TU Delft

Delft University of Technology

Efficient strategy synthesis for switched stochastic systems with distributional uncertainty

Gracia, Ibon; Boskos, Dimitris; Lahijanian, Morteza; Laurenti, Luca; Mazo, Manuel

DOI
10.1016/j.nahs.2024.101554

Publication date
2025

Document Version
Final published version

Published in
Nonlinear Analysis: Hybrid Systems

Citation (APA)
Gracia, I., Boskos, D., Lahijanian, M., Laurenti, L., & Mazo, M. (2025). Efficient strategy synthesis for switched stochastic systems with distributional uncertainty. *Nonlinear Analysis: Hybrid Systems*, *55*, Article 101554. https://doi.org/10.1016/j.nahs.2024.101554

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Green Open Access added to TU Delft Institutional Repository

'You share, we take care!' - Taverne project

https://www.openaccess.nl/en/you-share-we-take-care

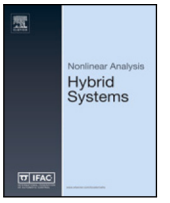
Otherwise as indicated in the copyright section: the publisher is the copyright holder of this work and the author uses the Dutch legislation to make this work public.

Nonlinear Analysis: Hybrid Systems 55 (2025) 101554

Contents lists available at ScienceDirect

Nonlinear Analysis: Hybrid Systems

journal homepage: www.elsevier.com/locate/nahs



Efficient strategy synthesis for switched stochastic systems with distributional uncertainty

Ibon Gracia [a,*], Dimitris Boskos [b], Morteza Lahijanian [a], Luca Laurenti [b], Manuel Mazo Jr. [b]

[a] University of Colorado Boulder, 3775 Discovery Dr, Boulder, 80303, CO, USA
[b] Delft University of Technology, Mekelweg 5, Delft, 2628, CD, Netherlands

ARTICLE INFO

Keywords:
Switched stochastic systems
Formal synthesis
Safe autonomy
Uncertain Markov decision processes
Wasserstein distance

ABSTRACT

We introduce a framework for the control of discrete-time switched stochastic systems with uncertain distributions. In particular, we consider stochastic dynamics with additive noise whose distribution lies in an ambiguity set of distributions that are ε−close, in the Wasserstein distance sense, to a nominal one. We propose algorithms for the efficient synthesis of distributionally robust control strategies that maximize the satisfaction probability of reach-avoid specifications with either a given or an arbitrary (not specified) time horizon, i.e., unbounded-time reachability. The framework consists of two main steps: finite abstraction and control synthesis. First, we construct a finite abstraction of the switched stochastic system as a *robust Markov decision process* (robust MDP) that encompasses both the stochasticity of the system and the uncertainty in the noise distribution. Then, we synthesize a strategy that is robust to the distributional uncertainty on the resulting robust MDP. We employ techniques from optimal transport and stochastic programming to reduce the strategy synthesis problem to a set of linear programs, and propose a tailored and efficient algorithm to solve them. The resulting strategies are correctly refined into switching strategies for the original stochastic system. We illustrate the efficacy of our framework on various case studies comprising both linear and non-linear switched stochastic systems.

1. Introduction

Switched stochastic systems are a ubiquitous class of control systems due to their capability to capture digitally controlled physical systems affected by noise. In this class, a controller can switch among a finite set of modes to achieve the desired behavior [1]. Systems that can be represented with such models are found in various real-world applications, including robotics [2] and cyber–physical systems [3]. These systems are often *safety-critical*, requiring formal guarantees of correctness, and their noise characteristics are *uncertain*, with statistical properties only available through the use of statistical estimation techniques and possibly subject to *distributional shifts* [4]. However, formal control synthesis and verification methods for switched stochastic systems typically assume exact and known noise characteristics. We aim to address this fundamental gap in the literature and focus on the following question: *how to derive formal guarantees for stochastic systems with uncertain noise characteristics?*

In this work, we introduce a formal framework for synthesis of distributionally robust control strategies for discrete-time switched stochastic systems with uncertain noise distribution under reach-avoid specifications. More precisely, we consider switched stochastic

* Corresponding author.
E-mail addresses: ibon.gracia@colorado.edu (I. Gracia), d.boskos@tudelft.nl (D. Boskos), morteza.lahijanian@colorado.edu (M. Lahijanian), l.laurenti@tudelft.nl (L. Laurenti), m.mazo@tudelft.nl (M. Mazo Jr.).

https://doi.org/10.1016/j.nahs.2024.101554
Received 25 February 2024; Accepted 26 October 2024
Available online 14 November 2024
1751-570X/© 2024 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

systems with general dynamics (possibly non-linear) with additive noise in each mode, and while the exact probability distribution of noise is unknown, we assume it belongs to a prescribed Wasserstein ambiguity set, i.e., the set of possible distributions defined as a ball of radius $\epsilon > 0$ around a nominal distribution under the Wasserstein distance [5,6]. For instance, such a set can be estimated using data-driven techniques, e.g., see [5]. The probabilistic reach-avoid specification can include an arbitrary or predefined time horizon to reach a target set. Such specifications aim to establish a lower bound on the probability of reaching a goal region while avoiding undesired configurations, and are the basis of logical properties such as syntactically-cosafe Linear Temporal Logic (sc-LTL) [7], LTL over finite traces (LTLf) [8], and Probabilistic Computation Tree Logic (PCTL) [9].

The framework consists of two main steps: (i) construction of a finite abstraction of the stochastic system in the form of an uncertain Markov decision process (MDP) [10,11], also known as *robust MDP*, and (ii) optimal robust strategy synthesis on the robust MDP. For the abstract construction, we first discretize the state space, and then use the nominal noise distribution (center of the ambiguity set) to compute bounds on the transition probabilities between the discrete regions. This results in a nominal Interval MDP (IMDP) [12,13] that contains the discretization error but not the distributional uncertainty. We then lift the ambiguity set from the noise space to the space of transition probabilities of the IMDP, and embed it as uncertainty, resulting in a robust MDP. Hence, the obtained robust MDP accounts for the distributional ambiguity in the original system as well as the discretization error.

Next, we synthesize an optimal strategy on the robust MDP that is robust with respect to the uncertain transition probabilities of this model. To that end, we first introduce a value iteration method to maximize the robust (worst-case) probability of satisfying the reach-avoid specification on the robust MDP. We prove that, for the unbounded-time specifications, the value iteration converges to the maximum worst-case probability of satisfaction. We then leverage recent results from distributionally robust optimization [4] to reduce value iteration to the solution of a set of linear programs, improving efficiency. Further, for unbounded-time properties, we show that stationary strategies are sufficient to obtain the optimal worst-case probability, based on which we introduce an algorithm for the extraction of such an optimal and robust strategy. We also formally prove the correctness of our framework for the original switched stochastic system. Finally, we illustrate the framework on four case studies including both linear and non-linear systems and for various ambiguity sets.

In short, the contributions of this work is fivefold:

- a framework for formal reasoning for distributionally-uncertain switched stochastic systems,
- an abstraction procedure to construct a finite robust MDP for nonlinear switched stochastic systems with additive uncertain noise,
- a robust strategy synthesis algorithm for robust MDPs with unbounded-time reachability specifications,
- an efficient robust optimization method via a dual formulation of linear programs, and
- a set of illustrative case studies and benchmarks for empirical evaluation of the framework.

1.1. Previous work

A preliminary version of this work appeared in [14], which the current manuscript extends in several aspects. First, in this paper, we present a new algorithmic approach to perform value iteration, which substantially improves the one in [14] in terms of computational efficiency: an improvement of two orders of magnitude is observed empirically. Furthermore, work [14] only considers finite-time reachability properties, whereas in this paper, we extend our results to unbounded-time reachability properties and prove convergence of value iteration and optimality of stationary strategies in this setting. We also introduce a method of extracting such a strategy, which is often ignored in the literature. Finally, we extend the experimental evaluations with a system subject to distributional shifts and an example of a system under a complex specification given as a temporal logic formula. With the exception of Sections 6.1.2, 6.3 and the corresponding(Appendices B and C), all the theorems and definitions remain the same as in [14]. Section 6.2 has been revised to account for the case of disturbances with a bounded support. The new case studies are presented in Sections 7.3 and 7.4. We also benchmark the new algorithm in Section 6.3 against the one in [14] to illustrate its computational superiority.

1.2. Related work

Formal verification and synthesis algorithms have been studied for switched stochastic systems either employing stochastic barrier functions [15] or abstractions in the form of *finite* Markov models [16–20]. Within the latter, many works (e.g., [3,13,21–23]) have employed *interval Markov decision processes* (IMDPs) as abstract model of choice, a class of Markov decision processes in which the transition probabilities belong to intervals [12,24] and that admits efficient control synthesis algorithms [3,13]. In the literature it is a common assumption that both the dynamics and noise distribution of the system are assumed to be exactly known. This in practice is an unrealistic assumption and often violated in the presence of, e.g., unmodeled dynamics, distributional shifts, or data-driven components.

More recently, researchers have proposed the use of machine learning algorithms, including neural networks and Gaussian processes, to synthesize formal control strategies when the dynamics are (partially) unknown or too complex to be modeled [25–27]. Yet, these works focus on the unknown dynamics and do not address the case of inexact knowledge of the probability distribution of stochastic variables in the dynamics. Another line of research tackles the opposite setting, in which the dynamics are known exactly but the probabilistic behavior of the system must be inferred from samples of the random terms [28,29]. Our setting is related to the latter in the sense that we consider ambiguity regarding the distribution of the random terms. However, our approach is general

with respect to the source of ambiguity, which encompasses the data-driven setting and those of distributionally robust control and robustness against distributional shifts.

Ambiguity sets are typically employed in distributionally robust optimization (DRO) to describe sets of probability distributions with respect to which one aims to make robust (optimal) decisions [30]. A typical way of defining an ambiguity set is as a set of probability distributions that are close to some nominal one. Depending on the employed measure of closeness, ambiguity sets can be constructed based on: moment constraints [31,32], statistical divergences [33], or optimal transport discrepancies [6,34,35] like the Wasserstein distance. Wasserstein ambiguity sets, as considered in this paper, constitute a particularly convenient description for ambiguous distributions in data-driven problems. The Wasserstein metric penalizes horizontal dislocations between distributions [36], providing a way to describe ambiguity sets that have finite-sample guarantees of containing the true distribution [37]. Furthermore, Wasserstein ambiguity sets enable the formulation of tractable DRO problems [5]. In [38] dynamic aspects of distributional uncertainty under optimal transport ambiguity are studied. In particular, [38] studies the evolution of Wasserstein ambiguity sets for systems with unknown state disturbance distribution. Also, [39] developed a risk-aware robot control scheme avoiding dynamic obstacles whose location is dictated by an ambiguous probability distribution.

A class of Markov processes that is closely related to robust MDPs are distributionally robust Markov decision processes (DR-MDPs) [40–42]. These are MDPs with transition probabilities dependent on uncertain parameters that lie in some ambiguity set. DR-MDPs and robust MDPs are substantially different as the latter do not consider any additional probabilistic structure over the ambiguous distributions to signify which uncertainty model is more likely to occur. Planning algorithms for complex specifications and diverse classes of robust Markov models have been already considered in the literature [10,13,19,43]. However, none of them addresses the main contribution of this work, which is how to combine these algorithms with optimal transport techniques to formally abstract and synthesize strategies for continuous-space dynamical systems with uncertain distribution of the noise. Moreover, the robust MDPs over which we synthesize strategies in this paper generalize the ones considered in [19,43]. These works assume that the transition probabilities of their robust MDPs vanish at the exact same states for all transition matrices. Here, we alleviate this restriction, which provides us substantial modeling flexibility as we can handle for instance data-driven ambiguity sets whose reference distribution, obtained from samples, is bounded even when the data-generating distribution may not be. This leads to the final key contribution of our paper, which is the rigorous synthesis of optimal strategies for infinite horizon reachability specifications for a broader class of robust MDPs. In addition, unlike [19,43], our approach does not require identifying and eliminating end components as a pre-processing step, which suffers from exponential complexity for general robust MDPs [44].

2. Basic notation

The set of non-negative integers is denoted by $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. Given a set A, $|A|$ denotes its cardinality. Given $\ell, m \in \mathbb{N}_0$ with $\ell \leq m$, we employ $[\ell : m]$ for the set $\{\ell, \ell+1, \dots, m\}$. We also use this notation when $m = \infty$ to denote the set $\{\ell, \ell+1, \dots\}$. For a separable metric space X, $\mathcal{B}(X)$ denotes its Borel σ-algebra and $\mathcal{D}(X)$ the set of probability distributions on $(X, \mathcal{B}(X))$. When X is discrete and $\gamma \in \mathcal{D}(X)$, $\gamma(x) := \gamma(\{x\})$ is the probability of the event described by the singleton $\{x\}$. Let $c : X \times X \to \mathbb{R}_{\geq 0}$ be a continuous cost function defined over the product space $X \times X$. The optimal transport discrepancy between two probability distributions $p, p' \in \mathcal{D}(X)$ is defined as

$$\mathcal{T}_c(p, p') := \inf_{\pi \in \Pi(p,p')} \int_{X \times X} c(x, y) d\pi(x, y), \tag{2.1}$$

where $\Pi(p, p')$ is the set of all transport plans between p and p', a.k.a. couplings, i.e., probability distributions $\pi \in \mathcal{D}(X \times X)$ with marginals p and p', respectively. Since cost c is non-negative, $\mathcal{T}_c$ provides a discrepancy measure between distributions in $\mathcal{D}(X)$. By continuity of c, there always exists a transport plan π for which the infimum in (2.1) is attained [45, Theorem 1.3].

Assume that X is equipped with a metric d. Given $s \geq 1$, we denote by $\mathcal{D}_s(X)$ the set of probability distributions on X with finite sth moment, i.e.,

$$\mathcal{D}_s(X) = \left\{ p \in \mathcal{D}(X) : \int_X d(x, y)^s dp(x) < \infty \text{ for some } y \in X \right\}.$$

The discrepancy $\mathcal{W}_s := (\mathcal{T}_{d^s})^{\frac{1}{s}}$ is then also a metric in the space $\mathcal{D}_s(X)$ coined as the s-Wasserstein distance [45]. Additionally, we use bold symbols, e.g., $\mathbf{v}$, to denote random variables and non-bold symbols to denote the respective realizations, i.e., v.

3. Problem formulation

Consider a discrete-time switched stochastic system given by:

$$\mathbf{x}_{k+1} = f_{u_k}(\mathbf{x}_k) + \mathbf{v}_k, \tag{3.1}$$

where $k \in \mathbb{N}$, $\mathbf{x}_k$ takes values in $\mathbb{R}^n$, $u_k \in U$, and $U = \{1, \dots, m\}$ is a finite set of *modes* or *actions*. For every $u \in U$, $f_u : \mathbb{R}^n \to \mathbb{R}^n$ is a possibly non-linear continuous function. The noise term $\mathbf{v}_k$ is an independent random variable that takes values in $\mathbb{R}^n$ with a distribution p_v^{true} that is identically distributed at each time step. We assume the exact noise distribution is unknown, but it is ε-close to a given (nominal) distribution as detailed below.

Assumption 3.1. The distribution p_v^{true} is ε-close (in the s-Wasserstein sense) to a known distribution $\widehat{p}_v \in \mathcal{D}_s(\mathbb{R}^n)$, which we call *nominal*, i.e., $p_v^{\text{true}} \in \mathcal{P}_v := \{p \in \mathcal{D}_s(\mathbb{R}^n) : \mathcal{W}_s(p, \widehat{p}_v) \leq \varepsilon\}$, where $\mathcal{W}_s$ is determined by the metric $d(x, y) = \|x - y\|$, where $\|\cdot\|$ is the Euclidean norm.

Intuitively, $\mathbf{x}_k$ is a stochastic process driven by the additive noise $\mathbf{v}_k$, whose distribution is uncertain but close to a nominal known distribution in the s-Wasserstein metric. Consequently, System (3.1) represents a large class of controlled stochastic systems with additive and uncertain noise. Such systems arise, for instance, in data-driven settings, where measure concentration results [37] can be employed to build a Wasserstein ambiguity set from data of $\mathbf{v}_k$ with high confidence [5], or in settings in which one wants to synthesize strategies able to cope with distributional shifts of the noise.

Let $\omega_x = x_0 \xrightarrow{u_0} x_1 \xrightarrow{u_1} \dots$ be a path (trajectory realization) of System (3.1) and denote by $\omega_x(k) = x_k$ the visited state at time k in the sequence ω_x. Further, we denote by ω_x^k the prefix of finite length $k+1$ of ω_x, and by $\text{last}(\omega_x^k)$ the last state of the finite path ω_x^k, i.e., x_k. We also denote by Ω_x^∞ the set of paths of infinite length and by Ω_x^{fin} the set of all paths with finite length, i.e, the set of prefixes $\omega_x^k = x_0 \xrightarrow{u_0} x_1 \xrightarrow{u_1} \dots \xrightarrow{u_{k-1}} x_k$ for all $k \in \mathbb{N}$. Given a finite path, a switching strategy selects an action, corresponding to a mode of System (3.1).

Definition 3.2 (*Switching Strategy*)**.** A *switching strategy* $\sigma_x : \Omega_x^{\text{fin}} \to U$ is a function that maps each finite path $\omega_x^k \in \Omega_x^{\text{fin}}$ to an action $u \in U$. We say that a strategy is *memoryless (Markovian)* if it only depends on $\text{last}(\omega_x^k)$ and k. Furthermore, if a memoryless strategy does not depend on k, we call it *stationary*.

For any $p_v \in \mathcal{P}_v$, $u \in U$, $X \in \mathcal{B}(\mathbb{R}^n)$, and $x \in \mathbb{R}^n$, let

$$T_{p_v}^u(X \mid x) = \int \mathbf{1}_X(f_u(x) + \bar{v}) p_v(\bar{v}) d\bar{v} \tag{3.2}$$

be the stochastic transition kernel induced by system (3.1) with noise fixed to p_v in mode $u \in U$, where $\mathbf{1}_X$ is the indicator function with $\mathbf{1}_X(x) = 1$ if $x \in X$, and $\mathbf{1}_X(x) = 0$ otherwise. From the definition of $T_{p_v}^u(X \mid x)$ it follows that, given a strategy σ_x, a noise distribution p_v, and an initial condition x_0, System (3.1) defines a stochastic process on the canonical space $(\mathbb{R}^n)^{\mathbb{N}}$ with the Borel sigma-algebra $\mathcal{B}((\mathbb{R}^n)^{\mathbb{N}})$ [46]. In particular, by the Ionescu-Tulcea theorem, the kernel $T_{p_v}^u$ generates a unique probability distribution $P_{p_v}^{x_0,\sigma_x}$ over the paths of system (3.1) originating from x_0, such that for each $k \in \mathbb{N}$

$$P_{p_v}^{x_0,\sigma_x}[\omega_x(0) \in X] = \mathbf{1}_X(x_0),$$
$$P_{p_v}^{x_0,\sigma_x}[\omega_x(k) \in X \mid \omega_x^{k-1}] = T_{p_v}^{\sigma_x(\omega_x^{k-1})}(X \mid \omega_x(k-1)).$$

In this paper we consider both finite-time and infinite-time probabilistic reach-avoid specifications for System (3.1).

Definition 3.3 (*Reach-avoid probability*)**.** For a time horizon $K \in \mathbb{N}_0 \cup \{\infty\}$, a bounded safe set X, a target region $X_{\text{tgt}} \subset X$ and an initial state $x_0 \in X$, the reach-avoid probability $P_{\text{reach}}(X, X_{\text{tgt}}, K \mid x_0, \sigma_x, p_v)$ is defined as:

$$P_{\text{reach}}(X, X_{\text{tgt}}, K \mid x_0, \sigma_x, p_v) :=$$
$$P_{p_v}^{x_0,\sigma_x}\left[\exists k \in [0 : K] \mid \omega_{\mathbf{x}}(k) \in X_{\text{tgt}} \ \wedge \ \forall k' < k, \ \omega_{\mathbf{x}}(k') \in X\right]. \tag{3.3}$$

We can now precisely formulate our problem, which is to synthesize control strategies that are robust to all distributions in the set $\mathcal{P}_v$.

Problem 3.4 (*Switching Strategy Synthesis*)**.** Consider the switched stochastic system (3.1), its corresponding ambiguity set $\mathcal{P}_v$, a bounded safe set X, and a target region $X_{\text{tgt}} \subset X$. Given an initial state $x_0 \in X$, a probability threshold $p_{\text{th}} \in [0, 1]$, and a horizon $K \in \mathbb{N}_0 \cup \{\infty\}$, synthesize a switching strategy σ_x such that, for all $p_v \in \mathcal{P}_v$,

$$P_{\text{reach}}(X, X_{\text{tgt}}, K \mid x_0, \sigma_x, p_v) \geq p_{\text{th}}. \tag{3.4}$$

Remark 3.5. Note that, while our focus on reach-avoid specifications in Problem 3.4, the proposed framework can be easily used for more complex logical specifications, such as scLTL, LTLf, and PCTL specification, which reduce to reachability problems. We illustrate this in an example in Section 7.4, where we synthesize a strategy for an LTLf specification.

3.1. Overview of the approach

To approach Problem 3.4, we construct a finite-state abstraction of System (3.1) in terms of a robust MDP as detailed in Section 5. In Section 6, we show how to synthesize an optimal strategy on the resulting abstraction. In particular, we prove that, while in the case of finite-time reachability specifications the optimal strategy is Markovian, the optimal strategy for an unbounded-time (infinite horizon) reachability property is stationary. We prove that in both cases an optimal strategy can be synthesized by solving a finite set of linear programs, thus guaranteeing efficiency. We then refine the resulting strategy into a switching strategy for System (3.1) and show how the robust MDP abstraction provides upper and lower bounds on the probability that System (3.1) satisfies the specification under the refined strategy.

4. Preliminaries

4.1. Robust Markov decision processes

Robust MDPs are a generalization of Markov decision processes in which the transition probability distributions between states are constrained to belong to an ambiguity set [10,47].

Definition 4.1 (*Robust MDP*)**.** A *robust Markov decision process* (robust MDP) is a tuple $\mathcal{M} = (Q, A, \Gamma)$, where

- Q is a finite set of states,
- A is a finite set of actions, and $A(q)$ denotes the set of available actions at state $q \in Q$,
- $\Gamma = \{\Gamma_{q,a}\}_{q\in Q, a\in A}$ is the set of possible transition probability distributions of $\mathcal{M}$, where $\Gamma_{q,a} \subseteq \mathcal{D}(Q)$ is the set of transition probability distributions for state–action pair $(q, a) \in Q \times A(q)$.[1]

A path of a robust MDP is a sequence of states $\omega = q_0 \xrightarrow{a_0} q_1 \xrightarrow{a_1} q_2 \xrightarrow{a_2} \dots$ such that $a_k \in A(q_k)$ and there exists $\gamma \in \Gamma_{q_k,a_k}$ with $\gamma(q_{k+1}) > 0$ for all $k \in \mathbb{N}$. We denote the ith state of a path ω by $\omega(i)$, a finite path of length $k+1$ by ω^k and the last state of a finite path ω^{fin} by $\text{last}(\omega^{\text{fin}})$. The set of all paths of finite and infinite length are denoted by Ω^{fin} and Ω, respectively.

Definition 4.2 (*IMDP*)**.** An *interval MDP* (IMDP) [3,21], also known as bounded-parameter MDP (BMDP) [12,24], is a class of robust MDP $\mathcal{I} = (Q, A, \Gamma)$ where Γ has the form

$$\Gamma_{q,a} = \{\gamma \in \mathcal{D}(Q) : \underline{P}(q, a, q') \le \gamma(q') \le \overline{P}(q, a, q') \text{ for all } q' \in Q\}, \tag{4.1}$$

for every $q \in Q$, $a \in A(q)$. The bounds $\underline{P}, \overline{P}$ are called transition probability bounds, for which it must hold that, for every state $q \in Q$ and action $a \in A(q)$, $0 \le \underline{P}(q, a, q') \le \overline{P}(q, a, q') \le 1$ and $\sum_{q'\in Q} \underline{P}(q, a, q') \le 1 \le \sum_{q'\in Q} \overline{P}(q, a, q')$.

The actions of robust MDPs and IMDPs are chosen according to a strategy σ which is defined below.

Definition 4.3 (*Strategy*)**.** A strategy σ of a robust MDP model $\mathcal{M}$ is a function $\sigma : \Omega^{\text{fin}} \to A$ that maps a finite path ω^k, with $k \in \mathbb{N}$, of $\mathcal{M}$ onto an action in $A(\text{last}(\omega^{\text{fin}}))$. If a strategy depends only on $\text{last}(\omega^{\text{fin}})$ and k, it is called a memoryless (Markovian) strategy, and if it only depends on $\text{last}(\omega^{\text{fin}})$, it is called stationary. The set of all strategies is denoted by Σ, and the set of all stationary strategies by Σ_s.

Given an arbitrary strategy σ, we are restricted to the set of robust Markov chains defined by the set of transition probability distributions induced by σ. To reduce this to a Markov chain, we define the adversary [12] (called "nature" in [10]), to be the function that assigns a transition probability distribution to each state–action pair.

Definition 4.4 (*Adversary*)**.** For a robust MDP $\mathcal{M}$, a (Markovian) adversary is a function $\xi : Q \times A \times \mathbb{N}_0 \to \mathcal{D}(Q)$ that, for each state $q \in Q$, action $a \in A(q)$ and time instant $k \in \mathbb{N}_0$, assigns an admissible distribution $\gamma_{q,a} \in \Gamma_{q,a}$. The set of all Markovian adversaries is denoted by Ξ_M.

For an initial condition $q_0 \in Q$, under a strategy and a valid adversary $\xi \in \Xi_M$, the robust MDP collapses to a Markov chain with a unique probability measure defined over its paths. With a small abuse of notation, we denote this measure by $P_\xi^{q_0,\sigma}$.

5. Robust MDP abstraction

To approach Problem 3.4, we start by abstracting System (3.1) into the IMDP $\hat{\mathcal{I}} = (Q, A, \hat{\Gamma})$ with the noise distribution fixed to the nominal one, $\hat{p}_v$. In this way, we embed the error caused by the state discretization into $\hat{\mathcal{I}}$. After that, we expand the set of transition probabilities $\hat{\Gamma}$ of $\hat{\mathcal{I}}$ to also capture the distributional ambiguity into the abstraction, obtaining the robust MDP $\mathcal{M} = (Q, A, \Gamma)$. Note that the sets of states Q and actions A are the same in $\hat{\mathcal{I}}$ and $\mathcal{M}$. Below, we describe how we obtain Q and A, and in Section 5.2 we consider the set of transition probability distributions Γ.

5.1. States and actions

The state space Q of $\mathcal{M}$ is constructed as follows: consider a set of non-overlapping regions $Q_{\text{safe}} = \{q_1, q_2, \dots, q_{|Q_{\text{safe}}|}\}$ partitioning the set X so that either $q \cap X_{\text{tgt}} = \emptyset$ or $q \cap (X \setminus X_{\text{tgt}}) = \emptyset$ for all $q \in Q_{\text{safe}}$. We denote by Q_{tgt} the subset of Q_{safe} for which $q \cap X_{\text{tgt}} = q$ and assume that it is a partition of Q_{tgt}. The states of the abstraction comprise of Q_{safe} and the unsafe region $q_u := \mathbb{R}^n \setminus X$, hence, $Q := Q_{\text{safe}} \cup \{q_u\}$. We index Q by $\mathcal{N} = \{1, \dots, N\}$, where $N := |Q|$ and denote the actions of the abstraction as $A := U$.

[1] Note that the sets of transition probability distributions of the robust MDP are independent for each state and action. This is known as *rectangular* property of the set of transition probability distributions [10,47]. Furthermore, unlike [19,43], this is the only assumption we impose regarding the structure of the robust MDP.

5.2. Transition probability distributions

5.2.1. Accounting for the discretization error

To capture the state discretization error in the abstraction, we first consider an IMDP abstraction of System (3.1) for a fixed distribution, namely, the nominal probability distribution $\widehat{p}_v$. Since this IMDP is constructed for the nominal distribution $\widehat{p}_v$, we call it "nominal" IMDP and use the notation $\widehat{\mathcal{I}} = (Q, A, \widehat{\Gamma})$. We note that building an IMDP abstraction of a stochastic system with disturbances of a known distribution is widely studied in the literature [3,13,25,48], and we report the full procedure here below for completeness. Construction of the state Q and action A spaces of $\widehat{\mathcal{I}}$ are described in Section 5.1. We now describe the set $\widehat{\Gamma}$ of $\widehat{\mathcal{I}}$. According to Definition 4.2, the set $\widehat{\Gamma}$ is defined by the transition probability bounds $\underline{P}$ and $\overline{P}$ of $\widehat{\mathcal{I}}$. To formally account for the discretization error, the bounds must satisfy for all $q \in Q_{\text{safe}}$, $q' \in Q$ and $a \in A = U$:

$$\underline{P}(q, a, q') \leq \min_{x \in q} T^a_{\widehat{p}_v}(q' \mid x) \quad \text{and} \quad \overline{P}(q, a, q') \geq \max_{x \in q} T^a_{\widehat{p}_v}(q' \mid x). \tag{5.1}$$

Since we are interested in the paths of system (3.1) that do not exit set X, we make state q_u absorbing, i.e.,

$$\underline{P}(q_u, a, q_u) = \overline{P}(q_u, a, q_u) = 1 \tag{5.2}$$

for all $a \in A$. In this way, we include the "avoid" part of the specification into the definition of the abstraction: the paths that reach q_u, will remain there forever and, therefore, will never reach the target set Q_{tgt}. Consequently, for each $q \in Q$ and $a \in A$, we obtain

$$\widehat{\Gamma}_{q,a} = \{\gamma \in \mathcal{D}(Q) : \underline{P}(q, a, q') \leq \gamma(q') \leq \overline{P}(q, a, q') \text{ for all } q' \in Q\}. \tag{5.3}$$

5.2.2. Accounting for the distributional uncertainty

Now, we expand the sets $\{\widehat{\Gamma}_{q,a}\}_{q \in Q, a \in A}$ of transition probabilities of $\widehat{\mathcal{I}}$ to also embed the distributional uncertainty into the abstraction. With this objective, we first define the cost

$$c(q, q') := \inf\{\|x - y\|^s : x \in q, y \in q'\}, \tag{5.4}$$

between any two states $q, q' \in Q$, where $\|\cdot\|$ and s are the same as for $\mathcal{W}_s$ in Assumption 3.1. The cost $c(q, q')^{\frac{1}{s}}$ is the minimum distance, in the sense of norm $\|\cdot\|$, between any pair of points in the regions q and q', respectively. Using this cost and the exponent s in $\mathcal{W}_s$, we define the optimal transport discrepancy $\mathcal{T}_c$ between distributions over Q as in (2.1).[2] Given a probability distribution $\gamma \in \mathcal{D}(Q)$ and $\epsilon \geq 0$, we denote by $\mathcal{T}_c^{\epsilon}(\gamma)$ the set of all distributions to which mass can be transported from γ incurring a c-transport cost lower than ϵ. Using the previous elements, we are finally able to define Γ.

Definition 5.1. The discrete uncertainty set Γ is defined for every state $q \in Q_{\text{safe}}$ and action $a \in A$ as

$$\Gamma_{q,a} := \bigcup_{\gamma \in \widehat{\Gamma}_{q,a}} \mathcal{T}_c^{\epsilon}(\gamma), \tag{5.5}$$

with $\epsilon = \varepsilon^s$. For state q_u and action $a \in A$, let $\Gamma_{q,a} := \widehat{\Gamma}_{q,a}$ (preserving the absorbing property of the unsafe state).

Each $\Gamma_{q,a}$ is the set of probability distributions over Q that are ϵ-close to $\widehat{\Gamma}_{q,a}$, in the sense of the optimal transport discrepancy $\mathcal{T}_c$. Note that the way in which we have defined Γ makes it possible that there exist two probability distributions $\gamma, \gamma' \in \Gamma_{q,a}$, for some $q \in Q$, $a \in A$, which have different support. Therefore, the robust MDPs we consider belong to a more general class than the ones in [19,43].

Once we have obtained the sets of transition probabilities Γ, along with the state Q and action A spaces, our robust MDP abstraction $\mathcal{M}$ is fully defined. The following proposition ensures that the abstraction captures all possible transition probabilities of System (3.1) to regions in the partition.

Proposition 5.2 (*Consistency of the Robust MDP Abstraction*). *Consider the robust MDP abstraction $\mathcal{M} = (Q, A, \Gamma)$ of System (3.1). Let $q \in Q_{\text{safe}}$, $a \in A$, $x \in q$, and $p_v \in \mathcal{P}_v$, and define $\gamma_{x,a} \in \mathcal{D}(Q)$ as*

$$\gamma_{x,a}(q') := T^a_{p_v}(q' | x)$$

for all $q' \in Q$. Then $\gamma_{x,a} \in \Gamma_{q,a}$.

The proof of Proposition 5.2 is given in Appendix A. The intuition behind Proposition 5.2 is that set $\Gamma_{q,a}$ contains the transition probabilities $\gamma_{x,a}$ obtained by starting from any $x \in q$, with $q \in Q_{\text{safe}}$ under $a \in A$ and for every $p_v \in \mathcal{P}_v$.

[2] Notice that, since c is also not a metric, the resulting optimal transport discrepancy $\mathcal{T}_c$ is not a distance.

Remark 5.3 (*Model Choice for the Abstraction*)**.** An alternative way to include the distributional ambiguity into the abstraction is to use an IMDP abstraction $\mathcal{I} = (Q, A, \Gamma^{\text{IMDP}})$, which has the same state Q and action A spaces as $\mathcal{M}$, and in which Γ^{IMDP} is defined by transition probability bounds that satisfy

$$\begin{aligned} \underline{P}^{\text{IMDP}}(q,a,q') &\leq \min_{p_v \in \mathcal{P}_v} \min_{x \in q} T^a_{p_v}(q' \mid x), \\ \overline{P}^{\text{IMDP}}(q,a,q') &\geq \max_{p_v \in \mathcal{P}_v} \max_{x \in q} T^a_{p_v}(q' \mid x) \end{aligned} \tag{5.6}$$

for all $q \in Q_{\text{safe}}$, $q' \in Q$ $a \in A$, and for which q_u is again absorbing. Therefore, for fixed $q \in Q$ and $a \in A$, set $\Gamma^{\text{IMDP}}_{q,a}$ of $\mathcal{I}$ is defined as

$$\Gamma^{\text{IMDP}}_{q,a} = \{\gamma \in \mathcal{D}(Q) : \underline{P}^{\text{IMDP}}(q,a,q') \leq \gamma(q') \leq \overline{P}^{\text{IMDP}}(q,a,q') \text{ for all } q' \in Q\}. \tag{5.7}$$

This choice of the abstraction model allows to use efficient synthesis algorithms for IMDPs [12,24] to solve Problem 3.4. By the definition of the transition probability bounds of $\mathcal{I}$ in (5.6), Γ^{IMDP} satisfies Proposition 5.2, effectively capturing all possible transition probabilities of System (3.1). However, IMDP $\mathcal{I}$ describes the uncertainty in a very loose way, i.e., set Γ^{IMDP} of $\mathcal{I}$ can be excessively large for many ambiguity sets $\mathcal{P}_v$. Intuitively, this is caused by $\Gamma^{\text{IMDP}}_{q,a}$ being only defined through decoupled interval constraints for every successor state $q' \in Q$, This is likely to result in more conservative solutions to the reach-avoid problem as we show in Section 7.

6. Robust strategy synthesis

Our goal is to synthesize a switching strategy $\sigma^*_{\mathbf{x}}$ for System (3.1) that maximizes (3.4). To capture the distributional uncertainty and the effect of quantization, we consider the proposed robust MDP abstraction $\mathcal{M}$. We synthesize the robustly maximizing strategy σ^* for the abstraction $\mathcal{M}$, which we then refine on the original system retaining formal guarantees of correctness. In Sections 6.1, 6.2, and 6.3, we show how an optimal strategy σ^* for the abstraction $\mathcal{M}$ can be efficiently computed via linear programming. Then, in Section 6.4, we prove the correctness of our strategy synthesis approach.

6.1. Robust dynamic programming

Recall that, in our robust MDP abstraction $\mathcal{M} = (Q, A, \Gamma)$, the uncertainties of $\mathcal{M}$ are characterized by an adversary ξ, which at each time step and given a path of $\mathcal{M}$ and an action, selects a feasible distribution from Γ (see Definition 4.4). Therefore, to be robust against all uncertainties, as common in the literature [10,11], we aim to synthesize a strategy σ^* such that, given a horizon $K \in \mathbb{N}_0 \cup \{\infty\}$,

$$\sigma^* \in \arg\max_{\sigma \in \Sigma} \inf_{\xi \in \Xi_M} P_{\text{reach}}(Q_{\text{safe}}, Q_{\text{tgt}}, K \mid q, \sigma, \xi), \tag{6.1}$$

for all $q \in Q$, where $P_{\text{reach}}(Q_{\text{safe}}, Q_{\text{tgt}}, K \mid q, \sigma, \xi)$ is defined as in (3.3) for System (3.1). We call σ^* the *optimal robust* or simply *optimal* strategy.

We denote by $\underline{p}^K$ and $\overline{p}^K$, respectively, the worst and best-case probabilities of the paths of $\mathcal{M}$ satisfying the reach-avoid specification under optimal strategy σ^*, i.e.,

$$\underline{p}^K(q) := \inf_{\xi \in \Xi_M} P_{\text{reach}}(Q_{\text{safe}}, Q_{\text{tgt}}, K \mid q, \sigma^*, \xi), \tag{6.2a}$$

$$\overline{p}^K(q) := \sup_{\xi \in \Xi_M} P_{\text{reach}}(Q_{\text{safe}}, Q_{\text{tgt}}, K \mid q, \sigma^*, \xi) \tag{6.2b}$$

for all $q \in Q$. In the following subsections, we show how to compute the previous quantities via robust dynamic programming. We distinguish the cases of finite ($K < \infty$) and infinite ($K = \infty$) horizon, since they require different treatment.

6.1.1. Finite horizon

The following proposition shows that the bounds in (6.2a) and (6.2b) and the optimal strategy in (6.1) can be obtained via dynamic programming.

Proposition 6.1. *Let $\underline{p}^K$ be as defined in (6.2a) and $k \in [0 : K-1]$. Then, it holds that*

$$\underline{p}^{k+1}(q) = \begin{cases} 1 & \text{if } q \in Q_{\text{tgt}} \\ \max_{a \in A} \min_{\gamma \in \Gamma_{q,a}} \sum_{q' \in Q} \gamma(q') \underline{p}^k(q') & \text{otherwise,} \end{cases} \tag{6.3}$$

with initial condition $\underline{p}^0(q) = 1$ for all $q \in Q_{\text{tgt}}$ and 0 otherwise. Furthermore, for each path ω^k with $k \in [0 : K-1]$, it holds that

$$\sigma^*(\omega^k) \in \arg\max_{a \in A} \left\{ \min_{\gamma \in \Gamma_{\text{last}(\omega^k),a}} \sum_{q' \in Q} \gamma(q') \underline{p}^{K-k-1}(q') \right\}. \tag{6.4}$$

Proof. The proof follows directly by applying [49, Theorem 1] to the setting of reachability. □

A consequence of Proposition 6.1 is that, in the finite horizon ($K < \infty$) case, there exists an optimal policy that is Markovian (note that the resulting strategy is time-dependent in general). Furthermore, once an optimal strategy σ^* is fixed, the upper bound $\overline{p}^K$ of P_{reach} in (6.2b) can be readily computed via the robust dynamic programming iterations

$$\overline{p}^{k+1}(q) = \begin{cases} 1 & \text{if } q \in Q_{\text{tgt}} \\ \max_{\gamma \in \Gamma_{q,\sigma^*}} \sum_{q' \in Q} \gamma(q') \overline{p}^k(q') & \text{otherwise,} \end{cases} \tag{6.5}$$

for all $k \in [0 : K-1]$, which is analogous to that in (6.3) and has initial condition $\overline{p}^0(q) = 1$ for all $q \in Q_{\text{tgt}}$ and $\overline{p}^0(q) = 0$ otherwise.

In Section 6.2 we show that the robust dynamic iterations in (6.3) and (6.5) boil down to solving a finite number of linear programs, but first, in the next subsection, we consider the infinite horizon case.

6.1.2. Infinite horizon

In the case of an infinite time horizon ($K = \infty$), we seek to bound the probability of the paths of $\mathcal{M}$ that with arbitrary lengths reach Q_{tgt} while remaining in Q_{safe}. To also synthesize an optimal strategy, we first show that the robust dynamic programming iterations in (6.3) converge to a fixed point. Then, using this result, we show that for the infinite horizon case, there exists an optimal stationary strategy, and we propose an approach to obtain it. All the proofs of the theorems in this section can be found in Appendix B.

We first show that the lower bound $\underline{p}^\infty$ on the infinite-horizon reachability probability obtained by (6.2a) for $K = \infty$ is a fixed-point of the robust dynamic programming operator in (6.3).

Theorem 6.2. *Let $\{\underline{p}^k\}_{k \in \mathbb{N}_0}$ be the infinite sequence defined recursively by (6.3) with initial condition $\underline{p}^0(q) = 1$ for all $q \in Q_{\text{tgt}}$ and 0 otherwise. Then $\{\underline{p}^k\}$ converges to $\underline{p}^\infty$, which is the least fixed-point of (6.3) on $\mathbb{R}^{|Q|}_{\geq 0}$.*

Theorem 6.2 guarantees that, to obtain the optimal reachability probabilities, we can simply run the robust dynamic programming iteration in (6.3) until convergence. We next address the problem of finding an optimal strategy. In particular, we are interested in determining a stationary optimal strategy that has the convenient feature of using the same decision rule at every time step. The following result establishes that stationary optimal strategies indeed exist.

Proposition 6.3. *The infinite horizon reachability problem (6.1) admits a stationary strategy $\sigma' \in \Sigma_s$.*

Since the mere existence of an optimal strategy is not sufficient for synthesis purposes, we seek conditions under which a stationary strategy that we will later explicitly construct is also optimal. The following result provides such optimality conditions, which generalize the corresponding requirements in the case of standard MDPs (see [50, Theorem 7.1.7]).

Proposition 6.4. *Consider the set $Q_{\text{reach}} := \{q \in Q : \underline{p}^\infty(q) > 0\}$ of states that have positive probability of reaching Q_{tgt} for some strategy, and let $\sigma^* \in \Sigma_s$ be a stationary strategy that, for each state $q \in Q$, satisfies*

$$\sigma^*(q) \in A^*(q) := \arg\max_{a \in A} \left\{ \min_{\gamma \in \Gamma_{q,a}} \sum_{q' \in Q} \gamma(q') \underline{p}^\infty(q') \right\} \tag{6.6a}$$

$$\lim_{k \to \infty} P^{q,\sigma^*}_{\xi}[\omega(k) \in Q_{\text{reach}}] = 0 \quad \text{for all } \xi \in \Xi_M. \tag{6.6b}$$

Then σ^* is optimal, i.e., it satisfies (6.1). Furthermore, condition (6.6a) is also necessary for every $\sigma \in \Sigma_s$ to be optimal.

The first condition of Proposition 6.4 imposes the requirement that σ^* can only pick actions that attain the maximum in the dynamic programming recursion in (6.3). The second requirement of the proposition is that under strategy σ^*, all paths should eventually exit Q_{reach} and remain outside that set forever, for every choice of the adversary and all initial conditions. A strategy that satisfies both these conditions is called *proper optimal*. We stress that, as we prove in Theorem 6.6, a proper optimal strategy always exists.

Remark 6.5. Restricting ourselves to stationary strategies comes at the price of enforcing σ^* to be proper optimal. In particular, the second condition required by a strategy σ^* to be optimal guarantees that, under σ^*, all states need to eventually exit Q_{reach} and remain outside forever. This additional condition is needed because there may be multiple strategies maximizing (6.6a) and some of them may introduce loops (cycles) with probability 1 in the graph of the resulting Markov chain. As a consequence, states that have nonzero probability of reaching Q_{tgt} for an optimal strategy may never achieve this under a strategy that only satisfies (6.6a).

We next extract an optimal proper stationary strategy by generalizing the approach employed in the proof of [51, Theorem 10.102] for MDPs. To this end, we recursively define the sets

$$Q^m := Q^{m-1} \cup \left\{ q \in Q \setminus Q^{m-1} : \exists\, a \in A^*(q) \text{ s.t. } \min_{\gamma \in \Gamma_{q,a}} \sum_{q' \in Q^{m-1}} \gamma(q') > 0 \right\}, \tag{6.7}$$

for each $m \in \{0, 1, \dots, m_{\max}\}$, where $Q^0 := Q_{\text{tgt}}$ and

$$m_{\max} := \min\{m \in \mathbb{N}_0 : Q^m = Q^{m-1}\}.$$

Each Q^m is the set of backward reachable states from Q_{tgt} in m steps by taking actions in A^*. Here reachability is interpreted in the worst-case sense where the underlying graph of the robust MDP depends on the worst-case choice of the adversary. Since Q is finite, the maximum number of backward steps to reach a state in Q from Q_{tgt} is bounded. Thus, $m_{\max}$ is always well-defined. The following result delineates how we can use the sets Q^m to extract a proper and optimal stationary strategy.

Theorem 6.6. *Let $A^*(q)$ be as given in (6.6a). Define a strategy $\sigma^* \in \Sigma_s$ with*

$$\sigma^*(q) \in \left\{ a \in A^*(q) : \min_{\gamma \in \Gamma_{q,a}} \sum_{q' \in Q^{m-1}} \gamma(q') > 0 \right\}$$

for all $q \in Q^m \setminus Q^{m-1}$ and $m \in \{1, \dots, m_{\max}\}$, and with arbitrary actions when $q \in Q_{\text{tgt}} \cup (Q \setminus Q_{\text{reach}})$. Then σ^* is proper and optimal.

Theorem 6.6 shows how to obtain an optimal stationary strategy via a backward reachability analysis. Note that, although this algorithm has to be employed once p^∞ has been obtained via robust dynamic programming, it converges within $|Q_{\text{reach}}| - |Q_{\text{tgt}}|$ iterations. Consequently, it is generally much faster than performing robust dynamic programming.

Having computed σ^*, we obtain the upper bound $\overline{p}^\infty$ in the reachability probability by iterating on (6.5) until we achieve convergence, to $\overline{p}^\infty$, starting from $\overline{p}^0(q) = 1$ for all $q \in Q_{\text{tgt}}$ and 0 otherwise. The proof is analogous to that of Theorem 6.2, and is omitted for simplicity.

6.2. Computation of robust dynamic programming via linear programming

We now show how for each state $q \in Q$ and time $k \in [0 : K-1]$, dynamic programming in (6.3) reduces to solving $|A|$ linear programs. In particular, Theorem 6.7 below guarantees that the inner problem in recursion (6.3) can be solved via linear programming. In what follows we explicitly consider $\underline{p}^k$, the upper bound $\overline{p}^k$ follows similarly. Before formally stating our result, we need to introduce some notations. Let $W \subseteq \mathbb{R}^n$ be a set containing the support of p_v^{true}, and for each $q \in Q$ and $a \in A$ denote

$$\overline{\mathcal{N}}^{q,a} := \{ i \in \mathcal{N} : \overline{P}(q, a, q_i) > 0 \}, \tag{6.8a}$$

$$\overline{\mathcal{N}}_W^{q,a} := \{ i \in \mathcal{N} : (f_a(q) + W) \cap q_i \neq \emptyset \}, \tag{6.8b}$$

where $\overline{P}(q, a, q_i)$ is the upper transition probability bound of the nominal IMDP as defined in (5.1) and (5.2). $\overline{\mathcal{N}}^{q,a}$ and $\overline{\mathcal{N}}_W^{q,a}$ represent respectively the set of indices of the discrete states that are reachable from q under action a for the nominal noise distribution, and for every distribution supported on W. Note that as the nominal distribution is also supported on W, it holds that $\overline{\mathcal{N}}^{q,a} \subseteq \overline{\mathcal{N}}_W^{q,a}$, and if $W = \mathbb{R}^n$, then $\overline{\mathcal{N}}_W^{q,a} = \mathcal{N}$. Intuitively, in Theorem 6.7, from each starting region q, we only consider transitions to states indexed by $\overline{\mathcal{N}}_W^{q,a}$ as those are the only states reachable by some distribution with support in W.

Theorem 6.7 (*Robust Dynamic Programming as a Linear Program*)**.** *Consider the robust dynamic programming (6.3) for the robust MDP $\mathcal{M} = (Q, A, \Gamma)$ and assume that the support of P_v^{true} is contained in the set $W \subseteq \mathbb{R}^n$. Then, for every $k \in [0 : K-1]$, $q \in Q$, and $a \in A$, the inner minimization problem in (6.3) is equivalent to the following linear program:*

$$\min_{\gamma_i, \widehat{\gamma}_j, \pi_{ij}} \sum_{i \in \overline{\mathcal{N}}_W^{q,a}} \gamma_i \underline{p}^k(q_i), \tag{6.9}$$

$$\text{s.t. } \underline{P}(q, a, q_j) \le \widehat{\gamma}_j \le \overline{P}(q, a, q_j) \qquad j \in \overline{\mathcal{N}}^{q,a} \tag{6.10a}$$

$$\sum_{j \in \overline{\mathcal{N}}^{q,a}} \widehat{\gamma}_j = 1 \tag{6.10b}$$

$$\pi_{ij} \ge 0, \qquad i \in \overline{\mathcal{N}}_W^{q,a}, j \in \overline{\mathcal{N}}^{q,a} \tag{6.10c}$$

$$\sum_{i \in \overline{\mathcal{N}}_W^{q,a}} \pi_{ij} = \widehat{\gamma}_j, \qquad j \in \overline{\mathcal{N}}^{q,a} \tag{6.10d}$$

$$\sum_{j \in \overline{\mathcal{N}}^{q,a}} \pi_{ij} = \gamma_i, \qquad i \in \overline{\mathcal{N}}_W^{q,a} \tag{6.10e}$$

$$\sum_{i \in \overline{\mathcal{N}}_W^{q,a}, j \in \overline{\mathcal{N}}^{q,a}} \pi_{ij} c(q_i, q_j) \le \varepsilon^s, \tag{6.10f}$$

where $\underline{P}$, $\overline{P}$, and c are defined in (5.1), (5.2), and (5.4), respectively, and s is given in Assumption 3.1.

Proof. We show that, for a fixed state $q \in Q$ and action $a \in A$, the set of transition probabilities $\Gamma_{q,a}$ defined in (5.5) is the polytope described by the linear Eqs. (6.10). First, assume that $W \equiv \mathbb{R}^n$ and consider the expressions in (6.10) for $\overline{\mathcal{N}}_W^{q,a} = \overline{\mathcal{N}}^{q,a} = \mathcal{N}$. Let $\gamma \equiv (\gamma_1, \dots, \gamma_N) \in \Gamma_{q,a}$. From the definition of $\Gamma_{q,a}$ in (5.5), there exist $\widehat{\gamma} \equiv (\widehat{\gamma}_1, \dots, \widehat{\gamma}_N) \in \widehat{\Gamma}_{q,a}$ and an optimal transport plan

$\pi \equiv (\pi_{ij})_{i,j=1,\dots,N}$ that transports mass from $\hat{\gamma}$ to γ with a cost $\mathcal{T}_c(\gamma,\hat{\gamma})$ smaller than ε^s. Consider now the set $\hat{\Gamma}_{q,a}$ as defined in (5.3) with the transition probability bounds $\underline{P}$ and $\overline{P}$ given by (5.1) and (5.2). Then $\hat{\gamma}$ satisfies the constraints (6.10a) and (6.10b). Since the optimal transport cost $\mathcal{T}_c(\gamma,\hat{\gamma})$ is attained by the transport plan π, it follows from (2.1) with $X \equiv Q$ and c as in (5.4) that

$$\mathcal{T}_c(\gamma,\hat{\gamma}) = \sum_{i,j\in\mathcal{N}} \pi_{ij} c(q_i,q_j).$$

Thus, since $\mathcal{T}_c(\gamma,\hat{\gamma})$ is less than ε^s, we deduce that γ, $\hat{\gamma}$, and π satisfy the linear constraints (6.10c)–(6.10f). Conversely, one can check along the same lines that for any $\gamma \equiv (\gamma_1,\dots,\gamma_N)$, $\hat{\gamma} \equiv (\hat{\gamma}_1,\dots,\hat{\gamma}_N)$, and $\pi \equiv (\pi_{ij})_{i,j=1,\dots,N}$ satisfying (6.10), it also holds that $\gamma \in \Gamma_{q,a}$.

In the general case where the support of p_v^{true} belongs to some known set $W \subseteq \mathbb{R}^n$, we can embed this constraint into the ambiguity sets of transition probabilities and their projections on the abstraction. In particular, we can describe the set of transition probabilities $\Gamma_{q,a}$ via transport plans that couple the regions reachable through the nominal distribution with the regions reachable by some distribution in the ambiguity set. From (6.8a) and (6.8b), these regions are indexed respectively by $\overline{\mathcal{N}}^{q,a}$ and $\overline{\mathcal{N}}^{q,a}_W$. We can therefore just substitute the set $\mathcal{N}$ by the previous ones, since $\hat{\gamma}$, γ, and π are zero for indices outside $\overline{\mathcal{N}}^{q,a}$, $\overline{\mathcal{N}}^{q,a}_W$, and $\overline{\mathcal{N}}^{q,a}_W \times \overline{\mathcal{N}}^{q,a}$, respectively. The proof is now complete. □

Intuitively, for each state $q \in Q$ and action $a \in A$, the Constraints (6.10a)–(6.10f) capture the union of the c-transport cost ambiguity balls in $\mathcal{D}(Q)$ that have radius ε^s and centers all possible distributions of the nominal IMDP. Specifically, Constraints (6.10a) and (6.10b) represent the distributions $\hat{\gamma}$ of the nominal IMDP, i.e., the set $\hat{\Gamma}_{q,a}$. Constraints (6.10c)–(6.10e) describe a transport plan π, i.e., a nonnegative measure on $Q \times Q$ (cf. (6.10c)), which has as its marginals the distribution $\hat{\gamma}$ of the nominal IMDP (cf. (6.10d)), and the target distribution γ (cf. (6.10e)), respectively. Finally, Constraint (6.10f) implies that transport cost to reach the target distribution γ is bounded by ε^s, namely, that γ belongs to the c-transport cost ambiguity ball of radius ε^s.

Remark 6.8. Theorem 6.7 guarantees that, similarly to *IMDP*s without distributional uncertainty [12,13], optimal policies can be computed by solving a set of linear programs. In particular, for every $k \in [0 : K-1]$ and $q \in Q$, we can solve the linear program in Theorem 6.7 for each $a \in A$ and take the action that maximizes the resulting value function. As a result, the computational complexity of synthesizing a strategy is polynomial in $|Q|$ and $|A|$. However, in the IMDP case the resulting linear program has substantially fewer variables and constraints compared to the problem in Theorem 6.7 (order of N for IMDPs, against order of N^2 for Theorem 6.7). Nevertheless, as we detail in Section 6.3, we can reformulate the LP in Theorem 6.7 to obtain a problem of substantially lower complexity, which can be solved more efficiently.

6.3. Dual reformulation of the robust dynamic programming algorithm

In this section, we make use of linear programming duality to obtain a computationally efficient solution to the linear program (6.9)–(6.10). The approach consists of formulating the dual of program (6.9)–(6.10), as common in the literature of robust MDPs [11,49,52], and then introducing a tailored algorithm that takes into account its structure to solve it efficiently. The following theorem provides the dual reformulation of the linear program (6.9)–(6.10).

Theorem 6.9. *Consider the optimization problem*

$$\max_{\mu\geq 0,\lambda} G(\lambda,\mu), \tag{6.11}$$

with

$$G(\lambda,\mu) := \sum_{j\in\overline{\mathcal{N}}^{q,a}} \min\{\underline{P}(q,a,q_j)(h_j(\mu)-\lambda), \overline{P}(q,a,q_j)(h_j(\mu)-\lambda)\} - \mu\epsilon + \lambda \tag{6.12a}$$

$$h_j(\mu) := \min_{i\in\overline{\mathcal{N}}^{q,a}_W} \{\underline{p}^k(q_i) + \mu c(q_i,q_j)\}. \tag{6.12b}$$

Then, the function G is concave, and the optimal value of problem (6.11)–(6.12) is the same as that of problem (6.9)–(6.10). Furthermore, for every $\mu \geq 0$, the maximization of G with respect to λ can be carried out by only evaluating it as λ ranges over the finite set $\{h_j(\mu)\}_{j\in\overline{\mathcal{N}}^{q,a}}$, i.e.,

$$\max_{\lambda} G(\lambda,\mu) = \max_{j\in\overline{\mathcal{N}}^{q,a}} \{G(h_j(\mu),\mu)\}.$$

The proof of Theorem 6.9 is given in Appendix C. Theorem 6.9 allows one to formulate Problem (6.9)–(6.10) as a concave maximization problem in two scalar variables, λ and μ. Furthermore, as Problem (6.11) is derived using duality, if we solve it with an iterative algorithm, then the algorithm can be stopped at any time to obtain a valid lower bound; thus, guaranteeing correctness. Note that $G(\lambda,\mu)$ is bivariate, and therefore its minimization does not fall within the class of dual optimization problems in [52], which also considers Wasserstein ambiguity sets. Nevertheless, our problem essentially reduces to the maximization of the univariate concave function $g(\mu) := \max_\lambda G(\lambda,\mu)$, whose values can be found by simply performing a finite search over the set $\{G(h_j(\mu),\mu)\}_{j\in\overline{\mathcal{N}}^{q,a}}$ by the last part of Theorem 6.9. Therefore, the maximum can be found efficiently by using standard tools from scalar convex optimization.

6.4. Correctness

In this section, we prove the correctness of our abstraction for System (3.1). We begin by refining the strategy σ^* to System (3.1). Let $J : \mathbb{R}^n \to Q$ map the continuous state $x \in \mathbb{R}^n$ to the corresponding discrete state $q \in Q$ of $\mathcal{M}$, i.e., for any $x \in \mathbb{R}^n$,

$$J(x) = q \iff x \in q. \tag{6.13}$$

Given a finite path $\omega^k_{\mathbf{x}} = \mathbf{x}_0 \xrightarrow{\mathbf{u}_0} \mathbf{x}_1 \xrightarrow{\mathbf{u}_1} \dots \xrightarrow{\mathbf{u}_{k-1}} \mathbf{x}_k$ of System (3.1), we define by

$$J(\omega^k_{\mathbf{x}}) = J(\mathbf{x}_0) \xrightarrow{\mathbf{u}_0} J(\mathbf{x}_1) \xrightarrow{\mathbf{u}_1} \dots \xrightarrow{\mathbf{u}_{k-1}} J(\mathbf{x}_k)$$

the corresponding finite path of the MDP abstraction. Consequently, a strategy σ^* for $\mathcal{M}$ is refined to a switched strategy σ^*_x for System (3.1) such that

$$\sigma^*_x(\omega^k_{\mathbf{x}}) := \sigma^*(J(\omega^k_{\mathbf{x}})). \tag{6.14}$$

The following theorem, which is a direct consequence of Theorem 6.7 and Theorem 2 in [23], ensures that the guarantees obtained for the robust MDP abstraction also hold for System (3.1).

Theorem 6.10 (*Correctness*). *Let $\mathcal{M}$ be a robust MDP abstraction of System (3.1), $\sigma^* \in \Sigma$ be an optimal strategy for $\mathcal{M}$, and $\sigma^*_{\boldsymbol{x}}$ be the corresponding switching strategy. Then, for every $q \in Q$, $x \in q$, and $p_v \in \mathcal{P}_v$, it holds that*

$$\overline{p}^K(q) \geq P_{\text{reach}}(X, X_{\text{tgt}}, K \mid x, \sigma^*_{\boldsymbol{x}}, p_v) \geq \underline{p}^K(q),$$

where $\underline{p}^K$ and $\overline{p}^K$ are defined in (6.2a) and (6.2b).

Theorem 6.10 guarantees that in order to solve Problem 3.4 we can synthesize an optimal strategy σ^* for a robust MDP abstraction of System (3.1) and then simply check if $\underline{p}^K(J(x))$ is greater than the given threshold p_{th}.

7. Experimental results

In this section, we evaluate our framework on several benchmarks. In Sections 7.1, 7.2 and 7.4, we consider systems with data-driven ambiguity sets built from M i.i.d. samples $\mathbf{v}^1, \dots, \mathbf{v}^M$ of the noise distribution p_v^{true}. Using these samples, we construct an ambiguity ball, whose center is the empirical distribution

$$\widehat{p}_v = \frac{1}{M} \sum_{i=1}^{M} \delta_{\mathbf{v}^i}, \tag{7.1}$$

of the data in [5], where $\delta_{\mathbf{v}^i}$ denotes the Dirac distribution that assigns unit mass to $\mathbf{v}^i$. As a metric we consider the 2-Wasserstein distance, which is often used in the literature [39]. The radius ϵ of the ambiguity set is a tuning parameter to hedge against the uncertainty coming from the imperfect knowledge of p_v^{true}. Then, in Section 7.3, we depart from the data-driven setting and consider the case where the nominal distribution is Gaussian and ϵ quantifies the degree of robustness against distributional shifts.

To synthesize the respective strategies, we run the dynamic programming algorithm for required time horizon $K \in \mathbb{N} \cup \{\infty\}$ and use the following "average error" to quantify the performance

$$e_{\text{avg}} := \frac{1}{N} \sum_{q \in Q} (\overline{p}^K(q) - \underline{p}^K(q)), \tag{7.2}$$

where $\overline{p}^K(q)$ and $\underline{p}^K(q)$ are the upper and lower bounds of probability of satisfaction, respectively, as defined in (6.3) and (6.5). A summary of the numerical results for all the case studies are shown in Table 1. The main highlight in Table 1 is the superiority of our dual algorithm of Section 6.3 for strategy synthesis when compared to using the standard linear programming routine *Linprog* (two orders of magnitude speed-up).

For all the case studies, the theoretical bounds on the reachability probabilities are validated empirically via 1000 Monte Carlo simulations starting from 1000 random initial conditions. To demonstrate that the theoretical bounds obtained with our approach are sound even in extreme cases, in Section 7.3 (Experiment #12), we consider remarkably large noise and show that the Monte Carlo probability values are still within the theoretical bounds.

All results are obtained by running a Matlab implementation of our framework on a single thread of an Intel Core i7 3.6 GHz CPU with 32 GB of RAM

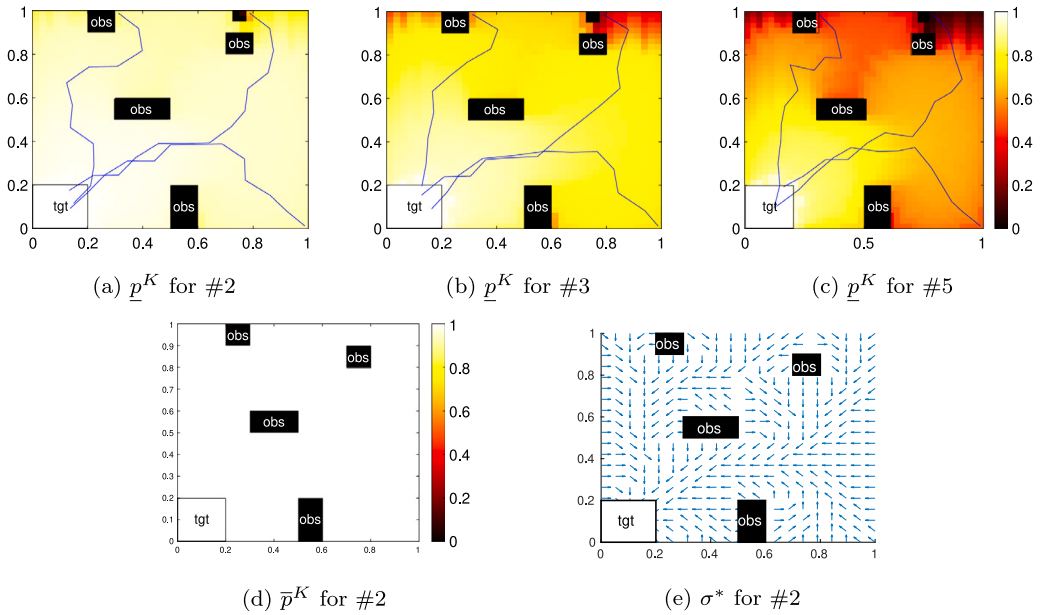
7.1. Linear system

We consider a discrete-time version of the unicycle model in [53], which is obtained using a first order Euler discretization with a time step $\Delta t = 1$. We fix the velocity of the vehicle to the constant value 1 and consider its orientation angle u as the control input leading to the following switched system:

$$x_{k+1} = x_k + \Delta t \left(\begin{bmatrix} \cos(u_k) \\ \sin(u_k) \end{bmatrix} + \mathbf{v}_k \right), \tag{7.3}$$

Table 1
Summary of the results obtained for all case studies. The label "robust MDP" denotes our proposed approach, while the label "IMDP" refers to the alternative approach pointed out in Remark 5.3.

#	Abstraction class	Spec.	$\lvert Q\rvert$	ϵ	e_{avg}	DP Alg.	Abst. time	Synth. time
1	Robust MDP	Reach-avoid	1601	5×10^{-3}	0.05	Linprog	6 s	10.1 h
2	Robust MDP	($K=40$)	1601	5×10^{-3}	0.05	Dual	6 s	76 s
3	Robust MDP	Reach-avoid	1601	10^{-2}	0.18	Linprog	6 s	10.4 h
4	Robust MDP	($K=40$)	1601	10^{-2}	0.18	Dual	6 s	94 s
5	Robust MDP	Reach-avoid	1601	1.5×10^{-2}	0.33	Linprog	6 s	11.4 h
6	Robust MDP	($K=40$)	1601	1.5×10^{-2}	0.33	Dual	6 s	93 s
7	IMDP	Reach-avoid	1601	5×10^{-3}	0.83	Linprog	6 s	1.2 h
8	IMDP	($K=40$)	1601	5×10^{-3}	0.83	Alg. in [12]	6 s	3.9 s
9	Robust MDP	Reach-avoid	1601	5×10^{-2}	0.25	Linprog	12 s	2.43 h
10	Robust MDP	($K=15$)	1601	5×10^{-2}	0.25	Dual	12 s	22 s
11	Robust MDP	Reach-avoid	3601	1.3×10^{-2}	0.13	Dual	35 s	19.5 min
12	Robust MDP	($K=\infty$)	3601	1.3×10^{-3}	0.47	Dual	64 s	42 min
13	Robust MDP	LTL_f	3601	5×10^{-3}	0.01	Dual	6 s	2.78 min



(a) $\underline{p}^K$ for #2 (b) $\underline{p}^K$ for #3 (c) $\underline{p}^K$ for #5

(d) $\overline{p}^K$ for #2 (e) σ^* for #2

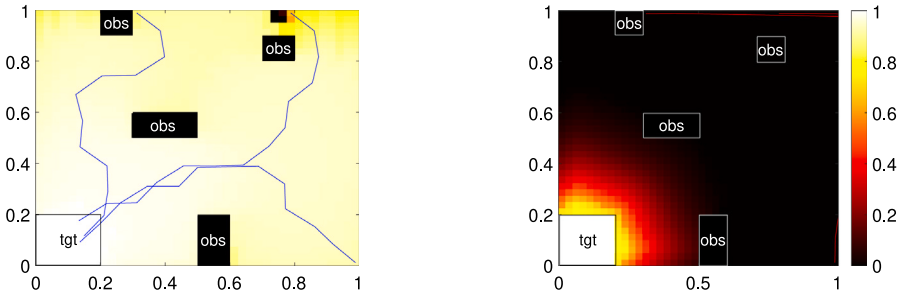
Fig. 1. Results of Experiments #1 – #6 in Table 1. In 1(a), 1(b), and 1(c), lower bound in the probability of reachability corresponding to Experiments #1 (and #2), #3 (and #4) and #5 (and #6), respectively. The plotted trajectories correspond to Monte Carlo simulations of System (7.3) taking samples from a distribution $\widetilde{p}_v \in \mathcal{P}_v$. In 1(d), upper bound in the probability of reachability corresponding to Experiments #1 – #6. In 1(e), optimal strategy of Experiments #1 (and #2). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

where the state of the system is the vehicle position $x_k \in \mathbb{R}^2$ and its control input u_k takes values in $U = \{0, \frac{2\pi}{8}, \dots, 7\frac{2\pi}{8}\}$.

The considered safe, unsafe (obstacle), and target sets are shown in Fig. 1(e). We partition the $[0,1]^2$ rectangle into a grid of $N = 1600$ regions. The ambiguity set is centered at an empirical distribution of $M = 10$ samples, which are drawn from a Gaussian mixture with two components, centered at $[-0.01, 0]$ and $[0.01, 0]$, respectively, and with the same covariance matrix $\text{diag}(2.5\times10^{-5}, 2.5\times10^{-5})$. As a result, the centers of both components are separated by a distance close to the size of the state discretization.

We obtained results for multiple values of the radius ϵ, shown in Table 1 (Experiments #1 – #6), and compare the performance of our dual approach to solve the inner problems in (6.3) to the one obtained by using Linprog. We observe that the average error e_{avg} increases as the ambiguity set grows. This means that, as expected, bigger ambiguity leads to more conservative bounds.

The bounds on the reachability probability for Experiments #1 – #6 are shown in Fig. 1, together with the vector field of the system in close loop with the optimal strategy. Since the obtained upper bound in the reachability probability is the same across Experiments #1 – #6 and the optimal strategy is practically the same, we only show these results in the case of Experiment #1 (equivalently #2).



(a) Robust MDP abstraction approach (b) IMDP abstraction approach

Fig. 2. Results of experiments #1 and #7 in Table 1. Lower bound in the probability of reachability. The trajectories in both figures correspond to Monte Carlo simulations taking samples from a distribution $\widetilde{p}_v \in \mathcal{P}_v$. The ones that satisfy the specification are presented in blue, while the ones that do not are presented in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

To compare our approach with the one that relies on IMDP abstractions described in Remark 5.3, we present the results obtained with the latter as Experiments #7 and #8 in Table 1. Observe that the average error between the bounds in reachability corresponding to Experiments #7 and #8 is way higher than the one corresponding to Experiments #1 and #2 despite the fact that ϵ is the same. This result highlights how our proposed approach, despite requiring a larger amount of time to perform strategy synthesis, is able to provide non-trivial satisfaction guarantees, unlike the approach from Remark 5.3. Furthermore, our approach is able to synthesize a strategy that satisfies the reachability task, in contrast to the approach based on the IMDP abstraction. We compare the obtained lower bounds on the reachability probability for experiments #1 (equivalently #2) and #7 (equivalently #8) in Fig. 2.

Note that Experiments #7 and #8 only differ in the algorithm used for strategy synthesis: whereas in Experiment #7 Linprog was used, a dedicated algorithm was employed in #8. The purpose of showing both results is to compare the computational improvement that stems from employing a dedicated synthesis algorithm for robust MDPs with this same benefit in the case of IMDPs.

7.2. Nonlinear system with 4 modes

For the second case study we consider the nonlinear system from [25,26] with dynamics:

$$x_{k+1} = x_k + \widetilde{f}_{u_k}(x_k) + \mathbf{v}_k. \tag{7.4}$$

Denoting by $x^{(i)}$ the ith component of the state, the map $\widetilde{f}_{u_k}$ is given by

$$\widetilde{f}_u(x) = \begin{cases} [0.5 + 0.2\sin(x^{(2)}), 0.4\cos(x^{(1)})]^T & \text{if } u = 1 \\ [-0.5 + 0.2\sin(x^{(2)}), 0.4\cos(x^{(1)})]^T & \text{if } u = 2 \\ [0.4\cos(x^{(2)}), 0.5 + 0.2\sin(x^{(1)})]^T & \text{if } u = 3 \\ [0.4\cos(x^{(2)}), -0.5 + 0.2\sin(x^{(1)})]^T & \text{if } u = 4. \end{cases} \tag{7.5}$$

The system has state $x_k \in \mathbb{R}^2$ and its control input u_k switches between the discrete values 1, 2, 3, and 4. The safe set is shown in Fig. 3(c). We discretize the rectangle $[-2, 2]^2$ into a uniform grid, which results in abstraction with $N = 1601$ states. The ambiguity set is centered at an empirical distribution of $M = 20$ samples, which are again drawn from a Gaussian mixture with two components, centered at $[-0.05, 0]$ and $[0.05, 0]$, respectively, with covariance matrix $\mathrm{diag}(4 \times 10^{-4}, 4 \times 10^{-4})$. The centers of both components are separated by a distance close to the size of the state discretization. The obtained results are shown in Fig. 3 and in Table 1 in the row corresponding to Experiments #9 and #10. As it can be observed, our approach is able to synthesize a robust control strategy that yields high satisfaction probabilities in a big portion of the state space, despite the size of the obstacles and the system being nonlinear.

7.3. Switched linear systems with 5 modes

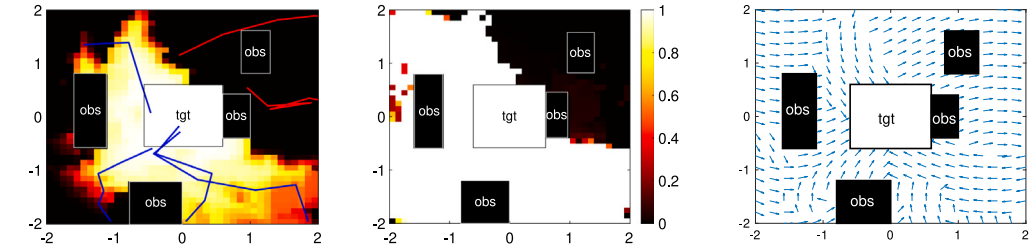
The third case study is a system that switches between 5 linear modes. Its dynamics are of the form

$$x_{k+1} = A_u x_k + \mathbf{v}_k, \tag{7.6}$$

where $U = \{1, \dots, 5\}$ and

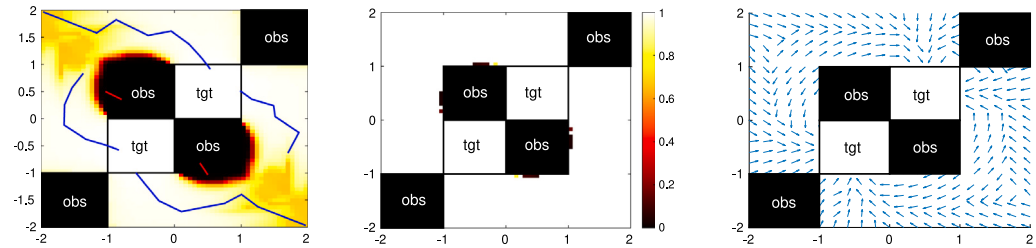
$$A_1 = \begin{pmatrix} 0.79 & 0.035 \\ 0 & 0.825 \end{pmatrix}, \; A_2 = \begin{pmatrix} 0.79 & 0.175 \\ 0 & 0.825 \end{pmatrix}, \; A_3 = \begin{pmatrix} 0.79 & 0 \\ 0.175 & 0.825 \end{pmatrix},$$

$$A_4 = \begin{pmatrix} 1 & 0.2 \\ -0.2 & 1 \end{pmatrix}, \; A_5 = \begin{pmatrix} 1 & -0.2 \\ 0.2 & 1 \end{pmatrix}.$$



(a) Lower bound on the reachability probability (b) Upper bound on the reachability probability (c) Optimal strategy

Fig. 3. Results of experiment #9 (and #10). The trajectories in Fig. 3(a) correspond to Monte Carlo simulations taking samples from a distribution $\widetilde{p}_v \in \mathcal{P}_v$. The ones that satisfy the specification are presented in blue, while the ones that do not are presented in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



(a) Lower bound on the reachability probability (b) Upper bound on the reachability probability (c) Optimal strategy

Fig. 4. Results of Experiment #11. The trajectories in Fig. 4(a) correspond to Monte Carlo simulations taking samples from a Gaussian distribution $\widetilde{p}_v \in \mathcal{P}_v$ at the boundary of the ambiguity set. The ones that satisfy the specification are presented in blue, while the ones that do not are presented in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The state space is defined in the same way as in the previous case studies, with the safe set as shown in Fig. 4(c). The rectangle $[-2,2]^2$ is discretized via a uniform grid, resulting in $N = 3601$ states. Unlike in the previous examples, the motivation for this case study is no longer a data-driven setting, but one of robustness against distributional shifts: the nominal distribution $\widehat{p}_v$ is known to be a Gaussian of zero mean and covariance $\mathrm{diag}(9 \times 10^{-4}, 9 \times 10^{-4})$, but we want to be robust against possible changes in this distribution. To this end, we pick an ambiguity set defined by $\epsilon = 0.0127$, with $s = 2$ and d being the 2-norm. This set contains, but is not limited to, Gaussian distributions with a covariance 1.7 times bigger than the nominal one, or off-diagonal terms bigger than 1/2 times the diagonal entries. Note that, while we truncate the p_v to 4 standard deviations, the ambiguity set contains also unbounded distributions.

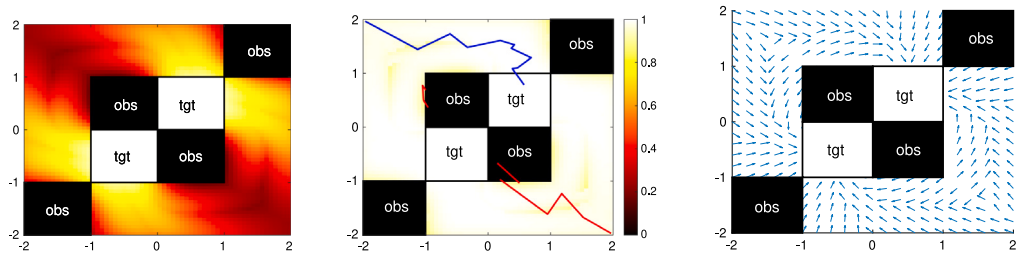
In this case, we obtained the abstraction by making use of the approach in [25]. The results correspond to Experiment #11 in Table 1.

The lower and upper bounds on the reachability probabilities are shown in Figs. 4(a) and 4(b). Additionally, the vector field of the system in closed loop with the optimal strategy is shown in Fig. 4(c).

The empirical results obtained via Monte Carlo simulations are always found within the bounds in the probability of reachability. Furthermore, the synthesized strategies achieve an average (across 1000 random initial conditions) empirical reachability probability of ≈ 1.

To show that the bounds in the probability of reachability are sound even when the empirical probability of reachability is smaller than 1, we increase the variance of the noise. The new nominal noise distribution has covariance $\mathrm{diag}(0.04, 0.04)$ truncated at 3 standard deviations and $\epsilon = 1.3 \times 10^{-3}$. The results correspond to Experiment #12 in Table 1. In this case, the average empirical probability of reachability was 0.802, a non-trivial result, and the empirical reachability probabilities for each initial condition were always found within the theoretical bounds. Note that the synthesis time corresponding to this case study is the longest among the cases in which our dual approach was used. The cause is that the variance of the nominal noise distribution is large. Therefore its support, corresponding to 3 standard deviations, is way bigger than in all the other experiments, which increases the computational complexity of the synthesis process. The average error e_{avg} is also noticeable bigger, due to the higher-variance noise. We plot the bounds in the probability of reachability together with the vector field of the closed-loop system in Fig. 5.

In both Experiments #11 and #12, the optimal stationary strategy was found via the approach proposed in Theorem 6.6, which took only a few seconds.

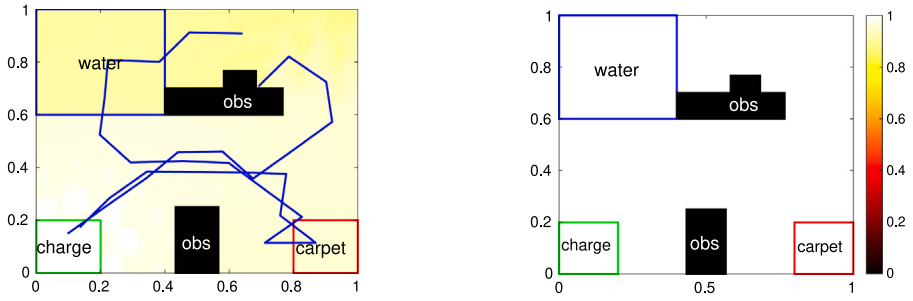


(a) Lower bound on the reachability probability

(b) Upper bound on the reachability probability

(c) Optimal strategy

Fig. 5. Results of Experiment #12. The trajectories in Fig. 5(b) correspond to Monte Carlo simulations taking samples from a Gaussian distribution $\widetilde{p}_v \in \mathcal{P}_v$ at the boundary of the ambiguity set. The ones that satisfy the specification are presented in blue, while the ones that do not are presented in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



(a) Lower bound on the reachability probability

(b) Upper bound on the reachability probability

Fig. 6. Bounds in the probability of reachability for Experiment #13. The trajectories in blue correspond to Monte Carlo simulations taking samples from a distribution $\widetilde{p}_v \in \mathcal{P}_v$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

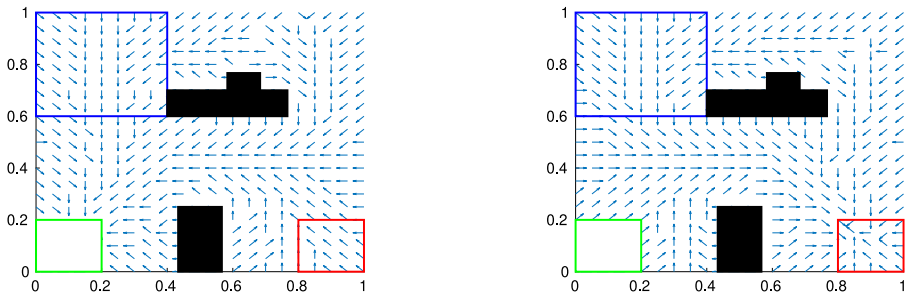
7.4. Complex specifications

In this section we showcase how our approach to solve unbounded reachability problems allows us to synthesize strategies that yield high probability of satisfying more complex specifications. Specifically, we consider the same setting as in case study Experiment #2, together with the following LTL$_f$ specification, borrowed from [54]: the system must reach a charge station while remaining safe and, if the system goes through a region with water, then it should first dry in a carpet before charging. The results are shown as Experiment #13 in Table 1. When compared with Experiment #2, the synthesis time is significantly higher. This happens due to the fact that the specification is more complex than just reachability. As a consequence, strategy synthesis requires solving a reachability problem on a bigger robust MDP, obtained by combining the abstraction with the *automaton* that represents the specification (for more details see [19]). The optimal stationary strategy was extracted via the approach proposed in Theorem 6.6 which, again, took only a few seconds.

Additionally, the bounds in the satisfaction probabilities are shown in Figs. 6 and the vector field of the system in closed loop with the optimal strategy is plotted in 7. Note that the strategy is composed by two stationary strategies, as can be observed in Fig. 7. The system follows the strategy "go to charge" when it is dry, and "go to carpet" when it is wet.

8. Conclusion

In this paper we presented a framework for the formal control of switched stochastic systems with additive, random disturbances whose probability distribution belongs to a Wasserstein ambiguity set. To this end, we derived a robust MDP abstraction of the original system and proposed an algorithm, termed robust dynamic programming, to synthesize robust strategies that maximize the probability of satisfying a (finite or unbounded time horizon) reach-avoid specification. The obtained results demonstrate the effectiveness of our approach in systems with both linear and nonlinear dynamics, and even under complex LTL$_f$ specifications. Our results also show the superiority of our abstraction approach with respect to leveraging directly IMDP abstractions, and the computational advantage of our synthesis algorithm with respect to using off-the-shelf linear programming solvers.



(a) Mode “go to charge” (b) Mode “go to carpet”

Fig. 7. Optimal strategy for the case of Experiment #13.

CRediT authorship contribution statement

Ibon Gracia: Validation, Writing – original draft, Writing – review & editing. **Dimitris Boskos:** Writing – original draft, Writing – review & editing. **Morteza Lahijanian:** Supervision, Writing – original draft, Writing – review & editing. **Luca Laurenti:** Writing – original draft, Writing – review & editing. **Manuel Mazo Jr.:** Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Ibon Gracia reports financial support was provided by National Science Foundation. Morteza Lahijanian reports financial support was provided by National Science Foundation. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported in part by the National Science Foundation, USA [grant number 2039062].

Appendix A. Proof of Proposition 5.2

To prove the proposition we will use two technical lemmas that link couplings and optimal transport discrepancies in the continuous and abstract space.

Lemma A.1 (*Induced Coupling on the Discrete Space*). *Consider the coupling $\pi \in \mathcal{P}(\mathbb{R}^n \times \mathbb{R}^n)$ with marginals p and p', a finite (measurable) partition Q of $\mathbb{R}^n$, and the induced distributions $\gamma, \gamma' \in D(Q)$ with $\gamma(q) := p(q)$ and $\gamma'(q) := p(q)$ for all $q \in Q$. Then $\nu \in \mathcal{P}(Q \times Q)$, defined by*

$$\nu(q, q') := \int_{q \times q'} d\pi(x, y) \tag{A.1}$$

is a coupling between γ and γ.

Proof. The proof follows directly from the fact that

$$\sum_{q' \in Q} \nu(q, q') = \sum_{q' \in Q} \int_{q \times q'} d\pi(x, y) = \int_{q \times \mathbb{R}^n} d\pi(x, y) = \gamma(q),$$

and analogously for the other marginal. □

Next, given two distributions on the continuous space $\mathbb{R}^n$ we establish bounds on the optimal transport discrepancy $\mathcal{T}_c$ of their induced distributions on Q, based on their s-Wasserstein distance in the continuous space.

Lemma A.2 (*Induced optimal transport discrepancy*). *Let $p, p' \in D_s(\mathbb{R}^n)$ and consider the induced distributions $\gamma, \gamma' \in D(Q)$ with $\gamma(q) := p(q)$ and $\gamma'(q) := p(q)$ for all $q \in Q$. Then for any $s \geq 1$ and $\varepsilon \geq 0$ it holds that*

$$\mathcal{W}_s(p, p') \leq \varepsilon \Rightarrow \mathcal{T}_c(\gamma, \gamma') \leq \varepsilon^s,$$

where c is given in (5.4).

Proof. Consider the map J in (6.13) and note that due to (5.4),

$$\|x - y\|^s \geq c(J(x), J(y)) \tag{A.2}$$

for all $x, y \in \mathbb{R}^n$. Let π be an optimal coupling for the s-Wasserstein distance $\mathcal{W}_s(p, p')$ and ν be the induced coupling on Q given by (A.1). Then we get from (2.1), (6.13), and (A.2) that

$$\begin{aligned}\mathcal{W}_s(p, p')^s &= \int_{\mathbb{R}^n \times \mathbb{R}^n} \|x - y\|^s d\pi(x, y) \geq \int_{\mathbb{R}^n \times \mathbb{R}^n} c(J(x), J(y)) d\pi(x, y) \\ &= \sum_{q,q' \in Q} c(q, q') \int_{q \times q'} d\pi(x, y) = \sum_{q,q' \in Q} c(q, q') \nu(q, q') \geq \mathcal{T}_c(\gamma, \gamma'),\end{aligned}$$

which implies the result. The last inequality follows from (2.1) and Lemma A.1, which asserts that ν is a coupling between γ and γ'. The proof is complete. □

The intuition behind Lemma A.2 is the following: if the s-Wasserstein distance between two distributions in $\mathbb{R}^n$ is at most ε, then the optimal transport discrepancy (based on c) between their induced distributions on Q is not more than ε^s.

Proof of Proposition 5.2. Let q, a, x, and p_v as given in the statement and define

$$\widehat{\gamma}_{x,a}(q') := T^a_{\widehat{p}_v}(q' \mid x)$$

for all $q' \in Q$. Then it follows from (5.1) and (5.3) that

$$\widehat{\gamma}_{x,a} \in \widehat{\Gamma}_{q,a}. \tag{A.3}$$

Next, we get from (3.2) and Assumption 3.1 that $T^a_{p_v}(\cdot \mid x)$ and $T^a_{\widehat{p}_v}(\cdot \mid x)$ are distributions in $\mathcal{D}_s(\mathbb{R}^n)$ and that

$$\mathcal{W}_s\left(T^a_{p_v}(\cdot \mid x), T^a_{\widehat{p}_v}(\cdot \mid x)\right) \leq \varepsilon.$$

Since the induced distributions of $T^a_{p_v}(\cdot \mid x)$ and $T^a_{\widehat{p}_v}(\cdot \mid x)$ on Q are $\gamma_{x,a}$ and $\widehat{\gamma}_{x,a}$, respectively, it follows from Lemma A.2 that $\mathcal{T}_c(\gamma_{x,a}, \widehat{\gamma}_{x,a}) \leq \varepsilon^s \equiv \epsilon$, namely, $\gamma_{x,a} \in \mathcal{T}^c_\epsilon(\widehat{\gamma}_{x,a})$. Thus, we deduce from (5.5) and (A.3) that $\gamma_{x,a} \in \Gamma_{q,a}$ and conclude the proof. □

Appendix B. Proofs of Section 6.1.2

We first introduce certain mappings and preliminary results that will used for the proofs of this section. Consider the discounted robust MDP with rewards $\widetilde{\mathcal{M}} := (Q, A, \widetilde{\Gamma}, r, \beta)$ as an extension of $\mathcal{M}$, where

$$r(q) := \begin{cases} 1 & \text{if } q \in Q_{\text{tgt}} \\ 0 & \text{otherwise} \end{cases}$$

is the reward function, $\beta \in (0, 1]$ is the discount factor and $\widetilde{\Gamma}$ is given by

$$\widetilde{\Gamma}_{q,a} := \begin{cases} \{\delta_{q_u}\} & \text{if } q \in Q_{\text{tgt}} \\ \Gamma_{q,a} & \text{otherwise} \end{cases}$$

for all $q \in Q$, $a \in A$. With a small abuse of notation we let Σ and Ξ_M denote, respectively, the sets of strategies and Markovian adversaries of $\widetilde{\mathcal{M}}$. We define the *value function* of a strategy $\sigma \in \Sigma$ and adversary $\xi \in \Xi_M$ as the *expected total reward* [50]

$$V_{\sigma,\xi,\beta}(q) := \mathbb{E}_{P^{q,\sigma}_{\xi}}\left[\sum_{k=0}^{\infty} \beta^k r(\omega(k))\right]$$

of $\widetilde{\mathcal{M}}$. Note that $V_{\sigma,\xi,1}(q) = P_{\text{reach}}(Q_{\text{safe}}, Q_{\text{tgt}}, \infty \mid q, \sigma, \xi)$, which is the desired reachability probability of $\mathcal{M}$ under σ and ξ. By (6.1) and (6.2a), this implies

$$\sup_{\sigma \in \Sigma} \inf_{\xi \in \Xi_M} V_{\sigma,\xi,1}(q) = \underline{p}^{\infty}(q) \tag{B.1}$$

for all $q \in Q$. We next establish that $V_{\sigma,\xi,\beta}$ is continuous at $\beta = 1$, which generalizes the corresponding well-known result for MDPs [50, Lemma 7.18].

Lemma B.1. *Given a strategy $\sigma \in \Sigma$ and an adversary $\xi \in \Xi_M$ of $\widetilde{\mathcal{M}}$, the value function $\beta \mapsto V_{\sigma,\xi,\beta}$ is continuous on $(0, 1]$.*

Proof. For each $q \in Q$, $V_{\sigma,\xi,\beta}(q)$ is the infinite sum of the functions

$$g_k(\beta) := \beta^k \mathbb{E}_{P^{q,\sigma}_{\xi}}[r(\omega(k))], \quad k \in \mathbb{N}_0,$$

which are uniformly bounded on $\beta \in (0, 1]$ by $\lambda_k := \mathbb{E}_{P^{q,\sigma}_{\xi}}[r(\omega(k))]$, $k \in \mathbb{N}_0$, respectively. Since $\sum_{k=0}^{\infty} \lambda_k = V_{\sigma,\xi,1}(q)$, which is finite, as it is the probability of reaching Q_{tgt} from q, $\sum_{k=0}^{\infty} g_k$ converges uniformly to $V_{\sigma,\xi,\beta}(q)$ by the Weierstrass M-test. Thus, since each g_k is continuous with respect to β on $(0, 1]$, the same holds also for $V_{\sigma,\xi,\beta}$. □

We identify the space of value functions $V : Q \to \mathbb{R}_{\geq 0}$ with $\mathcal{V} := \mathbb{R}_{\geq 0}^{|Q|}$ and write $V \geq (\leq, >, <) V'$ for $V, V' \in \mathcal{V}$ to indicate that the respective inequality holds component-wise. The limits of sequences $\{V^k\}$ in $\mathcal{V}$ are also interpreted component-wise. Let Ξ_s be the set of stationary adversaries of $\widetilde{\mathcal{M}}$. Namely, for every $q \in Q$ and $a \in A$, an adversary $\xi \in \Xi_s$ chooses a distribution $\gamma^{\xi}_{q,a} \in \widetilde{\Gamma}_{q,a}$. Given $\sigma \in \Sigma_s$, $\xi \in \Xi_s$, and $\beta \in (0, 1]$, we define the Bellman operator $T_{\sigma,\xi,\beta} : \mathcal{V} \longrightarrow \mathcal{V}$ of the auxiliary MDP $\widetilde{\mathcal{M}}$ by

$$T_{\sigma,\xi,\beta}[V](q) := r(q) + \beta \sum_{q' \in Q} \gamma^{\xi}_{q,\sigma(q)}(q') V(q')$$

for any $V \in \mathcal{V}$, $q \in Q$. We also define, given $\sigma \in \Sigma_s$, $T_{\sigma,\beta} : \mathcal{V} \longrightarrow \mathcal{V}$ with

$$T_{\sigma,\beta}[V](q) := r(q) + \min_{\gamma \in \widetilde{\Gamma}_{q,\sigma(q)}} \beta \sum_{q' \in Q} \gamma(q') V(q')$$

for any $V \in \mathcal{V}$, $q \in Q$. Finally, we define $T_{\beta} : \mathcal{V} \longrightarrow \mathcal{V}$ with

$$T_{\beta}[V](q) := r(q) + \max_{a \in A} \min_{\gamma \in \widetilde{\Gamma}_{q,a}} \beta \sum_{q' \in Q} \gamma(q') V(q')$$

for any $V \in \mathcal{V}$, $q \in Q$.

When $\beta = 1$, we simply write $T := T_1$, $T_{\sigma} := T_{\sigma,1}$ and $T_{\sigma,\xi} := T_{\sigma,\xi,1}$, respectively. Note that the sequence obtained by iterating on T with initial condition 0 is equivalent to performing robust dynamic programming as in Theorem 6.2. We can now proceed with the proofs of the results from Section 6.1.2.

Proof of Theorem 6.2. Consider the sequence of value functions $\{V^k\}_{k \in \mathbb{N}_0}$, defined recursively by $V^{k+1} = T[V^k]$ and $V^0 = 0$.

First, we prove that this sequence converges to a fixed point of T in $\mathcal{V}$. It is rather straightforward to show that the operator T is monotone, namely, that $T[V] \geq T[V']$ holds for all $V, V' \in \mathcal{V}$ with $V \geq V'$, which implies that $V^{k+1} \geq V^k$ for all $k \in \mathbb{N}_0$. One can also readily check that $\{V^k\}_{k \in \mathbb{N}_0}$ is bounded. Therefore, since Q is finite, the sequence converges uniformly to $V^{\infty} \in \mathcal{V}$. In addition, by following the exact same arguments as in [11, Proof of Theorem 3.2a], which establishes that discounted Bellman operators of robust MDPs are contractions, we deduce that T has Lipschitz modulus one and is therefore continuous. Thus, it follows that

$$V^{\infty} := \lim_{k \to \infty} V^{k+1} = \lim_{k \to \infty} T[V^k] = T[\lim_{k \to \infty} V^k] = T[V^{\infty}],$$

namely, V^{∞} is a fixed point of T. In order to show minimality of V^{∞}, let $V^* \in \mathcal{V}$ be an arbitrary fixed point of T. By monotonicity of T we get that $T[0] \leq T[V^*]$, and thus $V^1 \leq V^*$. By induction, we obtain that $V^k \leq V^*$ for all $k \in \mathbb{N}_0$, so $V^{\infty} \leq V^*$. Since V^* was chosen arbitrarily, V^{∞} must be the least fixed point of T in $\mathcal{V}$.

Next, we follow the same strategy as in the proof of [55, Theorem 3] to show that $V^{\infty} = \underline{p}^{\infty}$. To this end, note that $T_{\beta}[V] \leq T[V]$ for each $V \in \mathcal{V}$ and $\beta \in (0, 1)$. Denoting by T^k_{β} the k-fold composition of T_{β} with itself, we obtain by induction that $T^k_{\beta}[0] \leq T^k[0]$ for all $k \in \mathbb{N}_0$. Thus

$$\lim_{k \to \infty} T^k_{\beta}[0] \leq \lim_{k \to \infty} T^k[0] = V^{\infty}. \tag{B.2}$$

From the theory of robust MDPs [11, Theorem 3.2a], the operator T_{β} is a contraction mapping on $\mathcal{V}$ for all $\beta \in (0, 1)$. This implies that T^{β} has a unique fixed point V^{∞}_{β}, which by [11, Theorems 3.1 and 3.2b] satisfies

$$V^{\infty}_{\beta}(q) = \sup_{\sigma \in \Sigma} \inf_{\xi \in \Xi_M} V_{\sigma,\xi,\beta}(q) \tag{B.3}$$

for all $q \in Q$, and $\{T^k_{\beta}[0]\}_{k \in \mathbb{N}_0}$ necessarily converges to it. Thus, we deduce from (B.2) that $\sup_{\sigma \in \Sigma} \inf_{\xi \in \Xi_M} V_{\sigma,\xi,\beta}(q) \leq V^{\infty}(q)$ for all $q \in Q$ and $\beta \in (0, 1)$. From this fact and Lemma B.1 we get

$$\lim_{\beta \to 1} \sup_{\sigma \in \Sigma} \inf_{\xi \in \Xi_M} V_{\sigma,\xi,\beta}(q) = \sup_{\sigma \in \Sigma} \inf_{\xi \in \Xi_M} \lim_{\beta \to 1} V_{\sigma,\xi,\beta}(q) = \sup_{\sigma \in \Sigma} \inf_{\xi \in \Xi_M} V_{\sigma,\xi}(q) \leq V^{\infty}(q)$$

for all $q \in Q$. Hence, from the last inequality and (B.1) we obtain $\underline{p}^{\infty} \leq V^{\infty}$.

To also prove that $\underline{p}^{\infty} \geq V^{\infty}$, note that $V_{\sigma,\xi,\beta} \leq V_{\sigma,\xi}$ for all $\sigma \in \Sigma$, $\xi \in \Xi_M$, and $\beta \in (0, 1)$. Thus, we get again from (B.1) that $\sup_{\sigma \in \Sigma} \inf_{\xi \in \Xi_M} V_{\sigma,\xi,\beta}(q) \leq \underline{p}^{\infty}(q)$ for all $q \in Q$ and $\beta \in (0, 1)$. Combining this with (B.3) and the fact that each T_{β} is also monotone, which implies that the sequence $\{T^k_{\beta}[0]\}_{k \in \mathbb{N}_0}$ is upper bounded by its fixed point V^{∞}_{β}, we get that $T^k_{\beta}[0](q) \leq \underline{p}^{\infty}(q)$ for all $q \in Q$, $k \in \mathbb{N}_0$, and $\beta \in (0, 1)$. Thus, by continuity of each T^k_{β} with respect to $\beta \in (0, 1]$, which follows directly from continuity of T_{β}, we get

$$V^k = T^k[0] = \lim_{\beta \to 1} T^k_{\beta}[0] \leq \underline{p}^{\infty}$$

for all $k \in \mathbb{N}_0$, and taking the limit as $k \to \infty$ we obtain $V^{\infty} \leq \underline{p}^{\infty}$. Therefore, $V^{\infty} = \underline{p}^{\infty}$, which concludes the proof. □

Proof of Proposition 6.3. Consider first the discounted case with discount factor $\beta \in (0, 1)$ and denote, for all $q \in Q$, $\sigma \in \Sigma$, $V_{\sigma,\beta}(q) := \inf_{\xi \in \Xi_M} V_{\sigma,\xi,\beta}(q)$ and $V_{\beta}(q) := \sup_{\sigma' \in \Sigma} V_{\sigma',\beta}(q)$. By [11, Theorem 3.1], which establishes that Markovian strategies are sufficient for optimality and [10, Theorem 4], for every discount factor $\beta \in (0, 1)$, there exists a stationary strategy $\sigma' \in \Sigma_s$ that is optimal, i.e., $V_{\sigma',\beta} = V_{\beta}$. Let $\{\beta_m\}_{m \in \mathbb{N}_0}$ be a non-decreasing sequence that converges to 1. As in the proof of [50, Theorem 7.1.9], since Σ_s is finite, there exist a strategy $\sigma' \in \Sigma_s$ and a subsequence $\{\beta_{m_i}\}_{i \in \mathbb{N}_0}$ such that

$$V_{\sigma',\beta_{m_i}} = V_{\beta_{m_i}}, \ \forall i \in \mathbb{N}_0.$$

Since both $\beta \mapsto V_{\sigma',\beta}$ and $\beta \mapsto V_\beta$ are continuous with respect to $\beta \in (0,1]$ by Lemma B.1, it follows that for every $\sigma \in \Sigma$ and $q \in Q$,

$$\begin{aligned} V_\sigma(q) &= \lim_{\beta\to 1} V_{\sigma,\beta}(q) = \lim_{i\to\infty} V_{\sigma,\beta_{m_i}}(q) \\ &\le \lim_{i\to\infty} \max_{\sigma\in\Sigma} V_{\sigma,\beta_{m_i}}(q) = \lim_{i\to\infty} V_{\sigma',\beta_{m_i}}(q) = V_{\sigma'}(q). \end{aligned}$$

Thus, $\sigma' \in \Sigma_s$ is our desired optimal strategy for the undiscounted case, i.e., $V_{\sigma'} = \underline{p}^\infty$. □

Proof of Proposition 6.4. First, we show that any $\sigma' \in \Sigma_s$ that is optimal satisfies the condition (6.6a). Denote $V_{\sigma'}(q) := \inf_{\xi\in\Xi_M} V_{\sigma',\xi}(q)$ for all $q \in Q$. From optimality of σ' we get that $V_{\sigma'} = \underline{p}^\infty$. Additionally, since $\underline{p}^\infty$ and $V_{\sigma'}$ are respective fixed points of T and $T_{\sigma'}$, it follows that

$$\underline{p}^\infty = T_{\sigma'}[\underline{p}^\infty] \le \max_{\sigma\in\Sigma_s} T_\sigma[\underline{p}^\infty] = T[\underline{p}^\infty] = \underline{p}^\infty,$$

where the max should be interpreted in a component-wise fashion. This implies that

$$\sigma'(q) \in \arg\max_{a\in A} \min_{\gamma\in\widetilde{\Gamma}_{q,a}} \sum_{q'\in Q} \gamma(q')\underline{p}^\infty(q')$$

for all $q \in Q$. Therefore, we conclude that (6.6a) is a necessary condition for optimality of stationary strategies.

Next, we prove that conditions (6.6a) and (6.6b) are sufficient for optimality of stationary strategies. Let $\widetilde{\mathcal{M}}_{\sigma^*,\xi}$ be the time-varying Markov chain obtained by fixing the strategy and adversary of $\widetilde{\mathcal{M}}$ to σ^* and some $\xi \in \Xi_M$, respectively. Denote also by $\widetilde{M}^k_\xi \in \mathbb{R}^{|Q|\times|Q|}$ the transition matrix of $\widetilde{\mathcal{M}}_{\sigma^*,\xi}$ at time step k so that the ith row of $\widetilde{M}^k_\xi$ is a probability distribution from $\widetilde{\Gamma}_{q_i,\sigma^*(q_i)}$. Note that by the definition of the Bellman operators T and $T_{\sigma,\xi}$ at the beginning of this section, and the fact that $\underline{p}^\infty$ is a fixed point of T we have

$$T[\underline{p}^\infty] = \max_{\sigma\in\Sigma_s} \min_{\xi\in\Xi_s} T_{\sigma,\xi}[\underline{p}^\infty]$$

where the max and min are interpreted in a component-wise fashion. Thus, since σ^* satisfies condition (6.6a),

$$\underline{p}^\infty = T[\underline{p}^\infty] = \max_{\sigma\in\Sigma_s} \min_{\xi\in\Xi_s} T_{\sigma,\xi}[\underline{p}^\infty] = \min_{\xi\in\Xi_s} T_{\sigma^*,\xi}[\underline{p}^\infty] \le r + \widetilde{M}^1_\xi \underline{p}^\infty.$$

By applying inductively the same argument, we get

$$\underline{p}^\infty \le r + \left(\sum_{k=1}^{K-1} \widetilde{M}^k_\xi\right) r + \left(\prod_{k=1}^{K} \widetilde{M}^k_\xi\right) \underline{p}^\infty \quad \forall K \in \mathbb{N}, \xi \in \Xi_M.$$

Since the ith entry in the last term of this bound equals

$$\sum_{q'\in Q} P^{q_i,\sigma^*}_\xi[\omega(K) \in q']\underline{p}^\infty(q') = \sum_{q'\in Q_{\text{reach}}} P^{q_i,\sigma^*}_\xi[\omega(K) \in q']\underline{p}^\infty(q'),$$

the last term converges to 0 as $K \to \infty$ by the definition of Q_{reach} and condition (6.6b). Furthermore, the sum of the first two terms converges to $V_{\sigma^*,\xi}$. As a result, $\underline{p}^\infty(q) \le \inf_{\xi\in\Xi_M} V_{\sigma^*,\xi}(q)$ for all $q \in Q$. On the other hand, we have from (B.1) that $\underline{p}^\infty(q) \ge \inf_{\xi\in\Xi_M} V_{\sigma^*,\xi}(q)$ and we conclude that $\underline{p}^\infty(q) = \inf_{\xi\in\Xi_M} V_{\sigma^*,\xi}(q)$, namely, that σ^* is optimal. □

Proof of Theorem 6.6. We begin by proving that σ^* is well-defined. This is the case if σ^* is defined for all states, which is equivalent to having $Q^{m_{\max}} = Q_{\text{reach}}$. Let $\sigma' \in \Sigma_s$ be a stationary optimal strategy, namely, that $V_{\sigma'} = \underline{p}^\infty$ holds. Therefore, the sequence $\{V^k\}_{k\in\mathbb{N}_0}$ obtained via robust dynamic programming as in (6.3) with fixed strategy σ' converges monotonically to $\underline{p}^\infty$. Now we start an induction argument by assuming that for $k \in \mathbb{N}_0$, $V^k(q) > 0$ implies $q \in Q^k$, for $q \in Q$, which holds for $k = 0$. If the previous condition holds at iteration k, then for all $q \in Q \setminus Q^k$ we obtain that

$$\begin{aligned} V^{k+1}(q) &= \min_{\gamma\in\widetilde{\Gamma}_{q,\sigma'(q)}} \sum_{q'\in Q} \gamma(q')V^k(q') = \min_{\gamma\in\widetilde{\Gamma}_{q,\sigma'(q)}} \sum_{q'\in Q^k} \gamma(q')V^k(q') \\ &\le \max_{a\in A^*(q)} \min_{\gamma\in\widetilde{\Gamma}_{q,a}} \sum_{q'\in Q^k} \gamma(q')V^k(q') \le \max_{a\in A^*(q)} \min_{\gamma\in\widetilde{\Gamma}_{q,a}} \sum_{q'\in Q^k} \gamma(q'), \end{aligned}$$

where the first inequality follows from the last statement of Proposition 6.4. This result, together with monotonicity of $\{V^k\}_{k\in\mathbb{N}_0}$, guarantees that at iteration $k+1$, $V^{k+1}(q) > 0$ implies $q \in Q^{k+1}$ for all $q \in Q$. Thus the induction argument holds and we get that $\lim_{k\to\infty} Q^k = Q^{m_{\max}} = Q_{\text{reach}}$.

Next, to prove optimality of σ^*, it suffices to show that σ^* satisfies the two conditions of Proposition 6.4. Condition (6.6a) is satisfied directly by the construction of σ^*. In the following, we show that σ^* also satisfies condition (6.6b). The proof follows closely that of [51, Theorem 10.25]. Let $\widetilde{\mathcal{M}}_{\sigma^*,\xi}$ denote the time-varying Markov Chain corresponding to our strategy σ^* and an arbitrary adversary $\xi \in \Xi_M$. To show that the paths starting at any initial condition eventually exit Q_{reach} and remain outside forever with probability one, we equivalently show that its complement, namely, that Q_{reach} is visited infinitely often, has probability zero. Note that by definition of σ^* via backward reachability, all states $q \in Q_{\text{reach}}$ have nonzero probability of eventually reaching Q_{tgt} and thereafter q_u under σ^* and ξ. This implies that, for each $q \in Q_{\text{reach}}$, there exists a path fragment that reaches q_u and which has positive

probability since it is a finite fragment. Let $p > 0$ be a uniform lower bound on this probability across all states $q \in Q_{\text{reach}}$. Then, the probability of the event "Q_{reach} is visited at least k times, but the corresponding path fragment that reaches q_u is never taken" is upper bounded by $(1-p)^k$. Taking the limit as $k \to \infty$, we obtain that the probability that Q_{reach} is visited infinitely often is upper bounded by $\lim_{k\to\infty}(1-p)^k = 0$, meaning that all paths eventually reach $Q \setminus Q_{\text{reach}}$ and remain there forever with probability one. Since the choice of the adversary was arbitrary, we conclude that condition (6.6b) is also fulfilled, which completes the proof. □

Appendix C. Proof of Theorem 6.9

To facilitate the readability of the proof, we use the shorthand notation $p_i \equiv \underline{p}^k(q_i)$, $\underline{P}_j \equiv \underline{P}(q,a,q_j)$, $\overline{P}_j \equiv \overline{P}(q,a,q_j)$, $\overline{\mathcal{N}} \equiv \overline{\mathcal{N}}^{q,a}$ and $\overline{\mathcal{N}}_W \equiv \overline{\mathcal{N}}^{q,a}_W$.

Proof of Theorem 6.9. Concavity of G follows by standard arguments from convex analysis [56]. In particular, the functions h_j are concave as the pointwise minimum of affine functions and $\min\{\underline{P}_j(\lambda + h_j(\mu)), \overline{P}_j(\lambda + h_j(\mu))\}$ is in turn concave as the minimum of concave functions. Thus, G is also concave as the sum of concave functions.

To obtain (6.11), we first eliminate γ and $\widehat{\gamma}$ from (6.9)–(6.10) and get the equivalent problem

$$\begin{aligned}
\min_{\pi_{ij}} \quad & \sum_{i\in\overline{\mathcal{N}}_W, j\in\overline{\mathcal{N}}} \pi_{ij} p_i \\
\text{s.t.} \quad & \sum_{i\in\overline{\mathcal{N}}_W} \pi_{ij} \geq \underline{P}_j && \forall j \in \overline{\mathcal{N}} \\
& \sum_{i\in\overline{\mathcal{N}}_W} -\pi_{ij} \geq -\overline{P}_j && \forall j \in \overline{\mathcal{N}} \\
& \sum_{i\in\overline{\mathcal{N}}_W, j\in\overline{\mathcal{N}}} -c(q_i,q_j)\pi_{ij} \geq -\epsilon \\
& \sum_{i\in\overline{\mathcal{N}}_W, j\in\overline{\mathcal{N}}} \pi_{ij} = 1 \\
& \pi_{ij} \geq 0 && \forall i \in \overline{\mathcal{N}}_W, j \in \overline{\mathcal{N}}
\end{aligned}$$

By denoting $\alpha := (\alpha_1, \dots, \alpha_N)$, $\beta := (\beta_1, \dots, \beta_N)$, μ, and λ the corresponding dual variables of all but the last set of constraints, we obtain as in [57, Page 142] the dual linear problem

$$\begin{aligned}
\max_{\alpha,\beta,\mu\geq 0,\lambda} \quad & \sum_{j\in\overline{\mathcal{N}}} \left(\alpha_j \underline{P}_j - \beta_j \overline{P}_j\right) - \mu\epsilon + \lambda \\
\text{s.t.} \quad & \alpha_j - \beta_j - \mu c(q_i,q_j) + \lambda \leq p_i && \forall i \in \overline{\mathcal{N}}_W, j \in \overline{\mathcal{N}}.
\end{aligned}$$

Since the feasible set of the primal problem is nonempty and compact, strong duality holds (cf. [57, Theorem 4.4]). The constraints of the dual problem are equivalently written

$$-\alpha_j + \beta_j + \min_{i\in\overline{\mathcal{N}}_W} \{p_i + \mu c(q_i,q_j)\} - \lambda \geq 0 \qquad \forall j \in \overline{\mathcal{N}}$$

and since $\beta_j \geq 0$, they can be cast in the form

$$\beta_j \geq \max\{0, \alpha_j - \min_{i\in\overline{\mathcal{N}}_W} \{p_i + \mu c(q_i,q_j)\} + \lambda\} \qquad \forall j \in \overline{\mathcal{N}}.$$

Taking into account that the objective function of the dual problem is decreasing in each β_j, its optimal value is the same as that of the problem

$$\begin{aligned}
&\max_{\alpha,\mu\geq 0,\lambda} \sum_{j\in\overline{\mathcal{N}}} \left(\alpha_j \underline{P}_j - \max\{0, \alpha_j - \min_{i\in\overline{\mathcal{N}}_W} \{p_i + \mu c(q_i,q_j)\} + \lambda\}\overline{P}_j\right) - \mu\epsilon + \lambda \\
&= \max_{\alpha,\mu\geq 0,\lambda} \sum_{j\in\overline{\mathcal{N}}} \left(\alpha_j \underline{P}_j + \min\{0, -\alpha_j + h_j(\mu) - \lambda\}\overline{P}_j\right) - \mu\epsilon + \lambda \\
&= \max_{\alpha,\mu\geq 0,\lambda} \sum_{j\in\overline{\mathcal{N}}} \min\{\alpha_j \underline{P}_j, \alpha_j(\underline{P}_j - \overline{P}_j) + \overline{P}_j(h_j(\mu) - \lambda)\} - \mu\epsilon + \lambda,
\end{aligned} \tag{C.1}$$

where we used (6.12a) in the first equality. Next, since each $\overline{P}_j > 0$, the corresponding function

$$\alpha_j \mapsto \varphi_j(\alpha_j) := \min\{\alpha_j \underline{P}_j, \alpha_j(\underline{P}_j - \overline{P}_j) + \overline{P}_j(h_j(\mu) - \lambda)\}$$

consists of two affine branches that intersect at $h_j(\mu) - \lambda$. In particular, over the left segment $\alpha_j \mapsto \alpha_j \underline{P}_j$ where $\alpha_j \leq h_j(\mu) - \lambda$, the slope is non-negative, and it is non-positive over the right segment $\alpha_j \mapsto \alpha_j(\underline{P}_j - \overline{P}_j) + \overline{P}_j(h_j(\mu) - \lambda)$ where $\alpha_j \geq h_j(\mu) - \lambda$. Thus, the maximum of the function φ_j is attained at $h_j(\mu) - \lambda$ when $h_j(\mu) - \lambda \geq 0$ and equals $(h_j(\mu) - \lambda)\overline{P}_j$, and at 0 when $h_j(\mu) - \lambda < 0$ where it is equal to $(h_j(\mu) - \lambda)\underline{P}_j$. Namely, we have

$$\max_{\alpha_j \geq 0} \varphi_j(\alpha_j) = \begin{cases} (h_j(\mu) - \lambda)\overline{P}_j & \text{if } h_j(\mu) - \lambda \geq 0 \\ (h_j(\mu) - \lambda)\underline{P}_j & \text{if } h_j(\mu) - \lambda < 0 \end{cases}$$
$$= \min\{(h_j(\mu) - \lambda)\overline{P}_j, (h_j(\mu) - \lambda)\underline{P}_j\},$$

which by virtue of (C.1) implies the dual reformulation (6.11).

To prove the last assertion of the theorem, fix an arbitrary value of $\mu \geq 0$ and consider the function $\lambda \mapsto G_\mu(\lambda) \equiv G(\lambda, \mu)$ with G as given in (6.11). Notice that each min term in the expression of G_μ is a piecewise affine function of λ with two segments and a breakpoint at $h_j(\mu)$ (assuming without loss of generality that $\underline{P}_j \neq \overline{P}_j$). Since G_μ is the sum of piecewise affine functions, it is also piecewise affine and its breakpoints are included in the breakpoints of its components, i.e., in the set $\{h_j(\mu)\}_{j \in \overline{\mathcal{N}}}$. Taking further into account that G_μ is upper bounded by the optimal value of problem (6.11)–(6.12), which is finite by strong duality, its maximum is necessarily attained at some of its breakpoints. This concludes the proof. □

Data availability

No data was used for the research described in the article.

References

[1] G. Yin, C. Zhu, Hybrid Switching Diffusions: Properties and Applications, vol. 63, Springer, New York, 2010.
[2] R. Luna, M. Lahijanian, M. Moll, L.E. Kavraki, Asymptotically optimal stochastic motion planning with temporal goals, in: Algorithmic Foundations of Robotics XI, Springer, 2015, pp. 335–352.
[3] N. Cauchi, L. Laurenti, M. Lahijanian, A. Abate, M. Kwiatkowska, L. Cardelli, Efficiency through uncertainty: Scalable formal synthesis for stochastic hybrid systems, in: Proceedings of the 2019 22nd ACM International Conference on Hybrid Systems: Computation and Control, ACM, Montreal, QC, Canada, 2019, pp. 240–251, http://dx.doi.org/10.1145/3302504.3311805.
[4] H. Rahimian, S. Mehrotra, Distributionally robust optimization: A review, 2019, arXiv preprint arXiv:1908.05659.
[5] P. Mohajerin Esfahani, D. Kuhn, Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations, Math. Program. 171 (1) (2018) 115–166.
[6] R. Gao, A. Kleywegt, Distributionally robust stochastic optimization with Wasserstein distance, Math. Oper. Res. (2022).
[7] O. Kupferman, M.Y. Vardi, Model checking of safety properties, Form. Methods Syst. Des. 19 (2001) 291–314.
[8] G. De Giacomo, M.Y. Vardi, Linear temporal logic and linear dynamic logic on finite traces, in: IJCAI'13 Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence, Association for Computing Machinery, 2013, pp. 854–860.
[9] C. Baier, B. Haverkort, H. Hermanns, J.-P. Katoen, Model-checking algorithms for continuous-time Markov chains, Softw. Eng. IEEE Trans. 29 (6) (2003) 524–541.
[10] A. Nilim, L. El Ghaoui, Robust control of Markov decision processes with uncertain transition matrices, Oper. Res. 53 (5) (2005) 780–798.
[11] G.N. Iyengar, Robust dynamic programming, Math. Oper. Res. 30 (2) (2005) 257–280.
[12] R. Givan, S. Leach, T. Dean, Bounded-parameter Markov decision processes, Artificial Intelligence 122 (1–2) (2000) 71–109.
[13] M. Lahijanian, S.B. Andersson, C. Belta, Formal verification and synthesis for discrete-time stochastic systems, IEEE Trans. Autom. Control 60 (8) (2015) 2031–2045.
[14] I. Gracia, D. Boskos, L. Laurenti, M. Mazo, Jr., Distributionally robust strategy synthesis for switched stochastic systems, in: Proceedings of the 26th ACM International Conference on Hybrid Systems: Computation and Control, 2023, pp. 1–10.
[15] C. Santoyo, M. Dutreix, S. Coogan, A barrier function approach to finite-time stochastic system verification and control, Automatica 125 (2021) 109439.
[16] M. Lahijanian, S.B. Andersson, C. Belta, Approximate Markovian abstractions for linear stochastic systems, in: IEEE Conference on Decision and Control, CDC, IEEE, 2012, pp. 5966–5971.
[17] A. Lavaei, M. Zamani, From dissipativity theory to compositional synthesis of large-scale stochastic switched systems, IEEE Trans. Autom. Control (2022).
[18] M. Dutreix, J. Huh, S. Coogan, Abstraction-based synthesis for stochastic systems with omega-regular objectives, Nonlinear Anal. Hybrid Syst. 45 (2022) 101204.
[19] E.M. Wolff, U. Topcu, R.M. Murray, Robust control of uncertain Markov decision processes with temporal logic specifications, in: 2012 IEEE 51st IEEE Conference on Decision and Control, CDC, IEEE, 2012, pp. 3372–3379.
[20] J. Skovbekk, L. Laurenti, E. Frew, M. Lahijanian, Formal abstraction of general stochastic systems via noise partitioning, IEEE Control Syst. Lett. (2023).
[21] L. Laurenti, M. Lahijanian, A. Abate, L. Cardelli, M. Kwiatkowska, Formal and efficient synthesis for continuous-time linear stochastic hybrid processes, IEEE Trans. Autom. Control 66 (1) (2020) 17–32.
[22] J. Jackson, L. Laurenti, E. Frew, M. Lahijanian, Safety verification of unknown dynamical systems via gaussian process regression, IEEE CDC (2020).
[23] J. Jackson, L. Laurenti, E. Frew, M. Lahijanian, Formal verification of unknown dynamical systems via Gaussian process regression, 2021, arXiv preprint arXiv:2201.00655.
[24] X. Koutsoukos, D. Riley, Computational methods for reachability analysis of stochastic hybrid systems, in: International Workshop on Hybrid Systems: Computation and Control, Springer, 2006, pp. 377–391.
[25] S. Adams, M. Lahijanian, L. Laurenti, Formal control synthesis for stochastic neural network dynamic models, IEEE Control Syst. Lett. (2022).
[26] J. Jackson, L. Laurenti, E. Frew, M. Lahijanian, Strategy synthesis for partially-known switched stochastic systems, in: Proceedings of the 24th International Conference on Hybrid Systems: Computation and Control, 2021, pp. 1–11.
[27] R. Reed, L. Laurenti, M. Lahijanian, Promises of deep kernel learning for control synthesis, IEEE Control Syst. Lett. (2023).
[28] T.S. Badings, A. Abate, N. Jansen, D. Parker, H.A. Poonawala, M. Stoelinga, Sampling-based robust control of autonomous systems with non-gaussian noise, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2022, pp. 9669–9678.
[29] F.B. Mathiesen, L. Romao, S.C. Calvert, A. Abate, L. Laurenti, Inner approximations of stochastic programs for data-driven stochastic barrier function design, 2023, arXiv preprint arXiv:2304.04505.
[30] A. Shapiro, Distributionally robust stochastic programming, SIAM J. Optim. 27 (4) (2017) 2258–2275.
[31] E. Delage, Y. Ye, Distributionally robust optimization under moment uncertainty with application to data-driven problems, Oper. Res. 58 (3) (2010) 595–612.
[32] I. Popescu, Robust mean-covariance solutions for stochastic optimization, Oper. Res. 55 (1) (2007) 98–112.
[33] G.C. Calafiore, L.E. Ghaoui, On distributionally robust chance-constrained linear programs, J. Optim. Theory Appl. 130 (1) (2006) 1–22.

[34] J. Blanchet, K. Murthy, Quantifying distributional model risk via optimal transport, Math. Oper. Res. 44 (2) (2019) 565–600.
[35] J. Blanchet, K. Murthy, F. Zhang, Optimal transport-based distributionally robust optimization: Structural properties and iterative schemes, Math. Oper. Res. 47 (2) (2022) 1500–1529.
[36] F. Santambrogio, Optimal Transport for Applied Mathematicians, vol. 55, (58–63) Birkäuser, NY, 2015, p. 94.
[37] N. Fournier, A. Guillin, On the rate of convergence in Wasserstein distance of the empirical measure, Probab. Theory Related Fields 162 (3) (2015) 707–738.
[38] D. Boskos, J. Cortés, S. Martínez, High-confidence data-driven ambiguity sets for time-varying linear systems, IEEE Trans. Autom. Control (2023).
[39] A. Hakobyan, I. Yang, Wasserstein distributionally robust motion control for collision avoidance using conditional value-at-risk, IEEE Trans. Robot. 38 (2) (2021) 939–957.
[40] H. Xu, S. Mannor, Distributionally robust Markov decision processes, Adv. Neural Inf. Process. Syst. 23 (2010).
[41] I. Yang, A convex optimization approach to distributionally robust Markov decision processes with Wasserstein distance, IEEE Control Syst. Lett. 1 (1) (2017) 164–169.
[42] J.G. Clement, C. Kroer, First-order methods for Wasserstein distributionally robust MDP, in: International Conference on Machine Learning, PMLR, 2021, pp. 2010–2019.
[43] A. Puggelli, W. Li, A.L. Sangiovanni-Vincentelli, S.A. Seshia, Polynomial-time verification of PCTL properties of MDPs with convex uncertainties, in: International Conference on Computer Aided Verification, Springer, 2013, pp. 527–542.
[44] M. Weininger, T. Meggendorfer, J. Křetínskỳ, Satisfiability bounds for ω-regular properties in bounded-parameter Markov decision processes, in: 2019 IEEE 58th Conference on Decision and Control, CDC, IEEE, 2019, pp. 2284–2291.
[45] C. Villani, Topics in Optimal Transportation, vol. 58, American Mathematical Society, 2003.
[46] D. Bertsekas, S.E. Shreve, Stochastic Optimal Control: the Discrete-Time Case, vol. 5, Athena Scientific, 1996.
[47] W. Wiesemann, D. Kuhn, B. Rustem, Robust Markov decision processes, Math. Oper. Res. 38 (1) (2013) 153–183.
[48] M. Dutreix, S. Coogan, Specification-guided verification and abstraction refinement of mixed monotone stochastic systems, IEEE Trans. Autom. Control 66 (7) (2020) 2975–2990.
[49] L. El Ghaoui, A. Nilim, Robust solutions to markov decision problems with uncertain transition matrices, Oper. Res. 53 (5) (2005) 780–798.
[50] M.L. Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming, John Wiley & Sons, 2014.
[51] C. Baier, J.-P. Katoen, et al., Principles of Model Checking, MIT press Cambridge, 2008.
[52] S. Ramani, A. Ghate, Robust markov decision processes with data-driven, distance-based ambiguity sets, SIAM J. Optim. 32 (2) (2022) 989–1017.
[53] P. Tabuada, An approximate simulation approach to symbolic control, IEEE Trans. Autom. Control 53 (6) (2008) 1406–1418.
[54] M. Vazquez-Chanlatte, S. Jha, A. Tiwari, M.K. Ho, S. Seshia, Learning task specifications from demonstrations, in: Advances in Neural Information Processing Systems, vol. 31, 2018.
[55] D. Blackwell, Positive dynamic programming, in: Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, University of California Press Berkeley, 1967, pp. 415–418.
[56] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.
[57] D. Bertsimas, J.N. Tsitsiklis, Introduction to Linear Optimization, vol. 6, Athena scientific, 1997.