, Meissen C, Packard A (2016) Networks of Dissipative Systems: Compositional Certification of Stability, Performance, and Safety. SpringerBriefs in Electrical and Computer Engineering, Springer, DOI 10.1007/978-3-319-29928-0

Arcari E, Iannelli A, Carron A, Zeilinger MN (2023) Stochastic MPC with robustness to bounded parameteric uncertainty. IEEE Transactions on Automatic Control 68(12):7601–7615, DOI 10.1109/TAC.2023.3294868

Asarin E, Dang T, Girard A (2003) Reachability analysis of nonlinear systems using conservative approximation. In: Maler O, Pnueli A (eds) International Workshop on Hybrid Systems: Computation and Control (HSCC '03), Springer, pp 20–35, DOI 10.1007/3-540-36580-X_5

Åström KJ, Murray R (2021) Feedback Systems: An Introduction for Scientists and Engineers, 2nd edn. Princeton University Press

Aubin-Frankowski PC, Szabó Z (2020) Hard shape-constrained kernel machines. Advances in Neural Information Processing Systems 33:384–395

Axelrod CW (2013) Managing the risks of cyber-physical systems. In: 2013 IEEE Long Island Systems, Applications and Technology Conference (LISAT), IEEE, pp 1–6, DOI 10.1109/LISAT.2013.6578215

Badings T, Romao L, Abate A, Jansen N (2023) Probabilities are not enough: Formal controller synthesis for stochastic dynamical models with epistemic uncertainty. In: Proceedings of the AAAI Conference on Artificial Intelligence, vol 37, pp 14701–14710, DOI 10.1609/aaai.v37i12.26718

Bagnell JA, Farahmand Am (2015) Learning positive functions in a Hilbert space. In: NIPS Workshop on Optimization (OPT2015), vol 20, pp 3240–3255, URL https://www.opt-ml.org/papers/OPT2015_paper_45.pdf

Baier C, Katoen JP (2008) Principles of Model Checking. MIT Press

Banse A, Romao L, Abate A, Jungers RM (2023) Data-driven abstractions via adaptive refinements and a Kantorovich metric. In: 2023 62nd IEEE Conference on Decision and Control (CDC), IEEE, pp 6038–6043, DOI 10.1109/CDC49753.2023.10383513

Barajas NC, Rouchitsas A, Broo DG (2025) Examining human–robot interactions: Design guidelines for trust and acceptance. In: Human-Technology Interaction: Interdisciplinary Approaches and Perspectives, Springer, pp 117–137, DOI 10.1007/978-3-031-78357-9_4

Belta C, Yordanov B, Gol EA (2017) Formal Methods for Discrete-Time Dynamical Systems. Springer, DOI 10.1007/978-3-319-50763-7

Bengio Y (2025) The first international AI safety report: The international scientific report on the safety of advanced AI. SuperIntelligence-Robotics-Safety & Alignment 2(2), DOI 10.70777/si.v2i2.14755

Bengio Y, Cohen MK, Malkin N, MacDermott M, Fornasiere D, Greiner P, Kaddar Y (2024) Can a Bayesian oracle prevent harm from an agent? arXiv preprint DOI 10.48550/arXiv.2408.05284

Berlinet A, Thomas-Agnan C (2004) Reproducing Kernel Hilbert Spaces in Probability and Statistics. Springer, DOI 10.1007/978-1-4419-9096-9

Bestuzheva K, Besançon M, Chen WK, Chmiela A, Donkiewicz T, van Doornmalen J, Eifler L, Gaul O, Gamrath G, Gleixner A, Gottwald L, Graczyk C, Halbig K, Hoen A, Hojny C, van der Hulst R, Koch T, Lübbecke M, Maher SJ, Matter F, Mühmer E, Müller B, Pfetsch ME, Rehfeldt D, Schlein S, Schlösser F, Serrano F, Shinano Y, Sofranac B, Turner M, Vigerske S, Wegscheider F, Wellner P, Weninger D, Witzig J (2021) The SCIP optimization suite 8.0. arXiv preprint DOI 10.48550/arXiv.2112.08872

Bishop CM (2006) Pattern Recognition and Machine Learning. Information Science and Statistics, Springer

Blackmore L, Ono M, Bektassov A, Williams BC (2010) A probabilistic particle-control approximation of chance-constrained stochastic predictive control. IEEE Transactions on Robotics 26(3):502–517, DOI 10.1109/TRO.2010.2044948

Bobiti R, Lazar M (2018) Automated-sampling-based stability verification and doa estimation for nonlinear systems. IEEE Transactions on Automatic Control 63(11):3659–3674, DOI 10.1109/TAC.2018.2797196

Bogachev VI (2007) Measure theory. Springer, DOI 10.1007/978-3-540-34514-5

Bongartz D, Najman J, Sass S, Mitsos A (2018) MAiNGO – McCormick-based algorithm for mixed integer nonlinear global optimization. Tech. rep., RWTH Aachen University

Borkar VS (2012) Probability Theory: An Advanced Course. Universitext, Springer, DOI 10.1007/978-1-4612-0791-7

Boyd S, Vandenberghe L (2004) Convex Optimization. Cambridge University Press, DOI 10.1017/CBO9780511804441

Cao ME, Bloch M, Coogan S (2022) Efficient learning of hyperrectangular invariant sets using Gaussian processes. IEEE Open Journal of Control Systems 1:223–236, DOI 10.1109/OJCSYS.2022.3206083

Casablanca E, Schön O, Zuliani P, Soudjani S (2026) LUCID: Learning-enabled uncertainty-aware certification of stochastic dynamical systems. Accepted to AAAI Conference on Artificial Intelligence 2026

Cauchi N, Abate A (2018) Benchmarks for cyber-physical systems: A modular model library for building automation systems. IFAC-PapersOnLine 51(16):49–54, DOI 10.1016/j.ifacol.2018.08.009

Chekan JA, Langbort C (2023) Safety-aware learning-based control of systems with uncertainty dependent constraints. In: 2023 American Control Conference (ACC), pp 1264–1270, DOI 10.23919/ACC55779.2023.10156490

Chen Y, Li Y, Li S, Yin X (2025) Distributionally robust control synthesis for stochastic systems with safety and reach-avoid specifications. arXiv preprint DOI 10.48550/arXiv.2501.03137

Chew V (1966) Confidence, prediction, and tolerance regions for the multivariate normal distribution. Journal of the American Statistical Association 61(315):605–617, DOI 10.1080/01621459.1966.10480892

Chou Y, Chen X, Sankaranarayanan S (2017) A study of model-order reduction techniques for verification. In: Numerical Software Verification (NSV), Springer, pp 98–113, DOI 10.1007/978-3-319-63501-9_8

Chowdhury SR, Gopalan A (2017) On kernelized multi-armed bandits. In: International Conference on Machine Learning, PMLR, vol 70, pp 844–853

Cifuentes D, Parrilo PA (2017) Sampling algebraic varieties for sum of squares programs. SIAM Journal on Optimization 27(4):2381–2404, DOI 10.1137/15M1052548

Clarke EM, Wing JM (1996) Formal methods: State of the art and future directions. ACM Computing Surveys 28(4):626–643, DOI 10.1145/242223.242257

Cohen MH, Belta C, Tron R (2022a) Robust control barrier functions for nonlinear control systems with uncertainty: A duality-based approach. In: 2022 IEEE 61st Conference on Decision and Control (CDC), IEEE, pp 174–179, DOI 10.1109/CDC51059.2022.9992667

Cohen MK, Daulton S, Osborne MA (2022b) Log-linear-time Gaussian processes using binary tree kernels. Advances in Neural Information Processing Systems 35:8118–8129

Corral-Acero J, Margara F, Marciniak M, Rodero C, Loncaric F, Feng Y, Gilbert A, Fernandes JF, Bukhari HA, Wajdan A, et al. (2020) The 'digital twin' to enable

the vision of precision cardiology. European Heart Journal 41(48):4556–4564, DOI 10.1093/eurheartj/ehaa159

Corso A, Kochenderfer MJ (2020) Interpretable safety validation for autonomous vehicles. In: 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), IEEE, pp 1–6, DOI 10.1109/ITSC45102.2020.9294490

Corso A, Moss R, Koren M, Lee R, Kochenderfer M (2021) A survey of algorithms for black-box safety validation of cyber-physical systems. Journal of Artificial Intelligence Research 72:377–428

Cosner RK, Sadalski I, Woo JK, Culbertson P, Ames AD (2024) Generative modeling of residuals for real-time risk-sensitive safety with discrete-time control barrier functions. In: 2024 IEEE International Conference on Robotics and Automation (ICRA), IEEE, pp i–viii, DOI 10.1109/ICRA57147.2024.10611355

Dalrymple D, Skalse J, Bengio Y, Russell S, Tegmark M, Seshia S, Omohundro S, Szegedy C, Goldhaber B, Ammann N, et al. (2024) Towards guaranteed safe AI: A framework for ensuring robust and reliable AI systems. arXiv preprint DOI 10.48550/arXiv.2405.06624

De Moura L, Bjørner N (2008) Z3: An efficient SMT solver. In: Tools and Algorithms for the Construction and Analysis of Systems (TACAS '08), Springer, pp 337–340, DOI 10.1007/978-3-540-78800-3_24

Der Kiureghian A, Ditlevsen O (2009) Aleatory or epistemic? Does it matter? Structural safety 31(2):105–112, DOI 10.1016/j.strusafe.2008.06.020

Djelassi H, Mitsos A, Stein O (2021) Recent advances in nonconvex semi-infinite programming: Applications and algorithms. EURO Journal on Computational Optimization 9(100006), DOI 10.1016/j.ejco.2021.100006

Eaton R, Katupitiya J, Siew KW, Howarth B (2010) Autonomous farming: Modelling and control of agricultural machinery in a unified framework. International Journal of Intelligent Systems Technologies and Applications 8(1–4):444–457, DOI 10.1504/IJISTA.2010.03022

Edwards A, Giacobbe M, Abate A (2023) On the trade-off between efficiency and precision of neural abstraction. In: International Conference on Quantitative Evaluation of Systems (QEST '23), Springer, pp 152–171, DOI 10.1007/978-3-031-43835-6_12

Edwards A, Peruffo A, Abate A (2024) Fossil 2.0: Formal certificate synthesis for the verification and control of dynamical models. In: Proceedings of the 27th ACM International Conference on Hybrid Systems: Computation and Control (HSCC '24), ACM, DOI 10.1145/3641513.3651398

Elsayed-Aly I, Parker D, Feng L (2024) Distributional probabilistic model checking. In: Benz N, Gopinath D, Shi N (eds) NASA Formal Methods, Springer, pp 57–75, DOI 10.1007/978-3-031-60698-4_4

Emanuilov I (2017) Autonomous systems in aviation: Between product liability and innovation. SESAR Innovation Days 2017 7

Esterle K, Gressenbuch L, Knoll A (2020) Formalizing traffic rules for machine interpretability. In: 2020 IEEE 3rd Connected and Automated Vehicles Symposium (CAV), IEEE, pp 1–7, DOI 10.1109/CAVS51000.2020.9334599

Farahani SS, Majumdar R, Prabhu VS, Soudjani S (2019) Shrinking horizon model predictive control with signal temporal logic constraints under stochastic disturbances. IEEE Transactions on Automatic Control 64(8):3324–3331, DOI 10.1109/TAC.2018.2880651

Feng L, Kwiatkowska M, Parker D (2011) Automated learning of probabilistic assumptions for compositional reasoning. In: Giannakopoulou D, Orejas F (eds) Fundamental Approaches to Software Engineering, Springer, pp 2–17

Fiedler C, Scherer CW, Trimpe S (2021) Practical and rigorous uncertainty bounds for Gaussian process regression. In: 35th AAAI Conference on Artificial Intelligence (AAAI '21), Association for the Advancement of Artificial Intelligence, vol 35, pp 7439–7447, DOI 10.1609/aaai.v35i8.16912

Forejt V, Kwiatkowska M, Norman G, Parker D (2011a) Automated Verification Techniques for Probabilistic Systems, Springer, pp 53–113. DOI 10.1007/978-3-642-21455-4_3

Forejt V, Kwiatkowska M, Norman G, Parker D, Qu H (2011b) Quantitative multi-objective verification for probabilistic systems. In: Abdulla PA, Leino KRM (eds) Tools and Algorithms for the Construction and Analysis of Systems, Springer, pp 112–127, DOI 10.1007/978-3-642-19835-9_11

Gao S, Avigad J, Clarke EM (2012) δ-Complete decision procedures for satisfiability over the reals. In: International Joint Conference on Automated Reasoning (IJCAR '12), Springer, pp 286–300, DOI 10.1007/978-3-642-31365-3_23

Gao S, Kong S, Clarke EM (2013) dReal: An SMT solver for nonlinear theories over the reals. In: Bonacina MP (ed) International Conference on Automated Deduction (CADE '24), Springer, pp 208–214, DOI 10.1007/978-3-642-38574-2_14

Gao Y, Zhou C, Abate A, Johansson KH (2024) Adaptive task planning and formal control synthesis using temporal logic trees. In: Graf S, Pettersson P, Steffen B (eds) Real Time and Such: Essays Dedicated to Wang Yi to Celebrate His Scientific Career, Springer, pp 64–78, DOI 10.1007/978-3-031-73751-0_7

Garatti S, Campi MC (2024) Non-convex scenario optimization. Mathematical Programming 209:557–608, DOI 10.1007/s10107-024-02074-3

Garcia P, Bhowmik D, Stewart R, Michaelson G, Wallace A (2019) Optimized memory allocation and power minimization for FPGA-based image processing. Journal of Imaging 5(1):7, DOI 10.3390/jimaging5010007

Gardner J, Wooding B, Nejati A, Lavaei A (2025) TRUST: Stability and safety controller synthesis for unknown dynamical models using a single trajectory. In: Proceedings of the 28th ACM International Conference on Hybrid Systems: Computation and Control (HSCC '25), ACM, DOI 10.1145/3716863.3718036, URL 10.1145/3716863.3718036

Giacomo GD, Vardi MY (2013) Linear temporal logic and linear dynamic logic on finite traces. 23rd International Joint Conference on Artificial Intelligence (IJCAI '13) pp 854–860

Girard A, Pappas GJ (2009) Hierarchical control system design using approximate simulation. Automatica 45(2):566–571, DOI 10.1016/j.automatica.2008.09.016

Girard A, Pappas GJ (2011) Approximate bisimulation: A bridge between computer science and control theory. European Journal of Control 17(5–6):568–578, DOI 10.3166/ejc.17.568-578

Gracia I, Boskos D, Laurenti L, Mazo Jr M (2023) Distributionally robust strategy synthesis for switched stochastic systems. In: Proceedings of the 26th ACM International Conference on Hybrid Systems: Computation and Control (HSCC '23), 11, pp 1–10, DOI 10.1145/3575870.358712

Gretton A, Borgwardt KM, Rasch MJ, Schölkopf B, Smola A (2012) A kernel two-sample test. Journal of Machine Learning Research 13:723–773

Gugercin S, Antoulas AC (2004) A survey of model reduction by balanced truncation and some new results. International Journal of Control 77(8):748–766, DOI 10.1080/00207170410001713448

Gurobi Optimization, LLC (2024) Gurobi optimizer reference manual. URL https://www.gurobi.com

Haddad S, Monmege B (2018) Interval iteration algorithm for MDPs and IMDPs. Theoretical Computer Science 735:111–131, DOI 10.1016/j.tcs.2016.12.003

Haesaert S, Soudjani S (2021) Robust dynamic programming for temporal logic control of stochastic systems. IEEE Transactions on Automatic Control 66(6):2496–2511, DOI 10.1109/TAC.2020.3010490

Haesaert S, Soudjani S, Abate A (2017) Verification of general Markov decision processes by approximate similarity relations and policy refinement. SIAM Journal on Control and Optimization 55(4):2333–2367, DOI 10.1137/16M1079397

Haesaert S, Nilsson P, Vasile CI, Thakker R, Agha-mohammadi Aa, Ames AD, Murray RM (2018) Temporal logic control of POMDPs via label-based stochastic simulation relations. IFAC-PapersOnLine 51(16):271–276, DOI 10.1016/j.ifacol.2018.08.046

Hahn EM, Hartmanns A, Hermanns H, Katoen JP (2013) A compositional modelling and analysis framework for stochastic hybrid systems. Formal Methods in System Design 43:191–232, DOI 10.1007/s10703-012-0167-z

Hahn EM, Hashemi V, Hermanns H, Lahijanian M, Turrini A (2019) Interval Markov decision processes with multiple objectives: From robust strategies to Pareto curves. ACM Transactions on Modeling and Computer Simulation (TOMACS) 29(27):1–31, DOI 10.1145/3309683

Harapanahalli A, Jafarpour S, Coogan S (2023a) Contraction-guided adaptive partitioning for reachability analysis of neural network controlled systems. In: 2023 62nd IEEE Conference on Decision and Control (CDC), IEEE, pp 6044–6051, DOI 10.1109/CDC49753.2023.10383360

Harapanahalli A, Jafarpour S, Coogan S (2023b) Forward invariance in neural network controlled systems. IEEE Control Systems Letters 7:3962–3967, DOI 10.1109/LCSYS.2023.3341980

Harapanahalli A, Jafarpour S, Coogan S (2024) immrax: A parallelizable and differentiable toolbox for interval analysis and mixed monotone reachability in JAX. IFAC-PapersOnLine 58(11):75–80, DOI 10.1016/j.ifacol.2024.07.428

Hasanbeig M, Abate A, Kroening D (2019) Logically-constrained neural fitted Q-iteration. International Foundation for Autonomous Agents and Multiagent Systems (AAMAS '19), pp 2012–2014

Henriksen JG, Jensen J, Jørgensen M, Klarlund N, Paige R, Rauhe T, Sandholm A (1995) Mona: Monadic second-order logic in practice. In: Brinksma E, Cleaveland WR, Larsen KG, Margaria T, Steffen B (eds) Tools and Algorithms for the Construction and Analysis of Systems (TACAS '95), Springer, vol 1019, pp 89–110, DOI 10.1007/3-540-60630-0_5

Hewing L, Menner M, Tachatos N, Daners MS, Du Pasquier C, Lumpe TS, Shea K, Carron A, Zeilinger MN (2021) Volume control of low-cost ventilator with automatic set-point adaptation. In: 2021 European Control Conference (ECC), IEEE, pp 781–786, DOI 10.23919/ECC54610.2021.9655092

Himmelberg C (1975) Measurable relations. Fundamenta Mathematicae 87:53–72, DOI 10.4064/fm-87-1-53-72

Horst R, Tuy H (2013) Global Optimization: Deterministic Approaches. Springer, DOI 10.1007/978-3-662-03199-5

Hsu D, Kakade SM, Zhang T (2012) A tail inequality for quadratic forms of subgaussian random vectors. Electronic Communications in Probability 17:1–6, DOI 10.1214/ECP.v17-2079

Huang X, Kwiatkowska M, Wang S, Wu M (2017) Safety verification of deep neural networks. In: Majumdar R, Kunčak V (eds) Computer Aided Verification, Springer, pp 3–29, DOI 10.1007/978-3-319-63387-9_1

van Huijgevoort BC, Haesaert S (2022) Similarity quantification for linear stochastic systems: A coupling compensator approach. Automatica 144:110476, DOI 10.1016/j.automatica.2022.110476

Hüllermeier E, Waegeman W (2021) Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. Machine Learning 110:457–506, DOI 10.1007/s10994-021-05946-3

Ingham JF, Hutchings F, Zuliani P, Wang Y, Soudjani S, Taylor PN (2024) Vagus nerve stimulation: Laying the groundwork for predictive network-based computer models. arXiv preprint DOI 10.48550/arXiv.2406.02729

ISO 21448 (2022) ISO 21448:2022: Road vehicles — Safety of the intended functionality. International Organization for Standardization

ISO 34502:2022 (2022) Road vehicles — Test scenarios for automated driving systems — Scenario based safety evaluation framework. International Organization for Standardization

Jackson J, Laurenti L, Frew E, Lahijanian M (2020) Safety verification of unknown dynamical systems via Gaussian process regression. In: 2020 59th IEEE Conference on Decision and Control (CDC), IEEE, pp 860–866, DOI 10.1109/CDC42340.2020.9303814

Jafarpour S, Harapanahalli A, Coogan S (2023) Interval reachability of nonlinear dynamical systems with neural network controllers. In: Learning for Dynamics and Control Conference, PMLR, vol 211, pp 12–25

Jafarpour S, Liu Z, Chen Y (2025) Probabilistic reachability analysis of stochastic control systems. IEEE Transactions on Automatic Control pp 1–15, DOI 10.1109/TAC.2025.3566983

Jagtap P, Pappas GJ, Zamani M (2020a) Control barrier functions for unknown nonlinear systems using Gaussian processes. In: 2020 59th IEEE Conference on Decision and Control (CDC), IEEE, pp 3699–3704, DOI 10.1109/CDC42340.2020.9303847

Jagtap P, Soudjani S, Zamani M (2020b) Formal synthesis of stochastic systems via control barrier certificates. IEEE Transactions on Automatic Control 66(7):3097–3110, DOI 10.1109/TAC.2020.3013916

Jaynes ET, Kempthorne O (1976) Confidence intervals vs bayesian intervals. In: Harper WL, Hooker CA (eds) Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science, Springer, pp 175–257, DOI 10.1007/978-94-010-1436-6_6

Jiang J, Zhao Y, Coogan S (2022) Safe learning for uncertainty-aware planning via interval MDP abstraction. IEEE Control Systems Letters 6:2641–2646, DOI 10.1109/LCSYS.2022.3173993

Jobin A, Ienca M, Vayena E (2019) The global landscape of AI ethics guidelines. Nature Machine Intelligence 1(9):389–399, DOI 10.1038/s42256-019-0088-2

Kahneman D (2011) Thinking, Fast and Slow

Kalra N, Paddock SM (2016) Driving to safety: How many miles of driving would it take to demonstrate autonomous vehicle reliability? Transportation Research Part A: Policy and Practice 94:182–193, DOI doi.org/10.1016/j.tra.2016.09.010

Kamgarpour M, Summers S, Lygeros J (2013) Control design for specifications on stochastic hybrid systems. In: Proceedings of the 16th International Conference on Hybrid Systems: Computation and Control (HSCC '13), pp 303–312, DOI 10.1145/2461328.246137

Kanagawa M, Hennig P, Sejdinovic D, Sriperumbudur BK (2018) Gaussian processes and kernel methods: A review on connections and equivalences. DOI 10.48550/arXiv.1807.02582

Katoen JP, Zapreev IS, Hahn EM, Hermanns H, Jansen DN (2011) The ins and outs of the probabilistic model checker MRMC. Performance Evaluation 68(2):90–104, DOI 10.1016/j.peva.2010.04.001, advances in Quantitative Evaluation of Systems

Kattenbelt M, Kwiatkowska M, Norman G, Parker D (2010) A game-based abstraction-refinement framework for Markov decision processes. Formal Methods in System Design 36:246–280, DOI 10.1007/s10703-010-0097-6

Kazemi M, Soudjani S (2020) Formal policy synthesis for continuous-state systems via reinforcement learning. In: Integrated Formal Methods: 16th International Conference (IFM '20), Springer, pp 3–21, DOI 10.1007/978-3-030-63461-2_

Kazemi M, Majumdar R, Salamati M, Soudjani S, Wooding B (2024a) Data-driven abstraction-based control synthesis. Nonlinear Analysis: Hybrid Systems 52:101467, DOI 10.1016/j.nahs.2024.101467

Kazemi M, Perez M, Somenzi F, Soudjani S, Trivedi A, Velasquez A (2024b) Assume-guarantee reinforcement learning. In: Proceedings of the AAAI Con-

ference on Artificial Intelligence, vol 38, pp 21223–21231, DOI 10.1609/aaai.v38i19.30116

Keesman KJ (2011) System Identification: An Introduction. Springer, DOI 10.1007/978-0-85729-522-4

Khalil HK, Grizzle JW (2002) Nonlinear Systems, 3rd edn. Prentice Hall

Kim KKK, Shen DE, Nagy ZK, Braatz RD (2013) Wiener's polynomial chaos for the analysis and control of nonlinear dynamical systems with probabilistic uncertainties [historical perspectives]. IEEE Control Systems Magazine 33(5):58–67, DOI 10.1109/MCS.2013.2270410

Király FJ, Oberhauser H (2019) Kernels for sequentially ordered data. Journal of Machine Learning Research 20(31):1–45

Kirchhof M, Kasneci G, Kasneci E (2025) Reexamining the aleatoric and epistemic uncertainty dichotomy. In: The Fourth Blogpost Track at ICLR 2025

Klebanov I, Schuster I, Sullivan TJ (2020) A rigorous theory of conditional mean embeddings. SIAM Journal on Mathematics of Data Science 2(3):583–606, DOI 10.1137/19M1305069

Kochenderfer MJ, Holland JE, Chryssanthacopoulos JP (2012) Next-generation airborne collision avoidance system. Lincoln Laboratory Journal 19(1):17–33

Kuchar J, Drumm AC (2007) The traffic alert and collision avoidance system. Lincoln laboratory journal 16(2):277

Kupferman O, Vardi MY (2001) Model checking of safety properties. Formal Methods in System Design 19:291–314, DOI 10.1023/A:1011254632723

Kushner HJ (1967) Stochastic Stability and Control. Academic Press

Kwiatkowska M, Norman G, Parker D (2002) PRISM: Probabilistic symbolic model checker. In: Field T, Harrison PG, Bradley J, Harder U (eds) Computer Performance Evaluation: Modelling Techniques and Tools, Springer, pp 200–204, DOI 10.1007/3-540-46029-2_13

Kwiatkowska M, Norman G, Parker D (2007) Stochastic Model Checking, Springer, pp 220–270. DOI 10.1007/978-3-540-72522-0_6

Lacerda B, Parker D, Hawes N (2014) Optimal and dynamic planning for Markov decision processes with co-safe LTL specifications. In: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, pp 1511–1516, DOI 10.1109/IROS.2014.6942756

Lasserre JB (2001) Global optimization with polynomials and the problem of moments. SIAM Journal on optimization 11(3):796–817, DOI 10.1137/S1052623400366802

Laurenti L, Lahijanian M (2023) Unifying safety approaches for stochastic systems: From barrier functions to uncertain abstractions via dynamic programming. arXiv preprint DOI 10.48550/arXiv.2310.01802

Lavaei A, Somenzi F, Soudjani S, Trivedi A, Zamani M (2020a) Formal controller synthesis for continuous-space MDPs via model-free reinforcement learning. In: 2020 ACM/IEEE 11th International Conference on Cyber-Physical Systems (ICCPS), pp 98–107, DOI 10.1109/ICCPS48487.2020.00017

Lavaei A, Soudjani S, Zamani M (2020b) Compositional abstraction-based synthesis for networks of stochastic switched systems. Automatica 114(108827), DOI 10.1016/j.automatica.2020.108827

Lavaei A, Soudjani S, Zamani M (2021) Compositional abstraction-based synthesis of general MDPs via approximate probabilistic relations. Nonlinear Analysis: Hybrid Systems 39(100991), DOI 10.1016/j.nahs.2020.100991

Lavaei A, Soudjani S, Abate A, Zamani M (2022a) Automated verification and synthesis of stochastic hybrid systems: A survey. Automatica 146:110617, DOI 10.1016/j.automatica.2022.110617

Lavaei A, Soudjani S, Frazzoli E, Zamani M (2022b) Constructing MDP abstractions using data with formal guarantees. IEEE Control Systems Letters 7:460–465, DOI 10.1109/LCSYS.2022.3188535

Lavaei A, Perez M, Kazemi M, Somenzi F, Soudjani S, Trivedi A, Zamani M (2023) Compositional reinforcement learning for discrete-time stochastic control systems. IEEE Open Journal of Control Systems 2:425–438, DOI 10.1109/OJCSYS.2023.3329394

Lee D, Oberhauser H (2023) The signature kernel. arXiv preprint DOI 10.48550/arXiv.2305.04625

Lew T, Pavone M (2021) Sampling-based reachability analysis: A random set theory approach with adversarial sampling. In: Conference on Robot Learning, PMLR, vol 155, pp 2055–2070

Li Z, Meunier D, Mollenhauer M, Gretton A (2022) Optimal rates for regularized conditional mean embedding learning. Advances in Neural Information Processing Systems 35:4433–4445

Lindemann L, Jiang L, Matni N, Pappas GJ (2023) Risk of stochastic systems for temporal logic specifications. ACM Transactions on Embedded Computing Systems 22(3):1–31, DOI 10.1145/3580490

Liu C, Arnon T, Lazarus C, Strong C, Barrett C, Kochenderfer MJ (2021) Algorithms for verifying deep neural networks. Foundations and Trends® in Optimization 4(3–4):244–404, DOI 10.1561/2400000035

Lofberg J, Parrilo PA (2004) From coefficients to samples: A new approach to SOS optimization. In: Gramlich B, Miller D, Sattler U (eds) 2004 43rd IEEE Conference on Decision and Control (CDC), IEEE, vol 3, pp 3154–3159, DOI 10.1109/CDC.2004.1428957

Lopez BT, Slotine JJE (2023) Unmatched control barrier functions: Certainty equivalence adaptive safety. In: 2023 American Control Conference (ACC), IEEE, pp 3662–3668, DOI 10.23919/ACC55779.2023.10156496

Luckcuck M (2023) Using formal methods for autonomous systems: Five recipes for formal verification. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 237(2):278–292, DOI 10.1177/1748006X21103497

Luckcuck M, Farrell M, Dennis LA, Dixon C, Fisher M (2019) Formal specification and verification of autonomous robotic systems: A survey. ACM Computing Surveys 52(5):1–41, DOI 10.1145/3342355

Majumdar R, Mallik K, Rychlicki M, Schmuck AK, Soudjani S (2023a) A flexible toolchain for symbolic Rabin games under fair and stochastic uncertainties. In: Enea C, Lal A (eds) International Conference on Computer Aided Verification (CAV '23), Springer, pp 3–15, DOI 10.1007/978-3-031-37709-9_1

Majumdar R, Salamati M, Soudjani S (2023b) Neural abstraction-based controller synthesis and deployment. ACM Transactions on Embedded Computing Systems 22(5s):1–25, DOI 10.1145/3608104

Maler O, Nickovic D (2004) Monitoring temporal properties of continuous signals. In: International Symposium on Formal Techniques in Real-Time and Fault-Tolerant Systems, Springer, pp 152–166, DOI 10.1007/978-3-540-30206-3_12

Mallik K, Schmuck AK, Soudjani S, Majumdar R (2018) Compositional synthesis of finite-state abstractions. IEEE Transactions on Automatic Control 64(6):2629–2636, DOI 10.1109/TAC.2018.2869740

Marteau-Ferey U, Bach F, Rudi A (2020) Non-parametric models for non-negative functions. Advances in Neural Information Processing Systems 33:12816–12826

Mathiesen FB, Lahijanian M, Laurenti L (2024) IntervalMDP.jl: Accelerated value iteration for interval Markov decision processes. IFAC-PapersOnLine 58(11):1–6, DOI 10.1016/j.ifacol.2024.07.416, 8th IFAC Conference on Analysis and Design of Hybrid Systems (ADHS '24)

Mathiesen FB, Haesaert S, Laurenti L (2025) Scalable control synthesis for stochastic systems via structural IMDP abstractions. In: Proceedings of the 28th ACM International Conference on Hybrid Systems: Computation and Control (HSCC '25), 14, pp 1–12, DOI 10.1145/3716863.3718031

Matni N, Ames AD, Doyle JC (2024) A quantitative framework for layered multirate control: Toward a theory of control architecture. IEEE Control Systems Magazine 44(3):52–94, DOI 10.1109/MCS.2024.3382388

Mauroy A, Susuki Y, Mezic I (2020) The Koopman Operator in Systems and Control: Concepts, Methodologies, and Applications. Springer, DOI 10.1007/978-3-030-35713-9

McLachlan G, Peel D (2000) Finite Mixture Models. Wiley, DOI 10.1002/0471721182

Mehdipour N, Althoff M, Tebbens RD, Belta C (2023) Formal methods to comply with rules of the road in autonomous driving: State of the art and grand challenges. Automatica 152:110692, DOI 10.1016/j.automatica.2022.110692

Miner A, Parker D (2004) Symbolic Representations and Analysis of Large Probabilistic Systems, Springer, pp 296–338. DOI 10.1007/978-3-540-24611-4_9

Mollenhauer M, Mücke N, Sullivan T (2022) Learning linear operators: Infinite-dimensional regression as a well-behaved non-compact inverse problem. DOI 10.48550/arXiv.2211.08875

Monir N, Schön O, Soudjani S (2024) Lyapunov-based policy synthesis for multi-objective interval mdps. IFAC-PapersOnLine 58(11):99–106, DOI 10.1016/j.ifacol.2024.07.432

Muandet K, Fukumizu K, Sriperumbudur B, Schölkopf B (2017) Kernel mean embedding of distributions: A review and beyond. Foundations and Trends® in Machine Learning 10(1–2):1–141, DOI 10.1561/2200000060

Nazeri M, Badings T, Soudjani S, Abate A (2025) Data-driven yet formal policy synthesis for stochastic nonlinear dynamical systems. DOI 10.48550/arXiv.2501.01191

Nemmour Y, Kremer H, Schölkopf B, Zhu JJ (2022) Maximum mean discrepancy distributionally robust nonlinear chance-constrained optimization with finite-sample guarantee. In: 2022 IEEE 61st Conference on Decision and Control (CDC), pp 5660–5667, DOI 10.1109/CDC51059.2022.9993212

NHTSA (2022) Automated vehicles for safety. URL https://www.nhtsa.gov/vehicle-safety/automated-vehicles-safety, National Highway Traffic Safety Administration

NHTSA (2025) Standing general order on crash reporting for incidents involving ADS and Level 2 ADAS. URL https://www.nhtsa.gov/laws-regulations/standing-general-order-crash-reporting, National Highway Traffic Safety Administration

Norman G, Parker D, Zou X (2017) Verification and control of partially observable probabilistic systems. Real-Time Systems 53:354–402, DOI 10.1007/s11241-017-9269-4

Orwat C, Graefe A, Faulwasser T (2008) Towards pervasive computing in health care – A literature review. BMC Medical Informatics and Decision Making 8(26):1–18, DOI 10.1186/1472-6947-8-26

Pan G, Faulwasser T (2023) Distributionally robust uncertainty quantification via data-driven stochastic optimal control. IEEE Control Systems Letters 7:3036–3041, DOI 10.1109/LCSYS.2023.3290362

Papoulis A (1965) Random Variables and Stochastic Processes. McGraw Hill

Park J, Muandet K (2020) A measure-theoretic approach to kernel conditional mean embeddings. Advances in Neural Information Processing Systems 33:21247–21259

Park S, Serpedin E, Qaraqe K (2012) Gaussian assumption: The least favorable but the most useful [lecture notes]. IEEE Signal Processing Magazine 30, DOI 10.1109/MSP.2013.2238691

Parrilo PA (2003) Semidefinite programming relaxations for semialgebraic problems. Mathematical Programming 96:293–320, DOI 10.1007/s10107-003-0387-5

Penders J, Wang J, Bhowmik D, Di Nuovo A, Soranzo A, Rolph J, Popa I, Varrasi S, Lucas A, Conti D, Young J, Alboul L, Collins E, Aitken J, Law J, Cameron D (2018) Robots claiming space: Gauging public reaction using computer vision techniques. In: Towards Autonomous Robotic Systems (TAROS '18), Springer, vol 10965

Petersen DP, Middleton D (1962) Sampling and reconstruction of wave-number-limited functions in N-dimensional Euclidean spaces. Information and Control 5(4):279–323, DOI 10.1016/S0019-9958(62)90633-2

Pfeifer R, Iida F (2004) Embodied artificial intelligence: Trends and challenges. Embodied Artificial Intelligence: International Seminar pp 1–26, DOI 10.1007/978-3-540-27833-7_1

Pfister L, Bresler Y (2018) Bounding multivariate trigonometric polynomials. IEEE Transactions on Signal Processing 67(3):700–707, DOI 10.1109/TSP.2018.2883925

Prajna S, Papachristodoulou A, Seiler P, Parrilo PA (2005) SOSTOOLS and its Control Applications, Springer, pp 273–292. DOI 10.1007/10997703_14

Prajna S, Jadbabaie A, Pappas GJ (2007) A framework for worst-case and stochastic safety verification using barrier certificates. IEEE Transactions on Automatic Control 52(8):1415–1428, DOI 10.1109/TAC.2007.902736

Provan G (2024) Formal methods for autonomous vehicles. IT Professional 26(1):50–56, DOI 10.1109/MITP.2024.3356158

Pulch R (2021) 10 Model order reduction in uncertainty quantification, De Gruyter, pp 321–344. DOI 10.1515/9783110499001-010

Puterman ML (2014) Markov Decision Processes: Discrete Stochastic Dynamic Programming. John Wiley & Sons

Rahimi A, Recht B (2007) Random features for large-scale kernel machines. Advances in Neural Information Processing Systems 20

Ramos MA, Thieme CA, Utne IB, Mosleh A (2020) Human-system concurrent task analysis for maritime autonomous surface ship operation and safety. Reliability Engineering & System Safety 195:106697, DOI 10.1016/j.ress.2019.106697

Rasmussen CE, Williams CK (2006) Gaussian Processes for Machine Learning. Springer

Reed R, Laurenti L, Lahijanian M (2023) Promises of deep kernel learning for control synthesis. IEEE Control Systems Letters 7:3986–3991, DOI 10.1109/LCSYS.2023.3340995

Reis J, Cohen Y, Melão N, Costa J, Jorge D (2021) High-tech defense industries: developing autonomous intelligent systems. Applied Sciences 11(11):4920, DOI 10.3390/app11114920

Rizaldi A, Keinholz J, Huber M, Feldle J, Immler F, Althoff M, Hilgendorf E, Nipkow T (2017) Formalising and monitoring traffic rules for autonomous vehicles in Isabelle/HOL. In: Polikarpova N, Schneider S (eds) Integrated Formal Methods, Springer, pp 50–66, DOI 10.1007/978-3-319-66845-1_4

Romao L, Hota AR, Abate A (2023) Distributionally robust optimal and safe control of stochastic systems via kernel conditional mean embedding. In: 2023 62nd IEEE Conference on Decision and Control (CDC), pp 2016–2021, DOI 10.1109/CDC49753.2023.10383997

Sahinidis NV (1996) BARON: A general purpose global optimization software package. Journal of Global Optimization 8:201–205, DOI 10.1007/BF00138693

Salamati A, Lavaei A, Soudjani S, Zamani M (2024) Data-driven verification and synthesis of stochastic systems via barrier certificates. Automatica 159(C):111323, DOI 10.1016/j.automatica.2023.111323

Schnitzer Y, Abate A, Parker D (2024) Learning provably robust policies in uncertain parametric environments. arXiv preprints DOI 10.48550/arXiv.2408.03093

Schölkopf B, Smola AJ (2002) Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press

Schön O, Haesaert S, Soudjani S (2025) Formal control for uncertain systems via contract-based probabilistic surrogates. In: Prabhakar P, Vandin A (eds) Quantitative Evaluation of Systems and Formal Modeling and Analysis of Timed Systems (QEST), Springer, pp 62–82, DOI 10.1007/978-3-032-05792-1_4

Schuster I, Mollenhauer M, Klus S, Muandet K (2020) Kernel conditional density operators. In: International Conference on Artificial Intelligence and Statistics, PMLR, vol 108, pp 993–1004

Schweidtmann AM, Bongartz D, Grothe D, Kerkenhoff T, Lin X, Najman J, Mitsos A (2021) Deterministic global optimization with Gaussian processes embedded. Mathematical Programming Computation 13(3):553–581, DOI 10.1007/s12532-021-00204-y

Schön O, Zhong Z, Soudjani S (2024) Data-driven distributionally robust safety verification using barrier certificates and conditional mean embeddings. In: 2024 American Control Conference (ACC), pp 3417–3423, DOI 10.23919/ACC60939.2024.10644349

Segala R, Lynch N (1995) Probabilistic simulations for probabilistic processes. Nordic Journal of Computing 2(2):250–273

Seshadri P, Duncan A, Scillitoe A (2020) Bayesian polynomial chaos. In: Conference on Neural Information Processing Systems, URL https://ml4eng.github.io/camera_readys/33.pdf

Slotine JJE, Li W (1991) Applied Nonlinear Control. Prentice Hall

Smith RC (2024) Uncertainty Quantification: Theory, Implementation, and Applications. SIAM

Smola A, Gretton A, Song L, Schölkopf B (2007) A Hilbert space embedding for distributions. In: Hutter M, Servedio RA, Takimoto E (eds) Algorithmic Learning Theory, Springer, pp 13–31, DOI 10.1007/978-3-540-75225-7_5

So O, Serlin Z, Mann M, Gonzales J, Rutledge K, Roy N, Fan C (2024) How to train your neural control barrier function: Learning safety filters for complex input-constrained systems. In: 2024 IEEE International Conference on Robotics and Automation (ICRA), pp 11532–11539, DOI 10.1109/ICRA57147.2024.10610418

Song L, Zhang X, Smola A, Gretton A, Schölkopf B (2008) Tailoring density estimation via reproducing kernel moment matching. In: Proceedings of the 25th International Conference on Machine Learning (ICML '08), ACM, pp 992–999, DOI 10.1145/1390156.1390281

Song L, Huang J, Smola A, Fukumizu K (2009) Hilbert space embeddings of conditional distributions with applications to dynamical systems. In: Proceedings of the 26th Annual International Conference on Machine Learning (ICML '09), ACM, pp 961–968, DOI 10.1145/1553374.1553497

Soudjani S, Abate A (2013) Adaptive and sequential gridding procedures for the abstraction and verification of stochastic processes. SIAM Journal on Applied Dynamical Systems 12(2):921–956, DOI 10.1137/120871456

Steinwart I, Christmann A (2008) Support Vector Machines. Springer

Streett RS (1981) Propositional dynamic logic of looping and converse. In: Proceedings of the Thirteenth Annual ACM Symposium on Theory of Computing (STOC '81), pp 375–383, DOI 10.1145/800076.802492

Suilen M, Badings T, Bovy EM, Parker D, Jansen N (2025) Robust Markov Decision Processes: A Place Where AI and Formal Methods Meet, Springer, pp 126–154. DOI 10.1007/978-3-031-75778-5_7

Sun S, Zhang C, Yu G (2006) A Bayesian network approach to traffic flow forecasting. IEEE Transactions on Intelligent Transportation Systems 7(1):124–132, DOI 10.1109/TITS.2006.869623

Sutton RS, Barto AG (2018) Reinforcement Learning: An Introduction. MIT Press

Tabuada P (2009) Verification and Control of Hybrid Systems: A Symbolic Approach. Springer

Thimbleby H (2021) Fix IT: See and solve the problems of digital healthcare. Oxford University Press

Thorpe A, Lew T, Oishi M, Pavone M (2022) Data-driven chance constrained control using kernel distribution embeddings. In: Learning for Dynamics and Control Conference, PMLR, pp 790–802

Titsias M (2009) Variational learning of inducing variables in sparse Gaussian processes. In: Proceedings of the 12th International Conference on Artificial Intelligence and Statistics, PMLR, vol 5, pp 567–574

Tjoa E, Guan C (2020) A survey on explainable artificial intelligence (XAI): Toward medical XAI. IEEE Transactions on Neural Networks and Learning Systems 32(11):4793–4813, DOI 10.1109/TNNLS.2020.3027314

Toth C, Oberhauser H (2020) Bayesian learning from sequential data using gaussian processes with signature covariances. In: International Conference on Machine Learning, PMLR, vol 119, pp 9548–9560

Tuncali CE, Fainekos G, Prokhorov D, Ito H, Kapinski J (2019) Requirements-driven test generation for autonomous vehicles with machine learning components. IEEE Transactions on Intelligent Vehicles 5(2):265–280, DOI 10.1109/TIV.2019.2955903

Van Den Hof PM, Schrama RJ (1995) Identification and control – Closed-loop issues. Automatica 31(12):1751–1770, DOI 10.1016/0005-1098(95)00094-X

Van Huijgevoort B, Schön O, Soudjani S, Haesaert S (2023) SySCoRe: Synthesis via stochastic coupling relations. In: Proceedings of the 26th ACM International Conference on Hybrid Systems: Computation and Control (HSCC '23), 13, pp 1–11, DOI 10.1145/3575870.3587123

Vardi MY (1985) Automatic verification of probabilistic concurrent finite state programs. In: 26th Annual Symposium on Foundations of Computer Science (SFCS '19'85), IEEE, pp 327–338, DOI 10.1109/SFCS.1985.12

Veale M, Zuiderveen Borgesius F (2021) Demystifying the draft EU Artificial Intelligence Act – Analysing the good, the bad, and the unclear elements of the proposed approach. Computer Law Review International 22(4):97–112, DOI 10.9785/cri-2021-220402

Wajid R, Awan AU, Zamani M (2022) Formal synthesis of safety controllers for unknown stochastic control systems using Gaussian process learning. In: Learning for Dynamics and Control Conference, PMLR, pp 624–636

Wang C, Meng Y, Liu J, Smith S (2023) Stochastic control barrier functions with Bayesian inference for unknown stochastic differential equations. DOI 10.48550/arXiv.2312.12759

Wang L, Theodorou EA, Egerstedt M (2018) Safe learning of quadrotor dynamics using barrier certificates. In: 2018 IEEE International Conference on Robotics and Automation (ICRA), IEEE, pp 2460–2465, DOI 10.1109/ICRA.2018.8460471

Wang R, Sun Z, Haesaert S (2025) Unraveling tensor structures in correct-by-design controller synthesis. arXiv preprint DOI 10.48550/arXiv.2503.24085

Wang X, Knoedler L, Mathiesen FB, Alonso-Mora J (2024) Simultaneous synthesis and verification of neural control barrier functions through branch-and-bound verification-in-the-loop training. In: 2024 European Control Conference (ECC), IEEE, pp 571–578, DOI 10.23919/ECC64448.2024.10591251

Wells AM, Lahijanian M, Kavraki LE, Vardi MY (2020) LTLf synthesis on probabilistic systems. In: Electronic Proceedings in Theoretical Computer Science, EPTCS, vol 326, pp 166–181, DOI 10.4204/EPTCS.326.11

Wongpiromsarn T, Ghasemi M, Cubuktepe M, Bakirtzis G, Carr S, Karabag MO, Neary C, Gohari P, Topcu U (2023) Formal methods for autonomous systems. arXiv preprint DOI 10.1561/2600000029

Wooding B, Lavaei A (2024) IMPaCT: Interval MDP parallel construction for controller synthesis of large-scale stochastic systems. In: Hillston J, Soudjani S, Waga M (eds) Quantitative Evaluation of Systems and Formal Modeling and Analysis of Timed Systems (QEST '24), Springer, pp 249–267, DOI 10.1007/978-3-031-68416-6_15

Wooding B, Vahidinasab V, Soudjani S (2021) Operation and control of a population of active buildings at network level. In: Vahidinasab V, Mohammadi-Ivatloo B (eds) Active Building Energy Systems: Operation and Control, Springer, pp 193–218, DOI doi.org/10.1007/978-3-030-79742-3_8

Yin X, Gao B, Yu X (2024) Formal synthesis of controllers for safety-critical autonomous systems: Developments and challenges. Annual Reviews in Control 57:100940, DOI 10.1016/j.arcontrol.2024.100940

Zhang Z, Ma C, Soudijani S, Soudjani S (2024) Formal verification of unknown stochastic systems via non-parametric estimation. In: International Conference on Artificial Intelligence and Statistics, PMLR, vol 238, pp 3277–3285

Appendix A
Additional Proofs of Statements

A.1 Proof of Theorem 2.3

Proof. It is first established that (2.21) defines a sub-probability coupling between $\hat{\mathbf{t}}(\,\cdot\,|\hat{x},\hat{u})$ in (2.15) and $\mathbf{t}_\theta(\,\cdot\,|x,\hat{u})$ in (2.16) over $\mathcal{R}$ in (2.20). Verification of the conditions of Definition 2.8 is carried out and δ is derived. Subsequently, it is shown that (2.18) defines a valid control refinement. To simplify the notation, define the tuple $z := (\hat{x}, x, \hat{u})$, where $z \in \mathbb{Z}$ with $\mathbb{Z} := \{(\hat{x}, x, \hat{u}) | (\hat{x}, x) \in \mathcal{R} \text{ and } \hat{u} \in \hat{\mathbb{U}}\}$.

Definition 2.8(a)

It is shown that for all $z \in \mathbb{Z}$ and $\theta \in \Theta$, $\boldsymbol{\nu}_\theta$ is fully supported on $\mathcal{R}$, i.e., $\boldsymbol{\nu}_\theta(\hat{\mathbb{X}}\times\mathbb{X}|z) = \boldsymbol{\nu}_\theta(\mathcal{R}|z)$. First, $\boldsymbol{\nu}_\theta$ is integrated over $\hat{\mathbb{X}} \times \mathbb{X}$. From (2.21), it follows for all $z \in \mathbb{Z}$ and $\theta \in \Theta$ that

$$\begin{aligned}
\boldsymbol{\nu}_\theta(\hat{\mathbb{X}} \times \mathbb{X}|z) &= \int_{\hat{x}_+\in\hat{\mathbb{X}}} \int_{x_+\in\mathbb{X}} \int_{\hat{w}\in\hat{\mathbb{W}}} \int_{w\in\mathbb{W}} \delta_{\hat{\theta}^\top f(\hat{x},\hat{u})+\hat{w}}(d\hat{x}_+)\delta_{\gamma_\theta+w}(d\hat{w}) \\
&\qquad \times \delta_{\theta^\top f(x,u)+w}(dx_+)\, \min\{\mathcal{N}(dw|0,\Sigma), \mathcal{N}(dw|-\gamma_\theta,\Sigma)\}, \\
&= \int_{\hat{x}_+\in\hat{\mathbb{X}}} \int_{x_+\in\mathbb{X}} \int_{w\in\mathbb{W}} \delta_{x_+}(d\hat{x}_+)\delta_{\theta^\top f(x,u)+w}(dx_+) \\
&\qquad \times \min\{\mathcal{N}(dw|0,\Sigma), \mathcal{N}(dw|-\gamma_\theta,\Sigma)\},
\end{aligned} \tag{A.1}$$

where (2.20), (2.17), and the interface function $u \equiv \hat{u}$ have been used in the last step.

Next, (2.21) is integrated over $\mathcal{R}$ to obtain that for all $z \in \mathbb{Z}$ and $\theta \in \Theta$

$$\begin{aligned}\boldsymbol{v}_\theta(\mathcal{R}|z) &= \iiint_{\substack{w\in\mathbb{W}\\(\hat{x}_+,x_+)\in\mathcal{R}}} \delta_{x_+}(d\hat{x}_+)\delta_{\theta^\top f(x,u)+w}(dx_+)\min\{\mathcal{N}(dw|0,\Sigma),\mathcal{N}(dw|-\gamma_\theta,\Sigma)\},\\ &= \iiint_{\substack{w\in\mathbb{W}\\(\hat{x}_+,x_+)\in\hat{\mathbb{X}}\times\mathbb{X}}} \delta_{x_+}(d\hat{x}_+)\delta_{\theta^\top f(x,u)+w}(dx_+)\min\{\mathcal{N}(dw|0,\Sigma),\mathcal{N}(dw|-\gamma_\theta,\Sigma)\},\end{aligned}$$

where the choice of relation $\mathcal{R} := \{(\hat{x},x)\in\hat{\mathbb{X}}\times\mathbb{X}|x=\hat{x}\}$, interface function $u\equiv\hat{u}$, and offset $\gamma_\theta(x,u)=(\theta-\hat{\theta})^\top f(x,u)$ ensures that $(\hat{x}_+,x_+)\in\mathcal{R}$ for $(\hat{x},x)\in\mathcal{R}$. With (A.1), it follows that $\boldsymbol{v}_\theta(\mathcal{R}|z)=\boldsymbol{v}_\theta(\hat{\mathbb{X}}\times\mathbb{X}|z)$, verifying condition Definition 2.8(a).

Definition 2.8(b)

For measurable $S\subset\hat{\mathbb{X}}$, integration of $\boldsymbol{v}_\theta$ over $S\times\mathbb{X}$ yields

$$\begin{aligned}\boldsymbol{v}_\theta(S\times\mathbb{X}|z) &\le \int_{w\in\mathbb{W}}\int_{\hat{x}_+\in S}\delta_{\hat{\theta}^\top f(\hat{x},\hat{u})+w}(d\hat{x}_+)\min\{\mathcal{N}(dw|\gamma_\theta,\Sigma),\mathcal{N}(dw|0,\Sigma)\},\\ &\le \int_{w\in\mathbb{W}}\int_{\hat{x}_+\in S}\delta_{\hat{\theta}^\top f(\hat{x},\hat{u})+w}(d\hat{x}_+)\mathcal{N}(dw|0,\Sigma)=\hat{\mathbf{t}}(S|z),\end{aligned}$$

verifying condition Definition 2.8(b). The first inequality becomes an equality for unbounded $\mathbb{X}$.

Definition 2.8(c)

For measurable $S\subset\mathbb{X}$, integration of $\boldsymbol{v}_\theta$ over $\hat{\mathbb{X}}\times S$ yields

$$\boldsymbol{v}_\theta(\hat{\mathbb{X}}\times S|z)\le\int_{w\in\mathbb{W}}\int_{x_+\in S}\delta_{\theta^\top f(\hat{x},\hat{u})+w}(dx_+)\mathcal{N}(dw|0,\Sigma)=\mathbf{t}_\theta(S|z),$$

verifying condition (c). Hence, (2.21) constitutes a valid sub-probability coupling.

Proving δ in (2.22)

It is now verified that $\delta(\hat{x},\hat{u})$ in (2.22) satisfies $\boldsymbol{v}_\theta(\hat{\mathbb{X}}\times\mathbb{X}|z)\ge 1-\delta(\hat{x},\hat{u})$. For this, the objective is to find the relation between $\delta(\hat{x},\hat{u})$ and $\gamma_\theta(\hat{x},\hat{u})$. Observe that $\boldsymbol{v}_\theta(\hat{\mathbb{X}}\times\mathbb{X}|z)$ in (A.1) reduces to the minimum of two Gaussian distributions:

$$\boldsymbol{v}_\theta(\hat{\mathbb{X}}\times\mathbb{X}|z)\ge\int_{w\in\mathbb{W}}\min\{\mathcal{N}(dw|0,\Sigma),\mathcal{N}(dw|-\gamma_\theta,\Sigma)\},$$

using the fact that the integral of the Dirac delta measure over the whole domain is at most equal to one. In fact, the inequality becomes an equality if the spaces $(\hat{\mathbb{X}}, \mathbb{X})$ are unbounded. Due to the symmetry of the Gaussian distribution, it follows that

$$\boldsymbol{v}_\theta(\hat{\mathbb{X}} \times \mathbb{X}|z) = 2 \int_A \mathcal{N}(dw|0, \Sigma), \tag{A.2}$$

where $A = \{w \in \mathbb{W} \,|\, \gamma_\theta^\top \Sigma^{-1}(2w + \gamma_\theta) \leq 0\}$. This set represents a half-space derived from $\mathcal{N}(dw|0, \Sigma) \leq \mathcal{N}(dw|-\gamma_\theta, \Sigma)$. The computation in (A.2) can be interpreted as $\boldsymbol{v}_\theta(\hat{\mathbb{X}} \times \mathbb{X}|z) = 2\,\mathbb{P}(\gamma_\theta^\top \Sigma^{-1}(2w + \gamma_\theta) \leq 0)$ with w having Gaussian distribution with mean zero and covariance Σ. This amounts to integrating a multivariate Gaussian distribution over a half space. Note that in this case, $2\gamma_\theta^\top \Sigma^{-1} w$ is a one-dimensional normal random variable with zero mean and variance $4\gamma_\theta^\top \Sigma^{-1} \gamma_\theta$. Thus,

$$\boldsymbol{v}_\theta(\hat{\mathbb{X}} \times \mathbb{X}|z) = 2 \cdot \mathrm{cdf}\left(\frac{-\gamma_\theta^\top \Sigma^{-1} \gamma_\theta}{\sqrt{4\gamma_\theta^\top \Sigma^{-1} \gamma_\theta}} \right) = 2 \cdot \mathrm{cdf}\left(-\frac{1}{2} \sqrt{\gamma_\theta^\top \Sigma^{-1} \gamma_\theta} \right).$$

Therefore, it holds that $\boldsymbol{v}_\theta(\hat{\mathbb{X}} \times \mathbb{X}|z) \geq 1 - \delta(\hat{x}, \hat{u})$, with

$$\begin{aligned} \delta(\hat{x}, \hat{u}) &:= 1 - 2 \cdot \mathrm{cdf}\left(-\frac{1}{2} \sqrt{\zeta(\hat{x}, \hat{u})} \right), \\ \zeta(\hat{x}, \hat{u}) &:= \sup_{\theta \in \Theta} \left(\gamma_\theta(\hat{x}, \hat{u})^\top \Sigma^{-1} \gamma_\theta(\hat{x}, \hat{u}) \right). \end{aligned} \tag{A.3}$$

Note that a δ dependent on $(\hat{x}, \hat{u})$ is calculated to reduce the conservatism of the approach. Given the definition of γ_θ and the characterisation of Θ in (2.5), the computation of $\zeta(\hat{x}, \hat{u})$ is through the optimisation

$$\begin{aligned} \zeta(\hat{x}, \hat{u}) := \sup_{\theta} \;& f(\hat{x}, \hat{u})^\top (\theta - \hat{\theta}) \Sigma^{-1} (\theta - \hat{\theta})^\top f(\hat{x}, \hat{u}) \\ \text{s.t.} \quad & (\Delta\bar{\theta})^\top \Sigma_N^{-1} (\Delta\bar{\theta}) \leq n \cdot \chi^{-1}(1 - \alpha|n), \end{aligned} \tag{A.4}$$

where $\Delta\bar{\theta} := [\theta_1 - \hat{\theta}_1; \ldots; \theta_n - \hat{\theta}_n]$. This is a quadratically constrained quadratic program, which is convex and can be solved efficiently using interior point methods. Here, an upper bound for the optimal solution is computed, which is sufficient for the purpose of getting a correct bound for δ:

$$\begin{aligned} f(\hat{x}, \hat{u})^\top (\theta - \hat{\theta}) \Sigma^{-1} (\theta - \hat{\theta})^\top f(\hat{x}, \hat{u}) &= \Delta\bar{\theta}^\top (I_n \otimes f(\hat{x}, \hat{u})^\top)^\top \Sigma^{-1} (I_n \otimes f(\hat{x}, \hat{u})^\top) \Delta\bar{\theta}, \\ &\leq \left\| I_n \otimes f(\hat{x}, \hat{u})^\top \right\|^2 \left\| \Sigma^{-1} \right\| \left\| \Delta\bar{\theta} \right\|^2, \\ &= \| f(\hat{x}, \hat{u}) \|^2 \left\| \Sigma^{-1} \right\| \left\| \Delta\bar{\theta} \right\|^2, \\ &\leq \| f(\hat{x}, \hat{u}) \|^2 \left\| \Sigma^{-1} \right\| \frac{n \cdot \chi^{-1}(1 - \alpha|n)}{\lambda_{min}(\Sigma_N^{-1})}, \\ &= \| f(\hat{x}, \hat{u}) \|^2 \left\| \Sigma^{-1} \right\| \| \Sigma_N \| \cdot n \cdot \chi^{-1}(1 - \alpha|n). \end{aligned}$$

In the above, the following (in)equalities have been used:

- $x^\top Ax \leq \|A\| \|x\|^2$ and $x^\top Ax \geq \lambda_{min}(A) \|x\|^2$, for any vector x and symmetric positive semi-definite matrix A;
- $\|AB\| \leq \|A\| \cdot \|B\|$ and $\|A \otimes B\| = \|A\| \cdot \|B\|$ for any two matrices A and B; and
- $\|I_n\| = 1$.

Therefore, it follows that

$$\sqrt{\zeta(\hat{x}, \hat{u})} \leq \|f(\hat{x}, \hat{u})\| \sqrt{r}, \quad \text{with} \quad r := \left\|\Sigma^{-1}\right\| \|\Sigma_N\| \cdot n \cdot \chi^{-1}(1 - \alpha | n).$$

Substituting this into (A.3) completes the proof of the expression for δ and shows that condition (b) of Definition 2.8 is satisfied. Recall that condition (a) of Definition 2.8 holds by setting the initial states $\hat{x}_0 = x_0$ and that condition (c) is satisfied with $\varepsilon = 0$ since both systems use the same output mapping. It follows that $\widehat{\mathcal{M}} \preceq^{\delta}_{\varepsilon} \mathcal{M}(\theta)$ with $\delta(\hat{x}, \hat{u})$ in (2.22) and $\varepsilon = 0$.

Definition 2.9

It follows directly from $\widehat{\mathcal{M}} \preceq^{\delta}_{\varepsilon} \mathcal{M}(\theta)$ and Proposition 2.2 that there exists a valid refinement. A composed probability measure $\mathbf{t}^{\times}_{\theta}$ of the state mapping in (2.18) is derived using (2.9), which bounds the sub-probability coupling $\boldsymbol{\upsilon}_{\theta}$ in (2.21) from above for all $z \in \mathbb{Z}$ and $\theta \in \Theta$:

$$\begin{aligned}\mathbf{t}^{\times}_{\theta}(d\hat{x}_+ \times dx_+ | z) &= \int_{w \in \mathbb{W}} \int_{\hat{w} \in \hat{\mathbb{W}}} \delta_{\hat{\theta}^\top f(\hat{x}, \hat{u}) + w + \gamma_\theta}(d\hat{x}_+) \delta_{\theta^\top f(\hat{x}, \hat{u}) + w}(dx_+) \mathcal{N}(dw | 0, \Sigma), \\ &\geq \boldsymbol{\upsilon}_{\theta}(d\hat{x}_+ \times dx_+ | z).\end{aligned}$$

This concludes the proof of Theorem 2.3. □

A.2 Proof of Theorem 3.1

Proof. It is shown that $\widehat{\mathcal{M}}$ is in an SSR with $\mathcal{M}$ by proving that the conditions in Definition 3.3 hold for the monolithic case of $i \in \{1\}$ (equivalent to Definition 2.8). Condition (a) holds by setting the initial states $\hat{x}_0 = x_0$.

ε-Deviation: The proof of condition (c) is a trivial extension to the one provided by Lavaei et al. (2021, Theorem 4.3).

δ-Deviation: For condition (b), the sub-probability coupling $\boldsymbol{\upsilon}^i_{\theta^i}$ in (3.10) is completed to a probability kernel as follows:

$$\bar{\boldsymbol{v}}^i_{\theta^i}(d\hat{x}^i_+ \times dx^i_+) = \boldsymbol{v}^i_{\theta^i}(d\hat{x}^i_+ \times dx^i_+) + \frac{1}{1 - \boldsymbol{v}^i_{\theta^i}(\mathcal{R}_{x^i})} \times \left(\mathbf{t}^i_{\theta^i}(dx^i_+) - \boldsymbol{v}^i_{\theta^i}(\hat{\mathbb{X}}^i \times dx^i_+)\right)\left(\hat{\mathbf{t}}^i(d\hat{x}^i_+) - \boldsymbol{v}^i_{\theta^i}(d\hat{x}^i_+ \times \mathbb{X}^i)\right), \tag{A.5}$$

where the arguments of $\bar{\boldsymbol{v}}^i_{\theta^i}$, $\mathbf{t}^i_{\theta^i}$, and $\hat{\mathbf{t}}^i$ have been deprecated for conciseness. Refer to Proposition 2.1 for more details. The following lemma is established

Lemma A.1 *The sub-probability coupling $\boldsymbol{v}^i_{\theta^i}$ in (3.10) is maximal with respect to the transition kernels $(\hat{\mathbf{t}}^i, \mathbf{t}^i_{\theta^i})$ in (3.1) and (3.3), and relation $\mathcal{R}_{x^i}$ in (3.7), i.e., for the associated probability kernel $\bar{\boldsymbol{v}}^i_{\theta^i}$ (Proposition 2.1), $\bar{\boldsymbol{v}}^i_{\theta^i}(\mathcal{R}_{x^i}) = \boldsymbol{v}^i_{\theta^i}(\mathcal{R}_{x^i})$ holds.*

Proof. In the following proof, the superscript i is dropped. $\bar{\boldsymbol{v}}_\theta$ in (A.5) is integrated over $\mathcal{R}_x$, yielding

$$\boldsymbol{v}_\theta(\mathcal{R}_x) = \bar{\boldsymbol{v}}_\theta(\mathcal{R}_x) - \frac{\phi_\theta}{1 - \boldsymbol{v}_\theta(\mathcal{R}_x)}(\mathcal{R}_x), \quad \text{where} \tag{A.6}$$
$$\phi_\theta(\mathcal{R}_x) := \int_{\mathbb{X}} \left(\mathbf{t}_\theta(dx_+) - \boldsymbol{v}_\theta(\hat{\mathbb{X}} \times dx_+)\right)\left(\hat{\mathbf{t}}(dx_+) - \boldsymbol{v}_\theta(dx_+ \times \mathbb{X})\right).$$

Using the definitions of $\boldsymbol{v}_\theta, \mathbf{t}_\theta$, and $\hat{\mathbf{t}}$, this is expanded to get

$$\phi_\theta(\mathcal{R}_x) = \int_{\mathbb{X}} \int_{\mathbb{W}} \delta_{f_{\hat{\theta}}+w}(dx_+)\tilde{p}_w(dw|\theta) \int_{\mathbb{W}} \delta_{f_{\hat{\theta}}+\hat{w}}(dx_+)\tilde{p}_{\hat{w}}(d\hat{w}|\theta),$$

with the distributions

$$\tilde{p}_w(dw|\theta) := \min\left\{0,\ \mathcal{MW}_K(dw|\pi, \mu + \gamma_\theta, \Sigma) - \mathcal{MW}_K(dw|\hat{\theta})\right\},$$
$$\tilde{p}_{\hat{w}}(d\hat{w}|\theta) := \min\left\{\mathcal{MW}_K(d\hat{w}|\hat{\theta}) - \mathcal{MW}_K(d\hat{w}|\pi, \mu + \gamma_\theta,\ \Sigma), 0\right\}.$$

Since the distributions are disjoint and identically mapped onto $\mathbb{X}$, $\phi_\theta(\mathcal{R}_x) = 0$ holds, and it follows from (2.8) that $\bar{\boldsymbol{v}}_\theta(\mathcal{R}_x) = \boldsymbol{v}_\theta(\mathcal{R}_x)$. □

Analogous to the proof of Lavaei et al. (2021, Theorem 4.3), there exists a probability kernel $\bar{\boldsymbol{v}}_\theta = \prod_{i=1}^N \bar{\boldsymbol{v}}^i_{\theta^i}$ over $(\hat{\mathbb{X}} \times \mathbb{X}, \mathscr{B}(\hat{\mathbb{X}} \times \mathbb{X}))$ that couples $(\hat{\mathbf{t}}, \mathbf{t}_\theta)$ over $\mathcal{R}_x$. Using Lemma A.1, this yields $\bar{\boldsymbol{v}}_\theta(\mathcal{R}_x) = \prod_{i=1}^N \boldsymbol{v}^i_{\theta^i}(\mathcal{R}_{x^i})$. Hence, Proposition 2.1 yields that $\boldsymbol{v}_\theta := \prod_{i=1}^N \boldsymbol{v}^i_{\theta^i}$ defines a sub-probability coupling of $(\hat{\mathbf{t}}, \mathbf{t}_\theta)$ over $\mathcal{R}_x$, which satisfies all conditions for a valid sub-probability coupling of $(\hat{\mathbf{t}}, \mathbf{t}_{\theta^i})$ over $\mathcal{R}_x$:

$$\boldsymbol{v}_\theta(\hat{\mathbb{X}} \times \mathbb{X}) = \prod_{i=1}^N \boldsymbol{v}^i_{\theta^i}(\hat{\mathbb{X}}^i \times \mathbb{X}^i) = \prod_{i=1}^N \boldsymbol{v}^i_{\theta^i}(\mathcal{R}_{x^i}) = \boldsymbol{v}_\theta(\mathcal{R}_x),$$
$$\boldsymbol{v}_\theta(d\hat{x}_+ \times \mathbb{X}) = \prod_{i=1}^N \boldsymbol{v}^i_{\theta^i}(d\hat{x}^i_+ \times \mathbb{X}^i) \leq \prod_{i=1}^N \hat{\mathbf{t}}^i(d\hat{x}^i_+) = \hat{\mathbf{t}}(d\hat{x}_+),$$
$$\boldsymbol{v}_\theta(\hat{\mathbb{X}} \times dx_+) = \prod_{i=1}^N \boldsymbol{v}^i_{\theta^i}(\hat{\mathbb{X}}^i \times dx^i_+) \leq \prod_{i=1}^N \mathbf{t}^i_{\theta^i}(dx^i_+) = \mathbf{t}_\theta(dx_+).$$

Since $\boldsymbol{v}_\theta(\mathcal{R}_x) \geq \prod_{i=1}^{N}(1-\delta^i)$, it follows that $\delta = 1 - \prod_{i=1}^{N}(1-\delta^i)$. This concludes the proof of Theorem 3.1. □

A.3 Proof of Theorem 3.4

Proof. The proof is similar to Chapter 2, and showing that $\boldsymbol{v}_\theta$ in (3.10) is a sub-probability coupling of the systems (3.1) and (3.3) over relation $\mathcal{R}_x$ in (3.7) is straightforward. To compute the corresponding parameter δ, $\boldsymbol{v}_\theta$ is integrated over the identity relation $\mathcal{R}_x$. Relaxing the resulting expression using the trivial inequality $\min\left\{\sum_i f_i, \sum_i g_i\right\} \geq \sum_i \min\{f_i, g_i\}$ yields

$$\begin{aligned}\boldsymbol{v}_\theta(\mathcal{R}_x) \geq \sum_{k=1}^{K} \int_{\mathbb{W}} \int_{\mathbb{X}} \delta_{f_{\hat{\theta}}+w+\gamma_\theta}(dx)\delta_{f_\theta+w}(dx) \\ \times \min\left\{\pi_k \mathcal{N}(dw|\mu_k, \Sigma_k),\ \hat{\pi}_k \mathcal{N}(dw|\hat{\mu}_k - \gamma_\theta, \Sigma_k)\right\},\end{aligned}$$

where the dependence on (x, u) was omitted for conciseness. To reach a simpler formulation, the expression is expanded using the Cholesky decomposition $\Sigma_k = L_k L_k^\top$, where L_k is the lower triangular matrix of Σ_k. Rearranging yields

$$\boldsymbol{v}_\theta(\mathcal{R}_x) \geq \sum_{k=1}^{K} \int_{\mathbb{W}} \min\left\{\pi_k \mathcal{N}(dw|\beta_{k,\theta}, I),\ \hat{\pi}_k \mathcal{N}(dw|0, I)\right\},$$

with $\beta_{k,\theta} := \mu_k - \hat{\mu}_k + \gamma_\theta$ and utilising the choice of offset $\gamma_\theta(x, u) = f_\theta(x, u) - f_{\hat{\theta}}(x, u)$ for all (x, u, θ). By similar reasoning as van Huijgevoort and Haesaert (2022), the individual integrals are split into integrations over disjoint halfspaces $E_{k,\theta}, \hat{E}_{k,\theta}$ as follows:

$$\boldsymbol{v}_\theta(\mathcal{R}_x) \geq \sum_{k=1}^{K} \left[\pi_k \int_{w \in E_{k,\theta}} \mathcal{N}(dw|\beta_{k,\theta}, I) + \hat{\pi}_k \int_{w \in \hat{E}_{k,\theta}} \mathcal{N}(dw|0, I)\right].$$

For each component, the halfspaces are given by

$$\hat{E}_{k,\theta}: \quad \beta_{k,\theta}^\top w > \eta_{k,\theta}\beta_{k,\theta}^\top\beta_{k,\theta}, \qquad E_{k,\theta}: \quad \beta_{k,\theta}^\top w \leq \eta_{k,\theta}\beta_{k,\theta}^\top\beta_{k,\theta},$$

with $\eta_{k,\theta} := \frac{1}{2} - \frac{1}{\beta_{k,\theta}^\top\beta_{k,\theta}} \log \frac{\pi_k}{\hat{\pi}_k}$. Integration yields

$$\begin{aligned}
\boldsymbol{v}_\theta(\mathcal{R}_x|\theta) &\geq \sum_{k=1}^{K} \pi_k \int_{w:\,\beta_{k,\theta}^\top w \geq \eta_{k,\theta}\beta_{k,\theta}^\top\beta_{k,\theta}} \mathcal{N}(dw|0,I) + \hat{\pi}_k \int_{w:\,\beta_{k,\theta}^\top w < \eta_{k,\theta}\beta_{k,\theta}^\top\beta_{k,\theta}} \mathcal{N}(dw|\beta_{k,\theta},I), \\
&= \sum_{k=1}^{K} \pi_k \int_{w:\,\beta_{k,\theta}^\top w \leq -\eta_{k,\theta}\beta_{k,\theta}^\top\beta_{k,\theta}} \mathcal{N}(dw|0,I) + \hat{\pi}_k \int_{\hat{w}:\,\beta_{k,\theta}^\top \hat{w} < (\eta_{k,\theta}-1)\beta_{k,\theta}^\top\beta_{k,\theta}} \mathcal{N}(d\hat{w}|0,I), \\
&= \sum_{k=1}^{K} \pi_k \mathrm{cdf}\left(-\eta_{k,\theta}\left\|\beta_{k,\theta}\right\|\right) + \hat{\pi}_k \mathrm{cdf}\left((\eta_{k,\theta}-1)\left\|\beta_{k,\theta}\right\|\right),
\end{aligned}$$

where the invertability of the normal centred at zero was used in the first step. Hence, (3.14) is obtained by considering the supremum over all $\theta \in \Theta$. □

Remark A.1 The proof of Theorem 3.4 utilises the assumed pairwise identical covariance structure (see Subsection 3.3.3) to obtain planar hypersurfaces dividing the regions of the support where the component densities of the two noise distributions minimise each other. Relaxing this assumption would yield higher-order hypersurfaces with possibly more than two regions of mutual minimisation.

A.4 Proof of Proposition 4.1

Proof. The BT kernel in (4.6) can be rewritten as a sum of kernels:

$$\hat{k}_q(x,x') = \sum_{i=1}^{q} \kappa_i(x,x'), \tag{A.7}$$

where each kernel $\kappa_i(x,x')$ is defined as $\kappa_i(x,x') := \phi_i(x)^\top\phi_i(x)$, with the finite-dimensional feature maps

$$\phi_i(x) := \sqrt{\mathrm{w}_i}\left[\mathbb{1}(x \in \Xi_i(s))\right]_{s\in\mathbb{B}^i},$$

where each i^{th} summand of $\hat{k}_q$ was expanded with respect to all possible binary substrings of length i. Note that the binary substrings $s \in \mathbb{B}^i$ in $\kappa_i : \mathbb{X}\times\mathbb{X} \to [0,1]$ are unique and κ_i is clearly positive-definite. The associated RKHS $\mathscr{H}_{\kappa_i}$ has dimension 2^i. Since finite-dimensional spaces are automatically complete, the RKHS $\mathscr{H}_{\kappa_i}$ is simply the span of the functions $\kappa_i(x,\,\cdot\,)$, for all $x \in \mathbb{X}$. Every function $f_i \in \mathscr{H}_{\kappa_i}$ can therefore be written as a linear combination of the form

$$f_i(x) = \sum_{j=1}^{m} \alpha_j \kappa_i(x,x_j) = \sum_{j=1}^{m} \alpha_j \phi_i(x)^\top\phi_i(x_j),$$

for some $m \in \mathbb{N}$, $\alpha_j \in \mathbb{R}$, and $x_j \in \mathbb{X}$ (Berlinet and Thomas-Agnan, 2004). Since $\phi_i(x_j)$ places functional weight only on the partition indexed by $s = \Psi_i(x_j)$, every function $f_i \in \mathscr{H}_{\kappa_i}$ can equivalently be written as

$$f_i(x) = y_i^\top \phi_i(x),$$

for some coefficient vector $y_i \in \mathbb{R}^{2^i}$. Since the BT kernel is a weighted sum of kernels κ_i as shown in (A.7), the kernel sum rule (Berlinet and Thomas-Agnan, 2004, Theorem 5) implies that the corresponding RKHS $\mathscr{H}_{\hat{k}_q}$ consists of functions of the form

$$f(x) = y^\top \hat{\phi}_q(x),$$

with coefficient vector $y := [y_1; \ldots; y_q] \in \mathbb{R}^{\sum_{i=1}^m 2^i}$ and the concatenated feature map $\hat{\phi}_q(x) := [\phi_1(x); \ldots; \phi_q(x)] \in \mathbb{R}^{\sum_{i=1}^m 2^i}$. This coincides with the functional representation in (4.7) and confirms that the RKHS $\mathscr{H}_{\hat{k}_q}$ of $\hat{k}_q(x, x') = \langle \hat{\phi}_q(x), \hat{\phi}_q(x) \rangle_{\mathscr{H}_k}$ is $(\sum_{i=1}^q 2^i)$-dimensional. □

A.5 Proof of Proposition 4.2

Proof. The BT kernel in (4.6) can be expressed as a sum of positive-definite kernels:

$$\hat{k}_q(x, x') = \sum_{i=1}^q \mathrm{w}_i \sum_{s \in \mathbb{B}^i} \kappa_s(x, x'),$$

where each kernel κ_s is defined by

$$\kappa_s(x, x') := \mathbb{1}\left(\Psi_{|s|}(x) = s\right) \mathbb{1}\left(\Psi_{|s|}(x') = s\right).$$

Following similar reasoning as in the proof of Proposition 4.1, the associated RKHS is given by

$$\begin{aligned}
\mathscr{H}_{\kappa_s} &= \left\{ f_s \mid f_s = y_s \mathbb{1}\left(\Psi_{|s|}(\,\cdot\,) = s\right),\ y_s \in \mathbb{R} \right\}, \\
&= \left\{ f_s \mid f_s = y_s \mathbb{1}\left(\Psi_{|s|}(\,\cdot\,) = s\right) \mathbb{1}(s_{\leq |s|} = s),\ y_s \in \mathbb{R} \right\}, \\
&= \left\{ f_s \mid f_s = y_s \kappa_s(\,\cdot\,, x_s),\ y_s \in \mathbb{R},\ x_s \in \Xi_{|s|}(s) \right\}.
\end{aligned}$$

Thus, the norm of $f_s \in \mathscr{H}_{\kappa_s}$ reduces as follows:

$$\|f_s\|^2_{\mathscr{H}_{\kappa_s}} \equiv \langle f_s, f_s \rangle_{\mathscr{H}_{\kappa_s}} = y_s^2 \langle \kappa_s(\,\cdot\,, x_s), \kappa_s(\,\cdot\,, x_s) \rangle_{\mathscr{H}_{\kappa_s}} = y_s^2,$$

where $\kappa_s(x_s, x_s) = 1$ was used in the last step. According to Aronszajn's sums of kernels theorem (Berlinet and Thomas-Agnan, 2004, Theorem 5), the RKHS norm of a function $f \in \mathscr{H}_{\hat{k}_q}$ satisfies

$$\|f\|^2_{\mathscr{H}_{\hat{k}_q}} \leq \sum_{i=1}^q \mathrm{w}_i \sum_{s \in \mathbb{B}^i} \|f_s\|^2_{\mathscr{H}_{\kappa_s}},$$

concluding the proof. □

A.6 Proof of Theorem 4.1

Proof. Similar to the proof of Theorem 2 by Chowdhury and Gopalan (2017), the error bound ε_d is split into one term related to the noise and one to the statistical approximation error. For the data $\mathcal{D} = (X_N, Y_N)$, let $Y_d := [y_{1,d}, \ldots, y_{N,d}]^\top$ and recognise that it holds that $Y_d \equiv F_d + V_d$, with $F_d := [f_d(x^1), \ldots, f_d(x^N)]^\top$ and $V_d := [v_{1,d}, \ldots, v_{N,d}]^\top$. Via Cauchy-Schwarz, the error bound is decomposed:

$$\left|\hat{\mu}_{N,d}(s) - f_d(x)\right| \leq \left|\hat{k}_N(s)^\top \hat{C}_N^{-1} V_d\right| + \left|\hat{k}_N(s)^\top \hat{C}_N^{-1} F_d - f_d(x)\right|, \tag{A.8}$$

where $\hat{C}_N := [\hat{K}_N + \sigma_w^2 I_N]$. The noise term is bounded following similar steps to Fiedler et al. (2021, Proposition 2), to get

$$\left|\hat{k}_N(s)^\top \hat{C}_N^{-1} V_d\right| \leq \left\|\hat{C}_N^{-1} \hat{k}_N(s)\right\| \sqrt{N + 2\sqrt{N \log\left(\alpha^{-1}\right)} + 2\log\left(\alpha^{-1}\right)},$$

which yields the first term in the minimisation of $\varepsilon_{d,1}(s)$. An alternative bound can be derived by following the steps of Chowdhury and Gopalan (2017, Theorem 2):

$$|\hat{k}_N(s)^\top \hat{C}_N^{-1} V_d| \leq \sigma_w^{-1} \hat{\sigma}_{N,d}(s) \sqrt{V_d^\top \hat{K}_N \hat{C}_N^{-1} V_d}.$$

Hsu et al. (2012, Theorem 2.1) establish that

$$\mathbb{P}\left(\|\Sigma V_d\|^2 \leq \mathrm{Tr}(\Sigma) + 2\sqrt{\mathrm{Tr}(\Sigma^2)t} + 2\|\Sigma\| t\right) \geq 1 - e^{-t},$$

where $\Sigma := \hat{K}_N \hat{C}_N^{-1}$. Setting $t := \log\left(\alpha^{-1}\right)$ yields the second term in the minimisation of $\varepsilon_{d,1}(s)$. Depending on the setting, one or the other bound will provide a less conservative error $\varepsilon_{d,1}(s)$.

Next, some fixed points $x_s \in \mathbb{X}$ are selected for all $s \in \mathbb{S}$ such that $\bar{f}_d(s) := f_d(x_s)$. Expanding the second term in (A.8) with $\bar{f}_d(s)$ and Cauchy-Schwarz yields for every $s \in \mathbb{S}$ that for all $x \in \Xi(s)$

$$\left|\hat{k}_N(s)^\top \hat{C}_N^{-1} F_d - f_d(x)\right| \leq \left|\hat{k}_N(s)^\top \hat{C}_N^{-1} F_d - \bar{f}_d(s)\right| + \left|\bar{f}_d(s) - f_d(x)\right|. \tag{A.9}$$

The latter term is bounded from above via

$$\begin{aligned}
\left|\bar{f}_d(s) - f_d(x)\right| &= \left|\left(\phi(x_s) - \phi(x)\right)^\top f_d\right|, \\
&\leq \|\phi(x_s) - \phi(x)\|_{\mathcal{H}_k} \|f_d\|_{\mathcal{H}_k}, \\
&= \left|k(x_s, x_s) - 2k(x_s, x) + k(x, x)\right|^{\frac{1}{2}} \|f_d\|_{\mathcal{H}_k}, \\
&\leq \varepsilon_{d,2}(s).
\end{aligned}$$

Let $\Phi_N := [\phi(x_1), \ldots, \phi(x_N)]^\top$. With this and Cauchy-Schwarz, the prior term in (A.9) yields

$$\left|\hat{k}_N(s)^\top \hat{C}_N^{-1} F_d - \bar{f}_d(s)\right| \leq \left\|\hat{k}_N(s)^\top \hat{C}_N^{-1} \Phi_N - \phi(x_s)^\top\right\|_{\mathcal{H}_k} \|f_d\|_{\mathcal{H}_k},$$

which reduces to $\varepsilon_{d,3}(s)$, concluding the proof. □

Appendix B
SILP to LP Relaxation

This chapter provides further details on the finite-constraint relaxation of the SILP (5.20) to the LP (5.21). Let

$$\{x_0^{(1)},\dots,x_0^{(\hat{N}_0)}\}\subset\mathbb{X}_0,\quad \{x_u^{(1)},\dots,x_u^{(\hat{N}_u)}\}\subset\mathbb{X}_u,\quad X_{\hat{N}}:=\{x^{(1)},\dots,x^{(\hat{N})}\}\subset\mathbb{X}$$

be sampling lattices with a common $C_{\hat{N}}$ as in Lemma 5.1. Based on the constraint-tightening reasoning outlined in Subsection 5.8.2, it follows that the first constraint (5.20b), given by $\forall x_0\in\mathbb{X}_0:\ \phi_M(x_0)^\top b\le\eta$, holds if

$$\max_{\hat{N}}^{\mathbb{X}_0}\ge\frac{2}{C_{\hat{N}}+1}\eta+\frac{C_{\hat{N}}-1}{C_{\hat{N}}+1}\min_{\hat{N}}^{\mathbb{X}_0},$$

with the additional variables

$$\min_{\hat{N}}^{\mathbb{X}_0}\le\phi_M(x_0^{(i)})^\top b,\quad\text{and}\quad \max_{\hat{N}}^{\mathbb{X}_0}\ge\phi_M(x_0^{(i)})^\top b,\quad i=1,\dots,\hat{N}_0.$$

Renaming $\max_{\hat{N}}^{\mathbb{X}_0}=:\hat{\eta}$ yields the implication

$$\left.\begin{aligned}&\phi_M(x_0^{(i)})^\top b\le\hat{\eta},\\&\min_{\hat{N}}^{\mathbb{X}_0}\le\phi_M(x_0^{(i)})^\top b,\quad i=1,\dots,\hat{N}_0,\\&\hat{\eta}:=\tfrac{2}{C_{\hat{N}}+1}\eta+\tfrac{C_{\hat{N}}-1}{C_{\hat{N}}+1}\min_{\hat{N}}^{\mathbb{X}_0},\end{aligned}\right\}\Longrightarrow\forall x_0\in\mathbb{X}_0:\ \phi_M(x_0)^\top b\le\eta.$$

Similar reasoning for the constraint (5.20d) gives

$$\left.\begin{aligned}&\phi_M(x^{(i)})^\top(Hb-b)\le\hat{\Delta},\\&\min_{\Delta}\le\phi_M(x^{(i)})^\top b,\quad i=1,\dots,\hat{N},\\&\hat{\Delta}:=\tfrac{2}{C_{\hat{N}}+1}\left(c-\Re\bar{B}\kappa\right)+\tfrac{C_{\hat{N}}-1}{C_{\hat{N}}+1}\min_{\Delta},\end{aligned}\right\}\Longrightarrow\forall x\in\mathbb{X}:\ \phi_M(x)^\top(Hb-b)\le c-\Re\bar{B}\kappa.$$

For the remaining lower bounds in the constraints (5.20c) and (5.20b), the constraint-tightening reasoning outlined in Subsection 5.8.2 gives

$$\left.\begin{array}{l} \phi_M(x_u^{(i)})^\top b \geq \hat{\gamma}, \\ \max_{\hat{N}}^{\mathbb{X}_u} \geq \phi_M(x_u^{(i)})^\top b, \quad i = 1, \ldots, \hat{N}_u, \\ \hat{\gamma} := \frac{2}{C_{\hat{N}}+1}\gamma + \frac{C_{\hat{N}}-1}{C_{\hat{N}}+1}\max_{\hat{N}}^{\mathbb{X}_u}, \end{array}\right\} \Longrightarrow \forall x_u \in \mathbb{X}_u : \ \phi_M(x_u)^\top b \geq \gamma,$$

$$\left.\begin{array}{l} \phi_M(x^{(i)})^\top b \geq \hat{\xi}, \\ \max_{\hat{N}}^{\mathbb{X}} \geq \phi_M(x^{(i)})^\top b, \quad i = 1, \ldots, \hat{N}, \\ \hat{\xi} := \frac{C_{\hat{N}}-1}{C_{\hat{N}}+1}\max_{\hat{N}}^{\mathbb{X}}, \end{array}\right\} \Longrightarrow \forall x \in \mathbb{X} : \ \phi_M(x)^\top b \geq 0,$$

respectively. The LP (5.21) is obtained by substituting the relaxed constraints into the SILP (5.20).

List of Symbols

$\mathbb{B} = \{0, 1\}$	set of binary numbers
$\mathbb{N} = \{1, 2, \ldots\}$	set of positive natural numbers
$\mathbb{N}_0 = \{0, 1, 2, \ldots\}$	set of non-negative natural numbers
$\mathbb{R}$	set of real numbers
$\mathbb{R}_{\geq 0}$	set of non-negative reals
$A \times B$	Cartesian product of two sets A and B
I_n	identity matrix of size $n \times n$
$\text{diag}(x)$	diagonal matrix constructed from a vector x
$\text{Tr}(A)$	trace of a matrix A
$\mathbb{1}_A$	indicator function of a set A

Probability Theory: measure theory, probability distributions, expectation

$(\mathbb{X}, \mathscr{B}(\mathbb{X}), \mathbf{p})$	probability space
$\mathscr{B}(\mathbb{X})$	Borel σ-algebra on a sample space $\mathbb{X}$
$\mathbb{P}$	probability
$\mathscr{P}(\mathbb{X})$	set of all probability measures on $(\mathbb{X}, \mathscr{B}(\mathbb{X}))$
$\mathbb{E}$	expectation operator
δ_a	Dirac delta measure concentrated at a point a
$\mathscr{U}_{\mathbb{X}}$	uniform probability distribution over $\mathbb{X}$
$\mathcal{N}(dx\|\mu, \Sigma)$	Gaussian measure with mean μ and covariance Σ
$\mathcal{MN}(dx\|\pi, \mu, \Sigma)$	Gaussian mixture model with parameters (π, μ, Σ)
cdf $(x\|\mu, \Sigma)$	CDF of a Gaussian distribution with mean μ and covariance Σ

System: dynamics, data, controller

$\boldsymbol{\mathcal{M}}$	system model
$\mathbb{X}$	state space
$\mathbb{U}$	control input space
$\mathbb{W}$	process noise space

$\mathbb{Y}$	output space
$\mathbf{t}$	transition probability kernel
$\mathcal{D}$	data
$\boldsymbol{C}$	controller
ρ	control policy

Relations: simulation relation, composition, inter-system mappings

$\boldsymbol{C} \bowtie \boldsymbol{\mathcal{M}}$	closed-loop composition of $\boldsymbol{C}$ and $\boldsymbol{\mathcal{M}}$
$\widehat{\boldsymbol{\mathcal{M}}} \preceq_{\varepsilon}^{\delta} \boldsymbol{\mathcal{M}}$	(ε, δ)-SSR between $\boldsymbol{\mathcal{M}}$ and $\widehat{\boldsymbol{\mathcal{M}}}$
$\mathcal{R}$	relation
$\mathbf{i}$	interface function
$\mathbf{s}$	state mapping
$\boldsymbol{v}$	sub-probability coupling

Specifications: LTL, DFA, barrier

ψ	(temporal logic) specification
$\models$	satisfies
$\mathcal{L}$	labelling function
$\mathcal{A}_{\psi}$	DFA of a specification ψ
$\mathcal{B}$	control barrier certificate function

RKHS: kernel, feature map, mean embedding, ambiguity set

$k(x, x')$	positive symmetric kernel
ϕ	feature map
$\mathscr{H}$	(reproducing kernel) Hilbert space
$\langle \varphi, \varphi' \rangle_{\mathscr{H}}$	inner product of $\varphi, \varphi' \in \mathscr{H}$
$\lVert \cdot \rVert_{\mathscr{H}}$	RKHS norm
Ψ	(conditional) mean embedding
$C_{\mathfrak{R}}$	RKHS ambiguity set with radius $\mathfrak{R}$
$\mathfrak{R}$	RKHS ambiguity set radius

List of Abbreviations

ACAS X	Airborne Collision Avoidance System X
AI	Artificial Intelligence
AV	Autonomous Vehicle
B&B	Branch and Bound
BT	Binary Tree
BTGP	Binary-Tree Gaussian Process
CBC	Control Barrier Certificate
CBF	Control Barrier Function
CDCL	Conflict-Driven Clause Learning
CDF	Cumulative Distribution Function
CEGIS	CounterExample-Guided Inductive Synthesis
cf.	(Latin: confer) compare
CME	Conditional Mean Embedding
CPS	Cyber-Physical System
CPU	Central Processing Unit
DERs	Distributed Energy Resources
DFA	Deterministic Finite-state Automaton
DKL	Deep Kernel Learning
e.g.	(Latin: exempli gratia) for example
etc.	et cetera
EU	European Union
FFT	Fast Fourier Transform
gMDP	general Markov Decision Process
GMM	Gaussian Mixture Model
GP	Gaussian Process
GPS	Global Positioning System
IIA	Interval Iteration Algorithm
i.i.d.	independent, identically distributed

IMDP	Interval Markov Decision Process
IMC	Interval Markov Chain
IMU	Inertial Measurement Unit
ISO	International Organization for Standardization
JPEG	Joint Photographic Experts Group
KSOS	Kernel Sum Of Squares
LiDAR	Light Detection And Ranging
LP	Linear Program
LTL	Linear Temporal Logic
LTL_F	Finite-trace Linear Temporal Logic
MC	Markov Chain
ME	Mean Embedding
MDP	Markov Decision Process
MMD	Maximum Mean Discrepancy
MOR	Model Order Reduction
MPC	Model Predictive Control
MRI	Magnetic Resonance Imaging
NN	Neural Network
NP-complete	Nondeterministic Polynomial-time complete
PCA	Principal Component Analysis
pMDP	parametric Markov Decision Process
POMDP	Partially Observable Markov Decision Process
RAM	Random-Access Memory
RKHS	Reproducing Kernel Hilbert Space
RMDP	Robust Markov Decision Process
SAT	(Boolean) SATisfiability problem
SCADA	Supervisory Control And Data Acquisition
scLTL	syntactically co-safe Linear Temporal Logic
SDP	Semi-Definite Program
SILP	Semi-Infinite Linear Program
SIP	Semi-Infinite Program
SMT	Satisfiability Modulo Theories

SOS	Sum Of Squares
SOTIF	Safety Of The Intended Functionality
SSR	Sub-Simulation Relation
s.t.	such that
STL	Signal Temporal Logic
SVM	Support Vector Machine
TCAS	Traffic Alert and Collision Avoidance System
TD3	Twin-Delayed Deep Deterministic
US	United States
vs	versus

Curriculum Vitae

Oliver Schön was born on 11th August 1997 in Lennestadt-Altenhundem, and grew up in Cobbenrode, North Rhine-Westphalia, Germany. After completing secondary education at Gymnasium Maria Königin in 2015, he moved to Paderborn to study Mechanical Engineering at Paderborn University. He received his BSc and MSc degrees in 2018 and 2021, respectively, with a specialisation in Product Development and Control Engineering & Mechatronics. In 2020, he completed an exchange semester in Energy and Environmental Technology at the University of South-Eastern Norway, Porsgrunn.

During his Master's studies, he started working on projects at the intersection of first principles and data-driven learning, sparking a lasting fascination with applying machine learning to real-world dynamical systems. His project thesis explored how a-priori knowledge of physical dynamics models can be incorporated into Bayesian optimisation for the efficient tuning of engineering plants and processes. Building on this, his Master's thesis investigated physics-guided neural networks for identifying physically consistent dynamics models directly from data — culminating in his first scientific publication. Alongside these academic efforts, Oliver worked as a Student Trainee at Bosch Corporate Research in Stuttgart, where he applied these techniques to develop data-driven controllers for Lab-on-Chip devices used in COVID-19 PCR testing. He also served as a Research Assistant at the Heinz Nixdorf Institute in Paderborn.

In January 2022, Oliver moved to England to pursue his PhD in Computer Science at Newcastle University under the supervision of Professor Sadegh Soudjani. His research explores the intersection of formal methods and machine learning to enable safe and trustworthy control of complex stochastic systems under uncertainty. His work on uncertainty-aware correct-by-design kernel methods aims to bridge the gap between simulation and reality, supporting the certification and deployment of autonomous systems across mobility, healthcare, and smart infrastructure.

Throughout his doctoral studies, Oliver presented his work at leading international conferences, such as the IEEE Conference on Decision and Control (CDC) and the American Control Conference (ACC) — receiving several competitive student travel awards. He has been an invited speaker and visiting scholar at institutions including Stanford University, Eindhoven University of Technology, and Imperial College London. In 2023, his work was shortlisted for the IEEE CSS Technical Committee on Hybrid Systems Outstanding Student Paper Prize. Oliver has participated in multiple accelerator programmes and is the recipient of competitive scholarships and awards across science and engineering, including the Deutschlandstipendium (2015, 2016), the Erasmus+ European Scholarship (2020), and a PhD Studentship of the School of Computing at Newcastle University (2022).

List of Publications

Following is a list of the publications produced throughout the course of this PhD project.

Journal Papers

Oliver Schön, Birgit van Huijgevoort, Sofie Haesaert, and Sadegh Soudjani (2025). *"Bayesian Formal Synthesis of Unknown Systems via Robust Simulation Relations"*. In: *IEEE Transactions on Automatic Control (TAC)*, **Full Paper**, 70(3):1531–1546, 16 pages, DOI 10.1109/TAC.2024.3459308.

Oliver Schön, Zhengang Zhong, and Sadegh Soudjani (TBA). *"Kernel-Based Learning of Safety Barriers"*. Manuscript submitted for publication in *Journal of Artificial Intelligence Research (JAIR)*, 37 pages.

Conference Papers

Oliver Schön, Birgit van Huijgevoort, Sofie Haesaert, and Sadegh Soudjani (2022). *"Correct-by-Design Control of Parametric Stochastic Systems"*. In: *2022 IEEE 61st Conference on Decision and Control (CDC)*, Cancun, Mexico, 2022, pp 5580–5587, DOI 10.1109/CDC51059.2022.9992723.

Oliver Schön, Birgit van Huijgevoort, Sofie Haesaert, and Sadegh Soudjani (2023). *"Verifying the Unknown: Correct-by-Design Control Synthesis for Networks of Stochastic Uncertain Systems"*. In: *2023 62nd IEEE Conference on Decision and Control (CDC)*, Singapore, 2023, pp 7035–7042, DOI 10.1109/CDC49753.2023.10383307.

Oliver Schön, Shammakh Naseer, Ben Wooding, and Sadegh Soudjani (2024). *"Data-Driven Abstractions via Binary-Tree Gaussian Processes for Formal Verification"*. *IFAC-PapersOnLine*, 58(11):115–122, In: *The 8th IFAC Conference on Analysis and Design of Hybrid Systems (ADHS)*, Boulder, CO, USA, 2024, DOI 10.1016/j.ifacol.2024.07.434.

Negar Monir, **Oliver Schön**, and Sadegh Soudjani (2024). *"Lyapunov-Based Policy Synthesis for Multi-Objective Interval MDPs"*. IFAC-PapersOnLine, 58(11):99–106, In: *The 8th IFAC Conference on Analysis and Design of Hybrid Systems (ADHS)*, Boulder, CO, USA, 2024, DOI 10.1016/j.ifacol.2024.07.432.

Oliver Schön, Zhengang Zhong, and Sadegh Soudjani (2024). *"Data-Driven Distributionally Robust Safety Verification Using Barrier Certificates and Conditional Mean Embeddings"*. In: *American Control Conference (ACC)*, Toronto, ON, Canada, 2024, pp 3417–3423, DOI 10.23919/ACC60939.2024.10644349.

Oliver Schön, Sofie Haesaert, and Sadegh Soudjani (2025). *"Formal Control for Uncertain Systems via Contract-Based Probabilistic Surrogates"*. In: *Quantitative Evaluation of Systems and Formal Modeling and Analysis of Timed Systems (QEST)*, Aarhus, Denmark, 2025, pp 62–82, DOI 10.1007/978-3-032-05792-1_4.

Tool Papers and Competitions

Alessandro Abate, Henk Blom, Joanna Delicaris, Sofie Haesaert, Arnd Hartmanns, Birgit van Huijgevoort, Abolfazl Lavaei, Hao Ma, Mathis Niehage, Anne Remke, **Oliver Schön**, Stefan Schupp, Sadegh Soudjani, and Lisa Willemsen (2022). *"ARCH-COMP22 Category Report: Stochastic Models"*. In: *International Workshop on Applied Verification of Continuous and Hybrid Systems (ARCH)*, 90:113–141. EasyChair. DOI 10.29007/lsvc.

Birgit van Huijgevoort, **Oliver Schön**, Sadegh Soudjani, and Sofie Haesaert (2023). *"SySCoRe: Synthesis via Stochastic Coupling Relations"*. In: Proceedings of the *26th ACM International Conference on Hybrid Systems: Computation and Control (HSCC)*, San Antonio, TX, USA, 2023. ACM. DOI 10.1145/3575870.3587123.

Alessandro Abate, Henk Blom, Nathalie Cauchi, Joanna Delicaris, Sofie Haesaert, Birgit van Huijgevoort, Abolfazl Lavaei, Anne Remke, **Oliver Schön**, Stefan Schupp, Fedor Shmarov, Sadegh Soudjani, Lisa Willemsen, and Paolo Zuliani (2023). *"ARCH-COMP23 Category Report: Stochastic Models"*. In: *International Workshop on Applied Verification of Continuous and Hybrid Systems (ARCH)*, 96:126–150. EasyChair. DOI 10.29007/k7s6.

Ernesto Casablanca, **Oliver Schön**, Paolo Zuliani, Sadegh Soudjani (2026). *"LUCID: Learning-Enabled Uncertainty-Aware Certification of Stochastic Dynamical Systems"*. Accepted to *AAAI Conference on Artificial Intelligence*, Singapore, 2026.

Posters

Oliver Schön, Birgit van Huijgevoort, Sofie Haesaert, and Sadegh Soudjani (2023). *"Poster Abstract: Data-Driven Correct-by-Design Control of Parametric Stochastic Systems"*. In: Proceedings of the *26th ACM International Conference on Hybrid Systems: Computation and Control (HSCC)*, San Antonio, TX, USA, 2023. ACM, 2 pages. DOI 10.1145/3575870.3589547.