




# Distributional Reachability for Markov Decision Processes: Theory and Applications

Yulong Gao , *Member, IEEE*, Alessandro Abate , *Senior Member, IEEE*, Lihua Xie , *Fellow, IEEE*, and Karl Henrik Johansson , *Fellow, IEEE*

**Abstract**—We study distributional reachability for finite Markov decision processes (MDPs) from a control theoretical perspective. Unlike standard probabilistic reachability notions, which are defined over MDP states or trajectories, in this article reachability is formulated over the space of probability distributions. We propose two set-valued maps for the forward and backward distributional reachability problems: the forward map collects all state distributions that can be reached from a set of initial distributions, while the backward map collects all state distributions that can reach a set of final distributions. We show that there exists a maximal invariant set under the forward map and this set is the region where the state distributions eventually always belong to, regardless of the initial state distribution and policy. The backward map provides an alternative way to solve a class of important problems for MDPs: the study of controlled invariance, the characterization of the domain of attraction, and reach-avoid problems. Three case studies illustrate the effectiveness of our approach.

**Index Terms**—Distributional reachability, Markov decision processes (MDPs), probabilistic reachability, reach-avoid problems, set invariance.

## I. INTRODUCTION

### A. Motivation

**R**EACHABILITY is a fundamental problem in systems and control, as well as in formal verification, capturing spatial requirements on state trajectories of a given dynamical model: are solutions of the model remaining in given subsets of the state space? For stochastic dynamical systems, reachability analysis

is usually studied under probabilistic requirements over sets of solutions [1], [2]. For decision processes, where actions (inputs) can be chosen, the problem becomes that of optimizing such reachability probability via policy synthesis.

Much less investigated is a perspective focused on transient distributions: a Markov process can be studied under that lens, and accordingly reachability analysis can be investigated over the space of probability distributions of the model. This less studied perspective enables one to naturally extend the classical state-based reachability notions to new “distributional reachability.” In this article, we thus investigate distributional reachability problems for Markov decision processes (MDPs). We consider the following two problems.

- 1) *Forward distributional reachability*: which state distributions can be reached from an initial state distribution.
- 2) *Backward distributional reachability*: from which state distributions a final state distribution can be reached.

The study of distributional reachability encompasses both theoretical and practical motivations. Distributional reachability provides a new look to explore the fundamental properties of MDP models, e.g., distributional set invariance (as discussed later). In addition, computing reachable sets in the distribution space can be useful for many practical systems that can be specified under richer distribution-based requirements (rather than state-based specifications): we explore case studies in mobile robotics [3] and in models for pharmacokinetics from [4], [5].

Computing exact reachable sets in the probability space is, however, challenging and in general intractable. Even for finite MDP and polytopic reachable sets, the computational complexity of exact set manipulations in a space is exponential with respect to the space dimension. To the best of our knowledge, the state-of-the-art algorithms and tools in computational geometry cannot scalably support the computation of distributional reachability quantities needed in this article. Hence, we develop new efficient and scalable computational algorithms to make the distributional reachability usable in practice: this is an important practical achievement of this contribution.

### B. Related Work

MDP models have been extensively studied across disciplines, including control, operational research, formal methods, and reinforcement learning. There are numerous seminal papers and textbooks on MDPs, e.g., [6], [7], [8], [9], [10]. In the context of control theory, focus of much of the past work has been on optimal policies, such as feedback laws maximizing expected accrued rewards. The synthesis of optimal policies in a computationally efficient manner has been widely explored, recently thanks to the increasing popularity of reinforcement learning.

Manuscript received 22 June 2023; accepted 25 November 2023. Date of publication 11 December 2023; date of current version 28 June 2024. This work was supported in part by the Swedish Research Council Distinguished under Grant 2017–01078, in part by the Swedish Research Council International Postdoc under Grant 2021–06727, and in part by the Knut and Alice Wallenberg Foundation Wallenberg Scholar Grant. Recommended by Associate Editor Z. Shu. (Corresponding author: Yulong Gao.)

Yulong Gao is with the Department of Electrical and Electronic Engineering Imperial College London, SW7 2AZ London, U.K. (e-mail: yulong.gao@imperial.ac.uk).

Alessandro Abate is with the Department of Computer Science, University of Oxford, OX1 3QD Oxford, U.K. (e-mail: alessandro.abate@cs.ox.ac.uk).

Lihua Xie is with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798 (e-mail: elhxie@ntu.edu.sg).

Karl Henrik Johansson is with the Division of Decision and Control Systems, KTH Royal Institute of Technology, and with Digital Futures, 10044 Stockholm, Sweden (e-mail: kallej@kth.se).

Digital Object Identifier 10.1109/TAC.2023.3341282

The study of MDPs has been further extended to constrained or partially observable settings, see [11], [12], [13], [14]. In the following, we restrict our attention to relevant literature focusing on the important problem of MDP reachability.

Reachability analysis for MDPs or related stochastic models by and large refers to the computation of the likelihood that trajectories, initialized at given states, reach a set of states. Probabilistic reachability is fundamental in control theory [2], where it is used in probabilistic safety analysis, as well as in formal verification [15], where it is regarded as a core specification expressible in a probabilistic temporal logic [1]. Such reachability can be regarded either as a verification query, namely to check whether a set of states can be reached from other states with a prescribed probability under a given policy, or as a synthesis task, namely to design a policy that maximizes/minimizes the probability to reach a set of states. Next, we overview some cognate work, with focus on discrete-time models.

Over finite-space models, qualitative probabilistic verification problems (i.e., checking whether the probability is zero or one) have been classically studied using graph-theoretical algorithms [16], whereas quantitative verification problems (i.e., the computation of the actual probability of satisfaction) have been studied via recursive equations and value iteration for finite-horizon problems, or as the solution to systems of linear equations for infinite-horizon problems [1]. The literature has led to a few software tools for probabilistic model checking, such as MRMC [17], PRISM [15], and Storm [18]. Correspondingly, the extensive synthesis problem has been studied in the literature [1]. For example, probabilistic reachability has been explored by using a long-run average approach [19], which is partly related to the present contribution.

For continuous-space stochastic models, the verification problems have been studied, from basic notions of probabilistic safety/reachability [2], to reach-avoid (constrained reachability) [20], as well as more complex probabilistic properties [21], [22]. Related to our work, the authors of [23] investigated probabilistic invariance by means of approximations of transient probability distributions. Methods based on dynamic programming have been developed for solving the synthesis problem [2], [24], [25]. The relation between absorbing sets and probabilistic reachability has also been investigated [25], [26].

Unlike standard approaches to probabilistic reachability based on notions dealing with states and trajectories, this article investigates MDP reachability over the space of state distributions. The core idea is to represent and manipulate a distribution set, namely the convex hull of vertices in the distribution simplex, and to recursively perform reachability computations over this set. We show that a few standard probabilistic verification problems, such as the reach-avoid problem [20], can be studied using the proposed distributional reachability. Dual to reachability, probabilistic set invariance denotes the likelihood of trajectories to remain within a given set over a given time horizon [2], [27], and is in particular useful for probabilistic safety analysis. Instead of defining an invariant set over the state space, such a set can be defined over the distribution space and accounts for a region within which the state distributions should remain. The authors of [4] showed that under some conditions MDPs have a unique, nonempty, compact, invariant set of distributions. Algorithms exist to compute the maximal invariant set of distributions within a constraint set for Markov chains [28] and MDPs [29]. Similar to set invariance, safety constraints for stochastic systems can also be defined on the space of distributions, whereby a model is deemed to be safe if

the underlying transient distribution satisfies given constraints. A typical formulation is the MDP optimal control problem under constraints [30], [31], [32], for which performance-based control synthesis under safety constraints has been studied. In [33], temporal logic synthesis has been discussed for partially observable models. While the setup is quite different from ours, we remark that the temporal requirements over belief spaces involve conditional probability distributions, similar to those studied in this article.

Control of the probability distribution of an MDP is a classical stochastic control problem. The recent papers [34], [35], [36] investigated the optimal steering of a linear stochastic system to a final probability distribution. This problem is connected to the well-known optimal transport problem [37]. The distributional reachability analysis in this article can serve as a feasibility check for the problem of controlling a probability distribution of an MDP, or for the corresponding optimal transport problem.

The close connection between set invariance and backward reachability [38] is leveraged in this work: namely, we show that backward distributional reachability can solve the probabilistic set invariance problem for MDPs. It should be noted that the set invariance discussed here is different from the probabilistic controlled invariant set developed in [39], where invariance denotes whether the probability that state trajectories stay in a set is greater than a prescribed constant, regardless of the initial states.

### C. Contributions

In this article, we study the distributional reachability problem for discrete-time, finite-state MDPs. The main contributions are summarized as follows.

- 1) By considering a stochastic policy and the induced dynamics of the state distribution, we formulate the forward and backward distributional reachability problems. We correspondingly propose two set-valued maps:  $\mathcal{FR}$  and  $\mathcal{BR}$ . Given a set of state distributions  $\Pi$ , the set  $\mathcal{FR}(\Pi)$  collects all state distributions that can be reached from  $\Pi$  in one step, while  $\mathcal{BR}(\Pi)$  is the set of state distributions that can reach  $\Pi$  in one step. The computation of these sets is shown to leverage the manipulation of polytopes in the distributions simplex. If the given set of distributions is polytopic, both forward and backward reachable sets are polytopic. In order to mitigate the computational overhead related to set operations in high-dimensional spaces, we propose a sample-based method (see Algorithm 1) for underapproximating forward and backward reachable sets and prove that the approximations are tight in probability when the sample number tends to infinity (see Theorem 3.1).
- 2) We show that there always exists an  $\mathcal{FR}$ -invariant set. The maximal  $\mathcal{FR}$ -invariant set is the region where the state distributions eventually always belong to, regardless of the initial state distribution and the selected policy. We show that the maximal  $\mathcal{FR}$ -invariant set is smaller than the whole distribution space for some MDPs. We provide ways to compute  $\mathcal{FR}$ -invariant sets and give a condition on its uniqueness (see Theorem 4.1).

- 3) We revisit a class of important problems for MDPs: controlled invariance, domain of attraction, and reach-avoid problems. We show that all these problems can be reformulated and studied using  $\mathcal{BR}$ , and discuss computational ameliorations related to the proposed sample-based algorithm. We compare our approach with the state of the art through three case studies.

The rest of this article is organized as follows. Section II reviews preliminaries on MDPs. Section III investigates distributional reachability. Section IV is devoted to the characterization of  $\mathcal{FR}$ -invariant sets. In Section V, we study controlled invariance, domain of attraction, and reach-avoid problems using distribution reachability. The three case studies in Section VI illustrate the effectiveness of our approach and the scalability of our algorithm is tested in Section VII. Finally, Section VIII concludes this article.

## II. PRELIMINARIES

### A. Notation

Let  $\mathbb{N}$  denote the set of nonnegative integers and  $\mathbb{R}$  the set of real numbers. For some  $q, s \in \mathbb{N}$  and  $q < s$ , let  $\mathbb{N}_{\geq q}$  and  $\mathbb{N}_{[q,s]}$  denote the sets  $\{r \in \mathbb{N} \mid r \geq q\}$  and  $\{r \in \mathbb{N} \mid q \leq r \leq s\}$ , respectively. When  $\leq, \geq, <, >$  are applied to vectors, they are interpreted elementwise. For a set  $\mathbb{X}$ ,  $\mathcal{P}(\mathbb{X})$  denotes the space of probability distributions on  $\mathbb{X}$ . Given a set  $\{x_i\}_{i=1}^N$  with  $x_i \in \mathbb{R}^n$ ,  $\forall i \in \mathbb{N}_{[1,n]}$ ,  $\text{conv}(\{x_i\}_{i=1}^N)$  denotes the convex hull of  $\{x_i\}_{i=1}^N$ . For a set  $\mathbb{X}$ ,  $\text{cl}(\mathbb{X})$  denotes the closure of  $\mathbb{X}$  and  $\text{int}(\mathbb{X})$  denotes the interior of  $\mathbb{X}$ . Matrices of appropriate dimension with all elements equal to 1 and 0 are denoted by  $\mathbf{1}$  and  $\mathbf{0}$ , respectively.

*Definition 2.1:* A convex polytope  $\Pi \subset \mathbb{R}^n$  can be expressed in a (vertex) V-representation, i.e.,  $\Pi = \text{conv}(\{v_1, \dots, v_N\})$ ; or alternatively in a (face, or half-space) H-representation, i.e.,  $\Pi = \{z \in \mathbb{R}^n \mid Az \leq b\}$ , where  $v_i \in \mathbb{R}^n$ ,  $i = 1, \dots, N$ ,  $N \in \mathbb{N}$ ,  $A \in \mathbb{R}^{l \times n}$ ,  $b \in \mathbb{R}^l$ , and  $l$  is the number of half-spaces.

### B. Markov Decision Process

A finite MDP is described by a triple  $M = (\mathbb{X}, \mathbb{U}, T)$ , where:

- 1)  $\mathbb{X}$  is a finite state space, denoted  $\{1, 2, \dots, n\}$ ;
- 2)  $\mathbb{U}$  is a finite control space, denoted  $\{1, 2, \dots, m\}$ ;
- 3)  $T : \mathbb{X} \times \mathbb{U} \rightarrow \mathbb{R}$  is a transition probability, i.e.,  $T(y|x, u)$  assigns a probability from state  $x \in \mathbb{X}$  and control input  $u \in \mathbb{U}$ , to state  $y \in \mathbb{X}$ .

For each  $x \in \mathbb{X}$ ,  $\mathbb{U}_x \subseteq \mathbb{U}$  is the nonempty set consisting of the admissible control inputs when the state is  $x$ . For any  $x \in \mathbb{X}$  and  $u \in \mathbb{U}_x$ ,  $\sum_{y \in \mathbb{X}} T(y|x, u) = 1$ . A state distribution  $\pi \in \mathcal{P}(\mathbb{X})$  is a row vector in  $\mathbb{R}^n$  such that  $\sum_{x \in \mathbb{X}} \pi(x) = 1$ .

### C. Policy and Occupation Measure

A map  $\mu : \mathbb{X} \rightarrow \mathcal{P}(\mathbb{U})$  is called a one-step policy if for each  $x \in \mathbb{X}$ ,  $\mu(\cdot|x)$  assigns a probability distribution over  $\mathbb{U}_x$ , i.e.,  $\sum_{u \in \mathbb{U}_x} \mu(u|x) = 1$ . Each one-step policy induces a row stochastic matrix

$$P^\mu(y|x) = \sum_{u \in \mathbb{U}_x} T(y|x, u) \mu(u|x). \quad (1)$$

Denote with  $\bar{\mathcal{U}}$  the set of one-step policies. A policy  $\mu = \{\mu_0, \mu_1, \dots\}$  is a sequence of one-step policies, i.e.,  $\mu_k \in \bar{\mathcal{U}}$ ,

$\forall k \in \mathbb{N}$ . Denote by  $\mathcal{U}$  the set of policies. Next, let us define occupation measures.

*Definition 2.2:* A matrix  $Q \in \mathbb{R}^{n \times m}$  is said to be an occupation measure to the MDP  $M$  if  $Q \geq 0$  satisfies  $(Q\mathbf{1})^T \in \mathcal{P}(\mathbb{X})$ . Denote by  $\mathcal{O}$  the set of all occupation measures for  $M$ .

Given a state distribution  $\pi \in \mathcal{P}(\mathbb{X})$  and a one-step policy  $\mu \in \bar{\mathcal{U}}$ , there always exists an occupation measure  $Q \in \mathcal{O}$  such that

$$Q(x, u) = \mu(u|x) \pi(x). \quad (2)$$

Dually, given an occupation measure  $Q \in \mathcal{O}$ , one can recover a state distribution  $(Q\mathbf{1})^T \in \mathcal{P}(\mathbb{X})$  and a one-step policy  $\mu \in \bar{\mathcal{U}}$  with

$$\mu(u|x) = \begin{cases} \frac{Q(x, u)}{\sum_{v \in \mathbb{U}_x} Q(x, v)}, & \text{if } \sum_{v \in \mathbb{U}_x} Q(x, v) > 0 \\ \frac{1}{|\mathbb{U}_x|}, & \text{if } \sum_{v \in \mathbb{U}_x} Q(x, v) = 0 \text{ \& } u \in \mathbb{U}_x. \end{cases}$$

*Remark 2.1:* Occupational measures have been widely used. Their close connection to policies allows to reformulate MDP problems, e.g., a constrained MDP problem can be reformulated as a linear program [11], and its solution can be used to recover a policy. In our work, occupational measures will be used to handle the coupling between state distributions and policies, and to facilitate the formulation of distributional reachability analysis as a manipulation of polytopes in the distribution simplex (see Section III).

### D. State Distribution Dynamics

Given an initial state distribution  $\pi_0 \in \mathcal{P}(\mathbb{X})$  and a policy  $\mu \in \mathcal{U}$ , the state distribution evolves as

$$\pi_{k+1}(y) = \sum_{x \in \mathbb{X}} \pi_k(x) P^{\mu_k}(y|x) \quad (3)$$

where  $P^{\mu_k}$  is defined in (1). We rewrite (3) in vector form as

$$\pi_{k+1} = \pi_k P^{\mu_k}. \quad (4)$$

According to (4), the state distribution can be represented as  $\pi_k = \psi(k, \pi_0, \mu)$ , where  $\psi : \mathbb{N} \times \mathcal{P}(\mathbb{X}) \times \mathcal{U} \rightarrow \mathcal{P}(\mathbb{X})$  is

$$\psi(k, \pi_0, \mu) = \pi_0 P^{\mu_0} P^{\mu_1} \dots P^{\mu_{k-1}}. \quad (5)$$

## III. DISTRIBUTIONAL REACHABILITY

In this section we investigate both forward and backward distributional reachability and provide a way to recursively compute forward and backward reachable sets.

*Definition 3.1:* A state distribution  $\pi_f$  is reachable from the initial state distribution  $\pi_0$  at time step  $k$  if there exists a policy  $\mu \in \mathcal{U}$  such that  $\pi_f = \psi(k, \pi_0, \mu)$ .

*Problem 3.1:* (Forward distributional reachability) Given  $\Pi_0 \subseteq \mathcal{P}(\mathbb{X})$ , a set of initial state distributions, compute the set  $\text{FReach}(\Pi_0)$  equal to all the state distributions  $\pi_f$  reachable from  $\pi_0 \in \Pi_0$  at some time step  $k \in \mathbb{N}$ .

*Problem 3.2:* (Backward distributional reachability) Given  $\Pi_f \subseteq \mathcal{P}(\mathbb{X})$ , a set of final state distributions, compute the set  $\text{BReach}(\Pi_f)$  equal to all the state distributions  $\pi_0$  such that  $\pi_f \in \Pi_f$  is reachable from  $\pi_0$  at some time step  $k \in \mathbb{N}$ .

### A. Forward Reachable Sets

We first consider forward distributional reachability. Denote by  $\text{FReach}(\Pi_0, k)$  the  $k$ -step forward reachable set from



$\Pi_0$ . This set collects all the state distributions  $\pi_f$  reachable from  $\pi_0 \in \Pi_0$  at time step  $k$ . Then, the forward reachable set  $\text{FReach}(\Pi_0)$  is the union of the sets  $\text{FReach}(\Pi_0, k)$  overall  $k \in \mathbb{N}$ , i.e.,

$$\text{FReach}(\Pi_0) = \bigcup_{k \in \mathbb{N}} \text{FReach}(\Pi_0, k).$$

Define the map  $\mathcal{FR} : 2^{\mathcal{P}(\mathbb{X})} \rightarrow 2^{\mathcal{P}(\mathbb{X})}$  as

$$\mathcal{FR}(\Pi) = \left\{ \pi \in \mathcal{P}(\mathbb{X}) \left| \begin{array}{l} Q \in \mathcal{O}, (Q\mathbf{1})^T \in \Pi, \\ \forall y \in \mathbb{X}, \pi(y) = \\ \sum_{x \in \mathbb{X}} \sum_{u \in \mathbb{U}_x} T(y|x, u) Q(x, u) \end{array} \right. \right\} \quad (6)$$

where  $\Pi \subseteq \mathcal{P}(\mathbb{X})$ . Thus,  $\mathcal{FR}(\Pi)$  is the set of distributions that are reachable from  $\Pi$  in one step. The map (6) expresses the distribution dynamics in a tractable form, using occupational measures.

The next proposition shows how to recursively compute  $\text{FReach}(\Pi_0, k)$ .

**Proposition 3.1:** Given a set of initial state distributions  $\Pi_0 \subseteq \mathcal{P}(\mathbb{X})$ , the reachable set  $\text{FReach}(\Pi_0, k)$  evolves as

$$\text{FReach}(\Pi_0, k+1) = \mathcal{FR}(\text{FReach}(\Pi_0, k)) \quad (7)$$

with initialization  $\text{FReach}(\Pi_0, 0) = \Pi_0$ .

*Proof:* First, given a set  $\Pi \subseteq \mathcal{P}(\mathbb{X})$ , the set of distributions reachable from  $\Pi$  in one step is

$$\left\{ \pi \in \mathcal{P}(\mathbb{X}) \left| \begin{array}{l} \exists \pi' \in \Pi, \exists \mu \in \bar{\mathcal{U}}, \\ \text{s.t., } \forall y \in \mathbb{X}, \pi(y) = \sum_{x \in \mathbb{X}} \pi'(x) P^\mu(y|x) \end{array} \right. \right\}. \quad (8)$$

Note that the distribution dynamics (3) can be rewritten in the following form:

$$\begin{aligned} \pi(y) &= \sum_{x \in \mathbb{X}} \pi'(x) P^\mu(y|x) \\ &= \sum_{x \in \mathbb{X}} \pi'(x) \left( \sum_{u \in \mathbb{U}_x} T(y|x, u) \mu(u|x) \right) \\ &= \sum_{x \in \mathbb{X}} \sum_{u \in \mathbb{U}_x} T(y|x, u) \mu(u|x) \pi'(x) \end{aligned}$$

for some  $\mu \in \bar{\mathcal{U}}$ . It follows from (2) that the product of  $\mu(u|x)$  and  $\pi'(x)$  can be replaced by an occupation measure. Then, the set in (8) be rewritten as  $\mathcal{FR}(\Pi)$  in (6). That is, the set  $\mathcal{FR}(\Pi)$  collects all the one-step ahead state distributions given the set of current state distributions  $\Pi$ . Thus, the iteration (7) gives the  $k$ -step forward reachable set given the initial set  $\Pi_0$ . ■

**Example 3.1:** Consider the MDP with  $\mathbb{X} = \{1, 2, 3\}$ ,  $\mathbb{U} = \{1, 2, 3\}$ , and the transition probability, shown in Fig. 1(a). Given the initial distribution  $\pi_0 = [0.1 \ 0.2 \ 0.7]$  (asterisk), the forward reachable sets  $\text{FReach}(\pi_0, k)$  computed by the map  $\mathcal{FR}$  are shown in red in Fig. 2 for  $k = 1, 2, 3, 5$ .

### B. Backward Reachable Sets

Now let us consider backward distributional reachability. Denote by  $\text{BReach}(\Pi_f, k)$  the  $k$ -step backward reachable set from  $\Pi_f$ . This set collects all the state distributions  $\pi_0$  such that some state distribution  $\pi_f \in \Pi_f$  is reachable from  $\pi_0$  at time

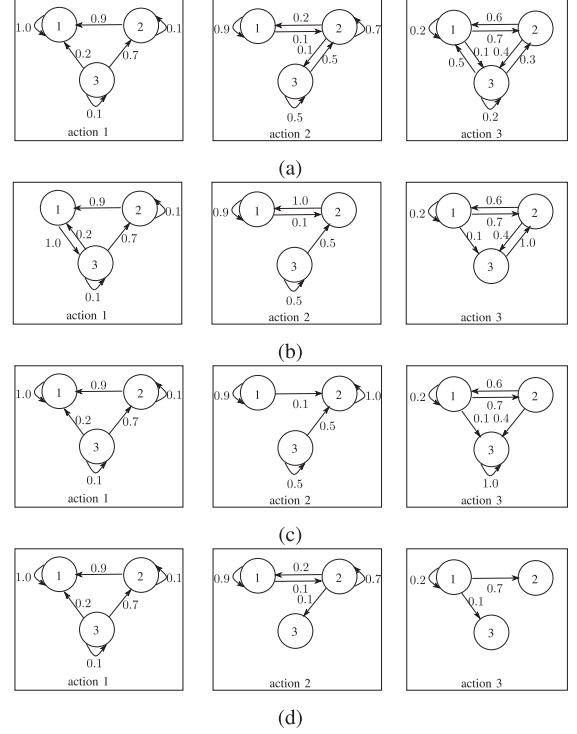


Fig. 1. Graphical syntax of the four MDPs employed in the examples, where  $\mathbb{X} = \{1, 2, 3\}$ ,  $\mathbb{U} = \{1, 2, 3\}$ .

step  $k$ . Then, the backward reachable set  $\text{BReach}(\Pi_f)$  is the union of the sets  $\text{BReach}(\Pi_f, k)$  overall  $k \in \mathbb{N}$ , i.e.,

$$\text{BReach}(\Pi_f) = \bigcup_{k \in \mathbb{N}} \text{BReach}(\Pi_f, k).$$

Define the map  $\mathcal{BR} : 2^{\mathcal{P}(\mathbb{X})} \rightarrow 2^{\mathcal{P}(\mathbb{X})}$

$$\mathcal{BR}(\Pi) = \left\{ (Q\mathbf{1})^T \in \mathcal{P}(\mathbb{X}) \left| \begin{array}{l} Q \in \mathcal{O}, \pi \in \Pi, \\ \forall y \in \mathbb{X}, \pi(y) = \\ \sum_{x \in \mathbb{X}} \sum_{u \in \mathbb{U}_x} T(y|x, u) Q(x, u) \end{array} \right. \right\} \quad (9)$$

where  $\Pi \subseteq \mathcal{P}(\mathbb{X})$ . Dual to the map  $\mathcal{FR}(\Pi)$ ,  $\mathcal{BR}(\Pi)$  encompasses the set of distributions that can reach  $\Pi$  in one step.

The next proposition shows how to recursively compute the set  $\text{BReach}(\Pi_f, k)$ .

**Proposition 3.2:** Given a set of final state distributions  $\Pi_f \subseteq \mathcal{P}(\mathbb{X})$ , the reachable set  $\text{BReach}(\Pi_f, k)$  evolves as

$$\text{BReach}(\Pi_f, k+1) = \mathcal{BR}(\text{BReach}(\Pi_f, k)) \quad (10)$$

with initialization  $\text{BReach}(\Pi_f, 0) = \Pi_f$ .

*Proof:* Similar to the proof of Proposition 3.1. The map  $\mathcal{BR}$  in (9) collects all the one-step state distributions that can reach the set of state distributions  $\Pi$ . The iteration (10) gives the  $k$ -step backward reachable set given the target set  $\Pi_f$ . ■

**Example 3.2:** Consider the MDP in Fig. 1. Given the target distribution  $\pi_f = [0 \ 0 \ 1]$ , the backward reachable sets  $\text{BReach}(\pi_f, k)$  are empty for all  $k \geq 1$ . Instead, if the target distribution is  $\pi_f = [0.5 \ 0.2 \ 0.3]$  (asterisk), the backward reachable sets  $\text{BReach}(\pi_f, k)$  computed by the map  $\mathcal{BR}$  are shown

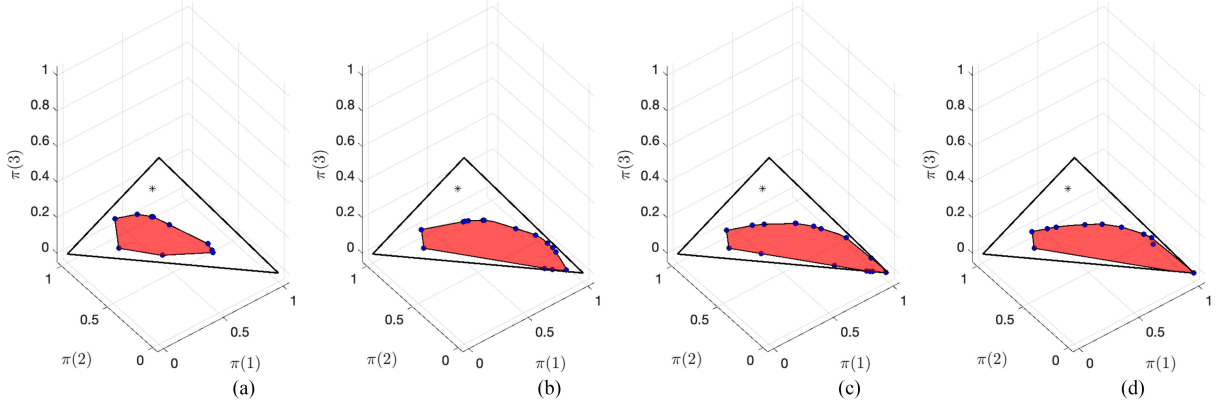


Fig. 2. Forward reachable sets  $\text{FReach}(\pi_0, k)$  (red) for Example 3.1 and the approximated forward reachable sets  $\widehat{\text{FReach}}(\pi_0, k)$  (convex hull of the blue dots) for Example 3.3. The sets are given for (a)  $k = 1$ ; (b)  $k = 2$ ; (c)  $k = 3$ ; and (d)  $k = 5$ .  $\pi_0$  is denoted by the asterisk (single point outside of the hull).

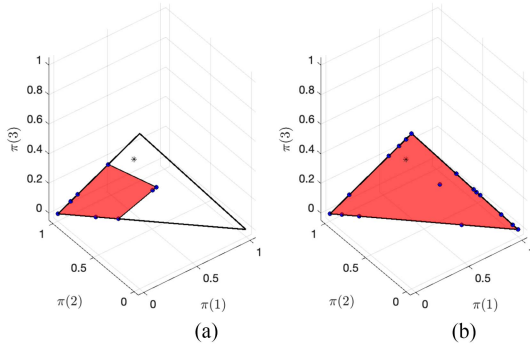


Fig. 3. Backward reachable sets  $\text{BReach}(\pi_f, k)$  (red) for Example 3.2 and the approximated backward reachable sets  $\widehat{\text{BReach}}(\pi_f, k)$  (convex hull of the blue dots) for Example 3.3. The sets are given for (a)  $k = 1$ ; (b)  $k = 2$ .  $\pi_f$  is denoted by the asterisk [point outside of the hull in (a)].

in Fig. 3. We see that the set  $\text{BReach}(\pi_f, 2)$  covers the whole distribution space  $\mathcal{P}(\mathbb{X})$ .

### C. Sample-Based Computation

In the following, we provide a sample-based procedure for approximating the reachable sets  $\mathcal{FR}(\Pi)$  and  $\mathcal{BR}(\Pi)$  when  $\Pi$  is a convex polytope. Note that the constraints on the MDP dynamics in terms of the occupation measures are linear. If the state distribution set  $\Pi$  is polytopic, it follows from (6) that the forward reachable set  $\mathcal{FR}(\Pi)$  is also polytopic. Dually, if the set  $\Pi$  is polytopic, it follows from (9) that the backward reachable set  $\mathcal{BR}(\Pi)$  is either empty or polytopic. The sets  $\mathcal{FR}(\Pi)$  in (6) and  $\mathcal{BR}(\Pi)$  in (9) are expressed as set projections from  $\mathbb{R}^{nm+n}$  to  $\mathbb{R}^n$ . When the MDP has a large number of states, these projections can be computationally heavy. To tackle this, next we present a sample-based approximation scheme as Algorithm 1.

The input to Algorithm 1 consists of two convex polytopes  $\Pi$  and  $\Gamma$  with  $\Pi \subseteq \mathcal{P}(\mathbb{X}) \subset \text{int}(\Gamma)$  and the number of samples  $N_s \in \mathbb{N}_{\geq 1}$ . In line 1, we select uniformly at random samples  $\{\pi_i^s\}_{i=1}^{N_s}$  in  $\mathbb{R}^n$  from  $\Gamma$ . Then, these samples are used to generate samples in  $\mathcal{FR}(\Pi)$  and  $\mathcal{BR}(\Pi)$ , line 3, by projecting  $\pi_i^s$  onto  $\mathcal{FR}(\Pi)$  and  $\mathcal{BR}(\Pi)$ . The output of Algorithm 1 is the two convex hulls of these projected samples, namely  $\widehat{\mathcal{FR}}_{N_s}(\Pi)$  and  $\widehat{\mathcal{BR}}_{N_s}(\Pi)$ .

#### Algorithm 1: Sample-based Reach Set Computation.

**Input:** two convex polytopes  $\Pi$  and  $\Gamma$  with

$\Pi \subseteq \mathcal{P}(\mathbb{X}) \subset \text{int}(\Gamma)$ ,  $N_s \in \mathbb{N}_{\geq 1}$

**Output:** approximated forward and backward reachable sets, denoted by  $\widehat{\mathcal{FR}}(\Pi)$  and  $\widehat{\mathcal{BR}}(\Pi)$

- 1: Select uniformly at random a group of samples  $\{\pi_i^s\}_{i=1}^{N_s}$  from  $\Gamma$ ;
- 2: **for**  $i = 1 : N_s$  **do**
- 3:   compute

$$\pi_i^{fs} = \underset{\pi \in \mathcal{FR}(\Pi)}{\text{argmin}} \|\pi - \pi_i^s\|^2, \quad (11)$$

$$\pi_i^{bs} = \underset{\pi \in \mathcal{BR}(\Pi)}{\text{argmin}} \|\pi - \pi_i^s\|^2; \quad (12)$$

4: **end for**

5: **return**

$$\widehat{\mathcal{FR}}_{N_s}(\Pi) = \text{conv} \left( \{\pi_i^{fs}, i \in \mathbb{N}_{[1, N_s]}\} \right), \quad (13)$$

$$\widehat{\mathcal{BR}}_{N_s}(\Pi) = \text{conv} \left( \{\pi_i^{bs}, i \in \mathbb{N}_{[1, N_s]}\} \right). \quad (14)$$

The sets  $\widehat{\mathcal{FR}}_{N_s}(\Pi)$  and  $\widehat{\mathcal{BR}}_{N_s}(\Pi)$  are inner approximations of  $\mathcal{FR}(\Pi)$  and  $\mathcal{BR}(\Pi)$ , respectively, for all  $N_s \in \mathbb{N}_{\geq 1}$ . The inclusion  $\widehat{\mathcal{FR}}_{N_s}(\Pi) \subseteq \mathcal{FR}(\Pi)$  follows from:

$$\begin{aligned} \text{conv} \left( \{\pi_i^{fs}, i \in \mathbb{N}_{[1, N_s]}\} \right) &\subseteq \mathcal{FR}(\text{conv}(\{\pi_i^s, i \in \mathbb{N}_{[1, N_s]}\})) \\ &\subseteq \mathcal{FR}(\Pi) \end{aligned}$$

and  $\widehat{\mathcal{BR}}_{N_s}(\Pi) \subseteq \mathcal{BR}(\Pi)$  similarly.

The following theorem states that such inner approximations become asymptotically tight in probability with respect to the increase in the number of samples  $N_s$ .

**Theorem 3.1:** Consider two convex polytopes  $\Pi$  and  $\Gamma$  with  $\Pi \subseteq \mathcal{P}(\mathbb{X}) \subset \text{int}(\Gamma)$  and an integer  $N_s \in \mathbb{N}_{\geq 1}$ , as the inputs to Algorithm 1. Let  $N_v^f$  and  $N_v^b$  be the number of the vertices of  $\mathcal{FR}(\Pi)$  and  $\mathcal{BR}(\Pi)$ , respectively. Then, there exist  $0 \leq \alpha_f, \alpha_b < 1$  such that

$$\Pr \left( \widehat{\mathcal{FR}}_{N_s}(\Pi) = \mathcal{FR}(\Pi) \right) \geq 1 - N_v^f \alpha_f^{N_s} \quad (15)$$

$$\Pr\left(\widehat{\mathcal{BR}}_{N_s}(\Pi) = \mathcal{BR}(\Pi)\right) \geq 1 - N_v^b \alpha_b^{N_s}. \quad (16)$$

*Proof:* See Appendix. ■

**Complexity of Algorithm 1:** The computational complexity of Algorithm 1 is linear in the number of samples  $N_s$ , and polynomial in the number of states  $n$  and of control inputs  $m$ . Projecting each sample  $\pi_i^s$  onto  $\mathcal{FR}(\Pi)$  [or  $\mathcal{BR}(\Pi)$ ] is a quadratic program with  $n + nm$  decision variables. The interior point method [40] can solve such quadratic program in  $\mathcal{O}((n + nm)^3)$ .

**Remark 3.1:** The selection of samples  $\pi_i^s \in \Gamma$  is important for reducing the conservativeness of the inner approximation of  $\widehat{\mathcal{FR}}_{N_s}(\Pi)$  for  $\mathcal{FR}(\Pi)$  [or  $\widehat{\mathcal{BR}}_{N_s}(\Pi)$  for  $\mathcal{BR}(\Pi)$ ]. A possible approach is to choose the set  $\Gamma$  to be a large hyperrectangle that contains  $\mathcal{P}(\mathbb{X})$  and to select the samples  $\pi_i^s$  not in  $\mathcal{P}(\mathbb{X})$ , but with projections onto  $\mathcal{FR}(\Pi)$  and  $\mathcal{BR}(\Pi)$  on the boundaries of  $\mathcal{FR}(\Pi)$  and  $\mathcal{BR}(\Pi)$ , respectively. Intuitively, the convex hulls of these projected samples are more likely to be close to  $\mathcal{FR}(\Pi)$  and  $\mathcal{BR}(\Pi)$ .

**Example 3.3:** Let us use Algorithm 1 for iteratively computing the approximated forward reachable sets  $\widehat{\text{FReach}}(\pi_0, k)$  of Example 3.1. The sets  $\widehat{\text{FReach}}(\pi_0, k)$  are denoted by the convex hull of the blue dots in Fig. 2. Similarly, the approximated backward reachable sets  $\widehat{\text{BReach}}(\pi_f, k)$  for Example 3.2 are the convex hull of the blue dots, shown in Fig. 3. We can see that the approximations in two cases are tight in the sense that  $\widehat{\text{FReach}}(\pi_f, k) = \text{FReach}(\pi_f, k)$  ( $k = 1, 2, 3, 5$ ) and  $\widehat{\text{BReach}}(\pi_f, k) = \text{BReach}(\pi_f, k)$  ( $k = 1, 2$ ).

### D. Comparison With Probabilistic Reachability

This section elaborates the difference between standard probabilistic reachability [1] and the proposed distributional reachability.

Probabilistic reachability in linear temporal logic [1] denotes the likelihood that state trajectories that reach a target set (a given set of states). In probabilistic computation tree logic (PCTL), instead, the emphasis is on the probabilistic measure associated to trajectories that satisfy given temporal requirements. For the latter case, consider an MDP  $M = (\mathbb{X}, \mathbb{U}, T)$  and a set of  $\mathbb{S} \subset \mathbb{X}$ . Then, given an initial state  $x_0 \in \mathbb{X}_0$  and a policy  $\mu$ , the probability to reach set  $\mathbb{S}$  is

$$\Pr_{x_0}^{\mu}(x \in \text{SPath}(x_0, \mu) \mid \exists k \geq 0, \text{s.t.}, x_k \in \mathbb{S})$$

where  $\text{SPath}(x_0, \mu)$  is the set of state trajectory  $x = x_0 x_1 \dots x_k x_{k+1} \dots$  starting  $x_0$  under the policy  $\mu$  that satisfies  $P^{\mu_k}(x_{k+1} | x_k) > 0, \forall k \in \mathbb{N}$ . And  $\Pr_{x_0}^{\mu}$  is the probability measure associated with the probability space over state trajectories, i.e.,  $(\text{SPath}(x_0, \mu), \sigma(\text{SPath}(x_0, \mu)), \Pr_{x_0}^{\mu})$ , where  $\sigma(\cdot)$  denotes the  $\sigma$ -algebra. In [1], it has been shown that  $\Pr_{x_0}^{\mu}$  is uniquely determined by the transition matrix  $T$ . This is different from the distributional reachability defined previously, which deals with the evolution of the transient probability distribution, rather than the behavior of state trajectories. Despite this difference, we find that some qualitative PCTL properties can be verified through distributional reachability. For example, consider a PCTL formula  $\Pr_{>0}(\Diamond \mathbb{S})$ , expressing whether from state  $x_0$ , under any policy, the probability of trajectories starting from  $x_0$  and eventually reaching  $\mathbb{S}$  is positive. This can be alternatively verified if there exists some  $k \geq 0$  such that the forward distributional reachable set  $\text{FReach}(e_{x_0}, k)$  is a subset of the

set  $\Pi_{\mathbb{S}} = \{\pi \in \mathcal{P}(\mathbb{X}) \mid \sum_{x \in \mathbb{S}} \pi(x) > 0\}$ . Further connections between PCTL model checking and distributional reachability are left for future work, see Section VIII.

The application of probabilistic reachability and distributional reachability are quite different. Whilst both of them can be used to ensure safety in general, safety constraints are defined in different ways: probabilistic reachability has been classically applied to the reach-avoid problem [20], which adds to a reachability goal a safety requirement to avoid a set of states deemed to be unsafe; distinctively, distributional reachability is useful to handle the safety constraints defined on state distributions [30], [31], [32]. We revisit variants of this problem in Section V-C.

### IV. $\mathcal{FR}$ -INVARIANT SETS

In this section, we characterize the sets invariant under the map  $\mathcal{FR}$ . Note that the maximal  $\mathcal{FR}$ -invariant set plays an important role for exploring the controllable distribution space for the MDP, i.e., the region within which all the distributions are reachable. We define the  $\mathcal{FR}$ -invariant set as follows.

**Definition 4.1:** For the MDP  $M$ , a set  $\Pi_{\text{FRInv}} \subseteq \mathcal{P}(\mathbb{X})$  is said to be  $\mathcal{FR}$ -invariant if  $\Pi_{\text{FRInv}}$  is the solution to the fixed-point equation

$$\mathcal{FR}(\Pi_{\text{FRInv}}) = \Pi_{\text{FRInv}}.$$

Definition 4.1 implies that for any  $\pi \in \Pi_{\text{FRInv}}$  and  $\mu \in \bar{\mathcal{U}}, \pi P^{\mu} \in \Pi_{\text{FRInv}}$ .

**Problem 4.1:** Given an MDP  $M$ , characterize the  $\mathcal{FR}$ -invariant set  $\Pi_{\text{FRInv}}$ .

Next, we describe how to solve Problem 4.1 through set expansions and set contractions.

#### A. Approach Based on Set Expansions

To characterize  $\mathcal{FR}$ -invariant sets through set expansions, let us first define the notion of equilibrium of an MDP.

**Definition 4.2:** A state distribution  $\bar{\pi}$  is an equilibrium of the MDP  $M$  if there exists a one-step policy  $\bar{\mu} \in \bar{\mathcal{U}}$  such that

$$\bar{\pi} = \bar{\pi} P^{\bar{\mu}}. \quad (17)$$

An equilibrium is also known as a stationary distribution in the literature [11].

**Definition 4.3:** The equilibrium set  $\mathbb{E}$  of the MDP  $M$  is the set of all equilibria of  $M$ .

A straightforward characterization of the equilibrium set  $\mathbb{E}$  is

$$\mathbb{E} = \left\{ (Q\mathbf{1})^T \in \mathcal{P}(\mathbb{X}) \mid \begin{array}{l} Q \in \mathcal{O}, \forall y \in \mathbb{X}, \\ \sum_{u \in \mathbb{U}_y} Q(y, u) = \\ \sum_{x \in \mathbb{X}} \sum_{u \in \mathbb{U}_x} T(y|x, u) Q(x, u) \end{array} \right\}. \quad (18)$$

Note that  $\mathbb{E}$  collects all the stationary distributions that can be attained by the MDP. It follows from (18) that  $\mathbb{E}$  necessarily is a polytope.

**Proposition 4.1:** The forward reachable sets from the equilibrium set  $\mathbb{E}$  have the following properties.

- 1)  $\text{FReach}(\mathbb{E}, k) \subseteq \text{FReach}(\mathbb{E}, k+1), \forall k \in \mathbb{N}$ .
- 2)  $\text{FReach}(\mathbb{E}, k)$  is a compact polytope,  $\forall k \in \mathbb{N}$ .
- 3)  $\lim_{k \rightarrow \infty} \text{FReach}(\mathbb{E}, k)$  exists and

$$\lim_{k \rightarrow \infty} \text{FReach}(\mathbb{E}, k) = \text{cl} \left( \bigcup_{k \in \mathbb{N}} \text{FReach}(\mathbb{E}, k) \right).$$



*Proof:*

- 1) According to the definition of equilibrium, it follows that:  $\bar{\pi} \in \mathcal{FR}(\bar{\pi}) \forall \bar{\pi} \in \mathbb{E}$ , which implies that  $\mathbb{E} \subseteq \mathcal{FR}(\mathbb{E})$ . Note that  $\mathcal{FR}$  is monotone: for any  $\Pi \subseteq \Pi' \subseteq \mathcal{P}(\mathbb{X})$ ,  $\mathcal{FR}(\Pi) \subseteq \mathcal{FR}(\Pi')$ . Thus, the iteration by  $\mathcal{FR}$  starting from  $\mathbb{E}$  ensures that  $\text{FReach}(\mathbb{E}, k) \subseteq \text{FReach}(\mathbb{E}, k+1)$ ,  $\forall k \in \mathbb{N}$ .
- 2) Since the initial set  $\mathbb{E}$  is a polytope, the iteration by the forward reachability map  $\mathcal{FR}$  ensures that the set  $\text{FReach}(\mathbb{E}, k)$  is a polytope,  $\forall k \in \mathbb{N}$ . The boundedness of  $\text{FReach}(\mathbb{E}, k)$  follows from that of  $\mathcal{P}(\mathbb{X})$ , since  $\mathcal{P}(\mathbb{X})$  is a simplex in  $\mathbb{R}^n$ . To prove the closure of  $\text{FReach}(\mathbb{E}, k)$ , we observe that the initial set  $\mathbb{E}$  is closed and that the map  $\mathcal{FR}$  preserves the closure. Thus, the set  $\text{FReach}(\mathbb{E}, k)$  is compact,  $\forall k \in \mathbb{N}$ .
- 3) Follows from that the sequence  $\{\text{FReach}(\mathbb{E}, k)\}_{k \in \mathbb{N}}$  is nondecreasing and the convergence of monotone set sequences [41]. ■

Denote  $\text{FReach}(\mathbb{E}, \infty) = \lim_{k \rightarrow \infty} \text{FReach}(\mathbb{E}, k)$ .

**Proposition 4.2:** For any MDP  $M$ , the infinite-time forward reachable set  $\text{FReach}(\mathbb{E}, \infty)$  is  $\mathcal{FR}$ -invariant.

*Proof:* The iteration and the convergence of the sequence  $\{\text{FReach}(\mathbb{E}, k)\}_{k \in \mathbb{N}}$  ensure that the set  $\text{FReach}(\mathbb{E}, \infty)$  is a solution to the fixed-point equation  $\Pi = \mathcal{FR}(\Pi)$ . Thus, it is  $\mathcal{FR}$ -invariant. ■

## B. Approach Based on Set Contractions

It is also possible to characterize the  $\mathcal{FR}$ -invariant set through set contractions.

**Definition 4.4:** For the MDP  $M$ , a set  $\Pi_{\text{FRInv}}^{\max} \subseteq \mathcal{P}(\mathbb{X})$  is said to be the maximal  $\mathcal{FR}$ -invariant set if it is  $\mathcal{FR}$ -invariant and all other  $\mathcal{FR}$ -invariant sets are its subsets.

We perform the forward distributional reachability analysis initialized from the whole distribution space  $\mathcal{P}(\mathbb{X})$ . Then, we have the following claims.

**Proposition 4.3:** The forward reachable sets from  $\mathcal{P}(\mathbb{X})$  have the following properties.

- 1)  $\text{FReach}(\mathcal{P}(\mathbb{X}), k+1) \subseteq \text{FReach}(\mathcal{P}(\mathbb{X}), k)$  and  $\text{FReach}(\mathcal{P}(\mathbb{X}), k)$  is nonempty,  $\forall k \in \mathbb{N}$ .
- 2)  $\text{FReach}(\mathcal{P}(\mathbb{X}), k)$  is a compact polytope,  $\forall k \in \mathbb{N}$ .
- 3)  $\lim_{k \rightarrow \infty} \text{FReach}(\mathcal{P}(\mathbb{X}), k)$  exists and

$$\lim_{k \rightarrow \infty} \text{FReach}(\mathcal{P}(\mathbb{X}), k) = \bigcap_{k \in \mathbb{N}} \text{FReach}(\mathcal{P}(\mathbb{X}), k).$$

*Proof:*

- 1) It is easy to show that  $\mathcal{FR}(\mathcal{P}(\mathbb{X})) \subseteq \mathcal{P}(\mathbb{X})$  and  $\mathcal{FR}(\Pi) \subseteq \mathcal{FR}(\Pi')$ , for any  $\Pi \subseteq \Pi' \subseteq \mathcal{P}(\mathbb{X})$ . Thus, the iteration by  $\mathcal{FR}$  starting from  $\mathcal{P}(\mathbb{X})$  ensures that  $\text{FReach}(\mathcal{P}(\mathbb{X}), k+1) \subseteq \text{FReach}(\mathcal{P}(\mathbb{X}), k)$ ,  $\forall k \in \mathbb{N}$ . Since each state has at least one admissible action,  $\mathcal{FR}(\pi)$  is nonempty for any  $\pi \in \mathcal{P}(\mathbb{X})$ . The iteration by  $\mathcal{FR}$  starting from  $\mathcal{P}(\mathbb{X})$  further implies that the set  $\text{FReach}(\mathcal{P}(\mathbb{X}), k)$  is nonempty,  $\forall k \in \mathbb{N}$ .
- 2) Since the sequence  $\{\text{FReach}(\mathcal{P}(\mathbb{X}), k)\}_{k \in \mathbb{N}}$  is monotone nonincreasing,  $\text{FReach}(\mathcal{P}(\mathbb{X}), k)$  is nonempty, and  $\mathcal{P}(\mathbb{X})$  is a simplex, we have that  $\text{FReach}(\mathcal{P}(\mathbb{X}), k)$  must be a compact polytope. This can be proven similarly to Proposition 4.1.

- 3) Follows from that the set sequence  $\{\text{FReach}(\mathcal{P}(\mathbb{X}), k)\}_{k \in \mathbb{N}}$  is nonincreasing, the convergence of monotone set sequences [41], and the compactness of  $\text{FReach}(\mathcal{P}(\mathbb{X}), k)$  (if it is not empty). ■

Let  $\text{FReach}(\mathcal{P}(\mathbb{X}), \infty) = \lim_{k \rightarrow \infty} \text{FReach}(\mathcal{P}(\mathbb{X}), k)$ .

**Proposition 4.4:** For any MDP  $M$ , the infinite-time forward reachable set  $\text{FReach}(\mathcal{P}(\mathbb{X}), \infty)$  is the maximal  $\mathcal{FR}$ -invariant set. Furthermore, it is compact.

*Proof:* The iteration and convergence of the sequence  $\{\text{FReach}(\mathcal{P}(\mathbb{X}), k)\}_{k \in \mathbb{N}}$  ensures that  $\text{FReach}(\mathcal{P}(\mathbb{X}), \infty)$  is a solution to the fixed-point equation  $\Pi = \mathcal{FR}(\Pi)$ . Thus,  $\text{FReach}(\mathcal{P}(\mathbb{X}), \infty)$  is an  $\mathcal{FR}$ -invariant set. It is also the maximal  $\mathcal{FR}$ -invariant set due to the nonincreasing feature of the sequence  $\{\text{FReach}(\mathcal{P}(\mathbb{X}), k)\}_{k \in \mathbb{N}}$ . Its compactness follows from the compactness of  $\text{FReach}(\mathcal{P}(\mathbb{X}), k)$  and the fact that the countable intersection of compact sets is also compact. ■

## C. Discussion on the Maximal $\mathcal{FR}$ -Invariant Set

Now we have that the set  $\text{FReach}(\mathcal{P}(\mathbb{X}), \infty)$  is the maximal  $\mathcal{FR}$ -invariant set, denoted by  $\Pi_{\text{FRInv}}^{\max}$ . This set is of great interest, in view of the following.

- 1) **Controllability:** Any state distribution in the maximal  $\mathcal{FR}$ -invariant set  $\Pi_{\text{FRInv}}^{\max}$  is reachable from some distribution under an admissible policy.
- 2) **End component:** Any distribution outside of  $\Pi_{\text{FRInv}}^{\max}$  is not reachable from the inside of  $\Pi_{\text{FRInv}}^{\max}$ .
- 3) **Absorbance:** For any initial distribution  $\pi_0$  and the implemented policy  $\mu$ , the state distribution will eventually enter  $\Pi_{\text{FRInv}}^{\max}$  and then stay there forever, thus it is important for the analysis of infinite-horizon problems.

In conclusion, the maximal  $\mathcal{FR}$ -invariant set can be thought of as a max end component [1] for the space of distributions of the MDP.

The following theorem provides criteria for evaluating the size of  $\Pi_{\text{FRInv}}^{\max}$  with respect to the whole state distribution space  $\mathcal{P}(\mathbb{X})$  and the equilibrium set  $\mathbb{E}$ .

**Theorem 4.1:** The following statements hold.

- a)  $\mathbb{E} \subseteq \Pi_{\text{FRInv}}^{\max} = \mathcal{P}(\mathbb{X})$  if and only if  $\mathbb{E} \subseteq \mathcal{FR}(\mathcal{P}(\mathbb{X})) = \mathcal{P}(\mathbb{X})$ .
- b)  $\mathbb{E} = \Pi_{\text{FRInv}}^{\max} = \mathcal{P}(\mathbb{X})$  if and only if  $\mathbb{E} = \mathcal{P}(\mathbb{X}) = \mathcal{FR}(\mathcal{P}(\mathbb{X}))$ .

If the fixed-point equation  $\Pi = \mathcal{FR}(\Pi)$  has a unique solution, then we have:

- c)  $\mathbb{E} \subseteq \Pi_{\text{FRInv}}^{\max} \subset \mathcal{P}(\mathbb{X})$  if and only if  $\mathbb{E} \subseteq \mathcal{FR}(\mathbb{E}) \subseteq \mathcal{FR}(\mathcal{P}(\mathbb{X})) \subset \mathcal{P}(\mathbb{X})$ ;
- d)  $\mathbb{E} = \Pi_{\text{FRInv}}^{\max} \subset \mathcal{P}(\mathbb{X})$  if and only if  $\mathbb{E} = \mathcal{FR}(\mathbb{E})$  and  $\mathcal{FR}(\mathcal{P}(\mathbb{X})) \subset \mathcal{P}(\mathbb{X})$ .

*Proof:* The assertions (a)–(b) directly follow from the definition of  $\text{FReach}(\mathcal{P}(\mathbb{X}), \infty)$  and  $\text{FReach}(\mathbb{E}, \infty)$ . Recall that both  $\text{FReach}(\mathcal{P}(\mathbb{X}), \infty)$  and  $\text{FReach}(\mathbb{E}, \infty)$  are solutions to the fixed-point equation  $\Pi = \mathcal{FR}(\Pi)$ . If  $\Pi = \mathcal{FR}(\Pi)$  has a unique solution, then  $\Pi_{\text{FRInv}}^{\max} = \text{FReach}(\mathcal{P}(\mathbb{X}), \infty) = \text{FReach}(\mathbb{E}, \infty)$ , from which the assertions (c)–(d) follow. ■

To ensure that the solution to the fixed-point equation  $\Pi = \mathcal{FR}(\Pi)$  is unique, a sufficient condition is that  $\mathcal{FR}$  is a contraction map. That is, there exists  $\lambda \in [0, 1)$  such that for any  $\pi, \pi' \in \mathcal{P}(\mathbb{X})$

$$H(\mathcal{FR}(\pi), \mathcal{FR}(\pi')) \leq \lambda \|\pi - \pi'\|.$$

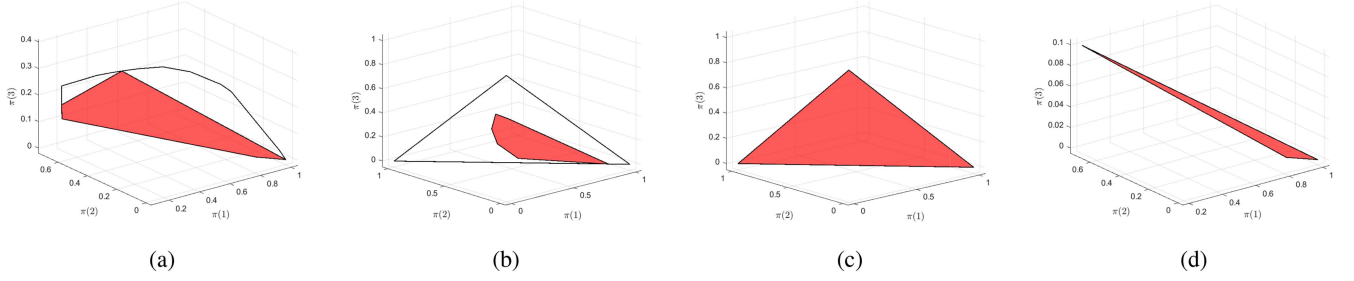


Fig. 4. Maximal  $\mathcal{FR}$ -invariant set  $\Pi_{\text{FRinv}}^{\max}$  and the equilibrium set  $\mathbb{E}$  for the four MDPs in Example 4.1. The sets with black edges and white faces are  $\Pi_{\text{FRinv}}^{\max}$  and the sets in red are  $\mathbb{E}$ . Note that for (a)  $\mathbb{E} \subset \Pi_{\text{FRinv}}^{\max} \subset \mathcal{P}(\mathbb{X})$ ; (b)  $\mathbb{E} \subset \Pi_{\text{FRinv}}^{\max} = \mathcal{P}(\mathbb{X})$ ; (c)  $\mathbb{E} = \Pi_{\text{FRinv}}^{\max} = \mathcal{P}(\mathbb{X})$ ; (d)  $\mathbb{E} = \Pi_{\text{FRinv}}^{\max} \subset \mathcal{P}(\mathbb{X})$ .

Here,  $\|\cdot\|$  denotes the vector norm and  $H(\cdot, \cdot)$  is the Hausdorff distance, defined as

$$H(\Pi, \Pi') = \max \left\{ \max_{\pi \in \Pi} \min_{\pi' \in \Pi'} \|\pi - \pi'\|, \max_{\pi' \in \Pi'} \min_{\pi \in \Pi} \|\pi - \pi'\| \right\}.$$

Note that the use of this contraction map is different from the study of contraction over sets in Section IV-B. Thanks to the completeness property of the Hausdorff distance in the space of compact sets, it then follows from the Banach fixed point theorem that  $\mathcal{FR}$  admits a unique fixed-point in the space of compact subsets of  $\mathcal{P}(\mathbb{X})$ .

*Remark 4.1:* Notice that first, the contraction condition is not necessary for many MDPs whose maximal  $\mathcal{FR}$ -invariant set is however unique [e.g., the same as  $\mathcal{P}(\mathbb{X})$ ]. Second, verifying if  $\mathcal{FR}$  is a contraction map is hard in practice, due to the difficulty related to calculating the Hausdorff distance. However, the contraction condition can be relaxed for Markov chains. It has been shown that if the stochastic transition matrix is regular, the limit distribution (the solution to  $\Pi = \mathcal{FR}(\Pi)$ ) is unique [42].

*Example 4.1:* Consider the four MDPs in Fig. 1, respectively. For these MDPs,  $\text{FReach}(\mathcal{P}(\mathbb{X}), \infty) = \text{FReach}(\mathbb{E}, \infty) = \Pi_{\text{FRinv}}^{\max}$ . The maximal  $\mathcal{FR}$ -invariant set  $\Pi_{\text{FRinv}}^{\max}$  and the equilibrium set  $\mathbb{E}$  for each MDP are shown in Fig. 4(a)–(d), where the sets with black edges and white faces are  $\Pi_{\text{FRinv}}^{\max}$  and the red sets are  $\mathbb{E}$ . Notice that these MDPs correspond to the four possible inclusion relations among  $\mathbb{E}$ ,  $\Pi_{\text{FRinv}}^{\max}$ , and  $\mathcal{P}(\mathbb{X})$ .

## V. APPLICATIONS OF BACKWARD DISTRIBUTIONAL REACHABILITY

In this section, we use backward distributional reachability to provide new solutions of three important problems: controlled invariance (see Section V-A), the characterization of domains of attraction and of escape sets (see Section V-B), and reach-avoid problems (see Section V-C). Although these three problems are different, they can be reformulated by using backward distributional reachability, which comes with computational convenience related to the proposed algorithm in Section III.

### A. Controlled Invariant Sets

We first introduce controlled invariant sets.

*Definition 5.1:* For the MDP  $M$ , a set  $\Pi_{\text{inv}} \subseteq \mathcal{P}(\mathbb{X})$  is said to be controlled invariant if for any  $\pi \in \Pi_{\text{inv}}$ , there exists a one-step policy  $\mu \in \mathcal{U}$  such that  $\pi P^\mu \in \Pi_{\text{inv}}$ .

*Remark 5.1:* The controlled invariant set in Definition 5.1 is different from the  $\mathcal{FR}$ -invariant set in Definition 4.1: the invariance property of controlled invariant sets is built on the existence

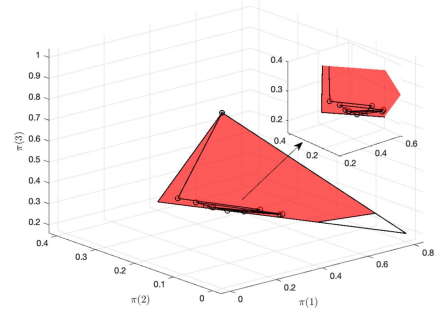


Fig. 5. Example 5.1: the maximal controlled invariant set  $\Pi_{\text{inv}}^{\max}$  (red) within  $\Pi$  (black edge and white face) and a distribution trajectory (circle line) starting from  $\pi_0 = [0 \ 0 \ 1]$  (asterisk).

of a policy, while the invariance property of  $\mathcal{FR}$ -invariant set holds under all policies. This difference suggests that computing a controlled invariant set requires backward distributional reachability, rather than forward distributional reachability. This is consistent with the methods for computing invariant sets in the control literature [38].

*Problem 5.1:* Given a compact set  $\Pi \subset \mathcal{P}(\mathbb{X})$ , find the maximal controlled invariant set  $\Pi_{\text{inv}}^{\max}$  within  $\Pi$ .

The following proposition shows that Problem 5.1 can be solved using backward distributional reachability.

*Proposition 5.1:* Given a compact set  $\Pi \subset \mathcal{P}(\mathbb{X})$

$$\Pi_{\text{inv}}^{\max} = \bigcap_{k \in \mathbb{N}} \text{IV}(\Pi, k)$$

where  $\text{IV}(\Pi, k)$  is recursively defined as

$$\text{IV}(\Pi, k+1) = \mathcal{BR}(\text{IV}(\Pi, k)) \cap \Pi \quad (19)$$

with  $\text{IV}(\Pi, 0) = \Pi$ .

*Proof:* First we have that

$$\mathcal{BR}(\Pi) \cap \Pi \subseteq \Pi \Rightarrow \text{IV}(\Pi, k+1) \subseteq \text{IV}(\Pi, k), \quad \forall k \in \mathbb{N}.$$

Since  $\Pi$  is compact, from the property of map  $\mathcal{BR}$  the set  $\text{IV}(\Pi, k)$  is compact. Thus,  $\lim_{k \rightarrow \infty} \text{IV}(\Pi, k) = \bigcap_{k \in \mathbb{N}} \text{IV}(\Pi, k)$  and it is the solution to fixed-point equation  $\Pi_{\text{inv}} = \mathcal{BR}(\Pi_{\text{inv}}) \cap \Pi$ . ■

*Example 5.1:* Consider the MDP in Fig. 1(a). Let  $\Pi = \{\pi \in \mathcal{P}(\mathbb{X}) \mid \pi(3) \geq 0.3, 2\pi(2) + \pi(3) \leq 1\}$ . The maximal controlled invariant set  $\Pi_{\text{inv}}^{\max}$  within  $\Pi$  is shown in Fig. 5. Starting from the initial distribution  $\pi_0 = [0 \ 0 \ 1]$  (asterisk), the trajectory of the transient distribution can be kept within  $\Pi_{\text{inv}}^{\max}$  under some policy, as shown by the dotted line.



## B. Domains of Attraction and Escape Sets

In this section, we show how to use backward distributional reachability to characterize the domain of attraction, as well as the escape set. Given an initial state  $x_0 \in \mathbb{X}$  and a policy  $\mu \in \mathcal{U}$ , the probability of reaching a set  $\mathbb{S} \subseteq \mathbb{X}$  at time step  $k$  is

$$\Pr_{x_0}^{\mu}(\mathbb{S}, k) = \sum_{x \in \mathbb{S}} \pi_k(x)$$

where  $\pi_k = \psi(k, e_{x_0}, \mu)$  and  $e_{x_0}$  is a unit vector with the  $x_0$ th element being 1, and all the other elements being 0.

**Definition 5.2:** The domain of attraction of a set  $\mathbb{S} \subset \mathbb{X}$ , denoted by  $\Lambda_{\mathbb{S}}$ , is the set of initial states from which the probability of reaching  $\mathbb{S}$  is positive under some policy, i.e.,

$$\Lambda_{\mathbb{S}} = \{x_0 \in \mathbb{X} \mid \exists \mu \in \mathcal{U}, \exists k \geq 0, \text{ s.t. } \Pr_{x_0}^{\mu}(\mathbb{S}, k) > 0\}.$$

The  $\alpha$ -domain of attraction of a set  $\mathbb{S} \subset \mathbb{X}$ , denoted by  $\Lambda_{\mathbb{S}}^{\alpha}$ , is the set of initial states from which the probability of reaching  $\mathbb{S}$  is no less than  $\alpha > 0$  under some policy, i.e.,

$$\Lambda_{\mathbb{S}}^{\alpha} = \{x_0 \in \mathbb{X} \mid \exists \mu \in \mathcal{U}, \exists k \geq 0, \text{ s.t. } \Pr_{x_0}^{\mu}(\mathbb{S}, k) \geq \alpha\}.$$

**Definition 5.3:** The escape set of a set  $\mathbb{S} \subset \mathbb{X}$ , denoted by  $\Gamma_{\mathbb{S}}$ , is the set of initial states from which the probability of reaching  $\mathbb{S}$  is zero under all possible policies, i.e.,

$$\Gamma_{\mathbb{S}} = \{x_0 \in \mathbb{X} \mid \Pr_{x_0}^{\mu}(\mathbb{S}, k) = 0 \forall \mu \in \mathcal{U} \text{ and } \forall k \in \mathbb{N}\}.$$

Let us define the following set of state distributions associated with  $\mathbb{S}$ :

$$\begin{aligned} \Pi_{\mathbb{S}} &= \left\{ \pi \in \mathcal{P}(\mathbb{X}) \mid \sum_{x \in \mathbb{S}} \pi(x) > 0 \right\} \\ \Pi_{\mathbb{S}}^{\alpha} &= \left\{ \pi \in \mathcal{P}(\mathbb{X}) \mid \sum_{x \in \mathbb{S}} \pi(x) \geq \alpha \right\}. \end{aligned} \quad (20)$$

The following proposition provides a characterization of the domain of attraction and of the escape set, using backward distributional reachability.

**Proposition 5.2:** Given a set  $\mathbb{S} \subset \mathbb{X}$ , the domain of attraction of  $\mathbb{S}$  can be characterized by

$$\Lambda_{\mathbb{S}} = \{x_0 \in \mathbb{X} \mid e_{x_0} \in \text{BReach}(\Pi_{\mathbb{S}})\}$$

and the  $\alpha$ -domain of attraction of  $\mathbb{S}$  is

$$\Lambda_{\mathbb{S}}^{\alpha} = \{x_0 \in \mathbb{X} \mid e_{x_0} \in \text{BReach}(\Pi_{\mathbb{S}}^{\alpha})\}.$$

The escape set of  $\mathbb{S}$  is

$$\Gamma_{\mathbb{S}} = \{x_0 \in \mathbb{X} \mid e_{x_0} \notin \text{BReach}(\Pi_{\mathbb{S}})\}.$$

*Proof:* Follows from the definition of BReach. ■

## C. Reach-Avoid

In this section, we show how to use distributional reachability analysis to solve reach-avoid problems over MDPs.

Given an initial distribution  $\pi_0$  and a policy  $\mu \in \mathcal{U}$ , the state trajectory of the MDP  $\mathcal{M}$  is a stochastic process defined over the probability space  $(\Omega, \mathcal{F}, \Pr_{\pi_0}^{\mu})$ , where  $\Omega = \mathbb{X}^{\infty} =: \mathbb{X} \times \mathbb{X} \times \dots$ ,  $\mathcal{F} = \sigma(\Omega)$  (the  $\sigma$ -algebra of  $\Omega$ ), and  $\Pr_{\pi_0}^{\mu}$  is defined by the distribution dynamics (5). Consider two disjoint sets  $\mathbb{S}, \mathbb{P} \subset \mathbb{X}$ . Define the hitting times of  $\mathbb{S}$  and  $\mathbb{P}$  as  $\tau_{\mathbb{S}} = \inf\{k \mid x_k \in \mathbb{S}\}$  and  $\tau_{\mathbb{P}} = \inf\{k \mid x_k \in \mathbb{P}\}$ , respectively. The probability of reaching  $\mathbb{S}$  while avoiding  $\mathbb{P}$  is  $\Pr_{\pi_0}^{\mu}(\tau_{\mathbb{S}} < \tau_{\mathbb{P}}, \tau_{\mathbb{S}} < \infty)$ . We will solve

two variants of the reach-avoid problem, introduced as the next two problems. The first is one infinite-horizon problem and defined via first hitting times [19], [20].

**Problem 5.2:** Find the set of initial states from which the probability of reaching  $\mathbb{S}$  while avoiding  $\mathbb{P}$  is positive under some policy or no less than  $\alpha$

$$\Upsilon_{\mathbb{S}, \mathbb{P}} = \left\{ x_0 \in \mathbb{X} \mid \exists \mu \in \mathcal{U}, \Pr_{e_{x_0}}^{\mu}(\tau_{\mathbb{S}} < \tau_{\mathbb{P}}, \tau_{\mathbb{S}} < \infty) > 0 \right\} \quad (21)$$

$$\Upsilon_{\mathbb{S}, \mathbb{P}}^{\alpha} = \left\{ x_0 \in \mathbb{X} \mid \exists \mu \in \mathcal{U}, \Pr_{e_{x_0}}^{\mu}(\tau_{\mathbb{S}} < \tau_{\mathbb{P}}, \tau_{\mathbb{S}} < \infty) \geq \alpha \right\}. \quad (22)$$

The second variant is a finite-horizon problem in the terminal hitting time [20].

**Problem 5.3:** Given an horizon  $N$  and an initial state distribution  $\pi_0$ , find the maximal probability of reaching  $\mathbb{S}$  at the terminal time step  $N$  while avoiding  $\mathbb{P}$ , i.e.,

$$r_{\pi_0, N}^*(\mathbb{S}, \mathbb{P}) = \max_{\mu \in \mathcal{U}} \Pr_{\pi_0}^{\mu}(\tau_{\mathbb{P}} > N, x_N \in \mathbb{S}).$$

**1) Characterization of  $\Upsilon_{\mathbb{S}, \mathbb{P}}$  and of  $\Upsilon_{\mathbb{S}, \mathbb{P}}^{\alpha}$ :** This section will provide a solution to Problem 5.2. Recall the set  $\Pi_{\mathbb{S}}$  in (20) and define the set of state distributions associated with  $\mathbb{P}$

$$\Pi_{\mathbb{P}} = \left\{ \pi \in \mathcal{P}(\mathbb{X}) \mid \sum_{x \in \mathbb{P}} \pi(x) = 0 \right\}. \quad (23)$$

**Proposition 5.3:** Consider the sets  $\mathbb{S}$  and  $\mathbb{P}$  as previous. Then, the following statements hold:

a)  $x_0 \in \Upsilon_{\mathbb{S}, \mathbb{P}}$  if  $e_{x_0} \in \text{RA}(\mathbb{S}, \mathbb{P})$ ;

b)  $x_0 \in \Upsilon_{\mathbb{S}, \mathbb{P}}^{\alpha}$  if  $e_{x_0} \in \text{RA}^{\alpha}(\mathbb{S}, \mathbb{P})$ ;

where  $\text{RA}(\mathbb{S}, \mathbb{P}) = \bigcup_{k \in \mathbb{N}} \text{RA}(\mathbb{S}, \mathbb{P}, k)$  and  $\text{RA}^{\alpha}(\mathbb{S}, \mathbb{P}) = \bigcup_{k \in \mathbb{N}} \text{RA}^{\alpha}(\mathbb{S}, \mathbb{P}, k)$  with

$$\text{RA}(\mathbb{S}, \mathbb{P}, k+1) = \mathcal{BR}(\text{RA}(\mathbb{S}, \mathbb{P}, k)) \cap \Pi_{\mathbb{P}}$$

$$\text{RA}^{\alpha}(\mathbb{S}, \mathbb{P}, k+1) = \mathcal{BR}(\text{RA}^{\alpha}(\mathbb{S}, \mathbb{P}, k)) \cap \Pi_{\mathbb{P}}$$

initialized as  $\text{RA}(\mathbb{S}, \mathbb{P}, 0) = \Pi_{\mathbb{S}}$  and  $\text{RA}^{\alpha}(\mathbb{S}, \mathbb{P}, 0) = \Pi_{\mathbb{S}}^{\alpha}$ .

*Proof:* Follows from the definition of BReach. ■

**2) Maximal Reach-Avoid Probability  $r_{\pi_0, N}^*(\mathbb{S}, \mathbb{P})$ :** Next let us solve Problem 5.3. According to the definition of  $\text{RA}(\mathbb{S}, \mathbb{P}, N)$ , we have that  $r_{\pi_0, N}^*(\mathbb{S}, \mathbb{P}) > 0$  if and only if  $\pi_0 \in \text{RA}(\mathbb{S}, \mathbb{P}, N)$ . This provides a way of finding the minimum horizon  $N^*$  such that  $r_{\pi_0, N^*}^*(\mathbb{S}, \mathbb{P}) > 0$ , i.e.,

$$N^* = \min \{N \in \mathbb{N} \mid \pi_0 \in \text{RA}(\mathbb{S}, \mathbb{P}, N)\}.$$

If  $\pi_0 \in \text{RA}(\mathbb{S}, \mathbb{P}, N)$ , then the maximal probability of reaching  $\mathbb{S}$  at the terminal time step  $N$  while avoiding  $\mathbb{P}$ , i.e.,  $r_{\pi_0, N}^*(\mathbb{S}, \mathbb{P})$ , is the optimum of the following LP:

$$\begin{aligned} \max_{Q_k, k \in \mathbb{N}_{0, N}} \quad & \sum_{x \in \mathbb{S}} \sum_{u \in \mathbb{U}_x} Q_N(x, u) \\ \text{s.t.} \quad & \sum_{u \in \mathbb{U}_y} Q_0(y, u) = \pi_0(y) \quad \forall y \in \mathbb{X} \end{aligned} \quad (24a)$$

$\forall k \in \mathbb{N}_{[0, N-1]} :$

$$\sum_{u \in \mathbb{U}_y} Q_{k+1}(y, u) = \sum_{x \in \mathbb{X}} \sum_{u \in \mathbb{U}_x} T(y|x, u) Q_k(x, u) \quad \forall y \in \mathbb{X} \quad (24b)$$

$$\forall k \in \mathbb{N}_{[0,N]} :$$

$$Q_k \in \mathcal{O} \quad (24c)$$

$$\sum_{x \in \mathbb{P}} \sum_{u \in \mathbb{U}_x} Q_k(x, u) = 0. \quad (24d)$$

Let  $\{Q_k^*\}_{k=0}^N$  be the optimal solution to (24). Using the property of occupation measures given in Section II, the optimal policy for Problem 5.3 is thus given by

$$\mu_k(u|x) = \begin{cases} \frac{Q_k^*(x, u)}{\sum_{v \in \mathbb{U}_x} Q_k^*(x, v)}, & \text{if } \sum_{v \in \mathbb{U}_x} Q_k^*(x, v) > 0 \\ \frac{1}{|\mathbb{U}_x|}, & \text{if } \sum_{v \in \mathbb{U}_x} Q_k^*(x, v) = 0 \text{ and } u \in \mathbb{U}_x. \end{cases}$$

## VI. CASE STUDIES

In this section, we will demonstrate how to use the proposed methods to solve a drug injection verification problem, a stochastic navigation problem, and a swarm deployment problem. The numerical experiments are run in MATLAB R2021b with YALMIP toolbox [43] and MOSEK toolbox [44] on a MacBook Pro laptop with Apple M1 chip and 8.0 GB Memory.

### A. Drug Injection Verification in Pharmacokinetics

We consider an MDP model for a pharmacokinetics system, as adapted from [4], [45], which consists of five states: plasma (Pl), interstitial fluid (IF), utilization and degradation (Ut), drug being injected (Dr), the drug being cleared (Cl), and “dummy” state (Re) (which allows to adjust the amount of drug being initially injected). As a slight deviation from the MDP model in Section II, the MDP model of the pharmacokinetics system is governed by two stochastic matrices  $P_{\text{normal}}$  and  $P_{\text{saturated}}$

$$P_{\text{normal}} = \begin{bmatrix} 0.94000 & 0.02634 & 0.02564 & 0.00780 & 0.00024 & 0 \\ 0 & 0.20724 & 0.48298 & 0.29624 & 0.01354 & 0 \\ 0 & 0.15531 & 0.42539 & 0.39530 & 0.02400 & 0 \\ 0 & 0.02598 & 0.10778 & 0.77854 & 0.0877 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

$$P_{\text{saturated}} = \begin{bmatrix} 0.9400 & 0.02425 & 0.02558 & 0.00809 & 0.00012 & 0 \\ 0 & 0.20728 & 0.48329 & 0.30257 & 0.00686 & 0 \\ 0 & 0.15540 & 0.42612 & 0.40627 & 0.01221 & 0 \\ 0 & 0.02653 & 0.11080 & 0.81776 & 0.04491 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Further introduce the set of all possible matrices in the convex combination of  $P_{\text{normal}}$  and  $P_{\text{saturated}}$ , denoted by  $\mathcal{S} = \{P \mid P = \lambda P_{\text{normal}} + (1 - \lambda) P_{\text{saturated}}, \lambda \in [0, 1]\}$ . The MDP selects nondeterministically any matrix within set  $\mathcal{S}$ : in other words, given an initial distribution  $\pi_0$ , the dynamics are  $\pi_{k+1} = \pi_k P_k$  for  $k > 0$ , where matrices can be selected as  $P_k \in \mathcal{S}$  at any time index  $k$ . The initial distribution is defined by  $\pi_0(\text{Dr}) = \alpha$ ,  $\pi_0(\text{Re}) = 1 - \alpha$ , and  $\pi_0(x) = 0$  for  $x \in \{\text{Pl}, \text{IF}, \text{Cl}, \text{Ut}\}$ , where  $\alpha$  is the amount of drug being initially injected. Following [4], [45], we set thresholds MEC =

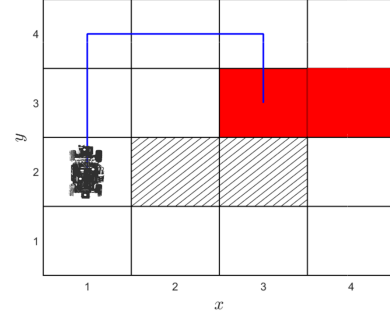


Fig. 6. Mobile robot in a grid world. Red colored blocks are target states and shadowed blocks are obstacles. In blue a feasible trajectory is shown.

0.13 and MTC = 0.20, and consider the set of atomic propositions  $\mathcal{AP}_d = \{\text{effective}, \text{nontoxic}, \text{cleared}\}$ . Considering  $\pi_k(\text{Ut})$ , namely the probability of the drug being in state Ut at time  $k$ , we define label effective as  $\pi_k(\text{Ut}) \geq \text{MEC}$ ; nontoxic as  $\pi_k(\text{Ut}) \geq \text{MTC}$ ; and cleared as  $\pi_k(\text{Cl}) \leq \epsilon$  for some given (small) value  $\epsilon > 0$ . The distribution-specified linear temporal logic formula of interest is  $\phi = \phi_1 \wedge \phi_2 \wedge \phi_3$ , with  $\phi_1 = \Box \text{nontoxic}$ ,  $\phi_2 = \Diamond(\text{effective} \wedge \bigcirc \text{effective})$ ,  $\phi_3 = \Diamond \text{cleared}$ . Here,  $\phi_1$  encodes the requirement that the drug level always stays in the safe zone;  $\phi_2$  stipulates the drug is eventually effective for at least two consecutive steps; and  $\phi_3$  specifies that the drug ought to be eventually cleared. More details on temporal logics refer to [1]. The problem of interest is that given the initial distribution  $\pi_0$  depending on  $\alpha$ , verify if the formula  $\phi$  is true for all the possible executions under the MDP model.

In [4] and [45], similar specifications were expressed as  $\omega$ -regular expressions and shown to be decidable over the MDP. However, these two works did not provide any model checking algorithm. The forward reachability in our work provides a way to solve the model checking problem. Let us define the distribution sets  $\Pi_{\text{nontoxic}} = \{\pi \in \mathcal{P}(\mathbb{X}) \mid \pi(\text{Ut}) \geq \text{MTC}\}$ ,  $\Pi_{\text{effective}} = \{\pi \in \mathcal{P}(\mathbb{X}) \mid \pi(\text{Ut}) \geq \text{MEC}\}$ , and  $\Pi_{\text{cleared}} = \{\pi \in \mathcal{P}(\mathbb{X}) \mid \pi(\text{Cl}) \leq \epsilon\}$ . Then, we have

- 1)  $\phi_1$  is true if  $\mathcal{FR}(\pi_0, k) \subseteq \Pi_{\text{nontoxic}}$  for all  $k \in \mathbb{N}$ ;
- 2)  $\phi_2$  is true if  $\mathcal{FR}(\pi_0, k) \subseteq \Pi_{\text{effective}}$  and  $\mathcal{FR}(\pi_0, k + 1) \subseteq \Pi_{\text{effective}}$  for some  $k \in \mathbb{N}$ ;
- 3)  $\phi_3$  is true if  $\mathcal{FR}(\pi_0, k) \subseteq \Pi_{\text{cleared}}$  for some  $k \in \mathbb{N}$ .

Based on these, we find that if  $\alpha \in [0.0540, 0.0590]$ , the formula  $\phi$  is true.

### B. Stochastic Navigation

We consider the stochastic navigation problem shown in Fig. 6, where a mobile robot moves in a  $4 \times 4$  grid world and has four possible actions: left, right, up, and down. Under each action, a transition to the chosen direction occurs with probability 0.80, whereas a transitions to each adjacent state in the chosen direction occurs with probability 0.10 (this is also known as “slippery” grid world). If a transition toward is elicited, then the robot remains in the present state. The states (3,3) and (4,3) are target states (the red grids in Fig. 6) and the states (2,2) and (3,2) are obstacles (the shadowed blocks in Fig. 6). The state (3,3) has only action “right,” which may induce a transition to (4,3) with probability 0.95 and another transition to itself with probability 0.05. The states (2,2) and (3,2) are absorbing

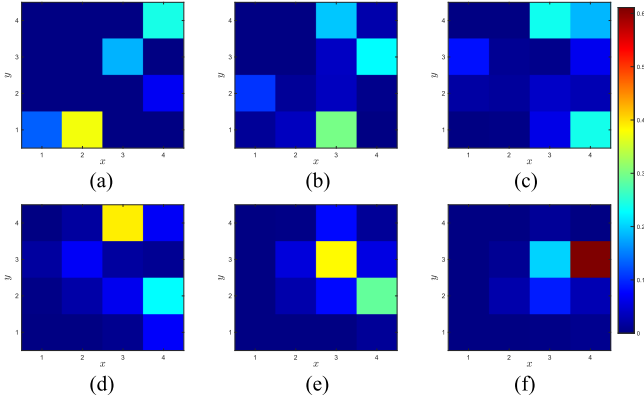


Fig. 7. Evolution of state distributions from an initial distribution such that, under a feasible policy, the probability of reaching  $\mathcal{S} = \{(3,3), (4,3)\}$  within five steps is greater than 0.80. (a)  $\pi_0$ . (b)  $\pi_1$ . (c)  $\pi_2$ . (d)  $\pi_3$ . (e)  $\pi_4$ . (f)  $\pi_5$ .

states, i.e., these states are invariant under all the actions. The stochastic navigation problem can be studied over an MDP model, where the state space  $\mathbb{X}$  is the set of the blocks in the grid world, the action space  $\mathbb{U} = \{\text{left, right, up, and down}\}$ , and the transition probability is defined based on the above-mentioned semantics. In the following, let  $\mathcal{S} = \{(3,3), (4,3)\}$  and  $\mathbb{P} = \{(2,2), (3,2)\}$ .

We first explore the domain of attraction for the target region by recursively implementing Algorithm 1 with 300 sample points. Using the characterization method in Section V-B, the set  $\Lambda_{\mathcal{S}}^{\alpha}$ , i.e., the  $\alpha$ -domain of attraction with  $\alpha = 0.80$ , is  $\Lambda_{\mathcal{S}}^{\alpha} = \{(1,2), (1,3), (1,4), (2,3), (2,4), (3,1), (3,3), (3,4), (4,1), (4,2), (4,3), (4,4)\}$ . The average time taken to compute  $\Lambda_{\mathcal{S}}^{\alpha}$ , averaged over 100 runs, is 65.82 s. Next, using the method in Section V-C, we implement Algorithm 1 (this time with only 30 samples) to characterize the set  $\Upsilon_{\mathcal{S},\mathbb{P}}^{\alpha}$  with  $\alpha = 0.80$ . We find that the states (1,4), (2,3), (2,4), (3,3), (3,4), (4,2), (4,3), and (4,4) are initial states from which the probability that the robot will eventually reach the target region  $\mathcal{S}$  while avoiding the obstacle region  $\mathbb{P}$  is no less than 0.80. That is, these states belong to  $\Upsilon_{\mathcal{S},\mathbb{P}}^{\alpha}$ . Since the number of samples used is much smaller than that employed earlier to compute the domain of attraction  $\Lambda_{\mathcal{S}}^{\alpha}$ , we obtain that the average time (over 100 runs) for computing  $\Upsilon_{\mathcal{S},\mathbb{P}}^{\alpha}$  is reduced to 4.10 s.

In addition to the computation of  $\Lambda_{\mathcal{S}}^{\alpha}$  and  $\Upsilon_{\mathcal{S},\mathbb{P}}^{\alpha}$ , our methods based on backward distributional reachability are able to also provide other initial distributions from which there exists a policy such that the specifications hold. An initial distribution that satisfies the problem of the domain of attraction (from previous), under a feasible policy, is exemplified in Fig. 7(a). Notice that, while the states (1,1) and (2,1) are not in the set  $\Lambda_{\mathcal{S}}^{\alpha}$  with  $\alpha = 0.80$ , the selected initial distribution in Fig. 7(a) assigns a positive probability to these two states. A feasible policy is obtained by solving the LP in (24) (without the constraints (24e) that are needed for the “avoid” requirement). The computation time taken to synthesize the feasible policy amounts to 1.70 s. Fig. 7(b)–(f) shows the evolution of the corresponding state distributions across the five-step time horizon. We additionally report that the probability of reaching  $\mathcal{S}$  assigned by the distribution  $\pi_5$  is 0.81, which is greater than the required  $\alpha = 0.80$ .

Similarly, an initial satisfying distribution for the reach-avoid problem is exemplified in Fig. 8(a). While the states (1,2) and

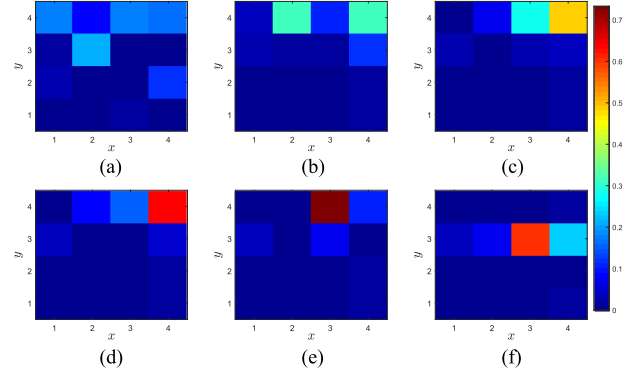


Fig. 8. Evolution of state distributions from an initial distribution such that, under a feasible policy, the probability of reaching  $\mathcal{S} = \{(3,3), (4,3)\}$  within five steps, while avoiding  $\mathbb{P} = \{(2,2), (3,2)\}$ , is greater than 0.80. (a)  $\pi_0$ . (b)  $\pi_1$ . (c)  $\pi_2$ . (d)  $\pi_3$ . (e)  $\pi_4$ . (f)  $\pi_5$ .

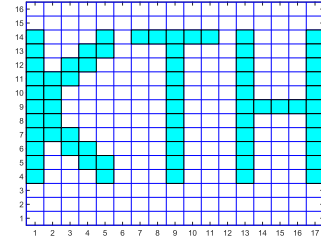


Fig. 9. Swarm deployment environment and target KTH.

(3,1) are not in the set  $\Upsilon_{\mathcal{S},\mathbb{P}}^{\alpha}$  with  $\alpha = 0.80$ , the initial distribution in Fig. 8(a) assigns a positive probability to these two states. Fig. 8(b)–(f) shows the evolution of the state distributions under the policy obtained by solving the LP in (24). The computation time to synthesize a feasible policy amounts to 1.60 s. We report that the probability of reaching  $\mathcal{S}$  assigned by the distribution  $\pi_5$  is 0.8390, which is greater than  $\alpha = 0.80$ ; moreover, as clear from the figures, we obtain that at each time step, the probability of colliding with the obstacles (i.e., the probability assigned by  $\pi_k$ ,  $k \in \mathbb{N}_{[0,5]}$ , to the set  $\mathbb{P}$ ) is actually equal to 0.

### C. Swarm Deployment Problem

Let us consider  $M$  agents evenly spaced over a  $17 \times 16$  grid world, as in Fig. 9. Initially, each square in the grid has one agent and thus the total number of agents is  $M = 17 \times 16 = 272$ . Each agent has five possible actions: left, right, up, down, and stay. We assume that all the agents sharing a given square will select the same action. Under each action in  $\{\text{left, right, up, down}\}$ , the agent moves to the adjacent square in the chosen direction with a probability 0.70, or stay in the same square with a probability 0.30. Under the action “stay,” the agent remains in the square with a probability equal to 1. To prevent excessive clustering of agents, we require that the expected number of agents in each square is at most 6, at any time step. The target of the problem is to provide a swarm deployment that takes the shape of KTH, as shown in Fig. 9. More specifically, KTH is deemed to be formed if more than 90% agents are eventually deployed over the squares of the acronym, and if the expected number of agents in each square is no less than 2 (this enforces agents to “spread” over the desired region).

This swarm deployment problem can be studied over an MDP model. The state space  $\mathbb{X}$  of the MDP corresponds to the set of the



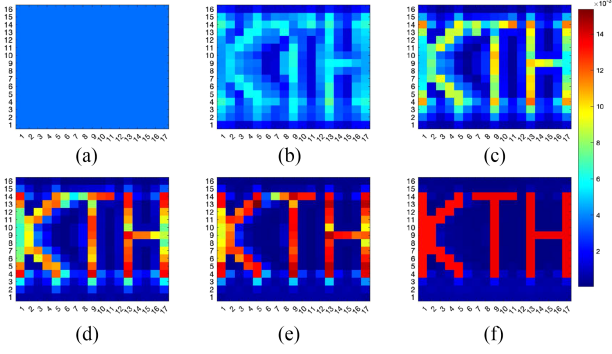


Fig. 10. Evolution of state distributions from an initial one such that, under a feasible policy, the formation of KTH is attained. (a)  $\pi_0$ . (b)  $\pi_1$ . (c)  $\pi_2$ . (d)  $\pi_3$ . (e)  $\pi_4$ . (f)  $\pi_5$ .

squares in the grid world (with dimension 272), the control space is  $\mathbb{U} = \{\text{left, right, up, down, stay}\}$ , and the transition probability  $T$  is defined according to the abovementioned semantics. The initial distribution is uniform, i.e.,  $\pi_0 = 1/M \times \mathbf{1}$ , where  $\mathbf{1}$  is a vector of 1's in  $\mathbb{R}^M$ . The space limitation associated to each square can be encoded as a constraint on the state distributions  $\pi_k$ , namely  $\pi_k \in \Pi_s$  where  $\Pi_s = \{\pi \in \mathcal{P}(\mathbb{X}) \mid \pi(x) \leq 6/M, \forall x \in \mathbb{X}\}$ . Denote by  $\mathbb{X}_f$  the set of squares corresponding to KTH of Fig. 9. The swarm deployment objective can then be expressed as the goal of steering the state distribution to the set  $\Pi_f = \{\pi \in \mathcal{P}(\mathbb{X}) \mid \sum_{x \in \mathbb{X}_f} \pi(x) \geq 0.90, \pi(x) \geq 2/M, \forall x \in \mathbb{X}_f\}$ . Notice that this objective is qualitatively quite different than that considered in the stochastic navigation problem.

In order to solve the swarm deployment problem using the MDP formulation, let us perform backward distributional reachability, namely compute  $\text{SD}(k+1) = \mathcal{BR}(\text{SD}(k)) \cap \Pi_s$ , initialized as  $\text{SD}(0) = \Pi_f$ . We recursively run Algorithm 1 to compute the backward reachable sets over the 272-dimensional state space, until the considered initial uniform distribution  $\pi_0$  belongs to  $\text{SD}(k)$ . We use 500 samples in Algorithm 1. Letting  $N = \min\{k \in \mathbb{N} \mid \pi_0 \in \text{SD}(k)\}$ , in this example we find that  $N = 5$ , which implies that the swarm objective can be achieved from  $\pi_0$  within five time steps. The average reach set computation time is 154.03 s. Given the horizon  $N = 5$ , we can thus formulate an optimization problem similar to (24) and compute a satisfying policy and corresponding state distributions. The computation time taken to synthesize a feasible policy is 4.70 s. Fig. 10 shows the evolution of the state distribution in time, from which we can see that the distribution constraints  $\Pi_s$  are satisfied, and that the KTH is formed over the distribution space.

We now empirically implement the obtained policy by sampling 1000 realizations of the model under the feasible policy. Each realization records the number of agents in each square over the horizon  $N = 5$ , which is an integer taking values from 0 to  $M = 272$ . We then take the average over 1000 realizations and obtain an empirical mean of the number of agents in each square at each time step, which is displayed in Fig. 11 across the five time steps. We can see that each square is occupied at most by four agents at each time step, which is less than the required max of six agents, and that the agents swarm to form the KTH, as required.

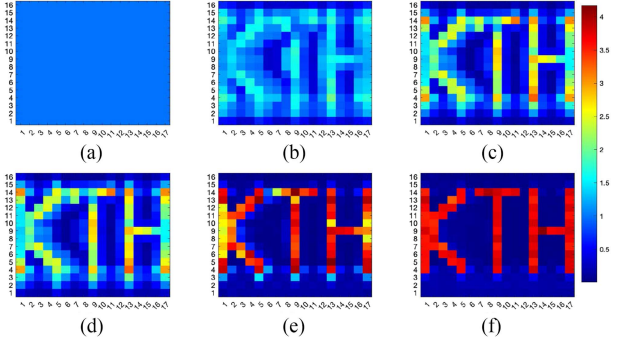


Fig. 11. Evolution of the average number of the agents per cell across time, obtained over 1000 realizations. (a)  $\bar{n}_0$ . (b)  $\bar{n}_1$ . (c)  $\bar{n}_2$ . (d)  $\bar{n}_3$ . (e)  $\bar{n}_4$ . (f)  $\bar{n}_5$ .

## VII. SCALABILITY AND APPROXIMATION QUALITY OF ALGORITHM 1

We test the scalability and approximation quality of Algorithm 1 under different MDPs, with increasing number of states  $n = 25, 100, 400, 1225$ , and a fixed number of actions  $m = 5$ . We run Algorithm 1 in MATLAB R2021b with YALMIP toolbox [43] and MOSEK toolbox [44] on a MacBook Pro laptop with Apple M1 chip and 8.0 GB Memory. We compute forward and backward reachable sets over multiple runs. Recall that the computation of reachable set is based on the polytope projection from  $\mathbb{R}^{nm+n}$  to  $\mathbb{R}^n$  [see (6) and (9)]. To the best of our knowledge, known tools in computational geometry, e.g., MPT3 [46], Qhull [47], and bensolve [48], are not usable to handle the cases in our experiments with spatial dimensions  $n + nm \geq 150$ .

Denote by  $\mathcal{FR}$  and  $\mathcal{BR}$  the exact forward and backward sets, respectively, and by  $\widehat{\mathcal{FR}}_{N_s}$  and  $\widehat{\mathcal{BR}}_{N_s}$  the corresponding approximate sets, respectively, from Algorithm 1 using  $N_s$  samples. Since computing the volume of the convex polytopes efficiently in high-dimensional spaces is hard, we define the following quantity to measure the approximation quality:

$$\rho_1 = \frac{1}{|\mathcal{I}_1|} \sum_{i \in \mathcal{I}_1} \frac{d_{1,i}^{\max} - d_{1,i}^{\min}}{D_{1,i}^{\max} - D_{1,i}^{\min}}$$

$$\rho_2 = \frac{1}{|\mathcal{I}_2|} \sum_{i \in \mathcal{I}_2} \frac{d_{2,i}^{\max} - d_{2,i}^{\min}}{D_{2,i}^{\max} - D_{2,i}^{\min}}$$

where

$$\begin{cases} d_{1,i}^{\max} = \max_{\pi \in \widehat{\mathcal{FR}}_{N_s}} e_i^T \pi, d_{2,i}^{\max} = \max_{\pi \in \widehat{\mathcal{BR}}_{N_s}} e_i^T \pi \\ d_{1,i}^{\min} = \min_{\pi \in \widehat{\mathcal{FR}}_{N_s}} e_i^T \pi, d_{2,i}^{\min} = \min_{\pi \in \widehat{\mathcal{BR}}_{N_s}} e_i^T \pi \\ D_{1,i}^{\max} = \max_{\pi \in \mathcal{FR}} e_i^T \pi, D_{2,i}^{\max} = \max_{\pi \in \mathcal{BR}} e_i^T \pi \\ D_{1,i}^{\min} = \min_{\pi \in \mathcal{FR}} e_i^T \pi, D_{2,i}^{\min} = \min_{\pi \in \mathcal{BR}} e_i^T \pi \end{cases}$$

and where  $e_i$  is a vector with the  $i$ th element equal to 1 and all the others set to 0. The index sets  $\mathcal{I}_1$  and  $\mathcal{I}_2$  are

$$\mathcal{I}_1 = \{i \in \mathbb{N}_{[1,n]} \mid D_{1,i}^{\min} \neq D_{1,i}^{\max}\}$$

$$\mathcal{I}_2 = \{i \in \mathbb{N}_{[1,n]} \mid D_{2,i}^{\min} \neq D_{2,i}^{\max}\}.$$

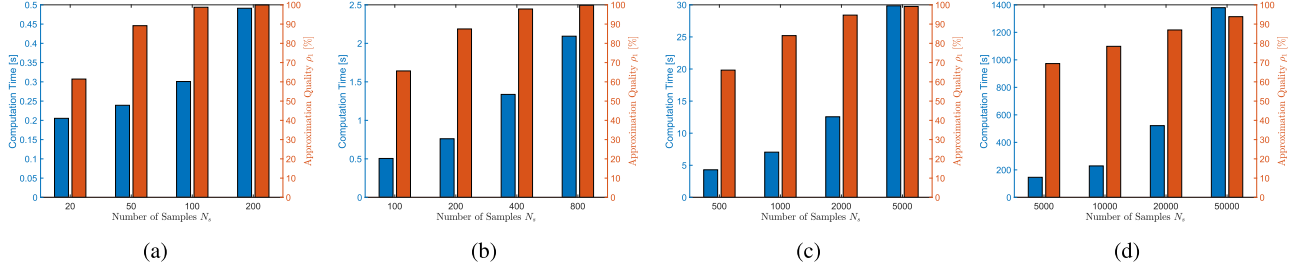


Fig. 12. Computation time and approximation quality  $\rho_1$  for forward reachable sets of different MDPs with increasing number of samples  $N_s$ . Here,  $n = \#$  states,  $n + nm =$  space dimension. (a)  $n = 25$ ,  $n + nm = 150$ . (b)  $n = 100$ ,  $n + nm = 500$ . (c)  $n = 400$ ,  $n + nm = 2400$ . (d)  $n = 1225$ ,  $n + nm = 7350$ .

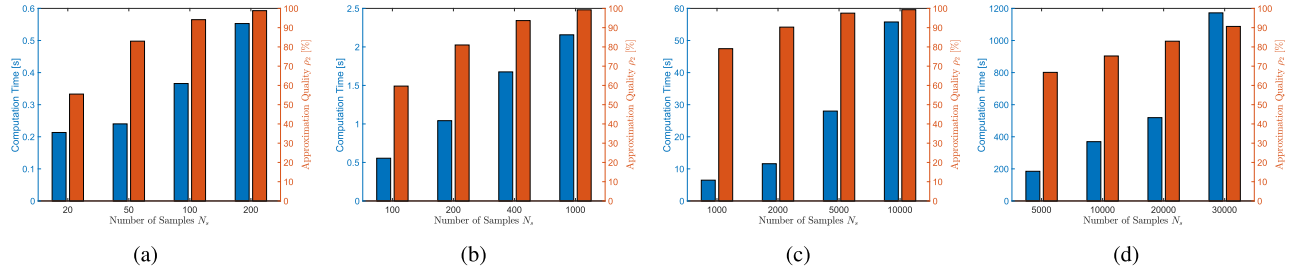


Fig. 13. Computation time and approximation quality  $\rho_2$  for backward reachable sets of different MDPs with increasing number of samples  $N_s$ . Here,  $n = \#$  states,  $n + nm =$  space dimension. (a)  $n = 25$ ,  $n + nm = 150$ . (b)  $n = 100$ ,  $n + nm = 500$ . (c)  $n = 400$ ,  $n + nm = 2400$ . (d)  $n = 1225$ ,  $n + nm = 7350$ .

The set  $\prod_{i=1}^n [D_{1,i}^{\min}, D_{1,i}^{\max}]$  is the smallest hyperrectangle that contains  $\mathcal{FR}$ , while  $\prod_{i=1}^n [d_{1,i}^{\min}, d_{1,i}^{\max}]$  is the smallest hyperrectangle that contains  $\widehat{\mathcal{FR}}_{N_s}$ . Similarly, the set  $\prod_{i=1}^n [D_{2,i}^{\min}, D_{2,i}^{\max}]$  is the smallest hyperrectangle that contains  $\mathcal{BR}$ , while  $\prod_{i=1}^n [d_{2,i}^{\min}, d_{2,i}^{\max}]$  is the smallest hyperrectangle that contains  $\widehat{\mathcal{BR}}_{N_s}$ .

The value of  $\rho_1$  and  $\rho_2$  quantify the average ratios of each edge of the corresponding two hyperrectangles, respectively, and is thus a good measure for the approximation quality for the forward and backward reachable sets from Algorithm 1. In particular,  $\rho_1$  and  $\rho_2$  can be efficiently computed, unlike the computation of volumes in high-dimensional spaces. Note that, if the sets  $\mathcal{FR}$  and  $\mathcal{BR}$  are not hyperrectangles,  $\rho_1 = 1$  and  $\rho_2 = 1$  is a necessary condition on tight approximation, that is,  $\widehat{\mathcal{FR}}_{N_s} = \mathcal{FR}$  and  $\widehat{\mathcal{BR}}_{N_s} = \mathcal{BR}$ .

Figs. 12 and 13 report the computation time and corresponding value of  $\rho$  for different MDPs, under different  $N_s$ . The quadratic programs in Algorithm 1 were solved by Yalmip [43] and Mosek [44]. Note that Algorithm 1 is able to perform set projections on spaces with a dimension up to  $nm + n = 2400$  very efficiently, and for 7350-dimensional spaces in a manageable time. We observe that the computation time is linear with respect to  $N_s$ . The approximation quality  $\rho_1$  and  $\rho_2$  increases with respect to  $N_s$ . In particular,  $\rho_1$  (or  $\rho_2$ ) can reach (almost) 1 for the MDPs with  $n = 25$ , 100, and 400, that is, we obtain a tight approximation between  $\mathcal{FR}$  and  $\widehat{\mathcal{FR}}_{N_s}$  (or between  $\mathcal{BR}$  and  $\widehat{\mathcal{BR}}_{N_s}$ ). For the MDP with  $n = 1225$ ,  $\rho$  can reach 90%, which implies a good approximation in a high-dimensional space. Thus, Algorithm 1 significantly expands the frontiers of the state of the art in reachability analysis: namely, it scales much better, and provides remarkably high-quality approximations

in high-dimensional spaces, which validates its usability for large-scale MDPs.

## VIII. CONCLUSION

We have investigated the forward and backward distributional reachability problems of finite MDPs. In order to solve the problems, we introduced the forward set-valued map  $\mathcal{FR}$  able to collect all the state distributions that can be reached from a set of initial distributions and  $\mathcal{BR}$  able to collect all the state distributions that can reach a set of final distributions. We have proved that there exists a maximal  $\mathcal{FR}$ -invariant set, which is the region where the state distributions eventually always belong to for any initial state distribution and any policy. We have revisited a number of important problems using  $\mathcal{BR}$ : controlled invariance, domain of attraction, and reach-avoid problems. Several examples have illustrated the effectiveness of our approach and its computational advantage.

There are quite a few directions that can be targeted as future work. A major extension is to tailor the forward distributional reachability approach to general PCTL specifications. We aim to reformulate trajectory-based probabilistic reachability as distributional reachability. Another interesting direction is the computation of quantities related to distributional reachability analysis with sample-based algorithms, with the goal of improved efficiency.

## APPENDIX PROOF OF THEOREM 3.1

To prove the approximation is asymptotically tight in probability, i.e., (15) and (16), we need to characterise the probability of samples in the set  $\Gamma$  whose projections onto the set

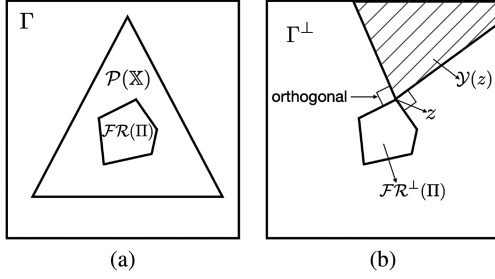


Fig. 14. (a) Sets  $\mathcal{FR}(\Pi) \subset \mathcal{P}(\mathbb{X}) \subset \text{int}(\Gamma)$  in row vector space. (b) Sets  $\mathcal{FR}^\perp(\Pi)$ ,  $\Gamma^\perp$ , and  $\mathcal{V}(z)$  [defined in (26)] where  $z$  is a vertex of  $\mathcal{FR}^\perp(\Pi)$ .

$\mathcal{FR}(\Pi)$  (or  $\mathcal{BR}(\Pi)$ ) are their vertices. Due to the uniform sampling in  $\Gamma$ , this is equivalent to characterise the subset of  $\Gamma$  whose projections onto  $\mathcal{FR}(\Pi)$  [or  $\mathcal{BR}(\Pi)$ ] are their vertices, which is addressed using the multiparametric quadratic program (mp-QP) in the following.

Consider  $\mathcal{FR}(\Pi) \subset \mathcal{P}(\mathbb{X}) \subset \text{int}(\Gamma)$  as shown in Fig. 14(a). Since the set  $\Pi$  is a convex polytope in the row vector space, it follows from (6) that  $\mathcal{FR}(\Pi)$  is also a convex polytope in the row vector space. Let  $\Gamma^\perp = \{y \mid y^T \in \Gamma\}$  and  $\mathcal{FR}^\perp(\Pi) = \{z \mid z^T \in \mathcal{FR}(\Pi)\}$ . The set  $\Gamma^\perp$  and  $\mathcal{FR}^\perp(\Pi)$  are convex polytopes in the column vector space, as shown in Fig. 14(b). Thus, we can denote  $\Gamma^\perp = \{y \in \mathbb{R}^n \mid Py \leq p\}$  and  $\mathcal{FR}^\perp(\Pi) = \{z \in \mathbb{R}^n \mid Qz \leq q\}$ , where  $P \in \mathbb{R}^{l_p \times n}$ ,  $p \in \mathbb{R}^{l_p}$ ,  $Q \in \mathbb{R}^{l_q \times n}$ , and  $q \in \mathbb{R}^{l_q}$ . Here,  $l_p$  and  $l_q$  are the number of half-spaces to define the sets  $\Gamma^\perp$  and  $\mathcal{FR}^\perp(\Pi)$ .

Consider an mp-QP

$$\begin{cases} \min_z \|y - z\|^2 \\ \text{s.t. } Qz \leq q \end{cases} \quad (25)$$

where  $z \in \mathcal{FR}^\perp(\Pi)$  is the decision variable and  $y \in \Gamma^\perp$  is the parameter. Given a  $z \in \mathcal{FR}^\perp(\Pi)$ , let  $\mathcal{V}(z)$  be the subset of  $\Gamma^\perp$  such that the optimal solution of mp-QP (25) is  $z$  for all  $y \in \mathcal{V}(z)$ .

Consider a  $z$  on the boundary of  $\mathcal{FR}^\perp(\Pi)$ . Let  $\mathcal{A}_z = \{i \in \mathbb{N}_{[1, l_q]} \mid Q^{(i)}z = q^{(i)}\}$ , where  $Q^{(i)}$  is the  $i$ th row of  $Q$  and  $q^{(i)}$  is the  $i$ th element of  $q$ . Let  $Q^{A_z}$  and  $q^{A_z}$  be the matrix and vector formed from  $Q$  and  $q$ , respectively, according to  $\mathcal{A}_z$ . The following lemma shows how to characterize  $\mathcal{V}(z)$ .

**Lemma A.1:** For any  $z$  on the boundary of  $\mathcal{FR}^\perp(\Pi)$ , assume that the matrix  $Q^{A_z}$  has full row rank. Then

$$\mathcal{V}(z) = \left\{ y \in \mathbb{R}^n \mid \begin{cases} Q(Q^{A_z})^T \Lambda (q^{A_z} - Q^{A_z}y) + Qy \leq q, \\ \Lambda^{-1} (q^{A_z} - Q^{A_z}y) \leq 0, \\ Py \leq p \end{cases} \right\} \quad (26)$$

where  $\Lambda = [Q^{A_z}(Q^{A_z})^T]^{-1}$ .

*Proof:* The mp-QP (25) can be solved by applying the Karush–Kuhn–Tucker conditions

$$z - y + Q^T \lambda = 0 \lambda \in \mathbb{R}^{l_q} \quad (27)$$

$$\lambda^{(i)} (Q^{(i)}z - q^{(i)}) = 0, \quad i = 1, \dots, l_q \quad (28)$$

$$\lambda \geq 0 \quad (29)$$

$$Qz - q \leq 0 \quad (30)$$

where  $\lambda$  is the nonnegative Lagrange multiplier. Since  $Q^{A_z}z = q^{A_z}$  and  $Q^{A_z}$  has full row rank, it follows from (28) to (29) that:

$$\lambda^{A_z} = \Lambda (Q^{A_z}y - q^{A_z}) \quad (31)$$

which implies that

$$z(y) = -(Q^{A_z})^T \lambda^{A_z} + y. \quad (32)$$

Then, the set  $\mathcal{V}(z)$  is characterized by substituting (32) and (33) into (30)–(31), which yields (26). ■

If  $z$  is a vertex of  $\mathcal{FR}^\perp(\Pi)$ , a graphical illustration of  $\mathcal{V}(z)$  is shown in Fig. 14(b).

*Proof of Theorem 3.1:* Let  $\mathcal{V}^f$  be the set of vertices and  $N_v^f$  be the number of its vertices for the set  $\mathcal{FR}(\Pi)$ . First of all, we have that

$$\widehat{\mathcal{FR}}_{N_s}(\Pi) = \mathcal{FR}(\Pi) \text{ if and only if } \mathcal{V}^f \subseteq \{\pi_i^{fs}, i \in \mathbb{N}_{[1, N_s]}\} \quad (33)$$

where  $\pi_i^{fs}$  are the corresponding projection of the sample  $\pi_i^s$ .

In order to evaluate the probability of  $\widehat{\mathcal{FR}}_{N_s}(\Pi) = \mathcal{FR}(\Pi)$ , we need to quantify the probability of  $\pi_i^{fs} \in \mathcal{V}^f$ . Equivalently, we need to find the subset of  $\Gamma^\perp$  from which the projection onto  $\mathcal{FR}^\perp(\Pi)$  is a vertex of  $\mathcal{FR}^\perp(\Pi)$ , whose transpose is a vertex of  $\mathcal{FR}(\Pi)$ .

For any vertex  $z$  of  $\mathcal{FR}^\perp(\Pi)$ , we have  $Q^{A_z}z = q^{A_z}$  and  $Q^{A_z} \in \mathbb{R}^{n \times n}$  has full rank. Then, the set  $\mathcal{V}(z)$  becomes

$$\mathcal{V}(z) = \left\{ y \in \mathbb{R}^n \mid \begin{cases} [(Q^{A_z})^T]^{-1} (y - z) \leq 0, \\ Py \leq p \end{cases} \right\}. \quad (34)$$

Since  $\mathcal{FR}(\Pi) \subset \mathcal{P}(\mathbb{X}) \subset \text{int}(\Gamma)$ ,  $z$  lies in the interior of  $\Gamma^\perp$ . Thus, we have that  $\mathcal{V}(z)$  has the following properties.

- 1)  $\mathcal{V}(z)$  is nonempty and in particular  $z$  is its vertex.
- 2)  $\mathcal{V}(z)$  has the same dimension with  $\Gamma^\perp$  and has nonempty interior.

According to Algorithm 1, let us assign  $\mu$  to be a uniform probability measure over  $\Gamma^\perp$ , i.e.,  $\int_{\Gamma^\perp} \mu(t) dt = 1$ . Thanks to the properties of  $\mathcal{V}(z)$ , we have

$$0 < \alpha(z) = \int_{\mathcal{V}(z)} \mu(t) dt \leq 1.$$

For a vertex  $\pi$  of  $\mathcal{FR}(\Pi)$ ,  $\alpha(\pi^T)$  is the probability that the projection of samples from  $\Gamma$  onto  $\mathcal{FR}(\Pi)$  is  $\pi$ .

Note that for two different vertices  $\pi_1$  and  $\pi_2$  of  $\mathcal{FR}(\Pi)$ , the interiors of  $\mathcal{V}(\pi_1^T)$  and  $\mathcal{V}(\pi_2^T)$  are disjoint. Then, under Algorithm 1, we have that

$$\begin{aligned} \Pr(\widehat{\mathcal{FR}}_{N_s}(\Pi) = \mathcal{FR}(\Pi)) \\ = \Pr(\mathcal{V}^f \subseteq \{\pi_i^{fs}, i \in \mathbb{N}_{[1, N_s]}\}) \end{aligned}$$



$$\begin{aligned} &\geq 1 - \sum_{\pi \in \mathcal{V}_f} \Pr\left(\pi \notin \{\pi_i^{fs}, i \in \mathbb{N}_{[1, N_s]}\}\right) \\ &= 1 - \sum_{\pi \in \mathcal{V}_f} (1 - \alpha(\pi^T))^{N_s} \end{aligned} \quad (35)$$

$$\geq 1 - N_v^f \alpha_f^{N_s} \quad (36)$$

where  $\alpha_f = \max_{\pi \in \mathcal{V}_f} \{1 - \alpha(\pi^T)\}$ . Since  $0 < \alpha(\pi^T) \leq 1$  for all  $\pi \in \mathcal{V}_f$ , we have that  $0 \leq \alpha_f < 1$ . Now we complete the proof of (15).

Following the similar steps, we can also prove (16).

## REFERENCES

- [1] C. Baier and J.-P. Katoen, *Principles of Model Checking*. Cambridge, MA, USA: MIT Press, 2008.
- [2] A. Abate, M. Prandini, J. Lygeros, and S. Sastry, "Probabilistic reachability and safety for controlled discrete time stochastic hybrid systems," *Automatica*, vol. 44, no. 11, pp. 2724–2734, 2008.
- [3] A. Jones, M. Schwager, and C. Belta, "Distribution temporal logic: Combining correctness with quality of estimation," in *Proc. IEEE 52nd Conf. Decis. Control*, 2013, pp. 4719–4724.
- [4] R. Chadha, V. A. Korthikanti, M. Viswanathan, G. Agha, and Y. Kwon, "Model checking MDPs with a unique compact invariant set of distributions," in *Proc. 8th Int. Conf. Quantitative Eval. Syst.*, 2011, pp. 121–130.
- [5] M. Agrawal, S. Akshay, B. Genest, and P. Thiagarajan, "Approximate verification of the symbolic dynamics of Markov chains," *J. ACM*, vol. 62, no. 1, pp. 1–34, 2015.
- [6] R. Bellman, "A Markovian decision process," *J. Math. mechanics*, vol. 6, no. 5, pp. 679–684, 1957.
- [7] M. L. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Hoboken, NJ, USA: John Wiley and Sons, Inc., 2014.
- [8] C. H. Papadimitriou and J. N. Tsitsiklis, "The complexity of Markov decision processes," *Math. Operations Res.*, vol. 12, no. 3, pp. 441–450, 1987.
- [9] D. Bertsekas, *Dynamic Programming and Optimal Control: Volume I*. Belmont, MA, USA: Athena Sci., 2012.
- [10] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.
- [11] E. Altman, *Constrained Markov Decision Processes*. Boca Raton, FL, USA: CRC Press, 1999.
- [12] K. J. Åström, "Optimal control of Markov processes with incomplete state information," *J. Math. Anal. Appl.*, vol. 10, pp. 174–205, 1965.
- [13] R. D. Smallwood and E. J. Sondik, "The optimal control of partially observable Markov processes over a finite horizon," *Operations Res.*, vol. 21, no. 5, pp. 1071–1088, 1973.
- [14] E. J. Sondik, "The optimal control of partially observable Markov processes over the infinite horizon: Discounted costs," *Operations Res.*, vol. 26, no. 2, pp. 282–304, 1978.
- [15] M. Kwiatkowska, G. Norman, and D. Parker, "Stochastic model checking," in *International School on Formal Methods for the Design of Computer, Communication and Software Systems*. Berlin, Germany: Springer, 2007, pp. 220–270.
- [16] M. Y. Vardi, "Automatic verification of probabilistic concurrent finite state programs," in *Proc. 26th Annu. Symp. Foundations Comput. Sci.*, 1985, pp. 327–338.
- [17] J.-P. Katoen, I. S. Zapreev, E. M. Hahn, H. Hermanns, and D. N. Jansen, "The ins and outs of the probabilistic model checker MRMC," *Perform. Eval.*, vol. 68, no. 2, pp. 90–104, 2011.
- [18] C. Dehnert, S. Junges, J.-P. Katoen, and M. Volk, "A storm is coming: A modern probabilistic model checker," in *Computer Aided Verification*, R. Majumdar and V. Kunčák, Eds., 2017, pp. 592–600.
- [19] D. Avila and M. Junca, "On reachability of Markov chains: A long-run average approach," *IEEE Trans. Autom. Control*, vol. 67, no. 4, pp. 1996–2003, Apr. 2022.
- [20] S. Summers and J. Lygeros, "Verification of discrete time stochastic hybrid systems: A stochastic reach-avoid decision problem," *Automatica*, vol. 46, no. 12, pp. 1951–1961, 2010.
- [21] A. Abate, J.-P. Katoen, and A. Mereacre, "Quantitative automata model checking of autonomous stochastic hybrid systems," in *Proc. 14th ACM Int. Conf. Hybrid Syst., Comput. Control*, Chicago, IL, 2011, pp. 83–92.
- [22] I. Tkachev and A. Abate, "Characterization and computation of infinite-horizon specifications over Markov processes," *Theor. Comput. Sci.*, vol. 515, pp. 1–18, 2014.
- [23] S. E. Z. Soudjani and A. Abate, "Precise approximations of the probability distribution of a Markov process in time: An application to probabilistic invariance," in *Proc. Int. Conf. Tools Algorithms Construction Anal. Syst.*, 2014, pp. 547–561.
- [24] I. Tkachev, A. Mereacre, J.-P. Katoen, and A. Abate, "Quantitative automata-based controller synthesis for non-autonomous stochastic hybrid systems," in *Proc. 16th ACM Int. Conf. Hybrid Syst., Computat. Control*, Boston, MA, 2013, pp. 293–302.
- [25] I. Tkachev, A. Mereacre, J.-P. Katoen, and A. Abate, "Quantitative model checking of controlled discrete-time Markov processes," *Inf. Computation*, vol. 253, no. 1, pp. 1–35, 2017.
- [26] I. Tkachev and A. Abate, "On infinite-horizon probabilistic properties and stochastic bisimulation functions," in *Proc. IEEE 50th Conf. Decis. Control Eur. Control Conf.*, 2011, pp. 526–531.
- [27] S. E. Z. Soudjani and A. Abate, "Probabilistic invariance of mixed deterministic-stochastic dynamical systems," in *Proc. 15th ACM Int. Conf. Hybrid Syst., Computat. Control*, Beijing, China, 2012, pp. 207–216.
- [28] D. Janak and B. Açikmeşe, "Maximal invariant set computation and design for Markov chains," in *Amer. Control Conf.*, 2019, pp. 1244–1249.
- [29] W. Wu, A. Arapostathis, and R. Kumar, "On non-stationary policies and maximal invariant safe sets of controlled Markov chains," in *Proc. 43rd IEEE Conf. Decis. Control*, 2004, pp. 3696–3701.
- [30] S.-P. Hsu, A. Arapostathis, and R. Kumar, "On optimal control of Markov chains with safety constraint," in *Amer. Control Conf.*, 2006, pp. 4516–4521.
- [31] M. El Chamie, Y. Yu, B. Açikmeşe, and M. Ono, "Controlled Markov processes with safety state constraints," *IEEE Trans. Autom. Control*, vol. 64, no. 3, pp. 1003–1018, Mar. 2019.
- [32] S. Akshay, B. Genest, and N. Vyas, "Distribution-based objectives for Markov decision processes," in *Pro. ACM/IEEE 33rd Annu. Symp. Log. Comput. Sci.*, 2018, pp. 36–45.
- [33] K. Leahy et al., "Control in belief space with temporal logic specifications using vision-based localization," *Int. J. Robot. Res.*, vol. 38, no. 6, pp. 702–722, 2019.
- [34] Y. Chen, T. T. Georgiou, and M. Pavon, "Optimal steering of a linear stochastic system to a final probability distribution, Part I," *IEEE Trans. Autom. Control*, vol. 61, no. 5, pp. 1158–1169, May 2016.
- [35] Y. Chen, T. T. Georgiou, and M. Pavon, "Optimal steering of a linear stochastic system to a final probability distribution, Part II," *IEEE Trans. Autom. Control*, vol. 61, no. 5, pp. 1158–1169, May 2016.
- [36] Y. Chen, T. T. Georgiou, and M. Pavon, "Optimal steering of a linear stochastic system to a final probability distribution-Part III," *IEEE Trans. Autom. Control*, vol. 63, no. 9, pp. 3112–3118, Sep. 2018.
- [37] G. Peyré and M. Cuturi, "Computational optimal transport: With applications to data science," *Foundations Trends in Mach. Learn.*, vol. 11, no. 5-6, pp. 355–607, 2019.
- [38] F. Blanchini and S. Miani, *Set-Theoretic Methods in Control*. Berlin, Germany: Springer, 2008.
- [39] Y. Gao, K. H. Johansson, and L. Xie, "Computing probabilistic controlled invariant sets," *IEEE Trans. Autom. Control*, vol. 66, no. 7, pp. 3138–3151, Jul. 2021.
- [40] Y. Ye and E. Tse, "An extension of Karmarkar's projective algorithm for convex quadratic programming," *Math. Program.*, vol. 44, pp. 157–179, 1989.
- [41] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*. Berlin, Germany: Springer, 2009.
- [42] E. Seneta, *Non-Negative Matrices and Markov Chains*. Berlin, Germany: Springer, 2006.
- [43] J. Löfberg, "YALMIP : A toolbox for modeling and optimization in MATLAB," in *Proc. CACSD Conf.*, 2004, pp. 284–289.
- [44] MOSEK ApS, "The MOSEK optimization toolbox for MATLAB manual. Version 10.0," 2022. [Online]. Available: <http://docs.mosek.com/9.0/toolbox/index.html>
- [45] V. A. Korthikanti, M. Viswanathan, G. Agha, and Y. Kwon, "Reasoning about MDPs as transformers of probability distributions," in *Proc. 7th Int. Conf. Quantitative Eval. Syst.*, 2010, pp. 199–208.
- [46] M. Herceg, M. Kvasnica, C. Jones, and M. Morari, "Multi-parametric toolbox 3.0," in *Eur. Control Conf.*, 2013, pp. 502–510.
- [47] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, "The quickhull algorithm for convex hulls," *ACM Trans. Math. Softw.*, vol. 22, no. 4, pp. 469–483, 1996.
- [48] A. Löhne and B. Weißing, "Equivalence between polyhedral projection, multiple objective linear programming and vector linear programming," *Math. Methods Operations Res.*, vol. 84, pp. 411–426, 2016.



**Yulong Gao** (Member, IEEE) received the B.E. degree in automation and the M.E. degree in control science and engineering from the Beijing Institute of Technology, Beijing, China, in 2013 and 2016, respectively, and the joint Ph.D. degree in electrical engineering from the KTH Royal Institute of Technology, Stockholm, Sweden, and Nanyang Technological University, Singapore, in 2021.

He is a Lecturer with the Department of Electrical and Electronic Engineering, Imperial College London, London, U.K. In 2019, he was a Visiting Student with the Department of Computer Science, University of Oxford, Oxford, U.K. From 2021 to 2022, he was a Researcher with KTH. Before joining Imperial College, he was a Postdoctoral Researcher at Oxford. His research interests include formal verification and control, machine learning, and applications to safety-critical systems.



**Alessandro Abate** (Senior Member, IEEE) received the Laurea in electrical engineering from the University of Padova, Padova, Italy, in 2002, and the M.Sc. and Ph.D. degrees in electrical engineering and computer sciences from University of California, Berkeley, Berkeley, CA, USA, in 2004 and 2007, respectively.

He is currently a Professor of verification and control with the Department of Computer Science, University of Oxford, Oxford, U.K., and a Fellow of the Alan Turing Institute for Data Sciences, London, U.K. He has been an International Fellow with CS Lab, SRI International, Menlo Park, CA, USA, and a Postdoctoral Researcher with the Department of Aeronautics and Astronautics, Stanford University, Stanford, CA, USA. From 2009 to 2013, he was an Assistant Professor with the Delft Centre for Systems and Control, TU Delft, Delft, The Netherlands.



**Lihua Xie** (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Newcastle, Australia, in 1992.

Since 1992, he has been with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, where he is currently a Professor and the Director of the Delta-NTU Corporate Laboratory for Cyber-Physical Systems and the Center for Advanced Robotics Technology Innovation. From 2011 to 2014, he was the Head of Division of Control and Instrumentation. From 1986 to 1989, he held teaching appointments with the Department of Automatic Control, Nanjing University of Science and Technology, Nanjing, China. His research interests include robust control and estimation, networked control systems, multiagent networks, localization, and unmanned systems.

Dr. Xie is an Editor-in-Chief for Unmanned Systems and has been an Editor for IET Book Series in Control and the Associate Editor for a number of journals, including IEEE TRANSACTIONS ON AUTOMATIC CONTROL, *Automatica*, IEEE TRANSACTIONS ON CONTROL SYSTEMS TECHNOLOGY, IEEE TRANSACTIONS ON NETWORK CONTROL SYSTEMS, and IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-II. From 2012 to 2014, he was an IEEE Distinguished Lecturer. He is the Fellow of Academy of Engineering Singapore, IFAC, and CAA.



**Karl Henrik Johansson** (Fellow, IEEE) received the M.Sc. and Ph.D. degrees in electrical engineering from Lund University, Lund, Sweden, in 1992 and 1997, respectively.

He is the Director of KTH Digital Futures, Stockholm, Sweden, and a Professor with the School of Electrical Engineering and Computer Science, KTH Royal Institute of Technology, Stockholm, Sweden. He has held visiting positions with the University of California, Berkeley, Berkeley, CA, USA; the California Institute of Technology, Pasadena, CA, USA; Nanyang Technological University, Singapore; the HKUST Institute of Advanced Studies, Hong Kong; and the Norwegian University of Science and Technology, Trondheim, Norway. His research interests include networked control systems, cyber-physical systems, and applications in transportation, energy, and automation.

Dr. Johansson has served on the IEEE Control Systems Society Board of Governors, the IFAC Executive Board, and is currently a Vice-President of the European Control Association Council. He was a recipient of several best paper awards and other distinctions. He was the recipient of Distinguished Professor with the Swedish Research Council and Wallenberg Scholar, and the Future Research Leader Award from the Swedish Foundation for Strategic Research and the triennial Young Author Prize from IFAC. He is a Fellow of the Royal Swedish Academy of Engineering Sciences.