Enhancing Data-Driven Stochastic Control via Bundled Interval MDP

Coppola, Rudi; Peruffo, Andrea; Romao, Licio; Abate, Alessandro; Mazo, Manuel

DOI
10.1109/LCSYS.2024.3417852

Publication date
2024

Document Version
Final published version

Published in
IEEE Control Systems Letters

Enhancing Data-Driven Stochastic Control via Bundled Interval MDP

Rudi Coppola, *Graduate Student Member, IEEE*, Andrea Peruffo, *Member, IEEE*, Licio Romao, *Member, IEEE*, Alessandro Abate, and Manuel Mazo Jr., *Senior Member, IEEE*

***Abstract*—The abstraction of dynamical systems is a powerful tool that enables the design of feedback controllers using a correct-by-design framework. We investigate a novel scheme to obtain data-driven abstractions of discrete-time stochastic processes in terms of richer discrete stochastic models, whose actions lead to non-deterministic transitions over the space of probability measures. The data-driven component of the proposed methodology lies in the fact that we only assume samples from an unknown probability distribution. We also rely on the model of the underlying dynamics to build our abstraction through backward reachability computations. The nondeterminism in the probability space is captured by a collection of Markov Processes, and we identify how this model can improve upon existing abstraction techniques in terms of satisfying temporal properties, such as safety or reach-avoid. The connection between the discrete and the underlying dynamics is made formal through the use of the scenario approach theory. Numerical experiments illustrate the advantages and main limitations of the proposed techniques with respect to existing approaches.**

***Index Terms*—Abstractions for control, Markov decision processes, scenario approach, stochastic control systems.**

I. Introduction

THE FRAMEWORK of control synthesis for dynamical systems usually includes a complex model, e.g., an ODE, coupled with a simple specification, e.g., stability, reachability, or invariance. These tasks are typically solved via a proxy approach as the construction of a Lyapunov function, or through numerical optimization methods. Alternatively, one can abstract a dynamical system to a finite-state representation, typically in terms of an automaton or Markov decision process, for which much more complex specifications can be solved [3], [21]. This process involves partitioning the state space into a finite set of regions, each represented by an abstract state, and computing transitions amongst abstract states, which is done by using a mathematical representation for the underlying dynamics. Actions in the abstract model correspond to control inputs (we refer the reader to [21] for more details). Whenever the original dynamics is stochastic, the transitions of its discrete representation are probabilistic.

A common modeling framework used in the stochastic context is provided by Markov decision processes (MDPs), which capture both the control synthesis task (i.e., policy synthesis) and the probabilistic nature of transitions. Richer frameworks, such as interval MDPs (IMDPs) [11], [13], are employed to describe uncertain transition probabilities. The evaluation of transition probabilities requires prior knowledge on the stochastic nature of the underlying system, e.g., by evaluating the integral of the stochastic kernel over partitions, and may be computationally expensive to obtain. As such, the use of samples for the construction of *data-driven* abstractions has recently gained attention [1], [2], [4], [5], [7], [8], [9], [10], [15], [17], [18] both for deterministic and stochastic systems. These approaches consider mostly black-box or grey-box models and construct an abstraction from collected trajectories. Several of these works provide probably approximately correct (PAC) guarantees through the application of the scenario approach [6], [19].

Related Works: In [2], [10], [15], Markov models are created using the scenario approach to evaluate transition probabilities in a stochastic dynamical model; in addition, [1] tackles also epistemic uncertainty in the dynamics. In [12], a notion of PAC alternating simulation relationship is defined by sampling one-step state transitions. Moving forward from one-step transitions are ℓ-complete models (i.e., memory-based approaches), as in [7], [8] for linear and nonlinear deterministic systems, and to synthesise controllers [9]. These methods have been applied on event-triggered control models [17], [18]. Further, in [4], [5] the construction of data-driven, memory-based models is equipped with an adaptive method to estimate the size of the needed memory from observations. With the exception of [2], the works cited above differ from ours in what is assumed to be known about the dynamics of the system. Section II-A we illustrate what knowledge of the dynamics we leverage for our construction.

Contributions: We revisit the approach presented in [2] to abstract a discrete-time dynamical system with additive noise as an IMDP, using techniques from the scenario approach, with the overall goal of studying reach-avoid control problems. In

Manuscript received 8 March 2024; revised 13 May 2024; accepted 4 June 2024. Date of publication 20 June 2024; date of current version 20 August 2024. Recommended by Senior Editor V. Ugrinovskii. *(Corresponding author: Rudi Coppola.)*

Rudi Coppola, Andrea Peruffo, and Manuel Mazo Jr. are with the Faculty of Mechanical Engineering, TU Delft, 2628 CD Delft, The Netherlands (e-mail: r.coppola@tudelft.nl).

Licio Romao is with the Department of Aeronautics and Astronautics, Stanford University, Stanford, CA 94305 USA.

Alessandro Abate is with the Department of Computer Science, University of Oxford, OX1 3QD Oxford, U.K.

Digital Object Identifier 10.1109/LCSYS.2024.3417852

doing so, we introduce an instance of Robust MDP [16] where the ambiguity set has a particular structure. Building upon the results therein, we present a new strategy to construct such an abstraction by incorporating nondeterminism in the transitions: this allows us to search for policies over a larger action space and, therefore, to synthesise controllers for a wider variety of scenarios, with in particular the attainment of the specification of interest with a possibly higher probability, if compared to [2].

II. Notation and Preliminaries

We denote by $\mathcal{P}(S)$ the set of all probability distributions on a discrete or continuous set S. Given a finite set S we denote its power set as 2^S. We denote the interval defined by $a, b \in \mathbb{R}$ as $[a, b]$. A *cover* of a set $\mathcal{X}$ is a finite collection of sets $\mathcal{T} = \{T_i\}_{i=1}^M$ such that each element of the collection is a subset of $\mathcal{X}$ and such that the union of the elements in the collection contains $\mathcal{X}$. A *partition* of a set $\mathcal{X}$ is a cover $\mathcal{Q} = \{Q_i\}_{i=1}^N$ such that the elements of the collection are pairwise disjoint.

A. Stochastic Difference Equations

Consider a stochastic control system represented by a stochastic difference equation, where the dynamics of the state $X_{k+1} \in \mathcal{X} \subset \mathbb{R}^n$ at time $k+1$ depends on a known function $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ of the previous state and input, and on the noise W_k. We formally define the model below.

Definition 1: Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and an independent and identically distributed random process $\{W_k(\omega) \in \mathbb{R}^n : k \in \mathbb{N}_0, \omega \in \Omega\}$. A *Stochastic Difference Equation* (SDE) with additive noise is a sequence of random variables (RVs) defined as

$$X_{k+1} = f(X_k, u_k) + W_k, \tag{1}$$

where $u_k : \mathbb{N}_0 \to U \subseteq \mathbb{R}^m$ and $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$.

Let us define a stochastic kernel $T : \mathbb{R}^n \times U \to \mathcal{P}(\mathbb{R}^n)$ to describe the distribution of X_{k+1} given x_k and u_k as

$$X_{k+1} \sim T(\cdot \mid x_k, u_k). \tag{2}$$

Furthermore we denote the next state under the *nominal dynamics* of the SDE without additive noise as

$$\hat{x}_{k+1} = f(x_k, u_k). \tag{3}$$

Our goal is to produce an abstraction for (1), where we assume to have full knowledge of the nominal dynamics $f(\cdot)$, whilst the distribution of the noise W_k, and hence the distribution of X_k, is unknown.

B. Reach-Avoid Specifications:

We focus on synthesising a controller for a SDE enforcing a reach-avoid specification $\varphi_{x_0}^K$ over a finite time horizon [1], [3]. Let $\mathcal{X}_G \subset \mathcal{X}$ be a *goal* set and let $\mathcal{X}_U \subset \mathcal{X}$ be an *unsafe* set. $\varphi_{x_0}^K$ is satisfied if the system initialized at x_0 reaches the goal set $\mathcal{X}_G$ within K steps, while avoiding the unsafe set $\mathcal{X}_U$. Given a controller $\phi : \mathbb{R}^n \times \mathbb{N}_0 \to U$ the state of the system at time k, X_k, is a RV, hence we denote the *probability* of satisfying a specification as $\mathbb{P}_\phi\{\varphi_{x_0}^K\}$.

Problem Statement: *Given a reach-avoid specification $\varphi_{x_0}^K$ and a SDE with* unknown *additive noise, compute a controller ϕ and a lower bound on the probability of satisfying $\varphi_{x_0}^K$.*

C. Markov Models

This problem can be tackled by means of formal abstractions represented by Markov models. We recall the notions required for our discussion and define, to the best of our knowledge, a new Markov model.

Definition 2 [13]: A *Markov Decision Process* (MDP) is a tuple $M = (S, \mathcal{A}, P, R)$ where S is a finite *set of states*, $\mathcal{A}$ is a set of *actions* where $\mathcal{A}(s)$ indicates the enabled actions in $s \in S$, $P : S \times \mathcal{A} \to \mathcal{P}(S)$ is a *transition probability function* and $R : S \to \mathbb{R}$ is a *reward function*.

Definition 3 [13]: An *Interval Markov Decision Process* (IMDP) is a tuple $M_\updownarrow = (S, \mathcal{A}, P_\updownarrow, R)$ where S, $\mathcal{A}$, and R are defined as in Definition 2, $P_\updownarrow : S \times \mathcal{A} \rightrightarrows \mathcal{P}(S)$ is an *uncertain transition probability function* such that for all s, s' and a there exists $0 \leq \underline{p} \leq \overline{p} \leq 1$ such that $P_\updownarrow(s, a)(s') = [\underline{p}, \overline{p}]$.

Definition 4: A *Bundled Interval Markov Decision Process* (bIMDP) is a tuple $M_\vdots = (S, \mathcal{A}, P_\vdots, R)$ where S, $\mathcal{A}$, and R are defined as in Definition 2, $P_\vdots : S \times \mathcal{A} \rightrightarrows \mathcal{P}(S)$ is an *uncertain transition probability function* such that for all s, s' and a there exists some $K \in \mathbb{N}$ such that $P_\vdots(s, a)(s') := \bigcup_{k=1}^K [\underline{p}_k, \overline{p}_k] \subseteq [0, 1]$.

The interested reader can find details on set valued probabilities in [20]. IMDPs and bIMDPs are instances of Robust MDPs where the *ambiguity set* has a special structure, see [16]. As the name suggests, bIMDPs can be thought of as a collection of IMDPs. Both can be thought of as collections of MDPs each represented by an instance of a transition probability function $P \in P_\vdots$ or $P \in P_\updownarrow$ respectively. For brevity, we present several useful notions for IMDPs; analogous observations hold for bIMDPs.

A deterministic time-varying policy for an IMDP is a function $\pi : S \times \mathbb{N}_0 \to \mathcal{A}$, with $\pi \in \Pi_{M_\updownarrow}$ being the admissible policy space [3]. A reach-avoid specification for an MDP $\varphi_s'^K$ given a goal set $S_G \subset S$ and unsafe set $S_U \subset S$ is defined analogously to a specification for a SDE, see [3]. Similarly, we denote the probability of satisfying a reach-avoid specification given a policy π and a fixed transition probability function $P \in P_\updownarrow$ as $\mathbb{P}_{\pi,P}\{\varphi_s'^K\}$. An optimal policy $\overline{\pi} \in \Pi_{M_\updownarrow}$ for the IMDP maximizes the worst-case probability of satisfying the specification with respect to all the possible transition probability functions coherent with the IMDP. Formally,

$$\overline{\pi} \in \arg\max_{\pi \in \Pi_{M_\updownarrow}} \min_{P \in P_\updownarrow} \mathbb{P}_{\pi,P}\{\varphi_s'^K\}. \tag{4}$$

Remark 1: Provably, the probability of satisfaction of a reach-avoid specification on an MDP can be equivalently expressed by computing the value function for a reward function defined as $R(s) = 1$ for all $s \in S_G$ and by making all states $s \in S_U$ absorbing, i.e., $P(s, a)(s) = 1$, see [3].

III. Finite-State Abstraction

Next, we describe the components required to construct a finite-state abstraction of a SDE: discretization of the state space, transitions among the abstract states, and the evaluation of the probability associated with each transition.

A. State Space Discretization

Let $\mathcal{Q} = \{Q_i\}_{i=1}^N$ be a partition of $\mathcal{X} \subset \mathbb{R}^n$ such that every Q is an n-dimensional convex polytope and let Q_0, the closure of

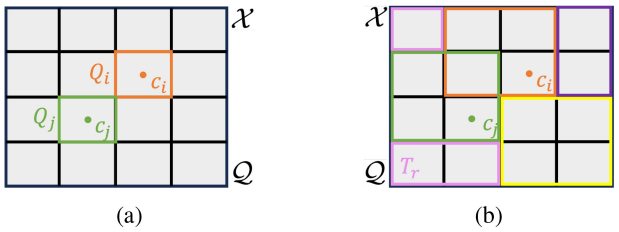


Fig. 1. (a) Partition $\mathcal{Q}$ of the domain of interest $\mathcal{X}$, where Q_i and Q_j are two elements of the partition, c_i and c_j are the respective reference points. (b) Cover of target sets $\mathcal{T}$ of the partition $\mathcal{Q}$. Each color represents a different target set.

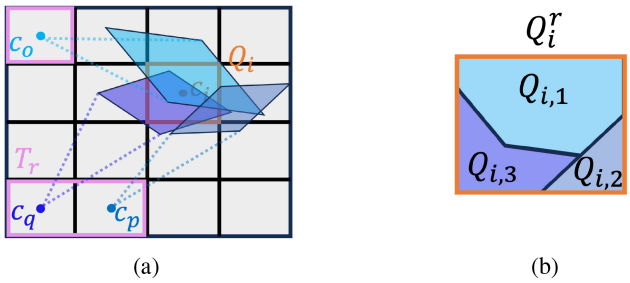


Fig. 2. (a) Pre(C_r) represented as the union of Pre(c_o), Pre(c_p), and Pre(c_q); $a_r \in \mathcal{A}(s_i)$ (b) Assuming the ordering $o < p < q$, the partition Q_i^r induced on Q_i by C_r.

$\mathbb{R}^n \setminus \mathcal{X}$, be a so-called *absorbing region*. We define an abstract state for each element of $\{Q_i\}_{i=0}^N$, yielding a set of $N+1$ discrete states $S = \{s_i\}_{i=0}^N$. We define the relation $R \subseteq \mathbb{R}^n \times S$ where $(x, s_i) \in R$ if and only if $x \in Q_i$, and the notation $R(x) := \{s : (x, s) \in R\}$ and $R^{-1}(s_i) := \{x : (x, s_i) \in R\} = Q_i$. Given a finite collection of *reference* points in $\mathcal{X}$ denoted by $\{c_i\}_{i=1}^N$ such that $R(c_i) = s_i$, we define a bijective map $\psi : S \setminus s_0 \to \{c_i\}_{i=1}^N$ as $\psi(s_i) = c_i$. For simplicity, the points $\{c_i\}_{i=1}^N$ represent the centers of each cell of a uniform grid, as shown in Figure 1a.

Remark 2: Without loss of generality, suppose that the goal set $\mathcal{X}_G$ and the unsafe set $\mathcal{X}_U$ align with the partition, meaning they can be represented as a union of elements from $\mathcal{Q}$. This allows to translate a specification from the concrete system $\varphi_{x_0}^K$ to an equivalent specification on a MDP $\varphi_s'^K$.

B. Actions

Below we define actions linking a single abstract state to (possibly) a *set* of abstract states, named target set.

Let $\mathcal{T} = \{T_i\}_{i=1}^M$ be a finite collection of *target sets* covering $\mathcal{Q}$, in particular $\mathcal{T} \subseteq 2^{\mathcal{Q}}$, see Figure 1b. Further, for every $i = 1, \dots, M$, let C_i be the set of reference states associated with T_i, that is $C_i = \bigcup_{Q \in T_i} \{\psi(R(x)) : x \in Q\}$.

The collection of target sets defines a set of M (arbitrary) elements $\mathcal{A} = \{a_r : r = 1, \dots, M\}$, termed *abstract actions*, where a_r is associated with the target set T_r as shown below. We construct the set of enabled actions at s_i as follows.

Action a_r is enabled at state s_i if for every state $x_k \in R^{-1}(s_i)$ there exists a control input u_k such that the next state under nominal dynamics (3) belongs to the set of reference points associated with T_r, or, in other words, $\hat{x}_{k+1} \in C_r$. Formally, we define the (nominal) *backward reachable set* of a point $x' \in \mathcal{X}$, and, with a slight abuse notation, the *backward reachable set* of a set $C_r \subset \mathcal{X}$ as

$$\text{Pre}(x') := \{x \in \mathcal{X} : \exists u \in U, f(x, u) = x'\} \tag{5}$$

$$\text{Pre}(C_r) := \bigcup_{c_j \in C_r} \text{Pre}(c_j). \tag{6}$$

We require that backward reachable sets of reference points, or a union of those, can contain regions Q_i.

Assumption 1: The backward reachable set of any reference point has a non-empty interior.

For instance, if the system's dynamics is linear, namely, $f(x_k, u_k) = Ax_k + Bu_k$, if A and B are invertible, and U has a non-empty interior then Assumption 1 holds, since (5) results in an affine transformation of U, see also [2].

Action a_r is enabled at state s_i iff $R^{-1}(s_i)$ is contained in Pre(C_r), as shown in Figure 2a:

$$Q_i = R^{-1}(s_i) \subseteq \text{Pre}(C_r) \iff a_r \in \mathcal{A}(s_i). \tag{7}$$

If Q_i satisfies (7) for some C_r, for $x \in Q_i$ there may exist multiple control inputs leading x to C_r, that is the sets $\{\text{Pre}(c_j)\}_{c_j \in C_r}$ need not be disjoint. We exploit the ordering of the partitioning sets $\{Q_i\}_{i=1}^N$ to assign a unique control input driving the state to C_r.

Let $c^* : \mathcal{X} \times \mathcal{A} \to \mathcal{X}$ be a function mapping a continuous state and abstract action to a continuous reference point indexed by the lowest integer, that is

$$c^*(x, a_r) := \arg\min_{c_j \in C_r} j \text{ s.t. } \quad x \in \text{Pre}(c_j). \tag{8}$$

Let us define a control law $u^* : \mathcal{X} \times \mathcal{A} \to U$ such that

$$u^*(x, a_r) \in \{u : f(x, u) = c^*(x, a_r)\}. \tag{9}$$

For every C_r, operation (8) naturally induces a partition on a set $R^{-1}(s_i)$ satisfying (7), defined as

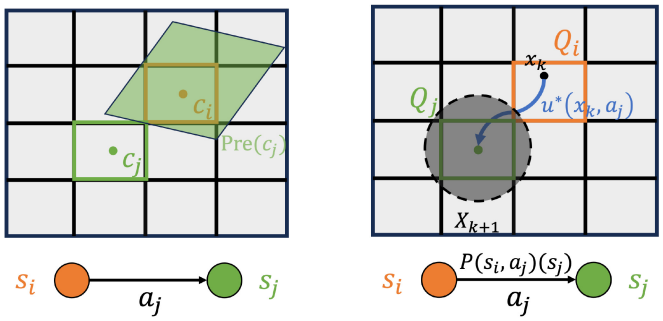
$$Q_i^r := \{X \subseteq Q_i : \forall x, x' \in X, \ c^*(x, a_r) = c^*(x', a_r)\}. \tag{10}$$

In other words, the partition $Q_i^r = \{Q_{i,\ell}\}_{\ell=1}^{L_r}$ is a collection of L_r sets defined by the points sharing the same next state under nominal dynamics and control law u^*, see Figure 2b.

C. Transition Probabilities

In the remaining part of this section, we recall the construction scheme proposed by [2] to abstract a SDE with additive noise to an MDP and introduce a shortcoming of such a procedure. Suppose that the collection of target sets $\mathcal{T}$ coincides with the partition $\mathcal{Q}$, more precisely, $T_r = \{Q_r\}$ for every $r = 1, \dots, N$. In this tailored setting we can simplify our discussion: every set C_r contains a single element, namely c_r, hence action a_r is enabled in the abstract state s_i if and only if $R^{-1}(s_i) \subseteq \text{Pre}(C_r) = \text{Pre}(c_r)$. Similarly, Q_i^r is the trivial partition and contains a single element, namely $Q_i^r = \{Q_i\}$ – cfr. (10) – as depicted in Figure 3a. A finite-state abstraction that describes this framework is an MDP $M = (S, \mathcal{A}, P, R)$, where given $x_k \in R^{-1}(s_i)$, an abstract action $a_j \in \mathcal{A}(s_i)$, and $u_k = u^*(x_k, a_j)$ the probability of transitioning to the abstract state s_j can be computed as:

$$P(s_i, a_j)(s_j) := \int_{R^{-1}(s_j)} T(dx_{k+1} | x_k, u_k). \tag{11}$$



(a) (b)

Fig. 3. (a) Consider $T_j = Q_j$. If action $a_j \in \mathcal{A}(s_i)$ then Pre$(c_j) \supseteq Q_i$. (b) For every $x_k \in Q_i$ there exists an input $u_k = u^*(x, a_j)$ driving the state to c_j. The shaded area represents the support of $T(dx_{k+1}|x_k, u_k)$.

Due to the noise being additive (1) and given the control law u^* we can express (11) as

$$P(s_i, a_j)(s_j) = \mathbb{P}\left\{\omega \in \Omega : c_j + W_k(\omega) \in R^{-1}(s_j)\right\}, \quad (12)$$

where we have used the fact that under nominal dynamics $f(x_k, u^*(x, a_j)) = c_j$. This situation is depicted in Fig. 3b.

D. Shortcomings and Motivating Example

One shortcoming of this approach is that it may lead to a significant *under-approximation* of the dynamics of the concrete system. Formally expressed in (7), if Q_i is not fully contained in Pre(c_j), a_j is not enabled for s_i. As such, one may have to exclude a large set of actions if the dynamics are not well aligned with the chosen partition.

Example 1: Consider the SDE given by

$$X_{k+1} = X_k - u_k + W_k, \quad (13)$$

where $X_k \in \mathbb{R}^2$, W_k is a RV taking values in $\mathbb{R}^2$, $u_k \in \mathcal{U} = [0, \eta] \times [\eta/2, 3\eta/2] \subset \mathbb{R}^2$ for some $\eta > 0$. The partition $\mathcal{Q}$ of $\mathcal{X}$ is a uniform grid where each set Q_i is a $\eta \times \eta$ box. Let $\mathcal{T} = \mathcal{Q}$. Consider a reference state c_j, the center of the box Q_j, and let us examine Pre(C_j) = Pre(c_j): by inverting the nominal dynamics we can characterize such set as

$$\text{Pre}(c_j) = \{x : \exists u \in \mathcal{U}, c_j + u = x\},$$

which represents a copy of Q_j with its center shifted by $[\eta/2, \eta]$, as shown in Figure 4. Considering only one reference point leads to an empty MDP: i.e., all abstract states s have an empty action set. If instead every target set comprises a pair of adjacent cells in the same row: it is easy to see that Q_i is included in Pre$(\{c_j, c_v\})$. This observation motivates the following section, containing our main contribution.[1]

IV. Uncertain Transition Probabilities

In contrast to Section III-C, let us now consider a general cover $\mathcal{T}$ of the partition $\mathcal{Q}$., where at least one of the M target sets, say T_r, contains more than one element of $\mathcal{Q}$; equivalently, C_r contains more than one reference state. Let

[1]A qualitatively different approach to mitigate the illustrated shortcoming is to consider the backward reachable set of reference polytopes, as in [1], instead of reference points.

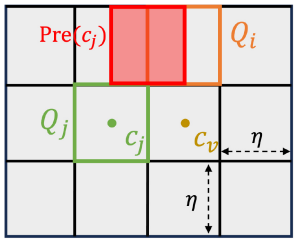
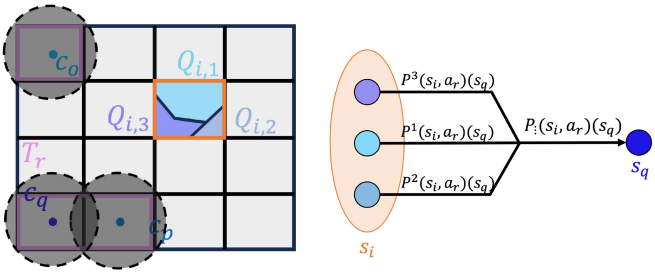


Fig. 4. The dynamics are misaligned with the partition.



(a) (b)

Fig. 5. (a) Computation of $P^\ell(s_i, a_r)(s_q)$ as per (14) for $\ell = 1, 2, 3$. (b) The uncertain transition probability function from a state s_i to s_q under action a_r is a set with L_r values.

$Q_i \subseteq$ Pre(C_r), as depicted in Figure 2a, and consider the non-trivial partition $Q_i^r = \{Q_{i,\ell}\}_{\ell=1}^{L_r}$ induced on Q_i by (8) and described by (10). We know that for every $\ell = 1, \ldots, L_r$ and for every $x \in Q_{i,\ell}$ there exists a control law $u^*(x, a_r)$ that drives the state to one of the reference points in C_r.

In this new setting, it is not possible to describe the transition from s_i under action a_r to a future abstract state $s_j = R(c)$ for $c \in T_r$ by a *single* transition probability function as in (12), but rather by a *set* of transition probability functions. Indeed, the probability of reaching s_j from s_i under action a_r depends on the actual continuous state $x \in Q_i$ from which the transition takes place.

In order to encompass this framework, we define an uncertain probability transition function $P_:$ which encapsulates all possible cases and captures the *nondeterminism* introduced by clustering multiple reference points. Consider an abstract state s_i, an action $a_r \in \mathcal{A}(s_i)$, the partitioning Q_i^r, and suppose that $x_k \in Q_{i,\ell}$ for some $\ell \in \{1, \ldots, L_r\}$: under the control input $u_k = u^*(x_k, a_r)$ the next state under nominal dynamics is the reference point $c^*(x, a_r) \in C_r$. Let us define

$$P^\ell(s_i, a_r)(s_j) := \int_{R^{-1}(s_j)} T(dx_{k+1}|x_k, u_k), \quad (14)$$

By enumerating $\ell = 1, \ldots, L_r$ we obtain a set of transition probability functions which describes all cases, namely $x \in Q_{i,1}, \ldots, x \in Q_{i,L_r}$. Accordingly, we define the bIMDP $M_: = (S, \mathcal{A}, P_:, R_:)$ where

$$P_:(s_i, a_r)(s_j) = \bigcup_{\ell=1}^{L_r} P^\ell(s_i, a_r)(s_j). \quad (15)$$

This is shown graphically in Figure 5a and Figure 5b.

Remark 3: The target sets can be selected arbitrarily. A simple choice is to select adjacent cells, creating 'neighborhoods' of increasing size.

V. PAC Probability Intervals via Sampling

Computing (14) is possible only when the distribution of the additive noise W_k is perfectly known. Additionally, even if it were known, computing explicitly the integral could be difficult or undesirable in certain cases. Instead we provide a lower and upper bound of $P^\ell(s_i, a_r)(s_j)$ using the *sampling-and-discarding* scenario approach proposed in [6] and improved in [19]. In particular, we adopt the framework presented in [2], under the following necessary assumption

Assumption 2 (Non-Degeneracy): For every k, W_k has a density with respect to the Lebesgue measure.

We summarize the results therein here. Let us collect a set of $Z \in \mathbb{N}$ i.i.d. samples of W_k, denoted $w_k^{(1)}, \dots, w_k^{(Z)}$ and define the quantities

$$Z_{s_j}^{\text{in}} = \left|\{w_k^{(i)} : w_k^{(i)} + c_j \in Q_j\}\right|, \quad Z_{s_j}^{\text{out}} = Z - Z_{s_j}^{\text{in}}.$$

In words, $Z_{s_j}^{\text{in}}$ is the number of samples $w_k^{(i)}$ which, when shifted by c_j, fall within region Q_j.

Theorem 1 (PAC Probability Intervals [2, Th. 1]): Given Z samples of the noise W_k, compute $Z_{s_j}^{\text{out}}$ and fix a confidence parameter β. It holds that

$$\mathbb{P}^Z\{\underline{p}_{j,\ell} \le P^\ell(s_i, a_r)(s_j) \le \overline{p}_{j,\ell}\} \ge 1 - \beta,$$

where $\underline{p}_{j,\ell} = 0$ if $Z_{s_j}^{\text{out}} = Z$, $\overline{p}_{j,\ell} = 1$ if $Z_{s_j}^{\text{out}} = 0$, and otherwise $\underline{p}_{j,\ell}$ and $\overline{p}_{j,\ell}$ are respectively the solutions of

$$\frac{\beta}{2Z} = \sum_{i=0}^{Z_{s_j}^{\text{out}}} \binom{Z}{i} \left(1 - \underline{p}_{j,\ell}\right)^i \underline{p}_{j,\ell}^{\,Z-i},$$

$$\frac{\beta}{2Z} = 1 - \sum_{i=0}^{Z_{s_j}^{\text{out}}-1} \binom{Z}{i} \left(1 - \overline{p}_{j,\ell}\right)^i \overline{p}_{j,\ell}^{\,Z-i}.$$

Theorem 1 allows us to provide an upper and lower bound on the individual transition probabilities $P^\ell(s_i, a_r, s_j)$.

We can then describe the resulting abstraction as a bIMDP $M'_{\updownarrow} = (S, \mathcal{A}, P'_{\updownarrow}, R_{\updownarrow})$ where

$$P'_{\updownarrow}(s_i, a_r)(s_j) := \bigcup_{\ell=1}^{L_r} \left[\underline{p}_{j,\ell}, \overline{p}_{j,\ell}\right]. \tag{16}$$

In order to leverage existing algorithms for value iteration on IMDPs, following the approach in [11], [13], we can embed (abstract) the resulting bIMDP into an IMDP $M_{\updownarrow} = (S, \mathcal{A}, P_{\updownarrow}, R)$ where the uncertain transition probability from s_i to state s_j under action a_r is defined as

$$P_{\updownarrow}(s_i, a_r)(s_j) = \left[\underline{p}_j, \overline{p}_j\right], \tag{17}$$

with $\underline{p}_j = \min P'_{\updownarrow}(s_i, a_r)(s_j)$ and $\overline{p}_j = \max P'_{\updownarrow}(s_i, a_r)(s_j)$.

It is obvious from (16) and (17) that the collection of MDPs described by $M'_{\updownarrow}$ is a subset of the MDPs described by $M_{\updownarrow}$. Indeed if $P \in P'_{\updownarrow}$ it implies that $P \in P_{\updownarrow}$. Let $\overline{\pi}$ denote the optimal policy for $M_{\updownarrow}$. It follows that

$$\min_{P \in P_{\updownarrow}} \mathbb{P}_{\overline{\pi},P}\{\varphi_s'^K\} \le \min_{P \in P'_{\updownarrow}} \mathbb{P}_{\overline{\pi},P}\{\varphi_s'^K\}.$$

This embedding allows us to employ existing tools to obtain a policy with a valid lower bound on the probability of satisfaction of the reach-avoid property for the bIMDP, see Remark 1. We compute the optimal policy $\overline{\pi}$ with respect to the IMDP, which in general differs from the optimal policy with respect to the bIMDP. We leave this for future work.

We conclude this section with the following theorem connecting the probability of satisfaction of the reach-avoid property on the IMDP given an optimal policy with the probability of satisfaction of the reach-avoid property on the underlying dynamical system by refining the policy to a time-varying feedback controller. The proof follows the rationale in [2, Theorem 2], and is omitted for brevity.

Theorem 2 (Adapted From [2]): Let $\overline{\pi}$ denote the optimal policy (4) for the IMDP obtained according to (17), and let $L_{\max} := \max_r L_r$. For $\alpha := \beta N M L_{\max}$, the controller $\phi := u^*(x, \overline{\pi}(R(x), k))$, and $x_0 \in R^{-1}(s)$ it holds

$$\min_{P \in P_{\updownarrow}} \mathbb{P}_{\overline{\pi},P}\{\varphi_s'^K\} \ge \eta \Rightarrow \mathbb{P}^Z\{\mathbb{P}_\phi\{\varphi_{x_0}^K\} \ge \eta\} \ge 1 - \alpha. \tag{18}$$

VI. Experimental Evaluation

We demonstrate our results on two systems. Our approach is suitable for nonlinear systems, however we focus on linear dyanamics to simplify the computation of (5). Our code is based upon [2], which for brevity is denoted as "single-target procedure" (STP) where the target sets are chose as in Section III-C, and has been modified in order to include multiple-target transitions (denoted as MTP) according to Section IV. We use PRISM [14] to compute optimal IMDP policies. The interval transition probabilities are computed from $Z = 2 \cdot 10^4$ samples and a confidence $\beta = 10^{-8}$.

Example 1 (Cont'd): We consider the dynamical model (13), over the domain $\mathcal{X} = [-25, 25]^2$, partitioned into 2500 square regions, where the goal set is the region $\mathcal{X}_G = [-25, 25] \times [-25, -24]$. The control input lies in the set $\mathcal{U} = [0, 1] \times [0.5, 1.5]$, and the noise follows a Gaussian distribution $W_k \sim \mathcal{N}(0, 0.15 \cdot I)$. Our goal is the computation of a control policy making the dynamics reach the goal in 50 time steps at most. As outlined in Example 1, the STP returns an IMDP with no actions enabled. In contrast, as argued at the end of Section III-D, if we define every target set as the union of two adjacent cells on the same row, the Pre set covers an entire cell region. Our procedure creates an IMDP equipped with 42391 transitions, and computes a policy whose lower bounds on the satisfaction probability is shown in Fig. 6, with a confidence (see (18)) $\alpha \simeq 0.12$.

Double Integrator: Let us consider the stochastic model

$$X_{k+1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} X_k + \begin{bmatrix} 0.5 & 0 \\ 0 & 1 \end{bmatrix} u_k + W_k, \tag{19}$$

where $W_k \sim \mathcal{N}(0, 0.15 \cdot I)$. The reach-avoid task is to reach the set $[-2, 2]^2$ in 5 time steps, while avoiding states $X \notin [-11, 11]^2$. The control input is limited by the set $[-2, 4] \times [-3, 3]$. We partition the domain $\mathcal{X} = [-11, 11]^2$ into square partitions, in five different configurations: 11x11, 15x15, 18x18, 20x20, and 25x25 regions. For the MTP the target sets are all the pairs of adjacent cells, vertically or horizontally. The complete results are reported in Table I, in terms of computational time, number of transitions of the resulting abstractions, and in percentage of states with a positive probability of reaching the goal set, along with the

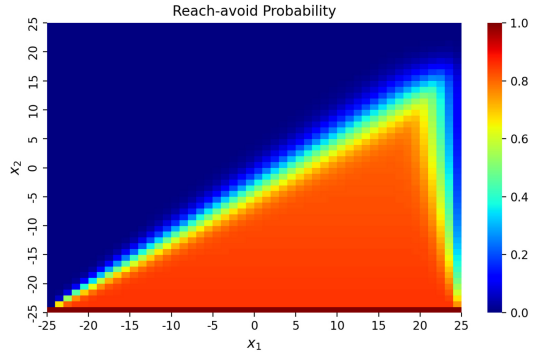


Fig. 6. Lower bound on the probability of reaching the goal set (represented by the lowest row of states) for Example 1.

TABLE I
COMPARISON BETWEEN THE STP AND MTP ABSTRACTIONS, IN TERMS OF NUMBER OF TRANSITIONS, COMPUTATIONAL TIME, AND PERCENTAGE OF STATES THAT HAVE A POSITIVE PROBABILITY TO REACH THE GOAL SET, AND CONFIDENCE α FOR THE MTP

Partition	Transitions		Time [s]		Reach [%]		$\alpha \cdot 10^3$
	STP	MTP	STP	MTP	STP	MTP	MTP
11^2	–	1907	–	7.9	–	61.9	0.5
15^2	–	5592	–	11.2	–	62.2	1.9
18^2	6423	26511	8.4	28.6	54.3	66.0	4.0
20^2	13155	52580	9.8	35.1	62.5	68.5	6.1
25^2	77507	262952	15.5	63.7	67.7	71.0	15

confidence α (see (18)) for the MTP approach. Due to the coarseness of the first two partitions (11x11 and 15x15) the STP returns an IMDP with no enabled actions, as motivated in Section III-D, while our new approach successfully returns a policy. For finer partitions, the STP returns smaller abstractions than the MTP; this is expected, as the MTP considers significantly more target sets – this is reflected in the higher time needed to construct the abstract models. In turn, the MTP yields a larger portion of states with a positive probability of reaching the goal set, thanks to the additional actions available. With finer partitioning, the difference between the STP and the MTP diminishes; the benefit of larger backward reachable sets in the MTP is offset by the smaller cell volume within the partition.

VII. DISCUSSION AND CONCLUSION

We have developed a novel abstraction procedure for discrete-time stochastic systems, exploiting nondeterministic transitions to generate finite-state abstract models. By allowing target sets to comprise multiple cells, rather than a single cell, we show that we can build an abstraction for a greater variety of situations, thus generalizing the scope of earlier results. Our experiments show that this flexibility comes at the cost of generating larger (in terms of transitions) models than the existing single-target approach, and hence introducing more behaviors in the abstraction. The computation of (6) may return a nonconvex set despite the arguments of the union being convex, complicating verifying whether the LHS of (7) holds. The selection of target sets and the embedding of an bIMDP into an IMDP affect the performance of our method: a deeper study of tailored algorithms for bIMDPs exploiting the structure of the uncertain transition function obtained by this scheme are matter of future efforts.

REFERENCES

[1] T. Badings, L. Romao, A. Abate, and N. Jansen, "Probabilities are not enough: Formal controller synthesis for stochastic dynamical systems with epistemic uncertainty," in *Proc. AAAI Conf. Artif. Intell.*, 2023, pp. 1–16.
[2] T. Badings et al., "Robust control for dynamical systems with non-Gaussian via formal abstractions," *J. Artif. Intell. Res.*, vol. 76, pp. 341–391, Jan. 2023.
[3] C. Baier and J.-P. Katoen, *Principles of Model Checking*. Cambridge, MA, USA: MIT, 2008.
[4] A. Banse, L. Romao, A. Abate, and R. Jungers, "Data-driven memory-dependent abstractions of dynamical systems," in *Proc. Learn. Dyn. Control Conf.*, 2023, pp. 891–902.
[5] A. Banse, L. Romao, A. Abate, and R. M. Jungers, "Data-driven abstractions via adaptive refinements and a Kantorovich metric," in *Proc. 62nd IEEE Conf. Decision Control (CDC)*, 2023, pp. 6038–6043.
[6] M. C. Campi and S. Garatti, "A sampling-and-discarding approach to chance-constrained optimization: Feasibility and optimality," *J. Optim. Theory Appl.*, vol. 148, no. 2, pp. 257–280, 2011.
[7] R. Coppola, A. Peruffo, and M. Mazo Jr., "Data-driven abstractions for verification of deterministic systems," 2022, *arXiv:2211.01793*.
[8] R. Coppola, A. Peruffo, and M. Mazo, "Data-driven abstractions for verification of linear systems," *IEEE Control Syst. Lett.*, vol. 7, pp. 2737–2742, 2023.
[9] R. Coppola, A. Peruffo, and M. Mazo Jr., "Data-driven abstractions for control systems," 2024, *arXiv:2402.10668*.
[10] M. Cubuktepe, N. Jansen, S. Junges, J.-P. Katoen, and U. Topcu. "Scenario-based verification of uncertain mdps," in *Proc. Int. Conf. Tools Algorithms Constr. Anal. Syst.*, 2020, pp. 287–305.
[11] T. Dean, R. Givan, and S. Leach. "Model reduction techniques for computing approximately optimal solutions for Markov decision processes," in *Proc. 13th Conf. Uncertain. Artif. Intell.*, 1997, pp. 124–131.
[12] A. Devonport, A. Saoud, and M. Arcak, "Symbolic abstractions from data: A PAC learning approach," in *Proc. 60th IEEE Conf. Decision Control (CDC)*, 2021, pp. 599–604.
[13] R. Givan, S. Leach, and T. Dean, "Bounded-parameter Markov decision processes," *Artif. Intell.*, vol. 122, no. 1, pp. 71–109, 2000.
[14] M. Kwiatkowska, G. Norman, and D. Parker, "PRISM 4.0: Verification of probabilistic real-time systems," in *Proc. Int. Conf. Comput.-Aided Verif.*, 2011, pp. 585–591.
[15] A. Lavaei, S. Soudjani, E. Frazzoli, and M. Zamani, "Constructing MDP abstractions using data with formal guarantees," *IEEE Control Syst. Lett.*, vol. 7, pp. 460–465, 2023.
[16] A. Nilim and L. El Ghaoui, "Robust control of Markov decision processes with uncertain transition matrices," *Oper. Res.*, vol. 53, no. 5, pp. 780–798, 2005.
[17] A. Peruffo and M. Mazo, "Data-driven abstractions with probabilistic guarantees for linear PETC systems," *IEEE Control Syst. Lett.*, vol. 7, pp. 115–120, 2023.
[18] A. Peruffo and M. Mazo Jr., "Sampling performance of periodic event-triggered control systems: A data-driven approach," *IEEE Trans. Control Netw. Syst.*, to be published.
[19] L. Romao, A. Papachristodoulou, and K. Margellos, "On the exact feasibility of convex scenario programs with discarded constraints," *IEEE Trans. Autom. Control*, vol. 68, no. 4, pp. 1986–2001, Apr. 2023.
[20] M. Stojaković, "Set valued probability and its connection with set valued measure," *Stat. Probab. Lett.*, vol. 82, no. 6, pp. 1043–1048, 2012.
[21] P. Tabuada, *Verification and Control of Hybrid Systems: A Symbolic Approach*. New York, NY, USA: Springer, 2009.