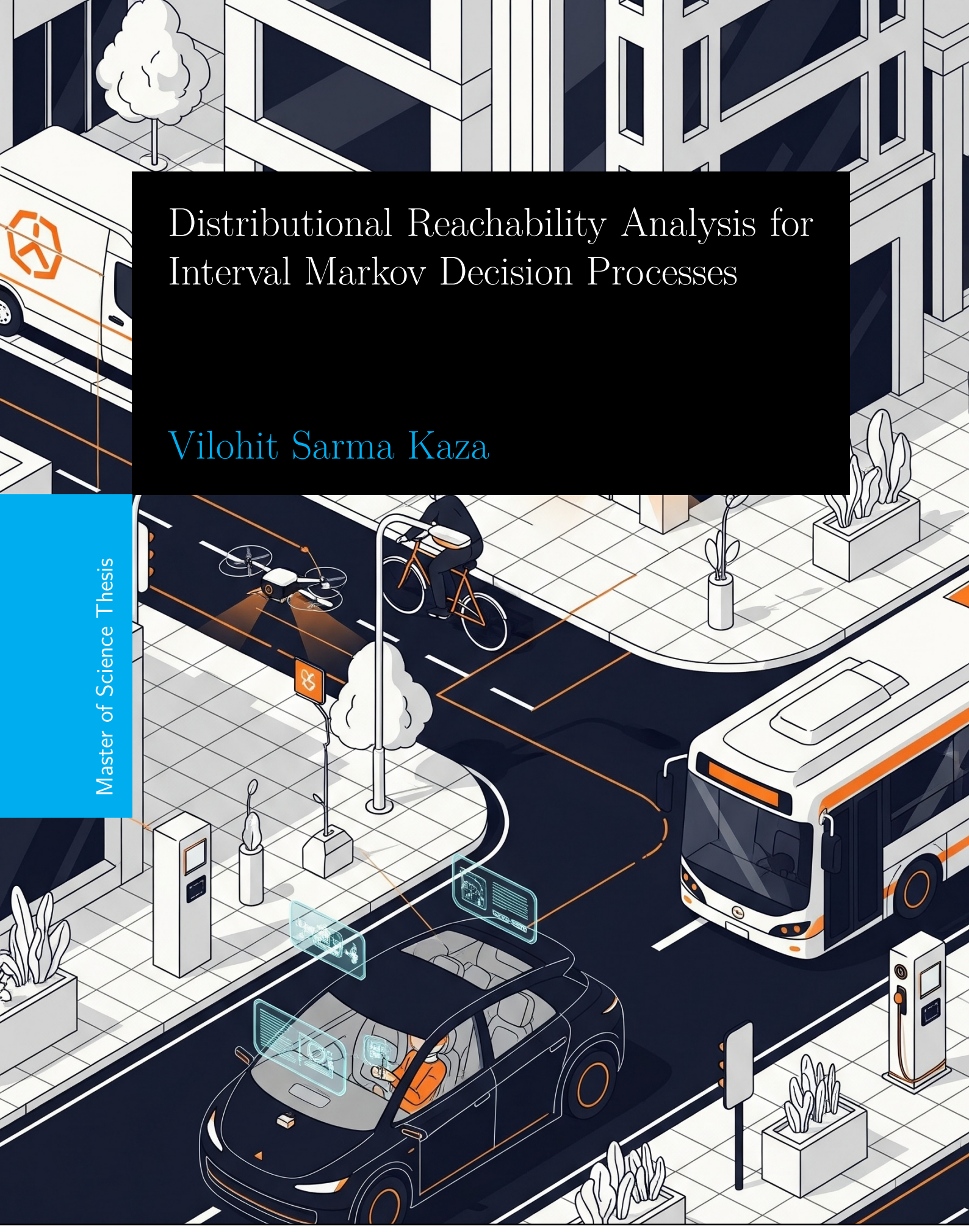


Distributional Reachability Analysis for Interval Markov Decision Processes

Vilohit Sarma Kaza

Master of Science Thesis

TU Delft Delft University of Technology

Delft Center for Systems and Control

Distributional Reachability Analysis for Interval Markov Decision Processes

Master of Science Thesis

For the degree of Master of Science in Computer and Embedded Systems Engineering at Delft University of Technology

Vilohit Sarma Kaza

August 20, 2025

Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS) and Mechanical Engineering (ME) · Delft University of Technology

Cover illustration generated using Gemini.

Delft University of Technology
Department of
Delft Center for Systems and Control (dcsc)

The undersigned hereby certify that they have read and recommend to the Faculty of Electrical Engineering, Mathematics and Computer Science (EEMCS) and Mechanical Engineering (ME) for acceptance a thesis entitled

Distributional Reachability Analysis for Interval Markov Decision Processes

by

Vilohit Sarma Kaza

in partial fulfillment of the requirements for the degree of

Master of Science Computer and Embedded Systems Engineering

Dated: August 20, 2025

Supervisor(s):

Dr. Luca Laurenti

Frederik Mathiesen

Reader(s):

Committee Chair: Dr. Manuel Mazo Jr.

Committee Member: Dr. Azita Dabiri

Abstract

The formal verification of multi-agent systems in safety-critical domains is challenged by the need to certify population-level behaviours, such as formation control, under environmental uncertainty. Traditional state-based verification techniques are often inadequate for expressing these system-wide objectives and face scalability limitations. This thesis addresses this gap by developing a distributional reachability framework that models the evolution of the system directly over the space of probability distributions, using Interval Markov Decision Processes (IMDPs) to capture model uncertainty. We introduce two complementary analysis algorithms to compute guaranteed bounds on the set of all reachable distributions: a forward method using occupation measures and McCormick relaxations, and a robust backward algorithm based on value iteration over a discretised distribution space. Case studies in swarm deployment demonstrate the efficacy of the framework in computing robust, set-based approximations of reachable distributions. Furthermore, results for the robust backward reachability algorithm are presented for a running example. This capability allows for the formal verification of complex distributional specifications and the synthesis of control policies with certified safety guarantees, establishing a computational foundation for designing certifiably safe, large-scale autonomous systems.

Contents

List of Figures

List of Tables

Preface & Acknowledgements

This thesis marks the end of a wild ride at Delft. Given the rise of autonomous systems, the topic of this thesis addresses an important overarching problem of robust safety-critical systems. I wanted to view this through a different lens which made the whole journey super interesting. The decision to return to academia after working professionally was a significant one, and the journey was made all the more demanding by switching specialisations and tackling a thesis in an entirely new field for me. This work represents the culmination of that challenging transition. It was a process that pushed my boundaries and taught me a great deal about resilience and the excitement of diving into the unknown.

Many people were a part of this journey. I would like to express my sincere gratitude to my thesis supervisors, Dr. Luca Laurenti and Frederik Mathiesen, for giving me the absolute freedom during my thesis and for letting me pick their brains with so many questions and discussions. Their invaluable guidance and insightful feedback were instrumental in shaping the direction of this thesis and navigating its complexities. Thank you guys for being so kind and patient toward me, even when I was running in a hundred directions and diving into so many rabbit holes. I could only imagine how frustrating it would have been to deal with me. There were many times when I was filled with self-doubt, but I did not know how to express it. It was only due to your kindness and encouragement that I could get past those feelings and do this work. I also thank Dr. Manuel Mazo Jr. and Dr. Azita Dabiri for agreeing to be a part of my thesis committee.

Finally, I would like to express a heartfelt thank you to my family and loved ones for their support at all times, and for always pushing me to do the things I wanted to do, even when I did not believe in myself. You guys are the best, and my writing capabilities do not do justice to what I feel about you all. In a journey that was mostly filled with lows than highs, I cannot imagine finishing my master's without your love and support.

Delft — Vilohit Sarma Kaza
August 20, 2025

To my parents...

उद्धरेदात्मनात्मानं नात्मानमवसादयेत् |
आत्मैव ह्यात्मनो बन्धुरात्मैव रिपुरात्मन : ||

Chapter 1

Introduction

Chapter Summary
This chapter introduces the problem of verifying reachability specifications in multi-agent systems operating under stochastic dynamics and model uncertainty. Safety and reachability are dual notions: verifying that a system avoids unsafe states (safety) can be reformulated as checking that it never reaches those states (reachability). Traditional state-based approaches are limited in multi-agent settings, where distribution-level objectives or specifications could be more relevant. To address this, we adopt Interval Markov Decision Processes (IMDPs), which use bounded transition probabilities to model uncertainty in transition probabilities. This allows us to reason about the evolution of agent distributions and forms the foundation for the distributional reachability analysis developed in this thesis.

1-1 Motivation

Autonomous systems are being increasingly deployed in safety-critical environments, under both stochastic dynamics and uncertain system environments. These include domains such as autonomous vehicles, multi-agent systems, air traffic management, and swarm robotics (Brambilla et al., 2013; Schranz et al., 2020). Such systems must operate safely, ensuring a minimum chance of an accident. An accident refers to a situation where a system behaves unexpectedly, resulting in undesirable or harmful outcomes instead of operating as intended or specified (Amodei et al., 2016). A common theme across domains is the need to guarantee that the system satisfies certain *reach-avoid* specifications, that it eventually reaches a target objective while avoiding unsafe states or situations along the way, despite model uncertainty and environmental disturbances.

Verification methods for stochastic systems can be classified into abstraction-free and abstraction-based approaches (Lavaei et al., 2022). Abstraction-free methods reason directly over the system dynamics, typically by constructing stochastic barrier functions, Lyapunov-like functions whose existence certifies safety (Santoyo et al., 2021), or by solving functional inequalities

(Prajna et al., 2007). These approaches often rely on restricted function classes (e.g., polynomials, piecewise linear, or sum-of-squares) to ensure computational tractability (Santoyo et al., 2021; Mazouz et al., 2024). Such structural limitations can lead to conservative results and hinder scalability, especially for high-dimensional or non-polynomial systems (Lavaei et al., 2022). In contrast, abstraction-based methods approximate the continuous system with a finite-state model, enabling the use of formal synthesis and model-checking tools for verification (Alur et al., 2000). Conceptually, both approaches can be unified under a dynamic programming framework, where abstraction-free methods yield conservative approximations of value functions computed over finite abstractions (Laurenti and Lahijanian, 2023). This perspective motivates the use of abstraction-based methods in this thesis.

A framework for reasoning about sequential decision making under uncertainty is the Markov Decision Process (MDP), which models the evolution of the system as a controlled stochastic process with a Markov property, which means that the future state of a system depends on the current state and the action taken and not the past states (Baier and Katoen, 2008). In classical MDP verification (Baier and Katoen, 2008; Forejt et al., 2011), one typically analyses properties such as reachability, which deals with determining whether the system can reach a desirable target set of states. This has been extensively studied through state-based formulations that focus on the behaviour of individual state-based trajectories (Baier and Katoen, 2008; Forejt et al., 2011). These verification methods have found considerable success in domains where the state space is small or where individual state-based outcomes are directly relevant (Kwiatkowska et al., 2011; Soudjani et al., 2015; Wooding and Lavaei, 2024).

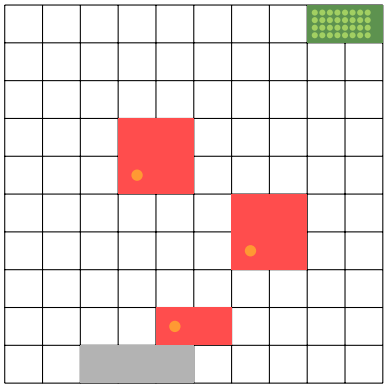
However, many modern applications, particularly in multi-agent or swarm systems, are no longer best described solely by individual state trajectories (Akshay et al., 2023). Instead, the relevant properties often pertain to the collective behaviour of populations of agents, where global system safety depends not on the state of any single agent but on how the distribution of agents evolves. This motivates a shift in viewpoint from reasoning over states to reasoning over state distributions, leading to the distributional perspective of reachability.

1-2 Motivating Example

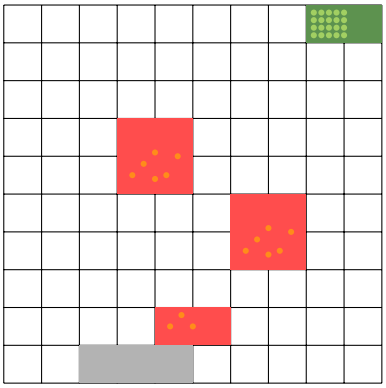
Modern autonomous systems, such as drone swarms or robotic fleets, require decision-making at multiple levels of abstraction. While low-level controllers govern actuation and trajectory tracking, higher-level planners reason about global safety, coordination, and specification satisfaction across large agent populations. In this thesis, we focus on this upper layer of abstraction, commonly referred to as *fleet-level planning* or *distributional coordination* (Schwarting et al., 2018), where the objective is not to track individual paths, but to reason about how aggregate distributions of agents evolve and whether they satisfy system-level safety and reachability constraints.

Consider a swarm of 100 drones operating on a discrete 10×10 grid, where each cell represents either a physical location or an abstract state. The drones aim to navigate from initial regions, such as warehouse launch pads, to designated delivery zones while avoiding restricted or hazardous areas, including no-fly zones, buildings, or sensitive airspace sectors. At each time step, a drone selects an action; however, environmental factors such as wind, sensor noise,

or actuator imprecision introduce stochasticity into the executed transitions. For instance, a drone intending to move north may instead drift northeast or remain stationary due to disturbances.



(a) Case 1: Most agents successfully reach target, few near unsafe regions.



(b) Case 2: More agents dispersed near unsafe regions.

Figure 1-1: Motivating example illustrating the difference between safe and unsafe distributions under identical marginal probabilities. The grey region indicates the launch pads, green denotes the target region, red indicates no-fly (unsafe) zones, and orange dots represent agents located near unsafe regions. Green dots represent agents that have reached the target.

Classical state-based analysis typically monitors the marginal probability of occupancy for each grid cell, i.e., the probability that an arbitrary agent resides in a particular state at a given time. For example, suppose that under a given policy, the marginal probability of a drone being in the target region (cells $(0, 8)$ to $(0, 9)$ in Figure 1-1) is 0.6, while the total marginal probability assigned to restricted regions is 0.02. Although these values quantify state-wise occupancy, they fail to capture the *global structure* of the agent distribution across the state space.

In fact, different population-level configurations can produce identical marginals but differ significantly in terms of safety. As illustrated in Figure 1-1:

- *Case 1*: The distribution is concentrated in the target region, with minimal presence near unsafe zones.
- *Case 2*: Although the marginal statistics are identical, the distribution is more dispersed, with agents clustered near several restricted zones. This can lead to localised safety violations and reduced mission effectiveness at the system level.

These scenarios highlight a fundamental limitation of marginal-based reasoning: many multi-agent safety specifications—such as bounding the number of agents near unsafe areas—are inherently *distributional*, and cannot be verified by state-wise probabilities alone. Such specifications depend on the joint occupancy pattern of agents across the state space.

Moreover, modelling each agent explicitly would require tracking the full joint state space of dimension $|S|^N$, which becomes intractable for large swarms (e.g., 100^{100} for 100 agents on a 100-cell grid). To address both expressiveness and scalability, we adopt a *distributional*

framework (Korthikanti et al., 2010), where the state of the system is represented as a probability distribution over the grid. This approach enables direct reasoning about the evolution of the population over the probability simplex $\Delta^{|S|}$, providing a compact and computationally feasible abstraction for verifying population-level reachability and safety under uncertainty. Although less expressive than full joint-state modelling, this perspective is well-suited to scenarios where aggregate behaviour governs system-level correctness.

This leads to the formal study of distributional reachability, characterising how probability distributions over state space evolve under both stochastic dynamics and policy choices (Korthikanti et al., 2010; Chadha et al., 2011; Akshay et al., 2018). In the distributional view, the system is no longer treated merely as a generator of individual random paths but rather as a transformer of distributions, mapping the current distribution of agents across the state space to future distributions through its probabilistic transition structure under the influence of a policy.

While the distributional perspective offers a natural modelling framework, practical verification is further complicated by the presence of model uncertainty. In real-world applications, exact transition probabilities are rarely known (Lavaei et al., 2023). They may instead be estimated from empirical data, subject to noise, or specified only approximately due to limited modelling capabilities called the epistemic uncertainty (Badings et al., 2023a). This uncertainty in the model parameters gives rise to the need for robust verification, where safety guarantees must hold across all plausible models consistent with the available information.

The interval Markov Decision Processes (IMDPs) framework provides a method to capture such transition uncertainty, where each transition probability is specified within a bounded interval rather than as a single deterministic value (Givan et al., 2000). In our motivating gridworld example, these intervals could represent uncertainty in wind estimation, variability in drone battery power that affects thrust, or unmodelled environmental disturbances. Designing controllers for such systems requires accounting for the worst-case behaviour across all transition realisations allowed by these intervals.

This motivates the central problem studied in this thesis: how to characterise the set of all state distributions that can be reached in finite time, starting from an uncertain initial distribution and evolving under both stochastic policies and adversarial transition uncertainty modelled via IMDPs. The focus is on developing verification and synthesis procedures that reason directly over reachable sets of distributions, accounting for both control decisions and worst-case realisations of transition uncertainties.

1-3 Contribution

The contributions of this thesis are centred on the development of a computational framework for verifying the distributional reachability of IMDPs with applications to multi-agent systems. We formally define the problem of forward distributional reachability, focusing on computing the set of all possible state distributions that arise from uncertain transitions and stochastic policies. A key contribution is the characterisation of reachable sets as convex sets that evolve through affine transformations, which are induced by both the policy and adversarial transition choices. We propose a recursive algorithm to compute these k-step reachable sets by propagating the convex sets under these uncertain transformations. We also provide

a value iteration-based backward reachability method to characterise the largest initial distribution set from which we are guaranteed to reach the target set of distribution in a given horizon. The practical application of this framework is demonstrated through a case study in multi-agent reachability problem and a demonstration of the robust backward reachability algorithm, illustrating how the developed methods can verify reachability properties under uncertainty.

1-4 Report Outline

The remainder of this report is structured to systematically develop the proposed framework. Chapter 2 provides a review of the relevant literature on probabilistic verification, robust control, and distributional analysis, situating this thesis within the current state-of-the-art. Chapter 3 establishes the necessary mathematical background, formally introducing Robust MDPs (RMDPs) and IMDPs and the distribution transformer perspective. The core contributions are presented in the subsequent chapters. Chapter 4 formulates and addresses the forward reachability problem, detailing a method for policy synthesis using occupation measures and convex relaxations. Complementing this, Chapter 5 develops the backward reachability problem, introducing a robust value iteration-based algorithm to compute the largest initial set of distributions that are guaranteed to reach a target set of distributions. Chapter 6 validates these methods through two distinct case studies: a multi-agent swarm deployment and a detailed analysis of the robust backward reachability algorithm. Finally, Chapter 7 concludes the thesis by summarising the contributions, discussing the limitations of the work, and proposing directions for future research.

Chapter 2

Related Work

Chapter Summary
This chapter reviews safety verification methods for stochastic systems, with an emphasis on abstraction-based techniques using MDPs and IMDPs. Existing approaches predominantly focus on verifying properties at the state level, assuming either exact or bounded transition probabilities. Since safety and reachability are dual concepts, many of these methods can also be adapted for reachability analysis. However, they generally reason about individual state trajectories and do not address the evolution of distributions over state space, particularly in the context of multi-agent systems under uncertainty. This review highlights this gap and motivates the need for distribution-level verification, which is the focus of this thesis.

In safety-critical domains such as autonomous driving and drone swarms, it is important to prevent unsafe states and quantify how likely the systems are to reach or avoid critical regions (so-called unsafe regions) under uncertainty (Abate et al., 2008). Safety ("nothing bad happens") and reachability ("will the system end up in some state(s)?") are dual concepts (Baier and Katoen, 2008). Classical MDPs provide a probabilistic framework for safety and reachability analysis using policy iteration, value iteration, maximal end component decomposition (Haddad and Monmege, 2018) and set invariance (Blanchini and Miani, 2007) to certify that 'unsafe' states are avoided with a specified confidence level. However, these methods typically assume that the transition kernels are either precisely known or are parametric with unknown but fixed parameters. This assumption limits their applicability to real-world systems (Jafarpour and Coogan, 2023), where the transition probabilities are often uncertain due to limited observational data or incomplete knowledge about the system environment or dynamics. RMDPs address this by treating transition probabilities as adversially perturbed within confidence sets, yielding worst-case guarantees via minmax optimization (Nilim and Ghaoui, 2003; Wiesemann et al., 2013). IMDPs, a special case of RMDPs, assume the transition probabilities to be bounded within an interval of values while solving the minmax optimisation problem for a given objective (Givan et al., 2000). These approaches evaluate safety by considering the state trajectories of the system. However, applying these methods to multi-agent systems is often computationally challenging. This is due to the need to evaluate

the joint state space of all the agents over time, which grows rapidly with the increase in the number of agents in the system. In contrast, distributional reachability of MDPs seeks to bound the probability of the distributions over the state space reaching a given distribution, which also ensures the safety of the system (Akshay et al., 2024; Gao et al., 2023). In this chapter, we present a survey of foundational methods for safety verification based on MDPs and IMDPs, and discuss their relation to probabilistic reachability. We then identify a gap in the existing literature about the lack of distribution-level reachability guarantees under uncertainty, thereby motivating the framework introduced in chapter 4.

2-1 Probabilistic Safety and Reachability

MDPs provide a framework for modelling probabilistic systems with non-determinism and for synthesizing control policies that optimise a given objective (Puterman, 1994; Baier and Katoen, 2008). In a reachability problem, e.g., guiding a robot through a 2-D grid world to a target set and some unsafe set of states, the objective is to *maximise* the probability of eventually reaching that set. In contrast, the safety problem would be to *maximise* the probability of never reaching unsafe sets or *minimise* the probability of leaving a set of safe states. By the duality of safety and reachability, solving the reachability problem also yields almost-sure safety guarantees, meaning that the system satisfies the safety specification with probability one (Abate et al., 2008).

A classical solution technique to solve the reachability problem is *value iteration*, which applies successive Bellman updates to converge to the unique reachability probabilities (Forejt et al., 2011). However, value iteration converges only in the limit and lacks a concrete stopping criterion: it provides upper and lower bounds that tighten over time, but one cannot guarantee in finite steps that the exact value has been reached (Haddad and Monmege, 2018). Although Chatterjee and Henzinger (2008) derive an upper bound on the number of iterations needed for convergence, this bound is overly pessimistic and impractical for exact value computation in large MDPs (Brázdil et al., 2025).

To overcome these limitations, more recent work has leveraged the underlying graph structure of MDPs to obtain finite-time guarantees. In particular, the decomposition of MDP graphs into *maximal end components* (MECs), subsets of states from which the system cannot escape, allows the isolation of recurrent behaviour and the reduction of the effective state space (Henzinger and Chatterjee, 2011; Wijs et al., 2014). These decomposition techniques improve convergence properties and algorithmic efficiency, but require full knowledge of the MDP graph in advance and are less suited for dynamically constructed or partially specified models.

Building on these insights, Haddad and Monmege (2018) propose an *interval iteration* algorithm that maintains explicit upper and lower bounds on reachability probabilities and refines two coupled sequences of value vectors. They prove convergence-rate bounds and propose an explicit stopping criterion that guarantees optimal policies, while still respecting the original MEC structure. This was extended by Brázdil et al. (2025) to support dynamic graphs with on-the-fly addition or deletion of transitions, while maintaining convergence guarantees.

Although much of the work presented until now focused on safety verification, reachability poses its dual: ensuring that the system can reach a target set with high probability. This perspective naturally leads to reach-avoid formulations, which have become central in recent

approaches to probabilistic reachability, particularly in systems where ensuring progress toward desired outcomes is as important as avoiding failure (Baier and Katoen, 2008; Forejt et al., 2011; Chatterjee and Henzinger, 2008; Haddad and Monmege, 2018). With increasing demands to verify and control high-dimensional stochastic systems, such as multi-agent networks or population-level models, classical state-based formulations of MDPs become computationally infeasible due to exponential state space growth. To address this, recent work has adopted a *distribution transformer* perspective (Gao et al., 2023; Akshay et al., 2024, 2023, 2018), modelling MDPs as transformers of probability distributions over states rather than as transition systems over individual states. This abstraction enables reasoning directly over population-level behaviours and synthesising policies to drive the system to a desired distribution with a certain confidence level. It also provides a more natural framework (Akshay et al., 2024) for verifying specifications of systems, especially in scenarios where safety and performance are defined over distributions rather than specific paths or configurations.

Korthikanti et al. (2010) propose analyzing MDPs by studying how the entire distribution over system states evolves over time under a policy, instead of focusing on individual trajectories. They develop a logic for expressing distribution-based properties and show that verification is undecidable in general, but becomes decidable under restrictions such as finite-memory or stateless policies. Chadha et al. (2011) build on this by identifying conditions under which, regardless of the policy used, the evolving distributions eventually enter and remain within a fixed, compact set. This behavior allows properties to be checked using fixed-point computations over this invariant set. Akshay et al. (2018) further extend this line of work by introducing distribution-based specifications for safety and reachability, and analyze the complexity of verifying whether such objectives can be achieved from a given initial distribution

Building on this, Akshay et al. (2023) developed a symbolic approach for safety verification using affine-invariant constraints over reachable distributions. By constructing inductive invariants and reducing verification to solving affine inequalities, the method achieves relative completeness within a fixed template. However, its connection to the Skolem problem renders the synthesis step undecidable in general, posing computational challenges.

In contrast, Gao et al. (2023) introduced set-valued maps to characterise forward and backward distributional reachability in finite Markov Decision Processes (MDPs), providing insights into invariant sets and domains of attraction.

The shift to reasoning over distributions in MDPs aligns with developments in control theory, where the focus has moved from analyzing individual trajectories to guiding the evolution of state distributions. While previous works addressed verification and synthesis in discrete models, similar ideas have been applied to systems with uncertainty, continuous dynamics, and limited data. The following approaches extend distributional reachability to such settings, particularly in control and multi-agent systems.

Chen et al. (2016a,b) study how to steer the state distribution of stochastic linear systems toward a desired target. Yang et al. (2022) addressed scalability in backward reachability analysis by developing algorithms that under-approximate backward reachable sets using constrained zonotopes. Their approach efficiently handles discrete-time uncertain linear and nonlinear systems, facilitating controller synthesis that ensures system correctness.

In the context of multi-agent systems with unknown dynamics, Meshkat Alsadat et al. (2024) developed the ODMU method, which uses limited data and side information to over-

approximate reachable sets and achieve near-optimal control. This approach is particularly relevant for real-time scenarios where system dynamics may change abruptly. Lastly, Lei et al. (2023) proposed a finite-time adaptive distributed optimisation framework for uncertain nonlinear multi-agent systems. Their two-stage approach combines an optimal estimator with an adaptive tracking controller, ensuring convergence to optimal solutions despite system uncertainties.

Together, these works illustrate a paradigm shift in formal methods, from exact, state-based analysis to scalable, distribution-based reasoning. It also lays the groundwork for incorporating robustness into the analysis, which is addressed next through robust and interval MDP frameworks.

Robust MDPs (RMDPs) and Interval MDPs (IMDPs) In real-world applications, safety-critical systems often operate under uncertain or partially known dynamics, where exact transition probabilities are difficult to obtain or inherently variable. To account for such uncertainty, the framework of *Robust* Markov Decision Processes (RMDPs) augments classical MDPs bounded by uncertainty sets, called ambiguity sets, over transitions. These models are useful in analysing and synthesising policies that optimise performance under worst-case realisations, providing formal guarantees for safety and reachability objectives.

Foundational work by (Iyengar, 2005) and (Nilim and El Ghaoui, 2005) formulated the RMDP problem as a minimax optimisation over uncertain transition kernels, offering tractable solutions under rectangular uncertainty sets. Building upon these formulations, (Wiesemann et al., 2013) unified various robustness models within a single framework, characterising exact and conservative conditions under which robust policies can be efficiently computed. More recent advances extend these formulations to settings with epistemic uncertainty, distributional robustness, and ambiguity sets based on Wasserstein distances (Mazumdar et al., 2024; Badings et al., 2023a), supporting both model-based verification and learning-based synthesis.

A subclass of RMDPs, known as Interval MDPs (IMDPs), specifies ambiguity sets via closed intervals. Introduced by (Givan et al., 2000), IMDPs abstract probabilistic systems with bounded noise or estimation error, enabling conservative guarantees for reachability. They introduce algorithms for IMDPs to maximise the upper and lower bounds of a *value function* for a state ordering, and also present methods to support policy evaluation and synthesis. However, these methods suffer from convergence issues due to the inherent nonlinearity introduced by interval uncertainty. To address this, (Haddad and Monmege, 2018) proposed the interval-iteration algorithm, providing finite convergence guarantees and explicit error bounds on safety probabilities.

Traditional IMDPs assume discrete action spaces and static uncertainty, limiting their applicability in continuous control settings. To overcome this, (Delimpaltadakis et al., 2023) introduced continuous-action IMDPs (caIMDPs), wherein transition probabilities are modelled as functions over continuous control variables. Their approach yields tighter safety bounds in structured domains but remains computationally expensive in high-dimensional settings. A complementary line of work by Jafarpour and Coogan (2023) studies continuous-action IMDPs from a dynamical systems perspective, proposing iterative refinement methods using contraction theory.

Further extensions to multi-objective settings and policy refinement are explored in works like (Hahn et al., 2017), which develop robust Pareto-optimal policy synthesis for multiple

reachability goals under interval uncertainty. Entropy-regularised techniques for IMDPs have also been proposed to strike a balance between conservatism and policy exploration (van Zutphen et al., 2024), particularly relevant in learning-driven control. Recent surveys consolidate the landscape of robust and interval-based decision models, emphasising their applicability to formal verification and AI systems under uncertainty (Suilen et al., 2024).

Collectively, these frameworks offer methods for modelling and verifying probabilistic systems under uncertainty. Although they provide strong guarantees for reachability and safety, computational scalability and policy expressiveness remain open challenges (Lavaei et al., 2022).

The methods described until now rely on *probabilistic model checking tools* to verify the specifications of stochastic systems. Tools such as `PRISM` (Kwiatkowska et al., 2011) and `Storm` (Hensel et al., 2022) support discrete and continuous-time Markov models using value iteration. For IMDP-specific tasks, tools like `bmdp-tool`, `IMPaCT` (Wooding and Lavaei, 2024), and `IntervalMDP.jl` (Mathiesen et al., 2024) exist. `IMPaCT` supports parallel controller synthesis and abstraction for stochastic systems, while `IntervalMDP.jl` focuses on GPU-accelerated reachability and reward optimisation for a given IMDP.

2-2 Abstraction Methods

The finite-state IMDP reachability algorithms presented above require a discrete model with finitely many states and explicit interval uncertainty. However, many safety-critical systems, such as swarm robots, stochastic hybrid processes, or neural network controllers, evolve over continuous or uncountable state spaces (Zamani et al., 2016). Abstraction techniques for probabilistic safety and reachability reduce systems with continuous or uncountable state spaces to finite-state models, such as MCs, MDPs, IMDPs, or IMCs, over-approximate dynamics and uncertainty up to quantifiable error. Early methods partitioned the state space uniformly and computed interval-valued transitions via Lipschitz-based estimates, thereby obtaining finite MCs whose error between the MC and the underlying system vanishes as the grid is refined (Abate et al., 2010). Such uniform abstractions, however, suffer from exponential growth in the number of regions as the dimension increases.

To mitigate this curse of dimensionality, subsequent work exploited structural properties of the system dynamics. Lahijanian et al. (2012) introduced an MC–based abstraction method for discrete-time linear stochastic systems with bounded noise, offering tighter abstraction error bounds than the approach proposed by Abate et al. (2010), under a specified threshold on the error bound. Lahijanian et al. (2015) extend this by developing abstractions based on IMCs and IMDPs for switched stochastic systems. Their approach involves partitioning the state space into polytopic regions and computing exact transition probability bounds through min-max operations that explicitly capture system uncertainty. Furthermore, they also support the verification of PCTL formulae. Their proposed refinement method to reduce the uncertainty in system modelling faces scalability challenges due to the exponential growth of partitioned regions, a consequence of the curse of dimensionality.

To reduce conservatism in abstraction error bounds,(Cauchi et al., 2019) proposed incorporating exact abstraction errors into the abstraction model through a convex optimisation

framework. Unlike earlier approaches, this method explicitly embeds error terms into the abstraction itself, thereby limiting the propagation of uncertainty in the transition probabilities and bounding error accumulation over time. Robust optimisation techniques are employed to synthesise control policies capable of handling specification uncertainty, and the verification of both csLTL and BLTL properties is formulated as an adversarial game between the controller and a worst-case environment (Cauchi et al., 2019; Lahijanian et al., 2015). However, the approach is currently limited to stochastic hybrid systems (SHS) with linear dynamics.

Extensions to continuous-time stochastic differential equations (SDEs) have proposed discretisations that depend on the system dynamics. Laurenti et al. (2020) discretised both time and space for switched diffusions with linear SDEs, constructing IMDPs whose discretisation errors are captured in closed form. Robust synthesis over these abstractions yields switching strategies that guarantee continuous-time safety while highlighting the trade-off between step size for time and the resolution of the discretised state space grid.

Finally, to further reduce the curse of dimensionality, Mathiesen et al. (2025) introduced *orthogonally decoupled* IMDPs (odIMDPs), in which transition bounds factor into marginal intervals along each coordinate. By exploiting decomposability in the original stochastic dynamics, odIMDPs store only per-coordinate marginals and perform robust value iteration. Empirical results demonstrate tighter probabilistic guarantees than standard IMDP abstractions on benchmarks of up to seven dimensions.

These abstraction techniques produce finite IMDP models, enabling reachability analysis over the IMDPs. However, they assume known noise models and static partitions, and they often suffer from the curse of dimensionality.

2-3 Data-driven Approaches

The abstraction methods reviewed in Section 2-2 typically assume that the distribution of system disturbances—modelled as process noise—is known and follows a Gaussian profile. While this simplifies the analysis, it limits the applicability of such approaches to real-world systems, where the noise distribution may be unknown, non-Gaussian, or derived from sparse and noisy measurements. Moreover, constructing finite abstractions using system identification techniques can be computationally demanding and challenging to scale in high-dimensional settings (Badings et al., 2022; Lavaei et al., 2023). Data-driven approaches address these gaps by constructing finite MDPs or IMDPs from observed trajectories, computing upper and lower transition bounds that hold with prescribed confidence levels.

For MDPs, Lavaei et al. (2023) propose a scenario-based approach to construct finite MDP abstractions for discrete-time stochastic systems with unknown dynamics. They collect state–action–next-state samples, then solve scenario convex programs to bound the probability that the learned MDP deviates from the true system by more than a user-specified confidence level. This yields a finite MDP with probabilistic guarantees on each transition probability.

Similarly, for IMDPs, Badings et al. (2022) partition the continuous state space into convex cells, draw disturbance samples, and solve a scenario program to compute interval transitions for an IMDP. Control actions at each abstract state are synthesised via backward reachability on a nominal dynamics model. However, this method accounts only for aleatoric uncertainty

(randomness) and does not account for epistemic uncertainty in system parameters. In contrast to previous approaches, such as the single-target abstraction procedure, where an action from an abstract state is allowed only if the backward reachable set of the nominal system lies entirely within a single target partition cell, Coppola et al. (2024) propose a multi-target abstraction method. In this approach, transitions are permitted to sets of neighbouring regions, rather than to a single cell. By allowing transitions to cover sets of nearby partition regions, the abstraction retains more feasible actions and better approximates the true system dynamics. While this improves abstraction tightness, all these approaches rely on probably approximately correct (PAC) guarantees, which may still lead to overly conservative bounds, especially in under-sampled regions. Furthermore, these methods have been demonstrated mainly for systems with linear dynamics.

As a nonparametric alternative, Skovbekk et al. (2024) extend the GP-based safety framework of Jackson et al. (Jackson et al., 2020). They fit Gaussian processes to transition data, derive RKHS-based error bounds on both regression and discretisation, and compute interval transitions for the resulting IMDP. While this accommodates complex, sub-Gaussian disturbances and unknown dynamics, it incurs cubic complexity in the number of samples and suffers from exponential growth in the finite model as dimensionality increases.

For non-linear stochastic systems with non-affine noise structure, Skovbekk et al. (2023) propose a disturbance partitioning scheme for non-linear systems with component-wise noise. By optimally binning the disturbance space guided by system monotonicity and clustering states with similar behaviour, they build a refinement-free IMC abstraction that tightens transition bounds and also addresses the state-space explosion problem. However, this method relies on monotonicity and independent-noise assumptions, and the optimality of the clustering strategy remains an open problem.

More recently, Reed et al. (2023) employed deep kernel learning (DKL) to obtain both a learned mean and rigorous variance estimates, then computed tight linear relaxations of these estimates over each region. The abstraction method presented requires only finitely many function value evaluations rather than full convex programs, improving scalability and preserving correctness on benchmarks of up to five dimensions.

In our motivating example from section 1-2, wind disturbances cannot be specified a priori, and only flight data are available. Data-driven abstraction methods, therefore, learn finite IMDP (or IMC) models directly from these samples, computing upper and lower transition bounds that hold with high confidence. Unlike classical abstraction methods, which assume Gaussian noise and manual partitions, these techniques infer both the disturbance model and the state grouping automatically, resulting in IMDPs on which the reachability algorithms from (2-1) can be applied to compute rigorous probability guarantees.

2-4 Discussion

The literature shows a recent shift of MDPs toward a ”distribution transformer” view for verifying multi-agent systems, as it better captures population-level objectives than traditional state-based methods (Akshay et al., 2024). In parallel, IMDPs and IMCs have been established as a standard framework for modelling the transition uncertainty inherent in these systems(Badings et al., 2023b; Lahijanian et al., 2015; Laurenti et al., 2020). However, a gap

exists at this intersection: prior distributional analysis has largely assumed precisely known dynamics, while work on IMDPs has focused on state-level reachability probabilities, not the geometry of reachable distribution sets. This thesis aims to bridge this divide by developing a computational framework specifically for the distributional reachability analysis of IMDPs, which has applications in multi-agent systems.

Chapter 3

Background

Chapter Summary
This chapter introduces the mathematical foundations for distributional reachability analysis. We motivate the need for a distribution-based view in settings like multi-agent navigation, where reasoning about populations is more relevant than individual trajectories. Core tools from set theory, probability, and convex analysis are presented, followed by a review of MDPs and their extensions, Robust and Interval MDPs, to model uncertainty in transitions. A running example is introduced and used in the rest of this report. We conclude with a geometric interpretation of system dynamics as transformations over probability distributions. These foundations support the reachability problem addressed in the next chapter.

3-1 Motivation for Distributional Reachability

This section motivates the central problem of this thesis and sets the stage for the modelling and theoretical tools introduced throughout this chapter. The next chapter formalises the problem and develops algorithms for distributional reachability verification.

Although safety verification has often been the primary objective in probabilistic systems, our focus here is on reachability, its dual. While safety requires that undesirable states be avoided at all times, reachability requires that a desired target set be reached at least once within the horizon. This distinction aligns with the classical formulation, where safety is an infinite-horizon property, whereas reachability is inherently finite-horizon (Summers and Lygeros, 2010; Baier and Katoen, 2008). This existential nature makes reachability a suitable model in many control and planning tasks (Yang et al., 2022), where the goal is not to avoid failure indefinitely, but to achieve success within a defined time horizon.

Moreover, focusing on reachability could allow us to pose richer and more flexible questions. In population-level systems, we often do not seek guarantees that every agent will succeed but that enough of them will, e.g., ”at least 90% of agents reach the target zone within 10 steps.”

Such questions are hard to pose or answer using only safety formulations. Reachability, in contrast, provides a direct framework for reasoning about such guarantees. To illustrate this, consider a swarm of autonomous agents deployed in a grid environment. Due to stochastic disturbances or partial observability, their movements are not deterministic; say, each agent begins in a known region and attempts to reach a goal zone, navigating uncertain terrain or weather conditions. The precise behaviour of each agent may vary, but we care about the overall distribution: what proportion of agents reach the goal, and what fraction might be lost to unsafe regions? These are questions of distributional reachability, where the goal is to compute or over-approximate the set of all state distributions that can arise under all admissible system evolutions.

In this setting, the key object of interest is the reachable set of distributions: all possible state distributions that can arise under all admissible transitions and policies within the time horizon. Understanding the structure of this set is essential for both verifying probabilistic specifications and synthesizing robust control strategies.

The central object of study in this chapter is the reachable set of distributions—the set of all probability distributions that the system can attain in a fixed number of steps, starting from a given initial distribution and under all valid policies and uncertainties. Rather than analyzing the system state by state, we take a geometric view of how distributions evolve under uncertain affine dynamics. Our goal is to understand the structure of this reachable set, identify when it is convex, and determine how it can be computed or bounded in practice.

This formulation allows us to ask and answer several key questions:

- How does one define and compute the one-step image of a distribution under uncertain transitions?
- How do such sets evolve over multiple steps, and what governs their shape?
- What role does policy selection play in shaping the boundary of the reachable set?
- Can these reachable sets be used to provide distribution-level guarantees for verification?

In the remainder of this chapter, we give precise definitions for these objects, develop the geometric tools needed to study them, and lay the foundation for the computational framework that follows.

3-2 Background

To characterise the distributional reachability of uncertain multi-agent systems, we rely on geometric tools from convex analysis, set theory, and reachability theory. These tools allow us to describe how sets of states evolve under stochastic dynamics under control policies. In this section, we summarise key concepts such as Minkowski sums, support functions, and reachable sets, all of which are foundational to the modelling and analysis presented in this chapter.

3-2-1 Computational Geometry

In analysing distributional reachability for multi-agent systems under uncertainty, it is essential to reason about how sets of possible state distributions evolve. Because exact computation of these reachable distributions is often infeasible in practice, especially in the presence of stochasticity and model uncertainty, we adopt a set-based approach that enables tractable approximations and formal verification.

We begin with the foundational notion of a *set*.

Definition 3.1 (Set (Boyd and Vandenberghe, 2004))**.** A set S is a collection of elements that share a common property or belong to a common space.

In our setting, these elements typically correspond to vectors in $\mathbb{R}^n$, such as state values, control inputs, or probability distributions. To support computational tractability and ensure desirable mathematical properties, we focus on sets that are (Boyd and Vandenberghe, 2004):

- **Compact**: bounded and closed, ensuring well-defined optimisation over the set.
- **Convex**: any convex combination of points in the set remains within the set, enabling efficient representation and computation.
- **Closed**: includes all its boundary points, ensuring robustness under limits and continuity.

Set Operations. To track how uncertainty evolves, we use basic geometric operations over sets. We define a couple of useful operations for our context below. The reader is referred to (Boyd and Vandenberghe, 2004, Chapter 2) for a detailed overview of these operations and other set properties.

Definition 3.2 (Scaled Set (Blanchini and Miani, 2007))**.** For $\lambda \geq 0$ and set $A \subseteq \mathbb{R}^n$, the scaled set is:

$$\lambda A = \{\lambda a \mid a \in A\}.$$

Definition 3.3 (Minkowski Sum (Berg et al., 2008))**.** The Minkowski sum of two sets $A, B \subseteq \mathbb{R}^n$ is defined as:

$$A \oplus B = \{a + b \mid a \in A,\ b \in B\}.$$

This operation is used to represent how disturbances expand the reachable set under a given policy. Similarly, scaling a set helps model the effect of scalar multiplication of vector elements.

These operations form the basis for computing reachable sets under uncertain transitions, which are central to our analysis.

Over-approximations. In the context of verification, it is often sufficient (and computationally preferable) to over-approximate the reachable set (Althoff and Frehse, 2016). That is, we seek a superset that *conservatively* contains all possible outcomes. If such an over-approximation avoids unsafe regions, we can guarantee the safety (or, in our case, reachability) of the true system.

Definition 3.4 (Convex Hull(Boyd and Vandenberghe, 2004))**.** The convex hull (conv) of a set C, denoted $\text{conv}(C)$, is the set of all convex combinations of points in C:

$$\text{conv}\, C = \left\{ \sum_{i=1}^{k} \theta_i x_i \;\middle|\; x_i \in C,\ \theta_i \geq 0,\ i = 1, \ldots, k,\ \sum_{i=1}^{k} \theta_i = 1 \right\}.$$

For any set C, $\text{conv}(C)$ can be seen as the smallest convex set containing C (Blanchini and Miani, 2007). Furthermore, we define a *polytope* P with vertices $(v_1, \ldots, v_n)$ as (Cauchi et al., 2019):

$$P = \text{conv}(v_1, \ldots, v_n)$$

We use $\text{conv}(\cdot)$ to form convex over-approximations of reachable distribution sets. This makes some of the computationally

Connection to System Dynamics. In our setup, each agent evolves under uncertain dynamics that include both deterministic and stochastic components. To formalise this, we first define the nominal (noise-free) part of the dynamics.

Definition 3.5 (Nominal Dynamics (Badings et al., 2023a))**.** Let the system evolve as:

$$x[k+1] = g(x[k], u[k]) + \xi[k],$$

where $x[k] \in \mathcal{X} \subset \mathbb{R}^n$, $u[k] \in \mathcal{U} \subset \mathbb{R}^p$, and $\xi[k]$ is a disturbance term. The *nominal dynamics* (i.e., with no disturbance) are:

$$x[k+1] = g(x[k], u[k]).$$

Understanding the nominal dynamics allows us to compute which states could have led to a given future state, a concept known as backward reachability.

Definition 3.6 (Backward Reachable Set (Badings et al., 2023a))**.** For a point $x' \in \mathcal{X}$, the backward reachable set under nominal dynamics is:

$$Back(x') = \{x \in \mathcal{X} \mid \exists u \in \mathcal{U} \text{ such that } x' = g(x, u)\}.$$

Backward reachable sets are especially useful in abstraction and policy synthesis, as they help identify feasible transitions from prior states and assist in constructing the transition bounds used in our IMDP model, as will be discussed in chapter 5.

3-2-2 Temporal Logic Specifications

To express the temporal goals of multi-agent systems, we adopt the standard syntax for bounded-horizon properties (see Chapter 5 Baier and Katoen, 2008). These properties describe how state trajectories interact with target and safe sets given a time horizon. For a discrete-time stochastic hybrid system in a state space S, with some measurable sets $P, Q \in S$ termed respectively 'safe' and 'target' sets, we define $\Box^{\leq T_s} P$ as the bounded-horizon safety property for all trajectories that start from safe set P and remain in it over a finite-time

horizon $k \in [0, T_s]$. Similarly, we define the bounded-horizon reachability, $\Diamond^{\leq T_d} Q$, if there is a $k \in [0, T_d]$ such that the state trajectory reaches Q in k steps. Finally, we define the reach-avoid specification as $P\mathbf{U}Q$ for the state trajectories to stay in a safe set P until the target set Q is reached (Lavaei et al., 2022). These logical forms enable formal reasoning about whether a given policy ensures distribution-level objectives, such as a minimum probability of reaching a target without violating safety.

3-2-3 Probability Theory

A probability distribution γ over a finite set S is a function $\gamma : S \to [0, 1]$, such that $\sum_{s \in S} \gamma(s) = 1$. Here, $\mathcal{D}(S)$ is the set of all probability distributions.

We define an *ambiguity set* as a subset of all distributions $\mathcal{D}(S)$. Further, an *interval* ambiguity set Γ is an ambiguity set where the distributions are lower and upper bounded by $\check{\gamma} : S \to [0, 1]$, and $\hat{\gamma} : S \to [0, 1]$, respectively, such that $\check{\gamma}(s) \leq \hat{\gamma}(s)$ for all $s \in S$ and $\sum_{s \in S} \check{\gamma}(s) \leq 1 \leq \sum_{s \in S} \hat{\gamma}(s)$. More specifically, an interval ambiguity set, $\Gamma \subset \mathcal{D}(S)$, is defined as:

$$\Gamma = \{\gamma \in \mathcal{D}(S) : \check{\gamma}(s) \leq \gamma(s) \leq \hat{\gamma}(s) \ \forall s \in S\}. \tag{3-1}$$

We denote the set of all interval ambiguity sets over S by $\mathbf{\Gamma}(S)$.

3-2-4 Markov Decision Processes

To model systems, *transition systems* are widely used (Baier and Katoen, 2008). These systems are represented as directed graphs, where the nodes are the *states* of the system, and the edges represent the *transitions*. *Transitions* specify how the system evolves.

Definition 3.7. (Transition System (Baier and Katoen, 2008)) A *transition system* can be defined as a tuple $(S, A, \rightarrow, I, AP, L)$, where S is a set of the states of the system, A is the set of actions, $\rightarrow \subseteq S \times A \times S$ is a transition relation, $I \subseteq S$ is a set of initial states, AP is a set of atomic propositions and $L : S \to 2^{AP}$ is a labelling function that maps states to a set of atomic propositions AP.

A transition $s \xrightarrow{\text{a}} s'$ describes the evolution of the system according to a transition $\rightarrow$ from a current state s to s' under an action $a \in A$. If there are multiple outgoing states from s, the next state of the system is chosen *non-deterministically*.

A transition system that has probabilistic transitions, instead of non-deterministic transitions, from one state to another is called a *Markov Chain (MC)*.

Definition 3.8. (Markov Chain (Haddad and Monmege, 2018)) A Markov chain is a tuple $\mathtt{M} = (S, \gamma, s_0)$ where S is a set of countable states $\gamma : S \times S \to [0, 1]$ is a probability transition function, where $\gamma_s(s')$ denotes the probability of a transition from state $s \to s'$ with $\sum_{s' \in S} \gamma_s(s') = 1$ for all $s \in S$, and $s_0 \in S$ is the initial state of the system.

MCs are limited in their ability to model systems interacting with uncertain or dynamic environments because they rely on fixed transition probabilities that represent a specific,

known environment. However, MCs cannot adapt to variations in the environment, rendering them ineffective when there is a non-determinism of choices involved.

A natural extension is to include non-determinism in the model. This results in a Markov Decision Process (MDP). Since there are different notations of MDPs like (Baier and Katoen, 2008; Puterman, 1994; Bellman, 1958) in the literature, here we present the one that is convenient for the rest of the report, which is based on (Haddad and Monmege, 2018).

Definition 3.9. (Markov Decision Process (Haddad and Monmege, 2018)) A *Markov Decision Process* (MDP) is a tuple $\mathcal{M} = (S, A, \gamma)$ where,

- S is a finite set of states,
- $A = \bigcup_{s \in S} A(s)$ where $A(s)$ is a non-empty finite set of actions for every state $s \in S$ with $A(s) \cap A(s') = \emptyset \ \forall s \neq s'$,
- and $\gamma : S \times A \to \mathcal{D}(S)$ is a partial probabilistic function defined for (s, a) iff $a \in A(s)$, where $\mathcal{D}(S)$ is the set of distributions over a finite set S such that for a distribution $\gamma_{s,a} \in \mathcal{D}(S)$, every mapping $\gamma_{s,a}(s') : S \times A \times S \to [0, 1]$ from S to the set $[0, 1]$ is such that $\sum_{s' \in S} \gamma_{s,a}(s') = 1$.

Remark. The notation of MDP in (Cauchi et al., 2019) is $\mathcal{M} = (S, A, P, AP, L)$ where S is a finite set of states, A is a finite set of actions, $P : S \times A \times S \to [0, 1]$ is a transition probability function, such that AP is a set of atomic propositions, and $L : S \to 2^{AP}$ is a labelling function that maps each state to possibly several labels of AP. The additional information of L, AP is useful in model checking or formal verification of systems of more complex properties, where the properties are usually defined using temporal logic specifications. For example, to verify whether a command is successfully sent over a communication channel, it is useful to have labels like "waiting for acknowledgement" or "data sent" (Baier and Katoen, 2008). Since these labels are not relevant in the context of this thesis, we omit L and AP; however, it is important to note that both notations represent the same concept in MDPs.

We define *support* of a distribution δ, $Supp(\delta) = \{s \in S \mid \delta(s) > 0\}$, where S is a finite set.

An MDP $\mathcal{M}$ evolves as follows: from a current state s, an action $a \in A(s)$ is chosen non-deterministically and the next state s' is chosen using the distribution $\gamma_{s,a}$ and the probability that s' is reached is given by $\gamma_{s,a}(s')$. We now define an *infinite path* of an MDP as a sequence $\rho = s_0 \xrightarrow{a_0} s_1 \xrightarrow{a_1} \dots$ where $s_i \in S$, $a_i \in A(s_i)$ and $s_{i+1} \in Supp(\gamma_{s_i,a_i}(\cdot))$ (Haddad and Monmege, 2018). Further, we use $\rho(i)$ to represent s_i. Similarly, a *finite path* of length $n + 1$ is a sequence $\rho_{\texttt{fin}} = s_0 \xrightarrow{a_0} s_1 \xrightarrow{a_1} \dots \xrightarrow{a_{n-1}} s_n$ ending in state $last(\rho_{\texttt{fin}}) = s_n$, a prefix of an infinite path ending in s_n. The sets of all finite and infinite paths are denoted by $\texttt{Paths}^{fin}$, $\texttt{Paths}$.

Definition 3.10. (Policy (Haddad and Monmege, 2018)) A *policy* of an MDP $\mathcal{M} = (S, A, \gamma)$ is a function $\pi : \texttt{Paths}^{fin} \to \mathcal{D}(A)$ such that $\pi(\rho_{\texttt{fin}})(a)$ is denoted as $\pi(a \mid \rho_{\texttt{fin}}) > 0$ only if $a \in A(last(\rho_{\texttt{fin}}))$.

A policy (also known as *strategy*) π is *deterministic* if $\pi(\rho_{\texttt{fin}})$ is such that only one action has a probability of 1, a so-called Dirac distribution, for all $\rho_{\texttt{fin}} \in \texttt{Paths}^{fin}$. Further, a policy is *stationary* or *memoryless* if the distribution $\pi(\rho_{\texttt{fin}})$ depends only on $last(\rho_{\texttt{fin}})$.

Given a policy π and an initial state $s_0 \in S$, an MDP is equivalent to an MC with the states as finite paths of $\texttt{Paths}^{fin}$ (Forejt et al., 2011). The probability measure over the paths of the MC starting in s will give us the probability measure over the set of all paths, $\texttt{Paths}$, of the MDP $\mathcal{M}$ capturing the behaviour of $\mathcal{M}$ from state s under policy π (Haddad and Monmege, 2018). These probabilistic semantics of MDPs form the foundation for verifying temporal logic properties such as reachability and safety, where one is interested in quantifying the probability of paths satisfying a given specification under a policy".

Given the dynamics of an uncertain system, one can *abstract* to reduce the complexity of a system by grouping certain behaviours or states into a simplified model that preserves essential properties for analysis. One such abstracted Once we have an MDP model for a system, we need to quantitatively ensure that the MDP satisfies the system properties such as 'reach a target region T while always being in safe states P within k steps'. This is particularly important in safety-critical applications, where system failures can incur significant costs, such as performance degradation or violations of operational constraints. To obtain such quantitative guarantees, probabilistic model checkers such as `PRISM` (Kwiatkowska et al., 2011) can be employed. Various quantitative properties, like the probability of reaching a target region, of the system, can be verified against the abstracted system. Two of the common methods to solve this problem are linear programming and value iteration. Value iteration performs better than linear programming, especially for systems with a larger number of states, and is, therefore, the generally preferred method (Brázdil et al., 2025).

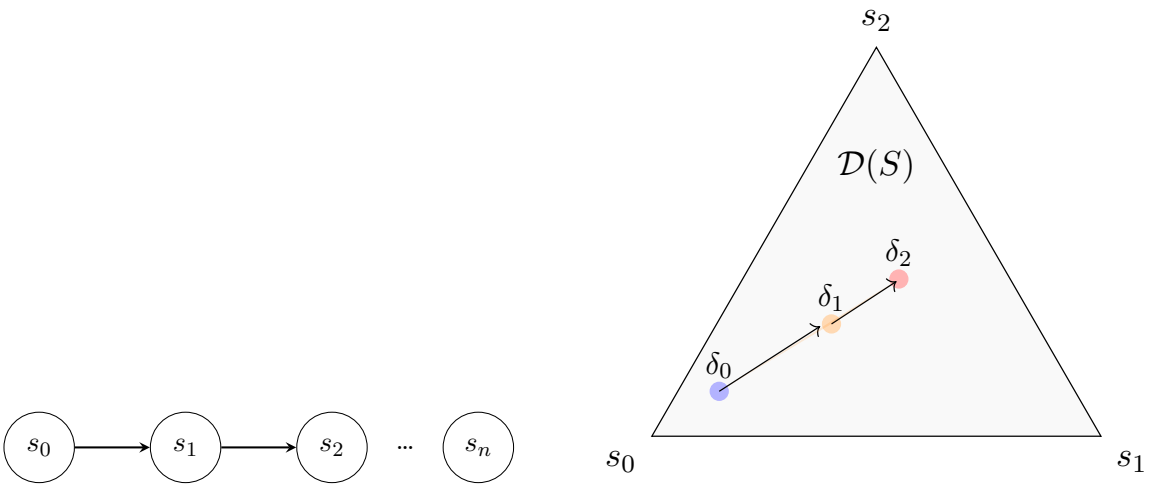
Distribution Transformers View Until now, we have defined MDPs as evolving over states, thereby generating random paths. Alternatively, an MDP can be interpreted as a *distribution transformer* (Korthikanti et al., 2010), where the system evolves over the space of probability distributions rather than over individual states as illustrated in Figure 3-1. In contrast to the *state-based* perspective, where we consider the trajectory of states, in the distribution transformer view, we look at the sequence of distributions generated by an MDP. This perspective could be useful for reasoning about the evolution of uncertainty over time and enables the analysis of population-level behaviours, policy synthesis over distributional goals, and compositional verification. This distributional perspective is particularly useful for reasoning about population-level behaviour, as it enables not only the analysis of how uncertainty evolves over time but also the synthesis of policies that shape the distribution of agents to achieve desired global objectives.

Formally, let $\mathcal{M} = (S, A, \gamma)$ be an MDP, where S is a finite set of states, A is a finite set of actions, and $\gamma : S \times A \to \mathcal{D}(S)$ is the transition kernel, with $\gamma_{s,a}(s')$ denoting the probability of transitioning to state $s' \in S$ from state $s \in S$ under action $a \in A(s)$. Let Π denote the set of admissible policies $\pi : S \to \mathcal{D}(A)$, and let $\delta_0 \in \mathcal{D}(S)$ be the initial state distribution.

For any policy $\pi \in \Pi$, we denote by $\delta_k^\pi \in \mathcal{D}(S)$ the state distribution at time step k, where $\delta_k^\pi(s)$ represents the probability of being in state $s \in S$ at time k under policy π. The distribution evolves according to the following update rule:

$$\delta_{k+1}^\pi(s') = \sum_{s \in S} \delta_k^\pi(s) \sum_{a \in A(s)} \gamma_{s,a}(s')\pi(a \mid s), \quad \forall s' \in S. \tag{3-2}$$

Equation 3-2 describes the evolution of the state distribution under a fixed policy π, with the dynamics at each step governed by the transition kernel γ and the chosen action. By iteratively



(a) State-based view. **(b)** Distribution transformer view.

Figure 3-1: Two perspectives on MDPs. Figure 3-1a shows the state based view. Figure 3-1b illustrates the distributional transformer view of an MDP. The system starts from an initial state distribution δ_0 and evolves under policy and uncertainty, producing reachable distributions δ_1, δ_2. The simplex $\mathcal{D}(S)$ represents all probability distributions over the state space $S = \{s_0, s_1, s_2\}$.

applying this rule, one obtains a sequence of distributions $\delta_0^\pi, \delta_1^\pi, \ldots, \delta_k^\pi$, each capturing the distribution of system states at time step k under policy π.

We now define the set of all distributions reachable at time step k from the initial distribution δ_0, under any admissible policy π:

$$\mathrm{R}_k(\delta_0) = \{\delta \in \mathcal{D}(S) \,|\, \exists \pi \in \Pi \text{ such that } \delta = \delta_k^\pi\}\,. \tag{3-3}$$

Connection between Switched Linear Systems and Distribution Transformers View Interpreting an MDP as a distribution transformer naturally connects to the theory of discrete-time switched stochastic systems. Consider a finite state space $S = \{1, \ldots, n\}$ and let the system state be represented by a probability distribution $\delta_k \in \Delta^n$, the n-dimensional probability simplex. At each step k, choosing an action $a \in A$ selects a corresponding stochastic transition matrix $P_a \in \mathbb{R}^{n \times n}$, where each row of P_a is a probability distribution over successor states. To make the switched system analogy precise, we consider here that the action set A is a "product action set," meaning the same set of actions is available in every state. The evolution of the system under deterministic action a_k is then given by the linear update:

$$\delta_{k+1} = \delta_k P_{a_k}. \tag{3-4}$$

This representation shows that an MDP can be seen as a *switched linear system* on the simplex, where each action determines which linear operator governs the evolution at that time step. The "switching signal" is the sequence of actions $\{a_k\}$ chosen by the policy.

If the policy is deterministic and stationary, then the switching law is fixed and the dynamics reduce to a time-invariant linear system on Δ^n. If the policy is deterministic but non-stationary, then the system corresponds to a time-varying linear system, where the switching sequence varies over time. If the policy is stochastic, then at each step the applied operator

is itself a convex combination of the P_a, weighted by the distribution of actions. In this case, the dynamics can be written as

$$\delta_{k+1} = \delta_k \left(\sum_{a \in A} \pi_k(a \mid \delta_k) P_a \right), \tag{3-5}$$

which highlights the bilinear nature of the system: the distribution update depends both on the current state δ_k and on the action probabilities prescribed by the policy.

Thus, the distribution transformer view reveals MDPs as special instances of switched (or bilinear, under stochastic or history-dependent policies) linear systems, where the 'transformers' are the stochastic matrices $\{P_a\}$ and the policy governs how the system switches between them. This perspective provides a natural bridge between control-theoretic tools for switched linear systems and distributional analyses of MDPs.

3-2-5 Robust Markov Decision Processes

In practical applications, the transition dynamics of an MDP are often not precisely known (Badings et al., 2023b). Uncertainty in the system model can arise from limited data, model approximation, or changing environments. To account for this, *Robust Markov Decision Process (RMDP)* extend the classical MDP model by introducing ambiguity sets that encode possible variations in transition probabilities. The goal is to synthesise policies that perform optimally under the worst-case realisations of these transitions.

Definition 3.11. (Robust Markov Decision Process) Formally, a Robust MDP is defined as a tuple $\mathcal{R} = (S, A, \Gamma)$, where:

- S is a finite set of states,
- $A = \bigcup_{s \in S} A(s)$, where $A(s)$ is a finite, non-empty set of actions available at state s, and $A(s) \cap A(s') = \emptyset$ for $s \neq s'$,
- $\Gamma = \{\Gamma_{s,a}\}_{s \in S, a \in A(s)}$, where each $\Gamma_{s,a} \in \mathbf{\Gamma}(S) \subseteq \mathcal{D}(S)$ is a convex, compact set of admissible transition distributions.

The set $\Gamma_{s,a}$ is called the *ambiguity set* for the pair (s, a) and encodes structured uncertainty about the transition distribution $\gamma_{s,a}$. Common choices for these sets include:

- *Interval-based sets* (Givan et al., 2000): each component of the transition distribution is bounded by known intervals.
- *Polytopic sets* (Chatterjee et al., 2024): where $\Gamma_{s,a}$ is a polytope in the probability simplex $\mathcal{D}(S)$.
- *Wasserstein balls* (Yang, 2017): neighbourhoods of empirical distributions under optimal transport distance metrics, useful in data-driven scenarios.

In our work, we adopt the standard (state-action) *rectangularity* assumption for the ambiguity sets $\Gamma = \{\Gamma_{s,a}\}_{(s,a)\in S\times A}$, where the transition uncertainty is decoupled across state-action pairs. That is, the adversary selects each transition distribution $\gamma_{s,a} \in \Gamma_{s,a}$ independently for each $(s,a) \in S \times A$. This structural assumption ensures the tractability of robust dynamic programming and enables a well-defined Bellman recursion in finite-horizon settings. A detailed discussion on rectangularity and its implications can be found in Suilen et al. (2024).

In RMDPs, we define policies and state trajectories analogously to those in standard MDPs. In robust reachability analysis, the objective is to compute the maximum (or minimum) probability of reaching a designated set of goal states $T \subseteq S$, starting from an initial state s_0, under all transition distributions $\gamma \in \Gamma$ and all policies π (Mazumdar et al., 2024). That is, we compute:

$$\sup_{\pi\in\Pi} \inf_{\gamma\in\mathbf{\Gamma}} \mathbb{P}^{\pi,\gamma}_{s_0} \left(\exists k \in \mathbb{N} : s_k \in T\right),$$

where Π is the set of policies, and $\mathbb{P}^{\pi,\gamma}_{s_0}$ denotes the probability measure over trajectories induced by policy π (see Chapter 10, Baier and Katoen, 2008) and transition kernel γ, starting from state s_0. We seek to apply a similar approach for the reachability analysis of RMDPs.

To formalize the inner minimization over uncertain transitions, we introduce an adversarial model of the environment (Suilen et al., 2024). The adversary selects, at each time step, a transition distribution from the allowed ambiguity set $\Gamma_{s,a}$ for each state-action pair $(s,a) \in S\times A$, with the objective of minimizing the controller's probability of success. This setup yields a two-player stochastic game between the controller and an adversary, where the controller chooses actions to maximize reachability and the adversary selects transitions to impede it. The following definition characterizes this adversarial selection mechanism.

Definition 3.12 (Adversary)**.** An *adversary* is a map $\gamma_{\text{adv}} : S \times A \to \mathcal{D}(S)$ such that for every $(s,a) \in S \times A$, $\gamma_{\text{adv},(s,a)} \in \Gamma_{s,a}$. At each time step, after the policy π selects an action $a \in A(s)$, the adversary selects a feasible transition distribution $\gamma_{\text{adv},(s,a)}$, determining how the probability mass evolves under the transition uncertainty.

Adversarial Semantics: Universal and Existential Variants The behaviour of the adversary in an RMDP can be interpreted through two distinct semantic models: *universal* and *existential* (Akshay et al., 2024). These models differ in how the transition uncertainty, represented by the selection of γ_{adv}, is quantified with respect to the specification being verified.

Under *universal semantics*, the adversary is assumed to select a transition distribution $\gamma_{\text{adv}} \in \Gamma$ in a worst-case manner at each step, independent of the policy. A specification φ is said to be satisfied if, for all admissible adversarial choices, there exists a policy π such that φ holds; formally, this corresponds to the quantification $\forall\gamma_{\text{adv}}\, \exists\pi\, \varphi$. This is the standard setting in *robust verification*, where correctness must be guaranteed regardless of how uncertainty is resolved. Techniques based on over-approximating the reachable sets of distributions and synthesising distributional invariants are commonly used in this context (Akshay et al., 2018).

In contrast, under *existential semantics*, the adversary is treated as part of the environment, and it is sufficient that there exists at least one admissible realisation γ_{adv} for which some policy π satisfies the specification; i.e., $\exists\gamma_{\text{adv}}\, \exists\pi\, \varphi$. This interpretation is prevalent in *policy*

synthesis and *feasibility analysis*, where the goal is to demonstrate that a favourable combination of transitions and policy exists. It allows for under-approximation techniques, such as sampling-based verification or template-based synthesis (Akshay et al., 2024).

The choice of semantics directly influences the conservativeness and computational complexity of verification procedures and will be reflected in the distributional reachability formulations considered in this thesis.

Classical solution techniques for robust reachability include game-theoretic formulations where an adversary selects transitions within the ambiguity sets (Iyengar, 2005; Nilim and El Ghaoui, 2005), robust policy iteration, and interval value iteration (Haddad and Monmege, 2018). We focus on robust reachability forms the foundational framework for addressing uncertainty in distributional forward analysis of multi-agent systems. Rather than computing expected outcomes, our objective is to identify the set of all reachable distributions under worst-case transitions and admissible policies.

This worst-case analysis guarantees that the synthesised policy achieves at least the computed reachability probability for any realisation of the transition dynamics in the ambiguity model. Such guarantees are particularly important in safety-critical settings where violating a constraint can lead to catastrophic outcomes.

State Dynamics under Distribution Transformer View Extending the distribution transformer perspective from section 3-2-4 to RMDPs, the evolution of the state distribution incorporates both the control policy and the adversarial transition selection. Following similar semantics as before, given a policy $\pi : S \to \mathcal{D}(A)$, an adversary $\gamma_{\text{adv}} : S \times A \to \mathcal{D}(S)$, and a state distribution $\delta_k^\pi \in \mathcal{D}(S)$ at time step k, the next state distribution $\delta_{k+1}^\pi \in \mathcal{D}(S)$ is given by:

$$\delta_{k+1}^\pi(s') = \sum_{s \in S} \delta_k^\pi(s) \sum_{a \in A(s)} \gamma_{\text{adv},(s,a)}(s')\pi(a \mid s), \quad \forall s' \in S. \tag{3-6}$$

This equation describes the population-level dynamics under a fixed policy, where the adversary selects transitions from the uncertainty sets to influence the evolution of the distribution over states. Like in subsection 3-2-4, we consider the distribution transformer view of RMDPs to approach the reachability problem as illustrated in Figure 3-2.

3-2-6 Interval Markov Decision Processes

RMDPs generalise the idea of an Interval Markov Decision Process (IMDP), where the ambiguity sets are interval-based. They provide a nice and tractable way to develop efficient algorithms to analyse systems with bounded uncertainty (Suilen et al., 2024) and are hence a useful way to model system uncertainty.

Definition 3.13 (Interval Markov Decision Processes)**.** We define an IMDP as a tuple $\mathcal{I} = (S, A, \Gamma)$ where,

- S is a finite set of states,
- $A = \bigcup_{s \in S} A(s)$ where every $A(s)$ is a non-empty finite set of actions available in state s with $A(s) \cap A(s') = \emptyset \ \forall \ s \neq s'$,

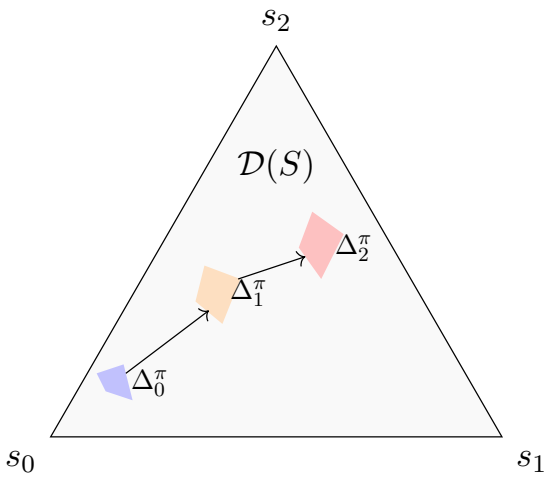


Figure 3-2: Distributional reachability of RMDPs. The system starts from an initial set of distributions Δ_0^π, and evolves into reachable sets Δ_1^π and Δ_2^π.

- $\Gamma = \{\Gamma_{s,a}\}_{s\in S, a\in A(s)}$ is the set containing sets of all transition probability distributions satisfying (3-1), with $\Gamma_{s,a} \in \mathbf{\Gamma}(S)$.

For an IMDP $\mathcal{I}$ and a state transition $s \xrightarrow{a} s'$ for $s', s \in S$ and $a \in A(s)$, let $\hat{\gamma}_{s,a}(s')$ and $\check{\gamma}_{s,a}(s')$ denote the upper and lower bounds, respectively, on the state distribution. For such a transition, $\Gamma_{s,a}$ is defined as:

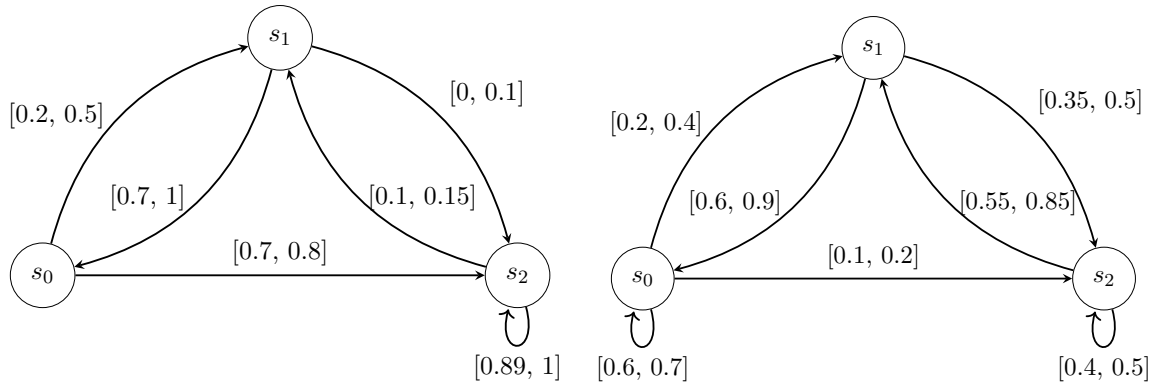
$$\Gamma_{s,a} = \{\gamma_{s,a} \in \mathcal{D}(S) : \check{\gamma}_{s,a}(s') \leq \gamma_{s,a}(s') \leq \hat{\gamma}_{s,a}(s') \ \forall s' \in S\}. \tag{3-7}$$

To illustrate how these concepts are applied in practice, we now introduce a running example that will serve as a reference throughout this work. While the distributional framework applies to general state spaces, we focus on a 3-state IMDP to allow for clear visualisations of both the evolving reachable sets and the propagation of uncertainty. This illustrative example captures the key structural properties of IMDPs under interval uncertainty and provides a concrete foundation for the subsequent theoretical developments.

Example 3.1 (Running Example)**.** We consider the following IMDP with $S = \{s_0, s_1, s_2\}$ and two available actions, $A = \{a_0, a_1\}$, in each state as shown in Figure 3-3.

For example, for state s_0 in Figure 3-3a, the transition probability intervals are: $\gamma_{s_0,a_0}(s_1) \in [0.2, 0.5]$, $\gamma_{s_0,a_0}(s_2) \in [0.7, 0.8]$.

Note. In some scenarios, based on the objective function for the problem statement, an ordering of the states becomes relevant wherein one can use the *order-* maximisation algorithm to obtain the feasible transition probability distributions. The reader is referred to the work by (Givan et al., 2000) for further information. Since we do not care about the order for our demonstrations, we need to consider the whole distribution polytope set as it is and propose methods to obtain the reachable sets.



(a) Under action 0. **(b)** Under action 1.

Figure 3-3: Running example of a 3-state IMDP under 2 possible actions.

Induced Interval Markov Chain

Building on the background of IMDPs, we now consider how the model behaves under a fixed policy. While IMDPs define uncertainty at the level of individual state-action pairs, a fixed policy induces probabilistic action selection in each state, requiring aggregation of transition uncertainties. This gives rise to an *induced* Interval Markov Chain (IMC), which captures the set of possible transition probabilities under the given policy. We formally define this induced model and describe how the corresponding bounds are computed.

Definition 3.14 (Induced IMC)**.** Let $\mathcal{I} = (S, A, \Gamma)$ be an IMDP and $\pi : S \to \mathcal{D}(A)$ a stationary policy on S. The *induced interval Markov chain* is

$$\begin{aligned} &\mathcal{I}^\pi = (S, \Gamma^\pi), \quad \Gamma^\pi = \{\Gamma^\pi_s\}_{s\in S}, \quad \text{where,} \\ &\Gamma^\pi_s = \Big\{\gamma^\pi_s \in \mathcal{D}(S) \;\Big|\; \sum_{a\in A(s)} \check{\gamma}_{s,a}(s')\,\pi(a\,|\,s) \le \gamma^\pi_s(s') \le \sum_{a\in A(s)} \hat{\gamma}_{s,a}(s')\,\pi(a\,|\,s),\ \forall s' \in S\Big\}. \end{aligned} \tag{3-8}$$

Interpretation and Proof of Inclusion The intuition behind this definition is as follows. Under a fixed policy π, the system behaves as a Markov chain with uncertainty only in the transition probabilities. At each state $s \in S$, the randomness in transitions arises due to both action selection via $\pi(a \mid s)$ and the uncertainty in the corresponding transition distributions $\Gamma_{s,a}$.

To derive the effective (induced) transition distribution γ^π_s, consider choosing arbitrary distributions $\gamma_{s,a} \in \Gamma_{s,a}$ for each $a \in A(s)$, and define the convex combination

$$\gamma^\pi_s := \sum_{a\in A(s)} \pi(a \mid s)\,\gamma_{s,a}. \tag{3-9}$$

We now prove that this γ^π_s lies in the set Γ^π_s as defined in (3-8).

Since each $\gamma_{s,a} \in \Gamma_{s,a}$, we have for all $s' \in S$:

$$\check{\gamma}_{s,a}(s') \le \gamma_{s,a}(s') \le \hat{\gamma}_{s,a}(s').$$

Multiplying these inequalities by $\pi(a \mid s) \geq 0$ preserves the inequality:

$$\pi(a \mid s)\check{\gamma}_{s,a}(s') \leq \pi(a \mid s)\gamma_{s,a}(s') \leq \pi(a \mid s)\hat{\gamma}_{s,a}(s').$$

Summing over all $a \in A(s)$, we obtain:

$$\sum_{a \in A(s)} \pi(a \mid s)\check{\gamma}_{s,a}(s') \leq \gamma_s^\pi(s') \leq \sum_{a \in A(s)} \pi(a \mid s)\hat{\gamma}_{s,a}(s'), \quad \forall s' \in S,$$

which proves $\gamma_s^\pi \in \Gamma_s^\pi$. Additionally, since γ_s^π is a convex combination of probability distributions, it is itself a valid distribution: $\gamma_s^\pi \in \mathcal{D}(S)$.

Collecting the vectors γ_s^π across all $s \in S$, the full set of transition matrices induced by policy π is:

$$\Gamma^\pi = \left\{ [\gamma_{s_1}^\pi, \ldots, \gamma_{s_n}^\pi]^\top \,\middle|\, \gamma_s^\pi \in \Gamma_s^\pi, \forall s \in S \right\}.$$

Hence, $\mathcal{I}^\pi = (S, \Gamma^\pi)$ characterises an IMC whose rows belong to convex combinations of the original ambiguity sets under the policy π.

We have now introduced IMDPs and shown how, under a fixed policy π, its transition uncertainty "collapses" to an IMC with row-sets Γ_s^π. In the next section, we will use these definitions to build the *reachable sets* over one or more steps. Concretely, starting from an initial convex set of state distributions, we form the one-step reachable set by applying every feasible transition matrix in Γ^π; algebraically, this is taking Minkowski sums of the interval rows we defined, which exactly captures how the lower- and upper-bounds on each transition combine. We then repeat this process to get the k-step reachable sets.

Example 3.2. (Running Example Continued) For the system described by Figure 3-3, we consider the following stationary randomized policy $\pi : S \to \mathcal{D}(A)$, assigning probabilities to the two actions at each state:

$$\pi = \begin{pmatrix} \pi(a_0 \mid s_0) & \pi(a_1 \mid s_0) \\ \pi(a_0 \mid s_1) & \pi(a_1 \mid s_1) \\ \pi(a_0 \mid s_2) & \pi(a_1 \mid s_2) \end{pmatrix} = \begin{pmatrix} 0.3 & 0.7 \\ 0.5 & 0.5 \\ 0.8 & 0.2 \end{pmatrix}.$$

The modelling and theoretical foundations presented in this chapter now allow us to define the reachability verification problem under transition uncertainty formally. In the next chapter, we develop the problem formulation, introduce key algorithms for policy verification and synthesis, along with an analysis of the methods presented.

Chapter 4

The Forward Reachability Problem

Chapter Summary
We formalise the core problem: how to compute and verify the set of all forward reachable state distributions in the presence of transition uncertainty. Building on the geometric and probabilistic models introduced earlier, we define the forward distributional reachability maps for IMDPs, and formulate the corresponding reach-avoid verification problem. The chapter presents a sampling-based algorithm for computing reachable sets and provides conditions under which convexity and approximation guarantees hold.

The motivating example in Chapter 1 established the need for distribution-level reasoning in fleet-level verification. Building upon the formalisms developed in Chapter 3, particularly the distribution transformer view of IMDPs, this chapter addresses the problem of forward distributional reachability as a verification task. The central objective is to characterise the set of state distributions that can arise from a given initial set, under a fixed policy and all admissible transition probabilities. Such an analysis is essential for verifying modern fleet-level specifications, including bounds on agent concentration or dispersion, which are predicates on the entire distribution and cannot be captured by state-wise marginals or individual trajectories.

The exact analysis of this problem is, however, computationally intractable. To this end, in this chapter, we introduce tractable approximation methods. We begin by formulating the forward reachable sets for IMCs and the approach to compute them. The formulation is then lifted to IMDPs where we provide a way to synthesise a policy. Finally, to mitigate computational growth and enhance scalability, a sampling-projection scheme is introduced.

4-1 Problem Formulation

Recall that our overarching goal is to understand and efficiently compute, for a given control policy, all the state-distribution trajectories that can arise when transitions are only known

within interval bounds. To that end, we turn to the central question of reachability by first formally stating the problem.

Let $\Delta_0 \subseteq \mathcal{D}(S)$ denote a convex polytope representing the initial uncertainty in the state distribution of the system, where $\mathcal{D}(S)$ is the set of all probability distributions over the finite state space S.

Definition 4.1 (One-step Reachable Set for an IMC)**.** Given an IMC $\mathcal{I}^\pi = (S, \Gamma^\pi)$ characterised by a transition uncertainty set as in (3-8), the one-step forward reachable set Δ_1^π under policy π from the initial set Δ_0 is defined as

$$\Delta_1^\pi := \{\delta_1 \in \mathcal{D}(S) \mid \delta_1 = \gamma\delta_0,\ \delta_0 \in \Delta_0,\ \gamma \in \Gamma^\pi\}\,. \tag{4-1}$$

This set captures all possible state distributions for $\mathcal{I}^\pi$ after one step, starting from the initial distribution set Δ_0 and the transition probabilities, $\gamma \in \Gamma^\pi$.

Definition 4.2 (Recursive k-step Reachable Set for an IMC)**.** The k-step reachable set $\Delta^{\pi,k}$ for an IMC $\mathcal{I}^\pi = (S, \Gamma^\pi)$ given by (3-8), is recursively defined as

$$\Delta_{k+1}^\pi := \{\delta_{k+1} \in \mathcal{D}(S) \mid \delta_{k+1} = \gamma\delta_k,\ \delta_k \in \Delta_k^\pi,\ \gamma \in \Gamma^\pi\}\,, \tag{4-2}$$

with base case $\Delta_0^\pi := \Delta_0$.

Definition 4.3 (Forward Reachable Set)**.** The forward reachable set for an IMC $\mathcal{I}^\pi = (S, \Gamma^\pi)$ given by (3-8), from the initial distribution set Δ_0 under a policy π is defined as

$$FR(\Delta_0) := \bigcup_{k=0}^{\infty} \Delta_k^\pi, \tag{4-3}$$

representing all state distributions that can be reached at any time step under the given uncertainty and policy.

Problem 4.1 (Forward Reachability for an IMC)**.** Given an IMDP $\mathcal{I} = (S, A, \Gamma)$ as defined in Definition 3.13, the initial convex set of state distributions Δ_0 and a fixed policy π that induces the IMC $\mathcal{I}^\pi = (S, \Gamma^\pi)$, compute or characterise the k-step reachable sets Δ_k^π and the forward reachable set $FR(\Delta_0)$. This involves accounting for the uncertainty captured in the interval transition probabilities Γ^π and propagating the set of possible distributions over time.

Challenges In our running example, the forward reachability problem concerns computing the set of all state distributions to which the system can evolve after k steps, starting from an initial distribution set. We assume that the system dynamics are known. When the system is deterministic and the initial distribution set is a polytope, the reachable sets remain polytopic at every time step. This is because linear transformations of polytopes yield polytopes, and the deterministic transition dynamics act as an affine map on the distribution space (Ziegler, 1995).

However, when there is uncertainty in the dynamics, such as interval-valued or polytopic transition kernels, the reachable sets may no longer be convex. Intuitively, under uncertain transitions, the system can evolve along multiple trajectories governed by different realisations

of the dynamics. The union of such images across time steps does not necessarily preserve convexity unless further assumptions (e.g., convexity of uncertainty sets and linearity of dynamics) are made (Blanchini and Miani, 2007). For instance, if the transition probabilities lie within known intervals but are selected adversarially at each step, the resulting reachable distributions can form non-convex sets due to this combinatorial branching. Nevertheless, in systems modelled using IMDPs and IMCs, where the uncertainty is bounded but not precisely known, it is common to compute outer convex approximations of the reachable set.

4-1-1 Approach

A typical approach is to propagate the convex hull of the reachable distributions at each step, effectively tracking the evolution of all possible states the system could occupy under the uncertainty model. This method, rooted in set-theoretic control (Blanchini and Miani, 2007), provides a computationally tractable way to reason about worst-case behaviour.

Definition 4.4 (Over-approximation of the Reachable Set)**.** Let $\mathcal{I}^\pi = (S, \Gamma^\pi)$ be the induced Interval Markov Chain (IMC) under a fixed policy π, where $\Gamma^\pi = \{\Gamma_s^\pi\}_{s \in S}$ is the set of row-wise transition ambiguity sets. Let the initial distribution set $\Delta_0^\pi \subseteq \mathcal{D}(S)$ be a convex polytope.

Then, for any $k \in \mathbb{N}$, the k-step reachable set $\Delta_k^\pi \subseteq \mathcal{D}(S)$, we recursively over-approximate the one-step reachable set as:

$$\hat{\Delta}_{k+1}^\pi := \operatorname{conv}\left\{\gamma\delta \mid \delta \in \Delta_k^\pi,\ \gamma \in \Gamma^\pi\right\}, \tag{4-4}$$

is a convex polytope for all k, where $\operatorname{conv}(\cdot)$ denotes the convex hull operator. 1 shows the steps involved in obtaining the over-approximation of the reachable set.

Algorithm 1 Convex Hull Propagation of Reachable Sets

1: **Input:** Initial set $\Delta_0 \subseteq \mathcal{D}(S)$, policy π, interval transition sets $\Gamma_{s,a}$
2: **for** $k = 0$ to $K - 1$ **do**
3: Compute one-step reachable distributions:

$$\Delta_{k+1}^\pi = \left\{ \sum_{s \in S} \delta(s) \sum_{a \in A(s)} \gamma_{s,a} \pi(a \mid s) \;\middle|\; \delta \in \Delta_k, \gamma_{s,a} \in \Gamma_{s,a} \right\}$$

4: Compute convex hull: $\hat{\Delta}_{k+1}^\pi = \operatorname{conv}(\Delta_{k+1}^\pi)$
5: Extract vertices of $\hat{\Delta}_{k+1}^\pi$ for next iteration
6: **end for**
7: **Output:** Approximated reachable set sequence $\{\hat{\Delta}_k^\pi\}_{k=1}^K$

While the convexity of the reachable sets holds at each step under fixed transitions and initial convex sets, this property does not automatically generalise to the IMDP setting where the policy and transition uncertainties interact simultaneously. In particular, when optimising over both the action-selection and the admissible transition distributions, the reachable set

at each step may no longer remain convex in general. To address this, and to retain computational tractability, we adopt a convex hull propagation approach that explicitly considers convexifications of the reachable sets at each stage. This enables us to work with outer approximations that remain polytopic, while still capturing the key aspects of the uncertainty propagation under IMDP dynamics. We now proceed to extend the reachability formulation to IMDPs by introducing occupation measures and corresponding convex relaxations.

4-2 Policy Synthesis for IMDPs

Until this section, we have looked at the distributional reachability properties of IMCs; however, these methods are not sufficient for IMDPs. In this section, we introduce the extension of the above methods to IMDPs. This problem is formulated as follows:

Problem 4.2 (Forward Reachability for an IMDP)**.** Given an IMDP $\mathcal{I} = (S, A, \Gamma)$ as defined in Definition 3.13, the initial convex set of state distributions Δ_0, compute or characterise the k-step reachable sets Δ_k and the forward reachable set $FR(\Delta_0)$. Furthermore, synthesise a policy $\pi : S \to \mathcal{D}(A)$ for the obtained $FR(\Delta_0)$ set.

4-2-1 Use of Occupation Measures

Definition 4.5. (Occupation Measure (Gao et al., 2023)) Given an MDP, $\mathbf{M}$ (or an IMDP, $\mathcal{I}$), with n states and m actions, we define the occupation measure as a matrix $Q \in \mathbb{R}^{n\times m}$, if $Q \geq 0$ with $(Q\mathbf{1})^T \in \mathcal{D}(S)$. In addition, $Q(s,a) = \delta(s)\pi(a \mid s)$. Denote by $\mathcal{O}$ the set of all occupation measures.

Intuition Occupation measures provide a compact way to describe the long-term behaviour of a controlled stochastic process (Altman, 1995). They serve as a bridge between policies and distributions. Instead of tracking the full trajectory or computing policies explicitly, they encapsulate the expected frequency of visiting state-action pairs under a given policy and initial distribution. Formally, the entry $Q(s,a)$ of the occupation measure represents the expected number of times the system occupies state s and takes action a over the planning horizon.

Extracting the policy Occupation measures couple policies and state distributions, thereby enabling one to extract either of them as follows (Gao et al., 2023):

$$\pi(a|s) = \begin{cases} \frac{Q(s,a)}{\sum_{a'\in A(s)} Q(s,a')} & \text{if } \sum_{a'\in A(s)} Q(s,a') > 0, \\ \frac{1}{|A(s)|} & \text{if } \sum_{a'\in A(s)} Q(s,a') = 0 \ \&\ a \in A(s). \end{cases} \tag{4-5}$$

Since occupation measures only quantify the number of times a state-action pair is chosen, the definition of occupation measure holds for IMDPs. Now we formulate the problem of forward distributional reachability of an IMDP $\mathcal{I} = (S, A, \Gamma)$ defined in Definition 3.13 as follows:

Definition 4.6. (Forward Distributional Reachability for IMDPs) Define the map $\mathcal{FR} : 2^{\mathcal{D}(S)} \to 2^{\mathcal{D}(S)}$ as

$$\mathcal{FR}(\Delta) = \left\{ \delta \in \mathcal{D}(S) \;\middle|\; \begin{array}{l} Q \in \mathcal{O}, (Q\mathbf{1})^T \in \Delta, \\ \forall s' \in S, \delta(s') = \sum_{s \in S} \sum_{a \in A(s)} \gamma_{s,a}(s') Q(s,a) \\ \text{with } \sum_{s' \in S} \gamma_{s,a}(s') = 1,\ \check{\gamma}_{s,a}(s') \leq \gamma_{s,a}(s') \leq \hat{\gamma}_{s,a}(s') \end{array} \right\}, \tag{4-6}$$

to obtain the set of one-step reachable distributions starting from a set Δ.

Challenges In the case of MDPs with fixed transitions, the formulation in (4-6) is a linear program in the occupation measure Q. However, for IMDPs, the transition probabilities $\gamma_{s,a}$ are uncertain and vary within given intervals, making them additional decision variables. This introduces bilinear terms involving $Q(s,a)\gamma_{s,a}(s')$, rendering the problem non-convex. To enable tractable computation of $\mathcal{FR}(\Delta)$ set, we apply linear relaxations to approximate these bilinear constraints using convex formulations.

4-2-2 Approach

McCormick Envelopes *McCormick relaxations* (McCormick, 1976) provide a convex approximation of bilinear terms, by introducing auxiliary variables and bounding constraints, the McCormick envelopes allow one to replace each bilinear term with a set of linear inequalities that tightly enclose the original non-convex term. This transformation enables the use of convex programming to over-approximate the set of reachable distributions.

In the context of distributional reachability, this method allows tractable computation of reachable sets, which are needed to verify whether a system can reach a desired distributional target under uncertainty. Related convex relaxations have also been used in stochastic control and verification, where efficient outer approximations are critical for scalable analysis (Bujarbaruah et al., 2021; Yang and Summers, 2022).

We now define an auxiliary variable called *transition* occupation measure to relax the non-convex program (4-6) using McCormick relaxations.

Definition 4.7. (Transition Occupation Measure) Given an IMDP $\mathcal{I} = (S, A, \Gamma)$, we define a transition occupation measure as a matrix $W \in \mathbb{R}^{n \times m \times n}$ such that $W \in [0,1]$ and

$$W(s,a,s') = \gamma_{s,a}(s') Q(s,a). \tag{4-7}$$

Using McCormick relaxations of a given function (Mitsos et al., 2009), we relax the bilinear term $Q(s,a)\gamma_{s,a}(s')$ in (4-6) to obtain the following *relaxed* convex forward reachable set

$\widehat{\mathcal{FR}}(\Delta)$, where vec$(\cdot)$ denotes a vectorised form of a matrix:

$$\widehat{\mathcal{FR}}(\Delta) = \left\{ \delta \in \mathcal{D}(S) \;\middle|\; \begin{array}{l} Q \in \mathcal{O},\ (Q\mathbf{1})^T \in \Delta,\ \mathrm{vec}(\sum_{s,a} W(s,a,s'))_{s' \in S} \in \Delta, \\ \sum_{s' \in S} W(s,a,s') = Q(s,a), \quad \forall s \in S,\ a \in A(s), \\ \delta(s') = \sum_{s \in S} \sum_{a \in A(s)} W(s,a,s'), \quad \forall s' \in S, \\ W(s,a,s') \geq \check{\gamma}_{s,a}(s') Q(s,a), \\ W(s,a,s') \geq \check{\gamma}_{s,a}(s') - (1 - Q(s,a))\hat{\gamma}_{s,a}(s'), \\ W(s,a,s') \leq \hat{\gamma}_{s,a}(s') Q(s,a), \\ W(s,a,s') \leq \hat{\gamma}_{s,a}(s') - (1 - Q(s,a))\check{\gamma}_{s,a}(s'), \\ \forall s' \in S \sum_{s' \in S} \gamma_{s,a}(s') = 1,\ \check{\gamma}_{s,a}(s') \leq \gamma_{s,a}(s') \leq \hat{\gamma}_{s,a}(s'). \end{array} \right\}, \tag{4-8}$$

Extending the argument from (Proposition 3.1 Gao et al., 2023), we say that the $k-$step reachable set starting from an initial ambiguity set Δ_0 can be recursively written as:

$$\widehat{FR}(\Delta_0, k+1) = \widehat{\mathcal{FR}}(FR(\Delta_0, k)). \tag{4-9}$$

4-2-3 Sampling-based Algorithm for Reachable Sets

Solving the linear program (4-8) for $k-$steps explodes quickly due to the exponential increase in the number of vertices of the polytopes at each step. Therefore, we use a sampling-based method to approximate the one-step polytope and repeat the process for k steps. We approximate one-step reachable sets via a sampling–projection scheme instantiated on the IMDP relaxation in (4-8). Given a convex set $\Pi \subset \mathcal{D}(S)$ and a sampling box $\Gamma \supset \mathcal{D}(S)$, draw N_s samples $\{\pi_i^s\}_{i=1}^{N_s} \subset \Gamma$ and solve, for each i,

$$\min_{\delta, Q, W, \gamma} \ \|\delta - \pi_i^s\|_2^2 \quad \text{s.t.} \quad (Q, W, \gamma, \delta) \text{ satisfy (4-8).} \tag{4-10}$$

The convex hull of the projected points, $FR_{N_s}(\Pi) = \mathrm{conv}\{\delta_i\}_{i=1}^{N_s}$, is a polytopic inner approximation with respect to the relaxed IMDP image, and multi-step sets follow from the recursion $FR(\Delta_0, k+1) = \widehat{\mathcal{FR}}(FR(\Delta_0, k))$ with $FR(\Delta_0, 0) = \Delta_0$ (4-9).

This procedure follows from Algorithm 1 of (Gao et al., 2023): uniform sampling in Γ, Euclidean projection onto $FR(\cdot)$, and convex hull aggregation. It provides inner approximations that become asymptotically tight in probability as N_s increases, and its complexity is linear in N_s with per-projection QP polynomial in n and m. Under McCormick relaxations of the bilinear terms $Q(s,a)\gamma_{s,a}(s')$, inner-ness is understood with respect to the relaxed feasible set (4-8).

Practical note. Choosing Γ slightly larger than $\mathcal{D}(S)$ and favouring samples whose projections land on the boundary of $FR(\Pi)$ improves hull sharpness while avoiding vertex enumeration.

Example 4.1. (Running Example Continued) We use the sampling-based algorithm for our running example to obtain the forward reachable set. We required 100 samples to compute the approximate forward reachable set instead of 500 points for the exact reachable set computation. It can be seen from Figure 4-1 that the convex hull obtained by considering the projected distributions provide a good approximation for the true convex hull.

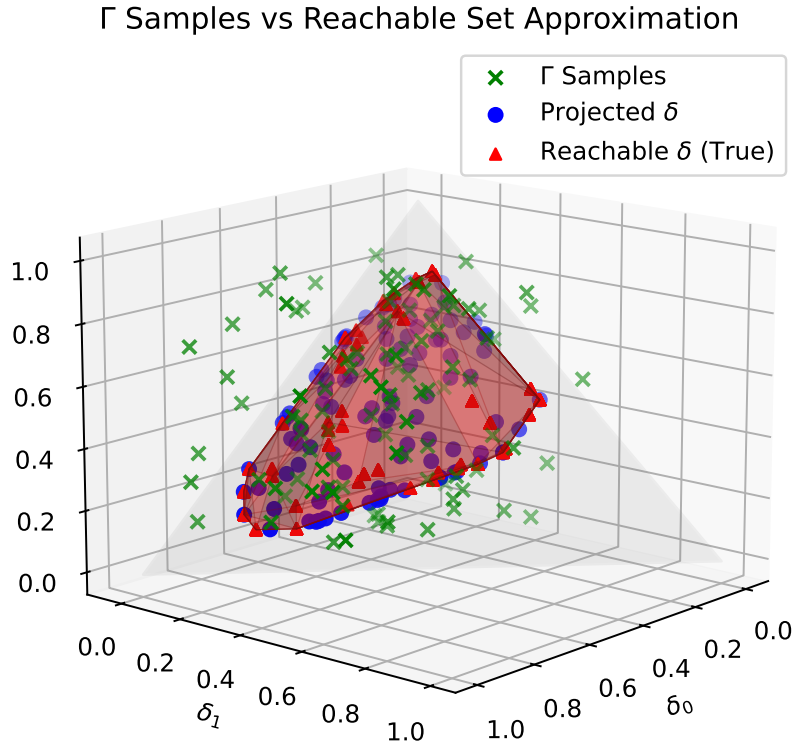


Figure 4-1: Sampling-based vs exact computation of the forward reachable set.

4-2-4 Effect of using Sampling-based algorithm with McCormick Envelopes

As discussed above, McCormick envelopes turn the non-convex $\mathcal{FR}(.)$ map from (4-6) into a convex outer-approximation $\widehat{\mathcal{FR}}(.)$. Hence, $\mathcal{FR}(\Pi) \subseteq \widehat{\mathcal{FR}}(\Pi)$ for any convex $\Pi \subset \mathcal{D}(S)$. Upon using the sampling-projection algorithm from subsection 4-2-3, random points from a bigger 'box' Γ onto the chosen one-step map and returns the convex hull of the projections. The output $FR_{N_s}(\Pi) \subseteq \widehat{FR}(\Pi)$ with asymptotic tightness in probability relative to $\widehat{FR}(\Pi)$ (extending the argumentation of Theorem 3.1 Gao et al., 2023 to the relaxed set $\widehat{FR}(\Pi)$). This yields an *inner-of-outer* approximation. As a result, forward-reachability claims based solely on FR_{N_s} can be optimistic, thereby addressing the existential variant of the problem to be solved. This difference between the sets can be expressed as:

$$d_H(FR_{N_s}, FR_{\text{true}}) \leq \underbrace{d_H(FR_{N_s}, \widehat{FR})}_{\text{sampling error}\to 0 \text{ in probability}} + \underbrace{d_H(\widehat{FR}, FR)}_{\text{relaxation gap (McCormick)}}, \tag{4-11}$$

where d_{H} is the Hausdorff distance (Blanchini and Miani, 2007).

This means that the relaxation gap can propagate across steps, thereby leading to approximations that are not tight. To use these methods for universal variants of the problem would require tighter guarantees on the approximation, which is not directly possible using the set-based methods presented above due to the non-linear nature of the true $\mathcal{FR}(.)$ map.

4-3 Computational Analysis

Let $n := |S|$, $m := \max_s |A(s)|$, $d := n - 1$. Denote by V_k the number of vertices of the k-step outer set $\hat{\Delta}_k^\pi$, by $E := |\operatorname{ext}(\Gamma^\pi)|$ the number of extreme transition matrices, and by $\mathrm{CH}(N, d)$ the cost of a convex hull of N points in $\mathbb{R}^d$.

1. **IMC convex hull propagation**: Step 4 in algorithm 1 $\hat{\Delta}_{k+1}^\pi$ using extreme maps yields at most EV_k candidate points. The per–step arithmetic cost is $\mathcal{O}(EV_k\, n^2)$ + $\mathrm{CH}(EV_k, d)$, with V_{k+1} potentially exponential in k.

2. **IMDP one–step relaxation**: The relaxed one–step program has $\Theta(n^2 m)$ decision variables (from Q, W, and δ) and a comparable number of linear constraints. The linear programs solve it in polynomial time. A rough bound is $\mathcal{O}((n^2 m)^3)$ per solve. Single steps are tractable for moderate n, m; repeating the solve across many steps becomes limited by the increasing complexity of the input sets.

3. **Sampling–based projection**: With N_s samples, the method solves N_s independent projections (each of the same order as in (ii)) and then takes one convex hull of N_s points in $\mathbb{R}^{n-1}$. The total cost is approximately $\mathcal{O}\big(N_s \times$cost of (ii)$\big)$ + cost of one convex hull on N_s points.

4-4 Discussion

This chapter established a tractable framework for forward distributional reachability under interval transition uncertainty. For fixed-policy IMCs, we propagated convex hulls to compute sound polytopic outer approximations of reachable sets. For the more general case of IMDPs, we lifted the problem using occupation measures and employed McCormick relaxations to define a computable one-step outer map and also synthesise a policy, which we complemented with a sampling-projection routine to generate efficient inner approximations.

These dual constructions directly support two modes of formal verification. The outer approximations can be used to prove universal safety: if the conservative outer set avoids an unsafe region, the true system is guaranteed to be safe. Conversely, the inner approximations provide existential reachability witnesses: if the inner set intersects a target, it confirms that the target is verifiably reachable.

From a computational perspective, the scalability of the IMC method is limited by vertex growth in the convex hulls, while the complexity of the IMDP relaxation is driven by the size of the underlying convex program. The sampling-based approach offers a scalable alternative

for exploration and finding existential evidence for a distribution to reach a target set of distributions.

It is important to note that this framework provides a forward analysis for answering ”what if” questions for a given IMDP. It does not provide ”must” guarantees that hold across all policies. Those challenges require backward reachability analysis, which serves as a natural complement to the analysis developed here and is discussed in the next chapter.

Chapter 5

The Backward Reachability Problem

Chapter Summary
We develop a set-based backward distributional reachability framework for IMDPs based on the framework from the previous chapter. Furthermore, we develop a robust backward reachability algorithm for IMCs to provide universal reachability guarantees and extend the formulation to IMDPs. The probability simplex is discretised, and a robust backward iteration to obtain the largest initial distribution set can be steered to a target distribution set within a finite horizon. The method highlights the accuracy–cost trade-off via mesh resolution and pre-computation. These methods complement the forward distributional analysis previously presented by delivering conservative guarantees that hold under transition uncertainty.

In the previous chapter, we formalised the distributional viewpoint via IMDPs, equipping us with uncertainty sets over transition kernels and a transformer on the probability simplex using forward distributional reachability. What is still missing is a method to characterise, under worst-case transition uncertainty, the largest set of initial distributions that are guaranteed to reach a target within a finite horizon. Forward images of distributions are ill-suited to this goal: outer approximations certify safety but are conservative for target attainment, while inner approximations are generally used for existential semantics. In this chapter, we address this gap through backward reachability on distribution sets, using monotone predecessor operators and fixed-point iteration to produce sound "must-reach" initial distribution sets. We present two complementary constructions:

1. a set-based relaxation using McCormick envelopes, an extension of the framework described in section 4-2,
2. a mesh-based method to perform robust value-iteration on the probability simplex.

Together, they provide rigorous, distribution-level guarantees aligned with the motivating example from section 1-2.

5-1 Problem Formulation

Due to the uncertainty present in the system, the number of feasible transition matrices could be infinite due to the continuous interval of state transition probability matrices. This means that the vertices of the convex hull of the 1-step reachable sets increases exponentially, making the problem computationally very expensive. Hence, we look to compute backward reachable sets under uncertainty, which can preserve the convexity of the computed sets (Blanchini and Miani, 2007). The problem is defined in terms of the computation of the backward reachable sets with the overarching idea of obtaining the largest set of initial distributions that are guaranteed to reach a target set of distributions.

Problem 5.1. (Backward Reachability Problem) Define the map $\mathcal{BR} : 2^{\mathcal{D}(S)} \to 2^{\mathcal{D}(S)}$ for an IMDP $\mathcal{I} = (S, A, \Gamma)$, as

$$\mathcal{BR}(\Delta) = \left\{ (Q\mathbf{1})^T \in \mathcal{D}(S) \;\middle|\; \begin{array}{l} Q \in \mathcal{O}, \delta \in \Delta, \\ \forall s' \in S, \delta(s') = \sum_{s \in S} \sum_{a \in A(s)} \gamma_{s,a}(s') Q(s,a) \\ \text{with } \sum_{s' \in S} \gamma_{s,a}(s') = 1,\ \check{\gamma}_{s,a}(s') \leq \gamma_{s,a}(s') \leq \hat{\gamma}_{s,a}(s') \end{array} \right\}, \tag{5-1}$$

which collects the set of distributions that can reach a target distribution set Δ in one step. Analogous to the definition of $FR(\Delta_0)$, we define the $k-$step backward reachable set, starting from a target distribution set Δ_f, to obtain the set of initial distributions δ_0 to reach a target distribution $\delta_f \in \Delta_f$ in $k-$ steps as follows:

$$BR(\Delta_f) = \bigcup_{k \in \mathbb{N}} BR(\Delta_f, k), \tag{5-2}$$

where $BR(\Delta_f, k) = \Delta_f^{\pi,k}$, which is recursively defined as:

$$\begin{aligned} \Delta_{f,1}^{\pi} &= \{\delta_{f-1} \in \mathcal{D}(S) : \delta_f = \gamma^{\pi_f} \delta_{f-1},\ \delta_f \in \Delta_f,\ \gamma^{\pi_f} \in \mathcal{I}^{\pi_f}\}, \\ \Delta_{f,k}^{\pi} &= \{\delta_k \in \mathcal{D}(S) : \delta_{k+1} = \gamma^{\pi_{k+1}} \delta_k,\ \delta_{k+1} \in \Delta_{f,k+1}^{\pi},\ \gamma^{\pi_{k+1}} \in \mathcal{I}^{\pi_{k+1}}\}, \end{aligned} \tag{5-3}$$

To perform the set-based computation to obtain the set of distributions per step, we use the McCormick relaxations using Equation 4-7 to obtain a *relaxed* backward reachability map $\widehat{\mathcal{BR}}(\Delta)$, the resulting set is similar to (4-8). Furthermore, to obtain a tighter over-approximation, we use the sampling-based method introduced in section 4-2-3.

5-2 Robust Backward Reachability

Due to the issues described in section 4-2-4 in terms of the tightness of the reachability guarantees, these methods are better suited for problems where we need optimistic reachability guarantees. To obtain robust distributional reachability guarantees using IMDPs, we introduce a *value iteration* based method (Bertsekas, 1995). We start by explaining the method for IMCs, using ideas from Algorithm 1, and later show its formulation for IMDPs. This approach transforms the infinite-state problem over the continuous simplex into a finite-state problem, allowing for the application of dynamic programming to find the "winning set" of initial distributions that can guarantee reachability of a target set T within a finite horizon K. Furthermore, we illustrate the proposed method using our running example 3.1 and some variants of it.

5-2-1 Proposed Method

The core idea is to partition the probability simplex $\mathcal{D}(S)$ into a finite collection of smaller, non-overlapping simplices, forming a mesh. We then define a binary value function over this mesh, where a value of 1 indicates that any distribution within a given mesh cell can be robustly driven to the target set, and 0 otherwise. We describe the method as follows:

1. **Discretisation of the Distribution Space** Let the probability simplex $\mathcal{D}(S) \subset \mathbb{R}^n$ be discretised into a regular simplicial mesh (triangulation) $M = \{S_i\}_{i=1}^{N_{\text{simp}}}$, where each $S_i = \text{conv}(\mathcal{V}_i)$ is a simplex (a cell) with n vertices $\mathcal{V}_i$, and $\bigcup_{i=1}^{N_{\text{simp}}} S_i = \mathcal{D}(S)$. This mesh serves as our finite abstract state space.

2. **Value Function** We define a time-dependent value function $V_k : \{1, \dots, N_{\text{simp}}\} \to \{0, 1\}$ for each time step $k \in \{0, \dots, K\}$. A cell S_i is considered part of the winning set at step k, denoted $V_k(i) = 1$, if starting from *any* distribution $\delta \in S_i$, the system is guaranteed to reach the target set T by time k (in reverse time).

3. **One-Step Robust Predecessor** For any set of winning cells $W \subseteq M$, we define the robust one-step predecessor set, $\text{Pre}_\forall(W)$, as the set of all distributions δ from which *all* possible next-step distributions land within W. Formally, for the IMC,

$$\text{Pre}_\forall(W) = \{\delta \in \mathcal{D}(S) \mid \text{Img}(\delta) \subseteq W\}$$

   where $\text{Img}(\delta) = \{\delta\gamma \mid \gamma \in \Gamma\}$ is the one-step image of δ. Our algorithm will operate on the mesh, identifying cells S_i that are entirely contained within this predecessor set.

4. **Robust Bellman Update** The value function is computed via a backward recursion. We initialise the winning set at step $h = 0$ as all cells that intersect the target set:

$$V_0(i) = 1 \quad \iff \quad S_i \cap T \neq \emptyset$$

   For subsequent steps $h = 0, \dots, K-1$, the value function is updated via a two-stage robust Bellman update. First, we compute an intermediate value, $\tilde{V}_{k+1}(i)$, which determines if a cell S_i can be robustly driven into the winning set of the previous step, $W_k = \{j \mid V_k(j) = 1\}$:

$$\tilde{V}_{k+1}(i) = \min_{j \in J_i} V_k(j)$$

   where $J_i = \{j \mid \text{Img}(S_i) \cap S_j \neq \emptyset\}$ is the precomputed set of successor cell indices for cell S_i. This intermediate value is 1 if and only if all successors of S_i are in the winning set W_k.

   The final value for the next step, $V_{k+1}(i)$, is then updated to ensure the winning set is monotonically non-decreasing:

$$V_{k+1}(i) = \max(\tilde{V}_{k+1}(i), V_k(i))$$

   This ensures that once a cell is marked as winning, it remains winning for all subsequent steps. The final winning set is the union of all cells that are winning at any step up to the horizon K.

5-2-2 Algorithm and Implementation

The practical implementation of this value iteration scheme involves two main phases: a one-time precomputation of successor sets and an iterative loop to compute the value function as outlined in Algorithm 2.

Algorithm 2 Robust Backward Reachability on a Mesh for an IMC

1: **Input:** IMC transition bounds Γ, target set T, horizon K, mesh $M = \{S_i\}$ specified by a grid size L.
Precomputation Phase:
2: **for** each cell $S_i = \text{conv}(\mathcal{V}_i) \in M$ **do**
3: $\{\text{ext}(\mathcal{P}_r)\}_{r=1}^{n}$ ▷ extreme points of the row transition polytopes
4: $P_i \leftarrow \bigcup_{v \in \mathcal{V}_i} \text{vert}(\text{Img}(v))$ ▷ one-step image vertices for the cell
5: $J_i \leftarrow \emptyset$.
6: **for** each cell $S_j = \text{conv}(\mathcal{V}_j) \in M$ **do**
7: **if** $P_i \cap S_j \neq \emptyset$ **then**
8: $J_i \leftarrow J_i \cup \{j\}$.
9: **end if**
10: **end for**
11: **end for**
Value Iteration Phase:
12: $V_0 \leftarrow$ array of size N_{simp}, where $V_0(i) = 1$ if $S_i \cap T \neq \emptyset$, else 0.
13: $V \leftarrow V_0$.
14: $W_{\text{total}} \leftarrow \{i \mid V_0(i) = 1\}$.
15: **for** $h = 0$ to $K - 1$ **do**
16: $V_{\text{next}} \leftarrow V$.
17: **for** $i = 1$ to N_{simp} **do**
18: **if** $V(i) = 0$ **then**
19: $\tilde{V}_{k+1}(i) \leftarrow \min_{j \in J_i} V(j)$.
20: $V_{\text{next}}(i) \leftarrow \max(\tilde{V}_{k+1}(i), V(i))$.
21: **end if**
22: **end for**
23: $V \leftarrow V_{\text{next}}$.
24: $W_{\text{total}} \leftarrow W_{\text{total}} \cup \{i \mid V(i) = 1\}$.
25: **end for**
26: **Output:** The total winning set of indices W_{total}.

Algorithm 2 can be extended to IMDPs by just considering the best action, defined by choosing the action which maximises the number winning cells in the next step or equivalently that gives the best value function in line 22.

5-2-3 Complexity Analysis of the Algorithm

The computational complexity of the algorithm is heavily dominated by the one-time precomputation phase, which builds the successor graph.

1. **Mesh Generation:** For a state space of size n, the probability simplex is an $(n-1)$-dimensional object. A regular simplicial mesh with L divisions along each edge results in $N_{\text{simp}} = L^{n-1}$ cells, and a total of $N_{\text{pts}} = \binom{L+n-1}{n-1}$ grid points. The number of cells grows polynomially with L but exponentially with the number of states n.

2. **Successor Set (J_i) Computation:** This is the most intensive step. For each of the N_{simp} cells, we must check for intersection with all other N_{simp} cells. The intersection check between two polytopes, $\text{conv}(P_i)$ and $\text{conv}(\mathcal{V}_j)$, with v_i and v_j vertices respectively, is formulated as a feasibility LP. This LP has $v_i + v_j$ variables and $n+2$ constraints (n for the equality of points, 2 for the sum-to-one constraints on the convex combination weights). The theoretical complexity of solving an LP is polynomial in the number of variables and constraints. Therefore, the total complexity of this phase is $O(N_{\text{simp}}^2 \cdot C_{LP})$, where C_{LP} is the cost of solving a single LP.

3. **Value Iteration:** The iterative part is computationally efficient. Each of the K steps involves a single pass through the N_{simp} cells. For each cell i, the check $\min_{j \in J_i} V(j)$ takes $O(|J_i|)$ time. The total complexity of this phase is therefore $O(K \cdot \sum_{i=1}^{N_{\text{simp}}} |J_i|)$, which is linear in the number of edges in the precomputed successor graph.

Note. For a higher number of states, even for a 3x3 slippery grid world modelled using interval transition probabilities (see section A-2 for a complete description of the model), the computational demand is massive. This is because the simplicial partitioning for higher dimensions is computationally demanding. For example, in the 3x3 grid world, consisting of 9 states, for $L = 2$, $N_{\text{simp}} = 256$ and $N_{\text{pts}} = 45$. This leads to an increase in the computation of the J_i set for all the points for even coarse grid levels. This is made worse due to the computation of convex hull of the vertices in $n-1$ dimensions. Furthermore, finer grid partitions lead to better accuracy in identifying regions in the distribution space which reach the target distribution set in the defined horizon. Hence, due to the significant increase in the computation times to obtain better results, we need to look at better ways to obtain the convex hull or the intersecting simplex indices thereof, possibly via parallelising the computation load over the grid obtained.

5-3 Guarantees of reach-avoid specifications

The distributional reachability analysis introduced in the previous sections provides a characterisation of how sets of state distributions evolve under fixed policies over a given number of steps. However, many practical verification problems, particularly in safety-critical applications, require reasoning not just about reachable distributions, but about whether these distributions satisfy specific temporal specifications such as reach-avoid objectives. In the distributional setting, these specifications require that the system reaches a designated target set of distributions while remaining within a safe set of distributions in all previous steps. Consequently, system evolution depends both on the choice of policy, which selects actions, and the choice of adversary, which selects feasible transition probabilities from the ambiguity sets at each step.

5-3-1 Satisfaction of a Reach-Avoid Specification

Let $H \subseteq \mathcal{D}(S)$ denote the safe set, and $T \subseteq \mathcal{D}(S)$ denote the target set in the distributional space. Given an initial distribution $\delta_0 \in \mathcal{D}(S)$, a trajectory $\{\delta_k\}_{k\in\mathbb{N}}$ satisfies the reach-avoid specification if there exists a finite time $\tau \in \mathbb{N}$ such that:

$$\forall k < \tau, \delta_k \in H \text{ and } \delta_\tau \in T.$$

Two variants of the distributional reach-avoid problem can be formulated:

- *Existential (optimistic) variant:* Determine whether there exists a policy $\pi \in \Pi$ and an adversary $\gamma_{\text{adv}} \in \Gamma_{\text{adv}}$ such that, for the given initial distribution δ_0, the induced trajectory satisfies the reach-avoid specification:

  $$\exists \pi \in \Pi,\ \exists \gamma_{\text{adv}} \in \Gamma,\ \exists \tau \in \mathbb{N} :\ \forall k < \tau,\ \delta_k \in H \text{ and } \delta_\tau \in T.$$

  This formulation checks whether there exists some combination of policy and favourable transition realisations that can satisfy the specification.

- *Robust (universal) variant:* Determine whether there exists a policy $\pi \in \Pi$ such that, for all adversaries $\gamma_{\text{adv}} \in \Gamma_{\text{adv}}$, the reach-avoid specification is satisfied:

  $$\exists \pi \in \Pi,\ \forall \gamma_{\text{adv}} \in \Gamma,\ \exists \tau \in \mathbb{N} :\ \forall k < \tau,\ \delta_k \in H \text{ and } \delta_\tau \in T.$$

  This variant guarantees that the reach-avoid property holds regardless of how the adversary resolves the transition uncertainty.

In the case of MDPs, both problems can be formulated over the forward and backward reachable sets previously introduced, as shown in (Gao et al., 2023). For the universal problem, one seeks the existence of the largest initial distribution set that lie within the backward reachable set of the target, while remaining inside the safe set during all intermediate steps. For the existential problem, one can check that forward reachable sets initiated from the initial distribution never leave the safe set before entering the target set.

5-4 Discussion

In this chapter, we developed a distributional framework for reachability analysis of IMDPs. Unlike classical state-based verification, this framework allows us to reason directly over the evolution of probability distributions under uncertainty. The convexity properties of reachable sets were established for the induced IMC case, and we extended the framework to IMDPs using occupation measures and transition occupation measures, allowing tractable relaxations via McCormick envelopes.

We further introduced backward reachable set computations, providing algorithms for both reachability and verification tasks under distributional specifications. The distinction between existential and universal problems was formalised, capturing both policy synthesis and robust verification goals.

In the following chapter, we apply the developed methods to concrete case studies that illustrate these methods, demonstrate verification of distributional reach-avoid specifications, and highlight the advantages and trade-offs of the proposed approaches.

Chapter 6

Case Studies

The preceding chapter established the formalism of distributional reachability for IMDPs, detailing the value-iteration–based backward operator, McCormick relaxations for bilinear terms, and the specification of the running example. The present chapter operationalises that framework through two case studies to examine algorithmic behaviour, approximation tightness, and robustness to model uncertainty. All computational experiments reported here were executed on a Dell G15 5520 equipped with an Intel Core i5-12500H processor and 16 GB RAM. The code for this is available on my GitHub repository[1].

6-1 Structured Swarm Deployment via Distributional Reachability

Inspired by the Swarm Deployment case study by Gao et al. (2023), we demonstrate a similar swarm deployment for IMDPs using the backward reachability set (5-1) using McCormick envelopes as described in section 5-1.

6-1-1 Problem Setup

We study a structured deployment task where $n = 100$ agents move in the state space S defined as 10×10 discrete grid. Each cell in the grid corresponds to a state $s \in S$, and the swarm state at any time is described by a distribution $\delta \in \mathcal{D}(S)$, representing the fraction of agents in each cell. Agents take actions from the set $A = \{\mathrm{N}, \mathrm{S}, \mathrm{E}, \mathrm{W}, \mathrm{Stay}\}$. Let $Y \subseteq S$ denote the fixed "UP" outline in the grid (see Figure 6-1).

We consider transition uncertainty with:

- Intended transition probabilities in $[0.75, 0.85]$,
- Slip transitions in $[0.10, 0.20]$,
- No static obstacles or explicitly unsafe states.

[1] https://github.com/vkaza52/thesisdatabase/tree/main/code

6-1-2 Reachability Specification

The objective is to steer the agent distribution into a *UP*-shaped region (i.e., the target set Y) within a finite horizon while satisfying two soft constraints:

- At least $\alpha = 0.9$ total probability mass must lie within Y at the final time.
- Each target state in Y must contain at least 2% agents.
- Finally, we add the constraint that at most 6% agents are present in any cell at any given point of time to ensure there is no congestion of space.

We express the above specifications as shown below:

$$\Delta_f \;=\; \Big\{\delta \in \mathcal{D}(S) \;\Big|\; \sum_{y\in Y} \delta(y) \;\geq\; \alpha, \quad \delta(y) \;\geq\; \tfrac{2}{|S|} \;\; \forall\, y \in Y, \quad \delta(s) \;\leq\; \tfrac{6}{|S|} \;\; \forall\, s \in S\Big\}, \quad \text{(6-1)}$$

For this specification, we want to synthesise a policy $\pi : S \to \mathcal{D}(A)$ for an IMDP $\mathcal{I} = (S, A, \Gamma)$, where Γ is as defined in Definition 3.13 and characterised by (3-7), based on the description above for the interval transition probabilities, using (4-5), to control the agents with respect to the distributions of the agents. In other words, we are looking at a high-level policy rather than an individual agent-level policy to reach the target distribution set.

6-1-3 Methodology

We use encode the recursive backward reachable set defined by (5-1), which are relaxed using McCormick envelopes as defined in (4-8) to provide convex relaxations on the bilinear program. This is encoded in a linear program as described below:

$$\begin{aligned}
\max \quad & \sum_{y\in Y} \delta_N(y) \\
\text{s.t.} \quad & \delta_0 = \tfrac{1}{n}\mathbf{1}, \\
& \delta_k \in \mathcal{D}(S) \;\; (k = 0, \ldots, N), \\
& \sum_{a\in A} Q_k(s,a) = \delta_k(s) \quad \forall s, \;\; k = 0, \ldots, N{-}1, \\
& \quad W_k(s,a,s') \geq \check{\gamma}_{s,a}(s') Q_k(s,a), \\
& \quad W_k(s,a,s') \geq \check{\gamma}_{s,a}(s') - (1 - Q_k(s,a))\hat{\gamma}_{s,a}(s'), \\
& \quad W_k(s,a,s') \leq \hat{\gamma}_{s,a}(s') Q_k(s,a), \\
& \quad W_k(s,a,s') \leq \hat{\gamma}_{s,a}(s') - (1 - Q_k(s,a))\check{\gamma}_{s,a}(s'), \forall (s,a,s'), \;\; k = 0, \ldots, N{-}1 \\
& \sum_{s'} W_k(s,a,s') = Q_k(s,a) \quad \forall (s,a), \;\; k = 0, \ldots, N{-}1, \\
& \delta_{k+1}(s') = \sum_{s\in S} \sum_{a\in A} W_k(s,a,s') \quad \forall s', \;\; k = 0, \ldots, N{-}1, \\
& \delta_k(s) \leq \tfrac{6}{n} \quad \forall s, \;\; k = 0, \ldots, N, \\
& \sum_{y\in Y} \delta_N(y) \geq \alpha, \;\; \delta_N(y) \geq \tfrac{2}{n} \;\; \forall y \in Y.
\end{aligned} \quad \text{(6-2)}$$

Given the specifications of the required final distribution, agent congestion constraints, we solve the problem using backward reachability computations over the given horizon. One can then extract the policy obtained using (4-5).

6-1-4 Results

A feasible policy was found within a horizon of $N = 6$ steps. Figure 6-1 shows the evolution of the distribution under this policy. The UP-shape becomes progressively clearer, with more than 90% of the agent mass eventually reaching the designated region. To synthesise the policy, on average over 5 runs, it takes about 44.37 seconds. Furthermore, upon increasing the slip probability from $[0.10, 0.20] \rightarrow [0.10, 0.25]$ while keeping the other constraints the same, the problem became infeasible. This shows the sensitivity of the method to uncertainties for a fixed horizon.

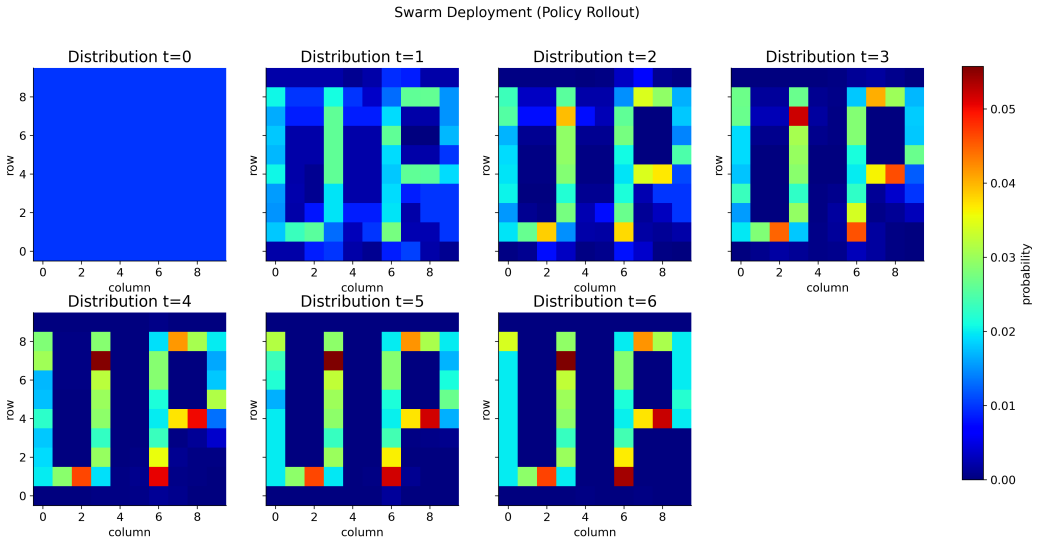


Figure 6-1: Evolution of swarm distribution over 6 time steps, forming the target UP-shape.

6-1-5 Limitations

While the approach demonstrates promising results for structured swarm deployment, several limitations remain:

1. **Outer Approximation Guarantees**: The backward reachable sets computed using McCormick relaxations are outer approximations. As a result, the synthesised policy may be overly *conservative* and not tight with respect to the true reachable set. Furthermore, due to the use of the sampling-based method, the inner approximations obtained might not be the tightest bound of the true set.
2. **Scalability:** The use of convex programming with high-dimensional variables (e.g., occupation measures, McCormick auxiliary variables) leads to computational overhead. For larger grids or longer horizons, solving the associated programs becomes increasingly expensive.

3. **Existential Guarantees:** The method described for this case study only provides existential guarantees, in the sense that it is an optimistic scenario of the existence of policy and corresponding transition probability matrix. This is also because of the outer approximations that are used to compute the set, which makes it hard to provide robust universal guarantees.

6-2 Value Iteration based Robust Backward Reachability

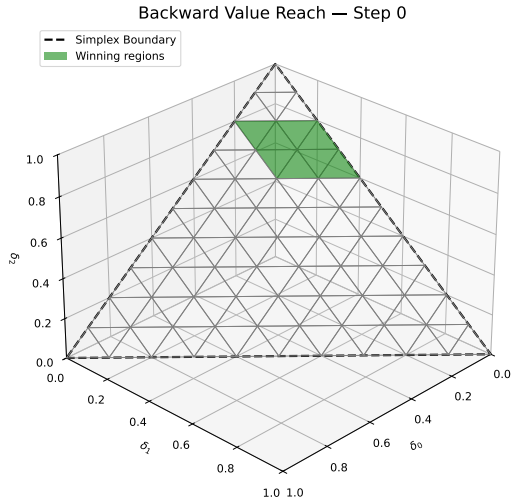
In this case study, we apply the value iteration-based algorithm described in Algorithm 2 to our running example 3.1 under action a_0 to obtain the largest initial set of distributions from which we are guaranteed to reach the target set of distributions under adversarial conditions. The primary objective of this analysis is to demonstrate the functionality of the algorithm and to visually illustrate the step-by-step computation of the robust winning set. By fixing a policy, in this instance, selecting action deterministically across all states, the IMDP from the running example is reduced to an IMC. Figure 6-2 shows the target distribution set considered for all the scenarios that follow for varying grid size L. Since we perform a backward iteration the step 0 for all the simulations is the same and is hence only shown once. Going forward, we only show the system evolution from step 1.

6-2-1 Running Example under Action a_0

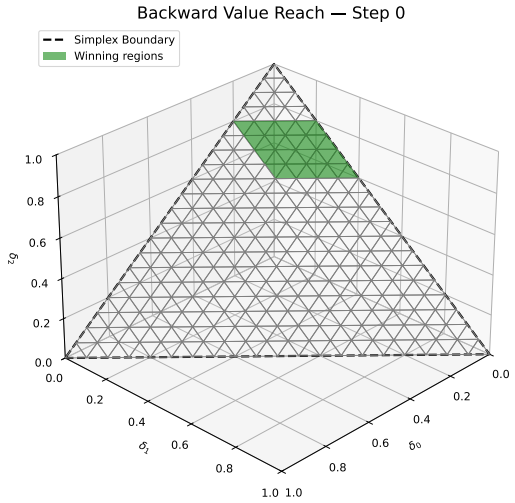
It can be observed from Figure 6-3 that the value iteration runs for the horizon $H = 5$ and $L = 60$, highlighting the winning cells at the end of the time horizon in *green* and intermediate winning regions are marked in *orange*. At each step, the algorithm recursively computes the set of simplices from which the target region can be reached. The algorithm stops when no new regions are found. To verify if the obtained largest initial distribution set holds, we use a forward analysis by sampling random simplices from the partitioned space and observing their trajectory over the time horizon. From the forward analysis shown in Figure 6-4, Figure 6-5, it can be seen that, indeed, the largest initial distribution obtained holds. All the cells which are part of the winning set obtained in the backward iteration reach the target set of distributions within the given horizon. Furthermore, it was also observed that for higher values of L, the defined winning sets were more accurate. Similar plots are provided for different grid sizes and transition probability matrices in the section A-3.

6-2-2 Analysis of the Computation Times

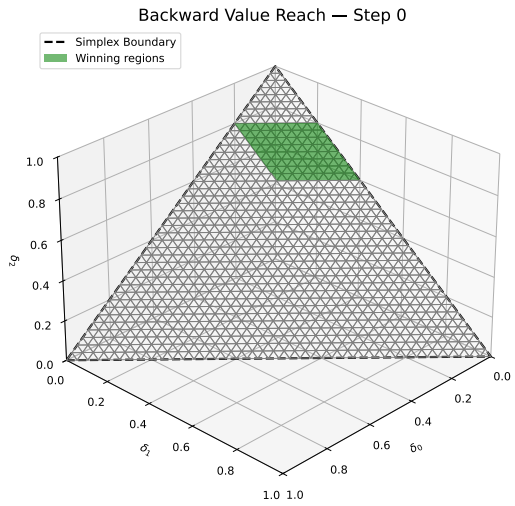
For the running example under action a_0, we provide the computation times of the pre-computation, value iteration and the forward analysis below. Table 6-1 reports the computation times for the different components of the algorithm as the grid level increases from 10 to 60. The pre-computation stage dominates the runtime, growing from approximately 8.2 s at grid level 10 to more than 760 s at $L = 60$. In contrast, the value iteration step remains negligible. The forward analysis also increases significantly with resolution, reaching about 30 s at $L = 60$. This aligns with the complexity analysis presented in subsection 5-2-3. Furthermore, this highlights the trade-off between accuracy and speed, since the winning cells are more accurate at a higher grid size but also take longer computation time. This suggests



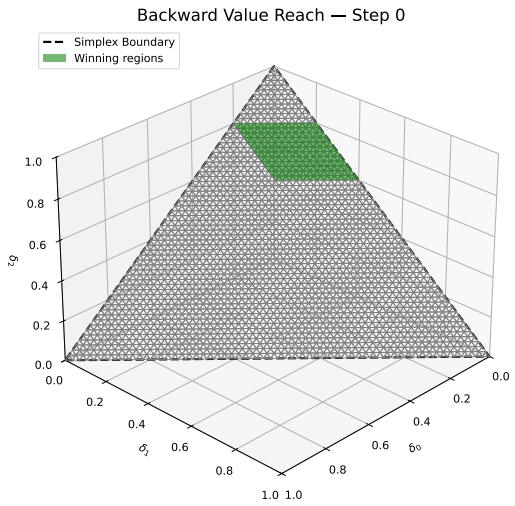
(a) $L = 10$



(b) $L = 20$

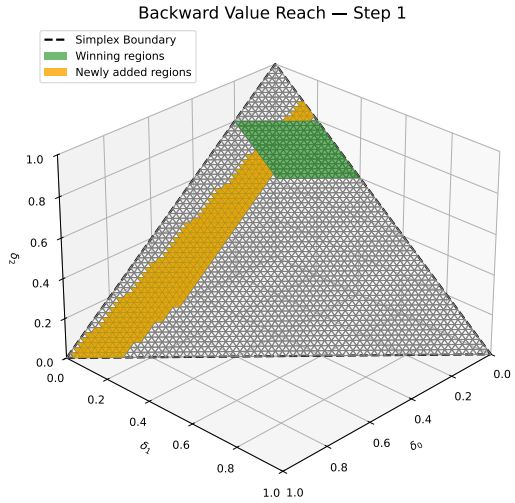


(c) $L = 40$

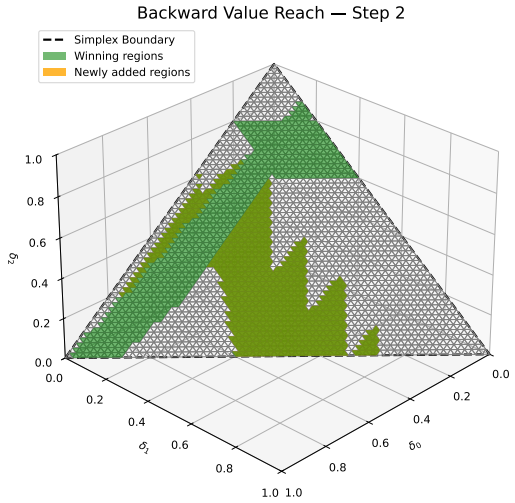


(d) $L = 60$

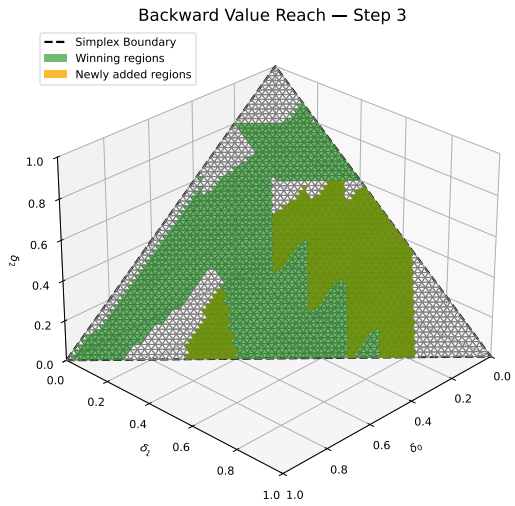
Figure 6-2: Target distribution for varying grid sizes.



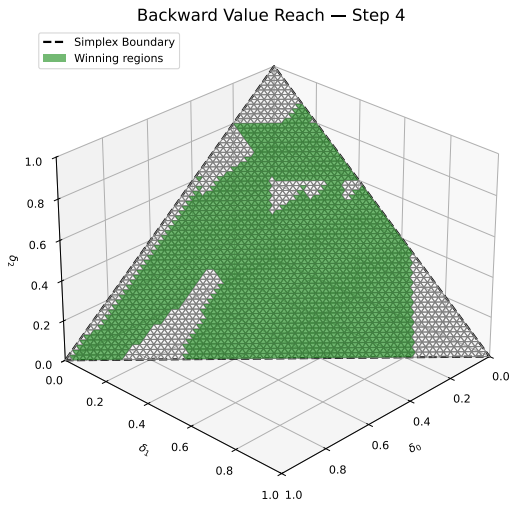
(a) $k = 1$



(b) $k = 2$

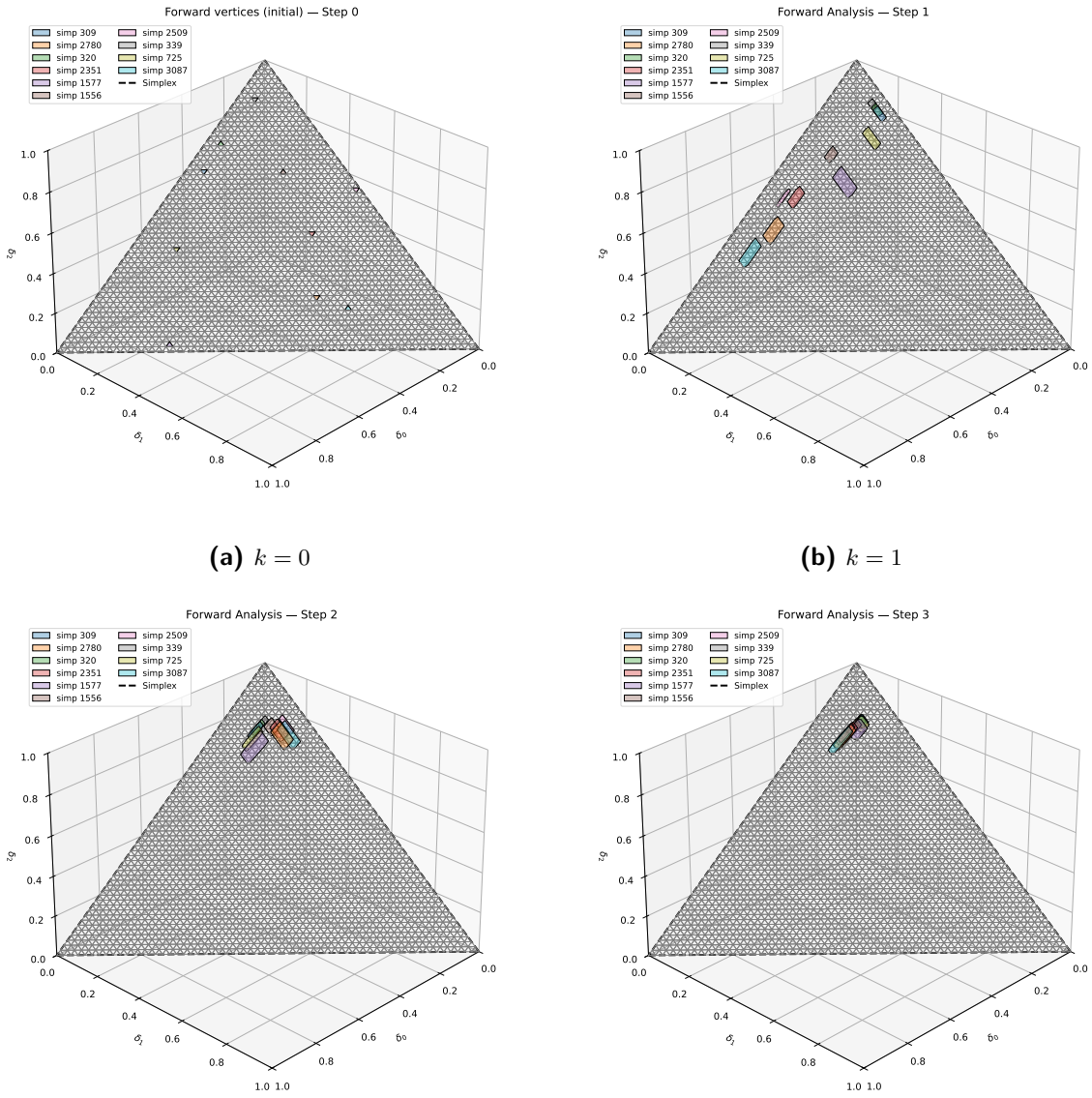


(c) $k = 3$



(d) $k = 4$

Figure 6-3: Backward reachability for $L = 60$ for the running example under action a_0.



(a) $k = 0$ **(b)** $k = 1$

(c) $k = 2$ **(d)** $k = 3$

Figure 6-4: Forward analysis for $L = 60$ for the running example under action a_0 until $k = 3$.

that for large-scale problems, optimising or parallelising the pre-computation phase would yield the largest performance gains, which is left as future work.

6-3 Discussion

The smaller grid and reduced population for the swarm deployment case study are chosen to keep computations tractable while still illustrating distributional reachability under interval uncertainty. Compared to standard MDP formulations such as those in (Gao et al., 2023), this

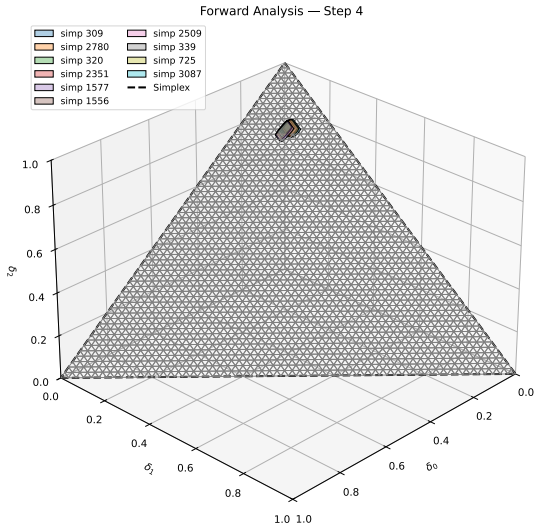


Figure 6-5: Forward analysis for $L = 60$ for the running example under action a_0 at $k = 4$

Table 6-1: Computation times (in seconds) for different grid levels ($H = 5$).

Grid Level	Precomputation	Value Iteration	Forward Analysis
10	8.18	0.0033	2.04
20	38.31	0.0099	4.09
40	221.16	0.0300	11.87
60	763.83	0.1025	30.69

IMDP setting introduces additional complexity due to the presence of interval uncertainty sets $\Gamma_{s,a}$, requiring joint reasoning over both policy randomness and adversarial uncertainty. This results in higher-dimensional convex feasibility problems involving auxiliary variables and additional constraints, making the problem computationally harder even for relatively small state spaces. Furthermore, in scenarios where data regarding the transition probabilities is unavailable, data-driven methods need to be integrated into the existing approach for the distributional reachability analysis.

Chapter 7

Conclusion & Future Work

Chapter Summary
This chapter summarises the thesis contributions, acknowledges its limitations, and proposes directions for future research. The thesis successfully addressed the problem of distributional reachability in multi-agent systems by developing a framework that models system evolution directly over probability distributions using IMDPs to capture uncertainty. The primary contributions include the formalisation of the distributional reachability problem, the analysis of the convex geometry of reachable sets, and the development of a computational framework for their computation. The work is limited by its focus on verification for fixed policies rather than synthesis, its finite-horizon formulation, the assumption of known uncertainty intervals, and the potential for conservatism in over-approximations. This research contributes to the shift from state-based verification to providing population-level safety guarantees for complex stochastic systems, offering a powerful method for analysis in uncertain environments.

7-1 Conclusion

This thesis addressed the challenging problem of distributional reachability for multi-agent systems operating under both stochastic dynamics and interval uncertainty. In many real-world scenarios—such as swarm robotics, air traffic control, and autonomous vehicle coordination—the safety and performance of the system are better described at the distributional level rather than at the level of individual agents or trajectories. Existing probabilistic verification techniques typically focus on state-based or path-based reachability, which often becomes computationally intractable or insufficiently expressive when applied to high-dimensional multi-agent systems.

To overcome these limitations, this thesis adopted a *distribution transformer view* of MDPs, wherein the evolution of the system is modelled directly as transformations over probability distributions on the state space. This perspective naturally accommodates population-level safety specifications and allows for compact representation even in systems with large numbers

of agents. However, real-world systems are rarely described by precisely known dynamics. To address this, the thesis extended the distributional framework to accommodate model uncertainty using IMDPs, where each transition probability is bounded within specified intervals to capture epistemic and aleatoric uncertainty.

The primary contributions of this thesis can be summarised as follows:

1. **Problem Formalisation :** The central problem of forward and backward distributional reachability was formally defined, focusing on computing or characterising the set of all possible state distributions that can arise under uncertain transitions and stochastic policies.
2. **Convex Geometry of Reachable Sets:** The reachable sets were formulated as convex sets evolving through affine transformations induced by both the policy and the adversarial choices of transition probabilities within the specified intervals. The mathematical framework leveraged convex analysis tools, including Minkowski sums, support functions, and polytopic representations, to analyse the evolution and structure of reachable sets.
3. **Computational Framework:** A recursive algorithm was proposed to compute k-step reachable sets by iteratively propagating convex sets under uncertain affine transformations. The computational framework provides both exact and over-approximate reachable sets, depending on the complexity of the underlying system and the degree of conservatism required.
4. **Case Studies:** The proposed framework was applied to representative case studies, including multi-agent reachability problem, a case study using the robust backward reachability problem for the running example, demonstrating how the developed methods can be employed to provide reachability guarantees under uncertainty.

Through these contributions, the thesis demonstrated that robust, distribution-level verification is both theoretically tractable and practically relevant for multi-agent systems, particularly when system-level specifications depend on aggregate behaviours rather than individual trajectories.

7-2 Limitations of the Current Work

Although the contributions presented in this thesis advance the state-of-the-art in distributional verification, several limitations remain that are important to acknowledge:

1. **Finite-Horizon Formulation:** The reachability problem was primarily treated in a finite-time horizon setting. Extending the framework to infinite-horizon formulations or probabilistic invariance problems would require additional theoretical developments.
2. **Known Uncertainty Intervals:** The interval uncertainty was assumed to be known. In practice, uncertainty may be time-varying, dependent on system state, or partially learned online. Incorporating such dynamic uncertainty models would increase realism but also complexity.

3. **Conservativeness in Over-Approximation:** The polytopic convex approximations used for reachable sets can become increasingly conservative over longer horizons, particularly when uncertainty propagates through multiple steps.

4. **Computational Scalability:** Whilst the use of convex polytopes enables tractable computation in small-to-moderate state spaces, scalability to very high-dimensional systems remains challenging due to the exponential growth of constraints in the polytope representation. Similar issues remain for the proposed value iteration method where finding the intersecting simplices remains a bottleneck.

5. **Limited Case Studies:** While a practical case study for the existential variant is provided, similar case studies for the value iteration-based backward reachability method were hard to compute due to the exponential increase in the computation time. However, improving the algorithm to obtain the intersecting simplices is expected to provide reasonable improvement in performance, which is left for future work.

7-3 Computational Aspects

A key contribution of this thesis is the development of a *computationally tractable framework* for distributional reachability under interval uncertainty, especially for the existential variant of the problem. In contrast to classical state-based approaches (e.g. value iteration for MDPs or interval iteration for IMDPs), which operate on scalar reachability probabilities, the proposed framework tracks convex sets of reachable distributions over the probability simplex, incorporating both policy stochasticity and adversarial interval uncertainty.

This formulation introduces additional computational complexity. Each one-step propagation involves Minkowski sums and convex hull operations over polytopes in $\mathbb{R}^{|S|}$. The worst-case complexity of convex hull computation grows exponentially with the number of vertices and the state space dimension. Consequently, after k steps, the size of the polytope representation may grow as $\mathcal{O}(V^k)$, where V denotes the number of vertices generated per step. The overall complexity is thus super-polynomial in both horizon length and state space size.

While symbolic IMDP tools (e.g. `IMPaCT`, `IntervalMDP.jl`) achieve faster runtimes by restricting to state-wise intervals, they cannot capture the full distributional evolution addressed here. The increased computational burden reflects the richer problem structure and the stronger guarantees provided at the distribution level.

In practice, the existential method remains feasible for problems with up to tens of states and moderate horizons, as demonstrated in the case study. Scaling to larger systems would require structural assumptions or approximate set representations that trade exactness for tractability.

7-4 Future Work

While the methods developed in this thesis form the foundations for robust distributional reachability analysis, they also open several promising directions for future research.

7-4-1 Data-Driven Uncertainty Quantification

The IMDP model employed here assumes known interval bounds on transition probabilities. In practice, such bounds are often estimated from empirical data. Data-driven abstraction methods, such as scenario-based convex programs (Badings et al., 2022; Lavaei et al., 2023), Gaussian Process regression (Skovbekk et al., 2024), or Bayesian non-parametric uncertainty sets (Reed et al., 2023), could be integrated into the distributional framework to automatically generate valid IMDP models with statistical guarantees.

Open questions remain regarding sample complexity and confidence bounds when estimating uncertainty sets for distributional specifications. Techniques from PAC-learning for stochastic systems (Lavaei et al., 2023) may serve as a foundation for addressing these challenges.

7-4-2 Scalability to High-Dimensional Systems

The proposed framework remains computationally feasible for moderate state spaces, but scaling to high-dimensional systems remains a key challenge.

Recent work on orthogonally decoupled IMDPs (odIMDPs) offers one promising direction by factorising uncertainty along state dimensions, thereby achieving linear complexity scaling under decomposability assumptions (Mathiesen et al., 2025). Zonotope-based methods (Yang et al., 2022) and set-based representations exploiting system structure could also serve to reduce conservatism and improve scalability.

Distributed and decentralised abstraction methods (Meshkat Alsadat et al., 2024; Coppola et al., 2024) that exploit local agent interaction structures may further enable tractable reachability computations in multi-agent systems.

7-4-3 Richer Formal Specifications

The current formulation focuses on finite-horizon reach-avoid problems. Extending the framework to handle richer formal specifications, such as those specified in temporal logics (e.g., PCTL, LTL) (Baier and Katoen, 2008), could enable verification of complex tasks.

Multi-objective policy synthesis under interval uncertainty has been studied for state-based IMDPs (Hahn et al., 2017), and extending such methods to distributional objectives could broaden the applicability of the framework.

7-4-4 Adversarial Learning and Game-Theoretic Extensions

Since IMDPs naturally encode worst-case adversarial models, there is potential for integrating adversarial reinforcement learning and game-theoretic approaches. Recent developments in distributionally robust reinforcement learning (Suilen et al., 2024) and Wasserstein-based ambiguity models (Mazumdar et al., 2024) could be extended to explicitly handle interval-based uncertainty models.

Game-theoretic models, such as Stackelberg formulations for robust hierarchical control (van Zutphen et al., 2024), may further enhance robustness against intelligent or adaptive adversaries that exploit weaknesses in stochastic systems.

7-4-5 Integration with Real-World Systems

Finally, applying these methods to real-world systems remains an important long-term goal. Multi-agent scenarios such as UAV swarm navigation under uncertain disturbances, traffic coordination in autonomous vehicle networks, warehouse robotics, and distributed sensing systems present natural testbeds for distributional reachability verification.

Bridging the gap between theoretical guarantees and real-world deployment will require advances in system identification, safety certification, and scalable verification toolchains (Wooding and Lavaei, 2024; Mathiesen et al., 2024).

7-5 Ethical Considerations

The framework for verifying multi-agent systems presented in this thesis has a significant dual-use nature, posing both societal benefits and ethical risks. Positively, these methods can enhance the safety and reliability of civilian autonomous systems. Applications include safer management of autonomous vehicle fleets, coordinated air traffic control for drone swarms, and effective deployment of robots for search and rescue operations. By providing formal guarantees for population-level behaviour, this research contributes to the trustworthy integration of autonomous technology into society.

However, the same principles carry profound ethical risks if applied to military contexts. The concepts could be used to optimise autonomous surveillance swarms or to guarantee the success of coordinated attacks by autonomous weapons systems. The ability to formally verify that a swarm can reach a target while avoiding defences could automate offensive military actions to a dangerous degree. This underscores a need for the scientific community to engage in a transparent discourse and for robust public oversight to prevent the misuse of such powerful technologies.

7-6 Closing Remarks

The methods developed in this thesis contribute to the growing body of research that moves beyond classical state-based verification toward distributional (population-level) safety guarantees for complex stochastic systems. The combination of interval transition probability modelling, convex geometric methods, and multi-agent reachability offers a powerful combination for both analysis and synthesis in uncertain environments. With continued development in the directions outlined above, distributional reachability has the potential to play a key role in the safe and scalable deployment of autonomous multi-agent systems.

Appendix A

Appendix

A-1 Statement on the use of AI tools

In this thesis, I used ChatGPT and Grammarly to enhance my language and better articulate my ideas. I have also used ChatGPT to help me code some helper functions required by my code.

A-2 Slippery Grid World for Stochastic Navigation

The primary goal in the stochastic navigation problem in a slippery grid world is to find the initial set of distributions from which the agent can reach the target region while avoiding the obstacle region with at least a given probability α. The actions available are {UP, DOWN, LEFT, RIGHT, STAY}. An action whose transition leads toward the wall renders the agent to stay in its cell with a probability of 1, while any other chosen action has an interval transition probability $[0.75, 0.85]$ along with a probability of slipping into cells perpendicular to the chosen action with a probability interval $[0.05, 0.20]$. Furthermore, it is assumed that the obstacle and target states are absorbing, i.e., once they hit those states, the only action possible is STAY with a probability of 1.

A-3 Plots for the Value Iteration based Backward Reachability from section 6-2

A-3-1 More results for running example under action a_0

In this section, we provide some more results for the running example under action a_0 for $L = 10, 40$ as shown in Figure A-2, Figure A-3, Figure A-4 and Figure A-5. It can be seen that as the grid size increases, the accuracy of the winning sets found also increases.

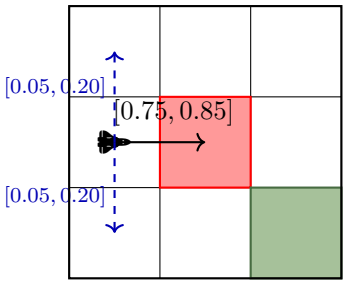
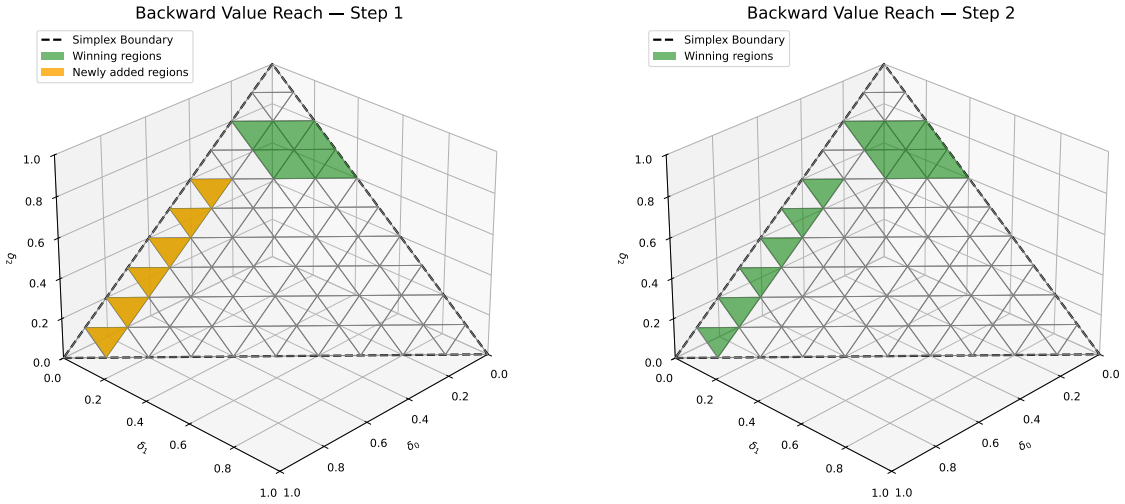


Figure A-1: 3×3 slippery grid used in the experiments. The red cell is an obstacle, and the green cell is the target state, both of which are considered to be absorbing states. Executing RIGHT from the middle-left cell succeeds with probability in $[0.75, 0.85]$ (solid arrow) and slips to the orthogonal neighbours with probabilities in $[0.05, 0.20]$ each (dashed arrows).

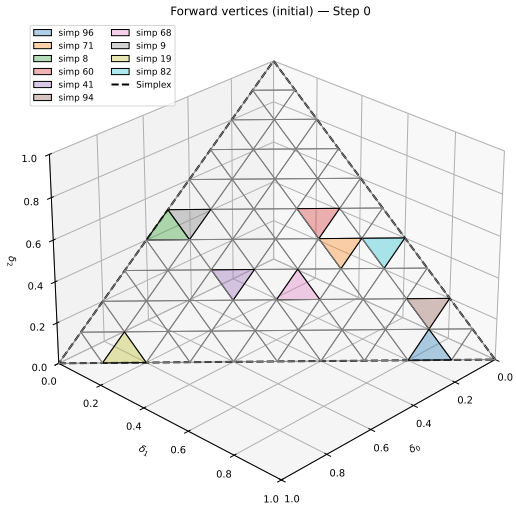


(a) $k = 1$ **(b)** $k = 2$

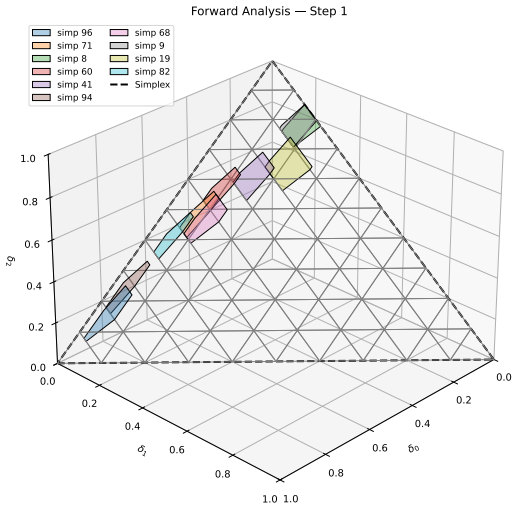
Figure A-2: Backward Iteration for running example under a_0 for $L = 10$.

A-3-2 Testing the results using different transition probability matrices

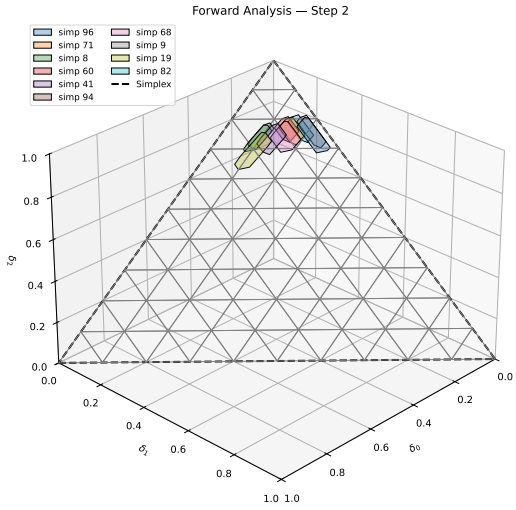
To further validate the results, we perform a similar analysis considering the following IMC in Figure A-6:



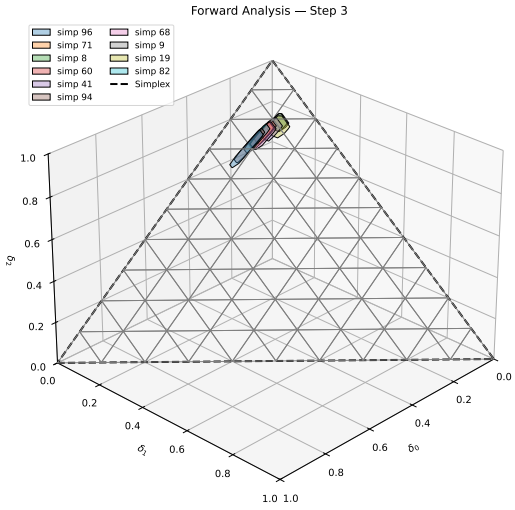
(a) $k = 0$



(b) $k = 1$

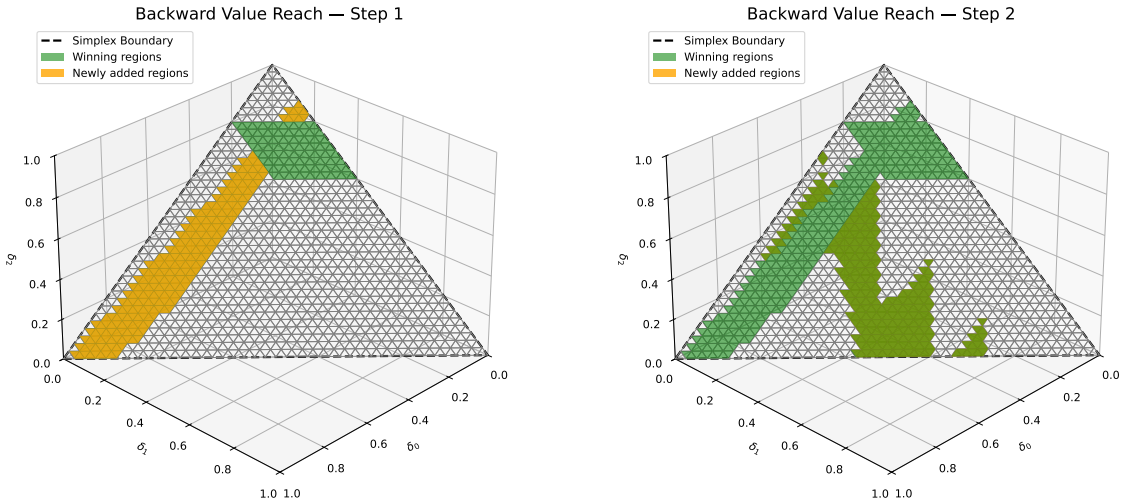


(c) $k = 2$



(d) $k = 3$

Figure A-3: Forward Analysis of the running example under a_0 for $L = 10$.



(a) $k = 1$ **(b)** $k = 2$

Figure A-4: Backward Iteration for running example under a_0 for $L = 40$.

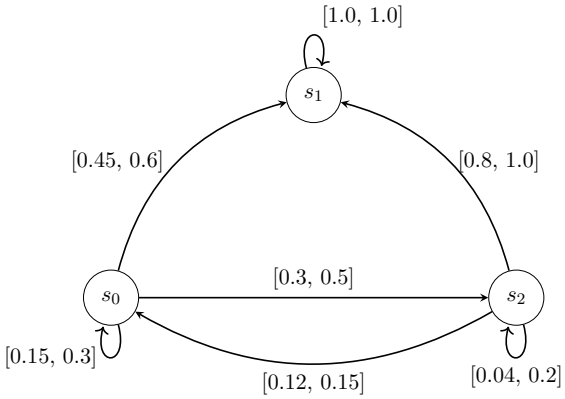
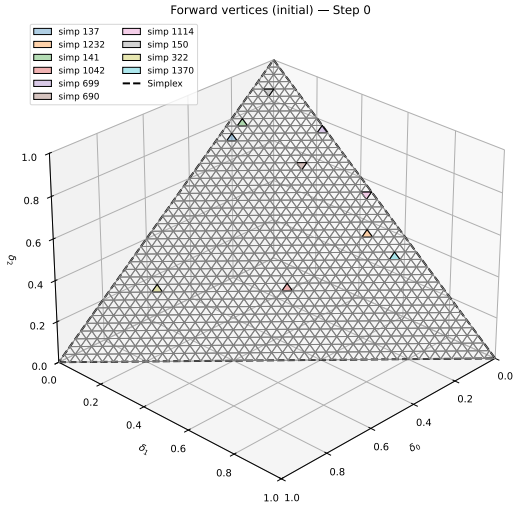
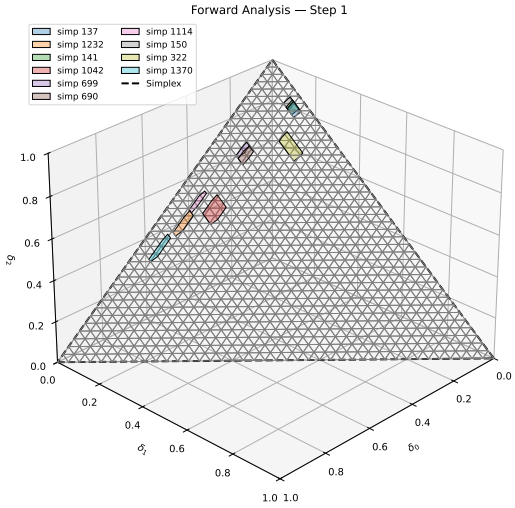


Figure A-6: IMC considered for illustration.

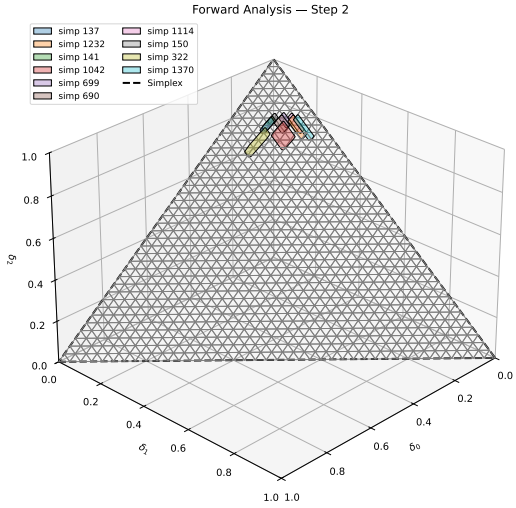
Table A-1 summarises the computation times for the IMC from Figure A-6 experiments with grid levels ranging from 10 to 60. As before, the pre-computation time is the dominant cost and grows sharply with resolution, reaching over 3000 s at grid level 60. The forward analysis also increases significantly, while the value iteration remains negligible. The significant increase in the pre-computation time compared to the previously reported one from the running example could be due to a larger number of extreme points for the defined transition probability matrices and also possibly due to running multiple experiments at the same time on the system, leading to a higher load on the processor. It can be seen that the distributions starting in the winning set reach the target at some point before the horizon ends. Since the IMC consists of an absorbing state $s2$, over time, if there is a feasible transition to $s2$ and hence the distribution moves towards the $s2$ distribution corner on the plane, and this can be observed in Figure A-8, Figure A-10.



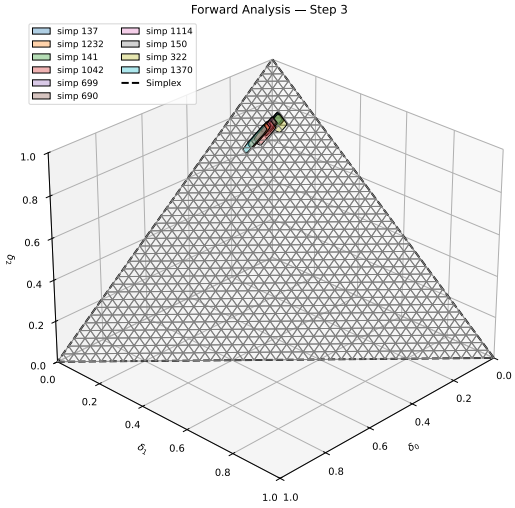
(a) $k = 0$



(b) $k = 1$



(c) $k = 2$

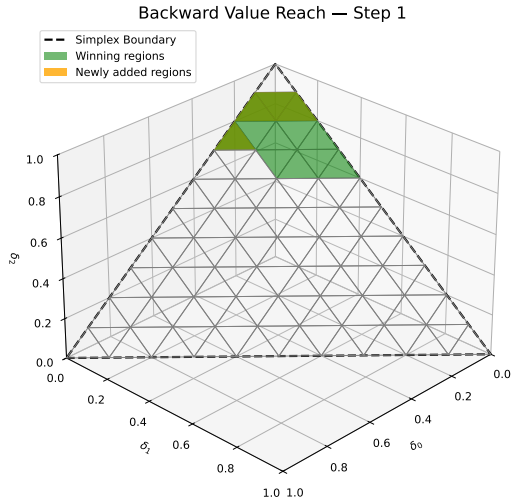


(d) $k = 3$

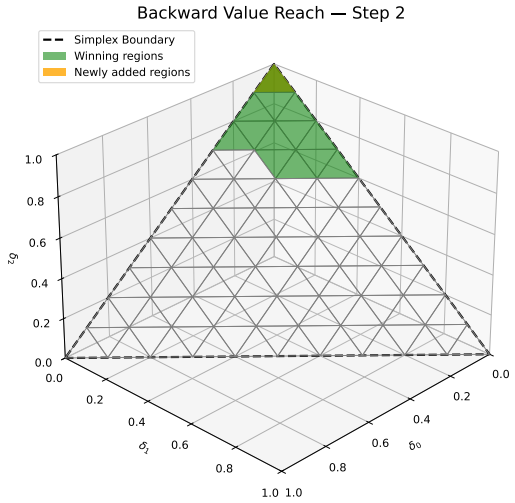
Figure A-5: Forward Analysis of the running example under a_0 for $L = 40$.

Table A-1: Computation times (in seconds) for IMC in Figure A-6 ($H = 5$).

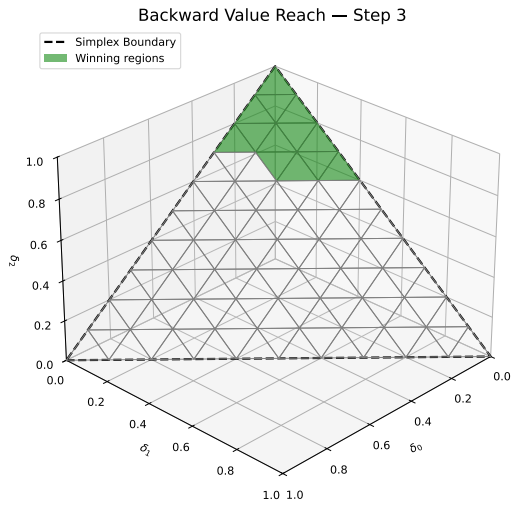
Grid Level	Precomputation	Value Iteration	Forward Analysis
10	20.74	0.0078	5.03
20	151.29	0.0093	10.73
40	1054.62	0.0631	31.96
60	3088.69	0.1016	47.93



(a) $k = 1$

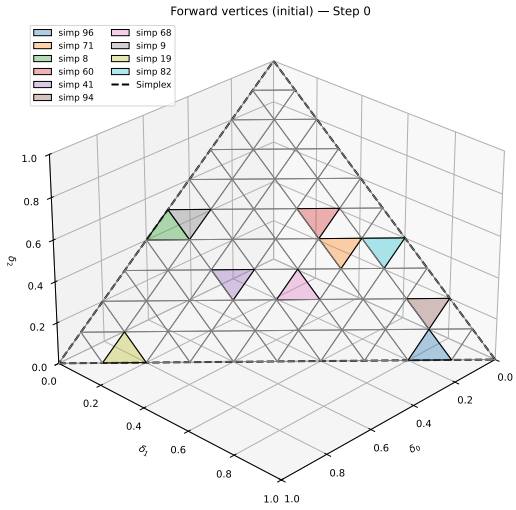


(b) $k = 2$

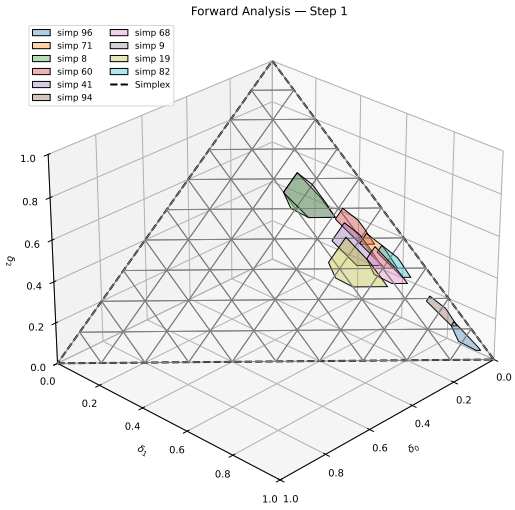


(c) $k = 3$

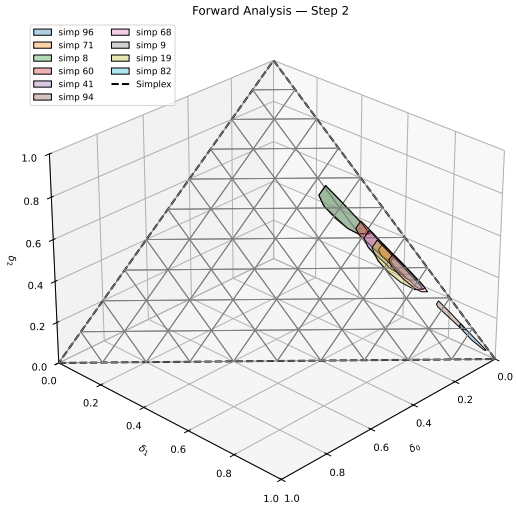
Figure A-7: Backward iteration for the illustration for $L = 10$.



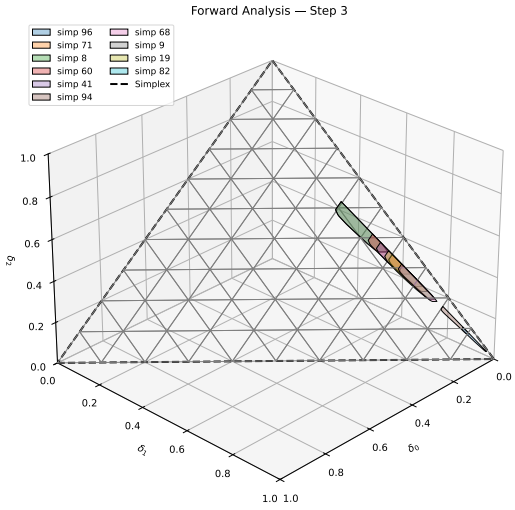
(a) $k = 0$



(b) $k = 1$

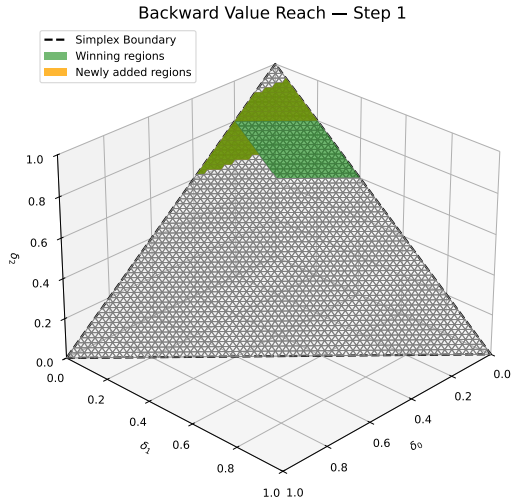


(c) $k = 2$

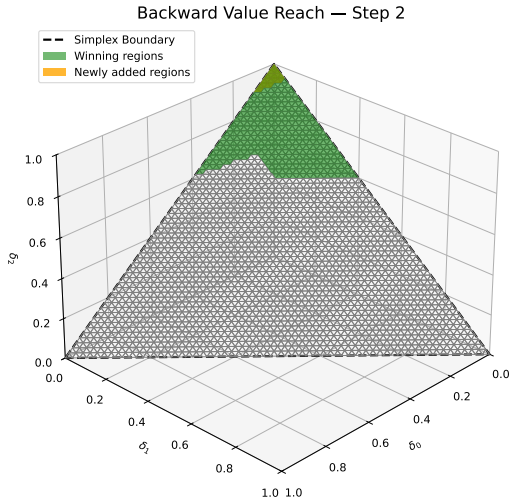


(d) $k = 3$

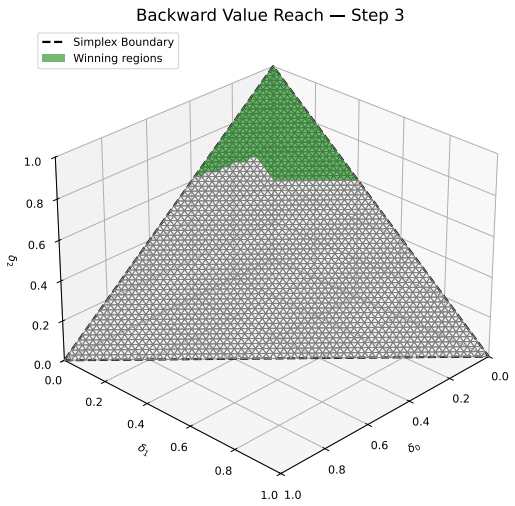
Figure A-8: Forward Analysis for the illustration for $L = 10$.



(a) $k = 1$

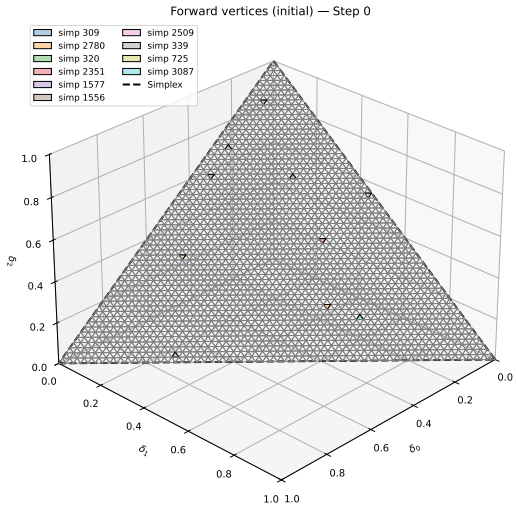


(b) $k = 2$

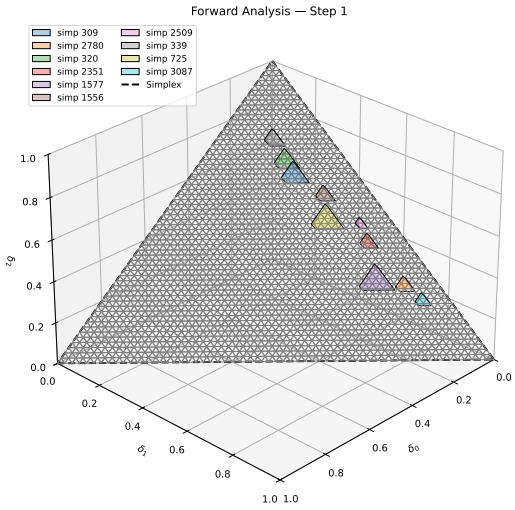


(c) $k = 3$

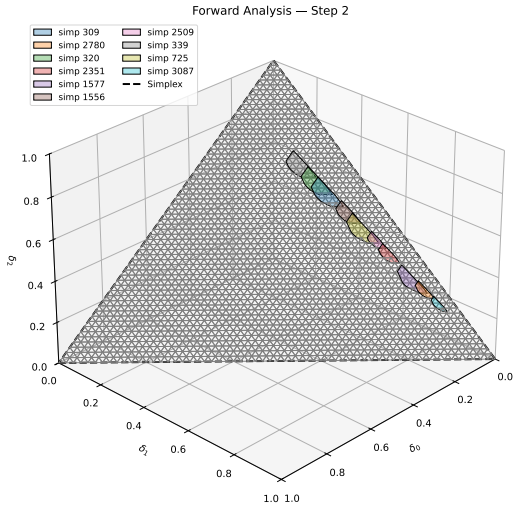
Figure A-9: Backward iteration for the illustration for $L = 60$.



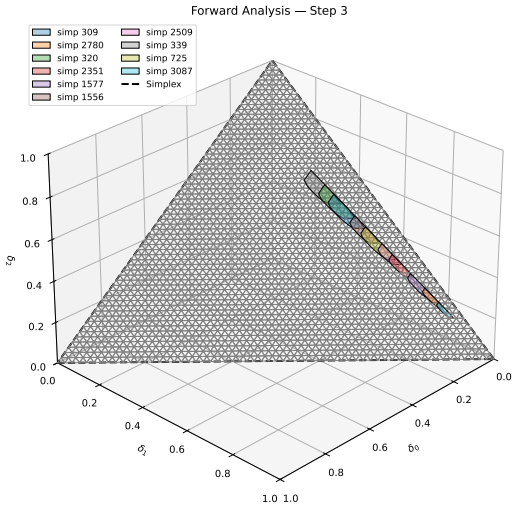
(a) $k = 0$



(b) $k = 1$



(c) $k = 2$



(d) $k = 3$

Figure A-10: Forward Analysis for the illustration for $L = 60$.

Bibliography

A. Abate, M. Prandini, J. Lygeros, and S. Sastry. Probabilistic reachability and safety for controlled discrete time stochastic hybrid systems. *Automatica*, 44(11):2724–2734, 2008. ISSN 0005-1098. doi: 10.1016/j.automatica.2008.03.027.

A. Abate, J.-P. Katoen, J. Lygeros, and M. Prandini. Approximate Model Checking of Stochastic Hybrid Systems. *European Journal of Control*, 16(6):624–641, 2010. ISSN 0947-3580. doi: 10.3166/ejc.16.624-641.

S. Akshay, B. Genest, and N. Vyas. Distribution-based objectives for Markov Decision Processes. In *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science*, LICS '18, page 36–45, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450355834. doi: 10.1145/3209108.3209185.

S. Akshay, K. Chatterjee, T. Meggendorfer, and Đ. Žikelić. MDPs as Distribution Transformers: Affine Invariant Synthesis for Safety Objectives. In C. Enea and A. Lal, editors, *Computer Aided Verification*, pages 86–112, Cham, 2023. Springer Nature Switzerland. ISBN 978-3-031-37709-9.

S. Akshay, K. Chatterjee, T. Meggendorfer, and Đ. Žikelić. Certified Policy Verification and Synthesis for MDPs under Distributional Reach-avoidance Properties. In K. Larson, editor, *Proceedings of the 33rd International Joint Conference on Artificial Intelligence, IJCAI 2024*, IJCAI International Joint Conference on Artificial Intelligence, pages 3–12. International Joint Conferences on Artificial Intelligence, 2024.

M. Althoff and G. Frehse. Combining Zonotopes and Support Functions for Efficient Reachability Analysis of Linear Systems. 12 2016. doi: 10.1109/CDC.2016.7799418.

E. Altman. *Constrained Markov Decision Processes*. PhD thesis, INRIA, 1995.

R. Alur, T. A. Henzinger, G. J. Pappas, and G. Lafferriere. Discrete Abstractions of Hybrid Systems. *Proceedings of the IEEE*, 88(106):971–984, 2000.

D. Amodei, C. Olah, J. Steinhardt, P. Christiano, J. Schulman, and D. Mané. Concrete problems in AI safety. *arXiv preprint arXiv:1606.06565*, 2016.

T. Badings, L. Romao, A. Abate, and N. Jansen. Probabilities are not enough: formal controller synthesis for stochastic dynamical models with epistemic uncertainty. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*, AAAI'23/IAAI'23/EAAI'23. AAAI Press, 2023a. ISBN 978-1-57735-880-0. doi: 10.1609/aaai.v37i12.26718.

T. Badings, L. Romao, A. Abate, D. Parker, H. A. Poonawala, M. Stoelinga, and N. Jansen. Robust Control for Dynamical Systems with Non-Gaussian Noise via Formal Abstractions. *Journal of Artificial Intelligence Research*, 76:341–391, Jan. 2023b. ISSN 1076-9757. doi: 10.1613/jair.1.14253.

T. S. Badings, A. Abate, N. Jansen, D. Parker, H. A. Poonawala, and M. Stoelinga. Sampling-Based Robust Control of Autonomous Systems with Non-Gaussian Noise. *Proceedings of the AAAI Conference on Artificial Intelligence*, 36(9):9669–9678, Jun. 2022. doi: 10.1609/aaai.v36i9.21201.

C. Baier and J.-P. Katoen. *Principles of Model Checking*. MIT Press, 2008.

R. Bellman. Dynamic Programming and Stochastic Control Processes. *Inf. Control.*, 1(3): 228–239, 1958. doi: 10.1016/S0019-9958(58)80003-0.

M. d. Berg, O. Cheong, M. v. Kreveld, and M. Overmars. *Computational Geometry: Algorithms and Applications*. Springer-Verlag TELOS, Santa Clara, CA, USA, 3rd ed. edition, 2008. ISBN 3540779736.

D. Bertsekas. *Dynamic Programming and Optimal Control*, volume 1. 01 1995.

F. Blanchini and S. Miani. *Set-Theoretic Methods in Control*. Birkhäuser Basel, 1st edition, 2007. ISBN 0817632557.

S. P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge university press, 2004.

M. Brambilla, E. Ferrante, M. Birattari, and M. Dorigo. Swarm robotics: a review from the swarm engineering perspective. *Swarm Intelligence*, 7(1):1–41, 2013.

T. Brázdil, K. Chatterjee, M. Chmelik, V. Forejt, J. Křetínskỳ, M. Kwiatkowska, T. Meggendorfer, D. Parker, and M. Ujma. Learning algorithms for verification of markov decision processes. *TheoretiCS*, 4, 2025.

M. Bujarbaruah, R. Spica, and C. N. Jones. Reachability analysis for safe learning of linear systems using zonotopes. *IEEE Control Systems Letters*, 5(5):1655–1660, 2021.

N. Cauchi, L. Laurenti, M. Lahijanian, A. Abate, M. Kwiatkowska, and L. Cardelli. Efficiency through uncertainty: scalable formal synthesis for stochastic hybrid systems. In *Proceedings of the 22nd ACM International Conference on Hybrid Systems: Computation and Control*, HSCC '19, page 240–251, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450362825. doi: 10.1145/3302504.3311805.

R. Chadha, V. A. Korthikanti, M. Viswanathan, G. Agha, and Y. Kwon. Model Checking MDPs with a Unique Compact Invariant Set of Distributions. In *QEST*, pages 121–130, 2011.

K. Chatterjee and T. A. Henzinger. *Value Iteration*, pages 107–138. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. ISBN 978-3-540-69850-0. doi: 10.1007/978-3-540-69850-0_7.

K. Chatterjee, E. Kafshdar Goharshadi, M. Karrabi, P. Novotnỳ, and D. Zikelic. Solving long-run average reward robust MDPs via stochastic games. In *33rd International Joint Conference on Artificial Intelligence*, 2024.

Y. Chen, T. T. Georgiou, and M. Pavon. Optimal Steering of a Linear Stochastic System to a Final Probability Distribution, Part I. *IEEE Transactions on Automatic Control*, 61(5): 1158–1169, 2016a. doi: 10.1109/TAC.2015.2457784.

Y. Chen, T. T. Georgiou, and M. Pavon. Optimal Steering of a Linear Stochastic System to a Final Probability Distribution, Part II. *IEEE Transactions on Automatic Control*, 61(5): 1170–1180, 2016b. doi: 10.1109/TAC.2015.2457791.

R. Coppola, A. Peruffo, L. Romao, A. Abate, and M. Mazo. Data-driven Interval MDP for Robust Control Synthesis. *ArXiv*, abs/2404.08344, 2024.

G. Delimpaltadakis, M. Lahijanian, M. Mazo Jr., and L. Laurenti. Interval Markov Decision Processes with Continuous Action-Spaces. In *Proceedings of the 26th ACM International Conference on Hybrid Systems: Computation and Control*, HSCC '23, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9798400700330. doi: 10.1145/3575870.3587117.

V. Forejt, M. Kwiatkowska, G. Norman, and D. Parker. *Automated Verification Techniques for Probabilistic Systems*, pages 53–113. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. ISBN 978-3-642-21455-4. doi: 10.1007/978-3-642-21455-4_3.

Y. Gao, A. Abate, L. Xie, and K. Johansson. Distributional Reachability for Markov Decision Processes: Theory and Applications. *IEEE Transactions on Automatic Control*, PP:1–16, 01 2023. doi: 10.1109/TAC.2023.3341282.

R. Givan, S. Leach, and T. Dean. Bounded-parameter Markov decision processes. *Artificial Intelligence*, 122(1):71–109, 2000. ISSN 0004-3702. doi: 10.1016/S0004-3702(00)00047-3.

S. Haddad and B. Monmege. Interval iteration algorithm for MDPs and IMDPs. *Theoretical Computer Science*, 735:111–131, 2018. ISSN 0304-3975. doi: 10.1016/j.tcs.2016.12.003. Reachability Problems 2014: Special Issue.

E. M. Hahn, V. Hashemi, H. Hermanns, M. Lahijanian, and A. Turrini. Multi-objective Robust Strategy Synthesis for Interval Markov Decision Processes. In N. Bertrand and L. Bortolussi, editors, *Quantitative Evaluation of Systems*, pages 207–223, Cham, 2017. Springer International Publishing. ISBN 978-3-319-66335-7.

C. Hensel, S. Junges, J.-P. Katoen, T. Quatmann, and M. Volk. The probabilistic model checker Storm. *Int. J. Softw. Tools Technol. Transf.*, 24(4):589–610, Aug. 2022. ISSN 1433-2779. doi: 10.1007/s10009-021-00633-z.

M. Henzinger and K. Chatterjee. Faster and Dynamic Algorithms For Maximal End-Component Decomposition And Related Graph Problems In Probabilistic Verification. In *Symposium on Discrete Algorithms (SODA)*, San Francisco, January 2011.

G. N. Iyengar. Robust Dynamic Programming. *Mathematics of Operations Research*, 30(2): 257–280, 2005. ISSN 0364765X, 15265471.

J. Jackson, L. Laurenti, E. Frew, and M. Lahijanian. Safety Verification of Unknown Dynamical Systems via Gaussian Process Regression. In *2020 59th IEEE Conference on Decision and Control (CDC)*, page 860–866. IEEE Press, 2020. doi: 10.1109/CDC42340.2020.9303814.

S. Jafarpour and S. Coogan. A Contracting Dynamical System Perspective toward Interval Markov Decision Processes. In *Proceedings of the IEEE Conference on Decision Control*. IEEE, 2023.

V. A. Korthikanti, M. Viswanathan, G. Agha, and Y. Kwon. Reasoning about MDPs as Transformers of Probability Distributions. In *2010 Seventh International Conference on the Quantitative Evaluation of Systems*, pages 199–208, 2010. doi: 10.1109/QEST.2010.35.

M. Kwiatkowska, G. Norman, and D. Parker. PRISM 4.0: Verification of Probabilistic Real-Time Systems. In G. Gopalakrishnan and S. Qadeer, editors, *Computer Aided Verification*, pages 585–591, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg. ISBN 978-3-642-22110-1.

M. Lahijanian, S. B. Andersson, and C. Belta. Approximate Markovian abstractions for linear stochastic systems. In *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*, pages 5966–5971, 2012. doi: 10.1109/CDC.2012.6426184.

M. Lahijanian, S. B. Andersson, and C. Belta. Formal Verification and Synthesis for Discrete-Time Stochastic Systems. *IEEE Transactions on Automatic Control*, 60(8):2031–2045, 2015. doi: 10.1109/TAC.2015.2398883.

L. Laurenti and M. Lahijanian. Unifying Safety Approaches for Stochastic Systems: From Barrier Functions to Uncertain Abstractions via Dynamic Programming. *ArXiv*, abs/2310.01802, 2023.

L. Laurenti, M. Lahijanian, A. Abate, L. Cardelli, and M. Kwiatkowska. Formal and Efficient Synthesis for Continuous-Time Linear Stochastic Hybrid Processes. *IEEE Transactions on Automatic Control*, PP:1–1, 02 2020. doi: 10.1109/TAC.2020.2975028.

A. Lavaei, S. Soudjani, A. Abate, and M. Zamani. Automated verification and synthesis of stochastic hybrid systems: A survey. *Automatica*, 146:110617, 2022. ISSN 0005-1098. doi: 10.1016/j.automatica.2022.110617.

A. Lavaei, S. Soudjani, E. Frazzoli, and M. Zamani. Constructing MDP abstractions using data with formal guarantees. *IEEE Control Systems Letters*, 7:460–465, 2023. doi: 10.1109/LCSYS.2022.3188535.

J. Lei, Y.-X. Li, and C. K. Ahn. Optimizing Multi-Agent Systems With Uncertain Dynamics: A Finite-Time Adaptive Distributed Approach. *IEEE Transactions on Signal and Information Processing over Networks*, 9:865–874, 2023. doi: 10.1109/TSIPN.2023.3338467.

F. B. Mathiesen, M. Lahijanian, and L. Laurenti. IntervalMDP.jl: Accelerated value iteration for interval markov decision processes. *IFAC-PapersOnLine*, 58(11):1–6, 2024. ISSN 2405-8963. doi: 10.1016/j.ifacol.2024.07.416. 8th IFAC Conference on Analysis and Design of Hybrid Systems ADHS 2024.

F. B. Mathiesen, S. Haesaert, and L. Laurenti. 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, page 1–12. ACM, May 2025. doi: 10.1145/3716863.3718031.

R. Mazouz, F. B. Mathiesen, L. Laurenti, and M. Lahijanian. Piecewise Stochastic Barrier Functions, 2024.

A. Mazumdar, Y. Hou, and R. Wisniewski. Distributionally Robust Safety Verification for Markov Decision Processes. *arXiv preprint arXiv:2411.15622*, 2024.

G. P. McCormick. Computability of global solutions to factorable nonconvex programs: Part I—Convex underestimating problems. *Mathematical programming*, 10(1):147–175, 1976.

S. Meshkat Alsadat, N. Baharisangari, and Z. Xu. Distributed on-the-fly control of multi-agent systems with unknown dynamics: Using limited data to obtain near-optimal control. In A. Abate, M. Cannon, K. Margellos, and A. Papachristodoulou, editors, *Proceedings of the 6th Annual Learning for Dynamics amp; Control Conference*, volume 242 of *Proceedings of Machine Learning Research*, pages 1440–1451. PMLR, 15–17 Jul 2024.

A. Mitsos, B. Chachuat, and P. I. Barton. McCormick-Based Relaxations of Algorithms. *SIAM Journal on Optimization*, 20(2):573–601, 2009. doi: 10.1137/080717341.

A. Nilim and L. El Ghaoui. Robust Control of Markov Decision Processes with Uncertain Transition Matrices. *Oper. Res.*, 53(5):780–798, Sept. 2005. ISSN 0030-364X. doi: 10.1287/opre.1050.0216.

A. Nilim and L. Ghaoui. Robustness in Markov Decision Problems with Uncertain Transition Matrices. In S. Thrun, L. Saul, and B. Schölkopf, editors, *Advances in Neural Information Processing Systems*, volume 16. MIT Press, 2003.

S. Prajna, A. Jadbabaie, and G. J. Pappas. A Framework for Worst-Case and Stochastic Safety Verification Using Barrier Certificates. *IEEE Transactions on Automatic Control*, 52(8):1415–1428, 2007. doi: 10.1109/TAC.2007.902736.

M. L. Puterman. Markov Decision Processes: Discrete Stochastic Dynamic Programming, 1994.

R. Reed, L. Laurenti, and M. Lahijanian. Promises of Deep Kernel Learning for Control Synthesis. *IEEE Control Systems Letters*, 7:3986–3991, 2023. doi: 10.1109/LCSYS.2023.3340995.

C. Santoyo, M. Dutreix, and S. Coogan. A barrier function approach to finite-time stochastic system verification and control. *Automatica*, 125:109439, 2021. ISSN 0005-1098. doi: 10.1016/j.automatica.2020.109439.

M. Schranz, M. Umlauft, M. Sende, and W. Elmenreich. Swarm Robotic Behaviors and Current Applications. *Frontiers in Robotics and AI*, 7, 04 2020. doi: 10.3389/frobt.2020.00036.

W. Schwarting, J. Alonso-Mora, and D. Rus. Planning and Decision-Making for Autonomous Vehicles. *Annual Review of Control, Robotics, and Autonomous Systems*, 1(Volume 1, 2018):187–210, 2018. ISSN 2573-5144. doi: 10.1146/annurev-control-060117-105157.

J. Skovbekk, L. Laurenti, E. Frew, and M. Lahijanian. Formal Abstraction of General Stochastic Systems via Noise Partitioning. *IEEE Control Systems Letters*, 7:3711–3716, 2023. doi: 10.1109/LCSYS.2023.3340621.

J. Skovbekk, L. Laurenti, E. Frew, and M. Lahijanian. Formal Verification of Unknown Dynamical Systems via Gaussian Process Regression, 2024.

S. E. Z. Soudjani, C. Gevaerts, and A. Abate. FAUST^2: Formal Abstractions of Uncountable-STate STochastic Processes. In C. Baier and C. Tinelli, editors, *Tools and Algorithms for the Construction and Analysis of Systems*, pages 272–286, Berlin, Heidelberg, 2015. Springer Berlin Heidelberg. ISBN 978-3-662-46681-0.

M. Suilen, T. Badings, E. M. Bovy, D. Parker, and N. Jansen. Robust markov decision processes: A place where AI and formal methods meet. In *Principles of Verification: Cycling the Probabilistic Landscape: Essays Dedicated to Joost-Pieter Katoen on the Occasion of His 60th Birthday, Part III*, pages 126–154. Springer, 2024.

S. Summers and J. Lygeros. Verification of discrete time stochastic hybrid systems: A stochastic reach-avoid decision problem. *Automatica*, 46(12):1951–1961, 2010. ISSN 0005-1098. doi: 10.1016/j.automatica.2010.08.006.

M. van Zutphen, G. Delimpaltadakis, W. Heemels, and D. Antunes. Predictable interval MDPs through entropy regularization. In *2024 IEEE 63rd Conference on Decision and Control (CDC)*, page 6659–6664. IEEE, Dec. 2024. doi: 10.1109/cdc56724.2024.10886853.

W. Wiesemann, D. Kuhn, and B. Rustem. Robust Markov Decision Processes. *Mathematics of Operations Research*, 38(1):153–183, 2013. ISSN 0364765X, 15265471.

A. Wijs, J.-P. Katoen, and D. Bošnački. GPU-Based Graph Decomposition into Strongly Connected and Maximal End Components. In A. Biere and R. Bloem, editors, *Computer Aided Verification*, pages 310–326, Cham, 2014. Springer International Publishing. ISBN 978-3-319-08867-9.

B. Wooding and A. Lavaei. IMPaCT: A parallelized software tool for imdp construction and controller synthesis with convergence guarantees. In *Proceedings of the 27th ACM International Conference on Hybrid Systems: Computation and Control*, HSCC '24, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400705229. doi: 10.1145/3641513.3652532.

I. Yang. A convex optimization approach to distributionally robust Markov decision processes with Wasserstein distance. *IEEE control systems letters*, 1(1):164–169, 2017.

L. Yang, H. Zhang, J.-B. Jeannin, and N. Ozay. Efficient Backward Reachability Using the Minkowski Difference of Constrained Zonotopes. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 41(11):3969–3980, 2022.

X. Yang and T. H. Summers. Tight interval MDP abstraction of uncertain linear systems. In *2022 ACM/IEEE 13th International Conference on Cyber-Physical Systems (ICCPS)*, pages 47–58. IEEE, 2022.

M. Zamani, M. Rungger, and P. M. Esfahani. Approximations of stochastic hybrid systems: A compositional approach. *IEEE Transactions on Automatic Control*, 62(6):2838–2853, 2016.

G. M. Ziegler. Lectures on Polytopes. *Graduate Texts in Mathematics*, 152, 1995.

Glossary

List of Acronyms

conv	convex hull
MC	Markov Chain
MDP	Markov Decision Process
RMDP	Robust Markov Decision Process
IMDP	Interval Markov Decision Process
IMC	Interval Markov Chain