Data-driven Abstractions for Verification of Deterministic Systems

RUDI COPPOLA*, ANDREA PERUFFO*, and MANUEL MAZO JR., TU Delft, Netherlands

A common technique to verify complex logic specifications for dynamical systems is the construction of symbolic abstractions: simpler, finite-state models whose behaviour mimics the one of the systems of interest. Typically, abstractions are constructed exploiting an accurate knowledge of the underlying model: in real-life applications, this may be a costly assumption. By sampling random \( t \)-step trajectories of an unknown system, we build an abstraction based on the notion of \( t \)-completeness. We newly define the notion of probabilistic behavioural inclusion, and provide probably approximately correct (PAC) guarantees that this abstraction includes all behaviours of the concrete system, for finite and infinite time horizon, leveraging the scenario theory for non convex problems. Our method is then tested on several numerical benchmarks.

Additional Key Words and Phrases: verification, data-driven, abstractions

ACM Reference Format:
Rudi Coppola, Andrea Peruffo, and Manuel Mazo Jr. 2023. Data-driven Abstractions for Verification of Deterministic Systems. 1, 1 (March 2023), 13 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Data-driven modeling and analysis is living a new renaissance with the advances in machine-learning and artificial intelligence enabled by unprecedented computing power. The field of system’s verification, aiming at providing formal performance and safety guarantees is not alien to this trend. Recent work has been focused on the use of collected data from a system to derive directly (with no model involved) barrier functions certifying invariance [2, 23], or finite abstractions to verify and synthesize controllers [12, 19, 22]. A popular approach is to employ scenario-based optimization techniques to derive probably approximately correct (PAC) guarantees on the performance metric of interest.

The use of scenario-based optimization requires independent samples, generated from the probability distribution that drives the system’s uncertainty. A (simpler) special case considered in many works [12, 17, 20, 26, 27] is that of deterministic systems (for which a model is not available), in which the only uncertainty is their initialization. In such a system the independence of samples can be derived by sampling the initial state in an independent fashion, e.g. typically a uniform distribution is selected for compact initial sets. However, when employing data to construct finite abstractions, or verifying barrier functions, the necessary data are individual transitions of the system: thus, to acquire independent samples of transitions implies acquiring samples from independent trajectories. This means that to collect data to learn e.g. a barrier function for a set larger than the initial one, one may have to discard large portions of trajectories to retain independence of samples. By using one-step transitions, the scenario-based approach provides guarantees for the satisfaction of one-step properties, e.g. decrease of a function across transitions, provided that transitions are indeed sampled independently. However, one is typically interested on inferring long (even infinite) horizon specifications from one-step properties. These approaches can reason only about finite horizon notions, as for infinite horizon properties the probability of satisfaction becomes trivially zero.

Contributions. Addressing this limitation is the main objective of this work. We consider deterministic systems with unknown dynamics and uncertainty in their initialization. Our approach provides a construction of data-driven finite abstractions, built on the notions of behavioural inclusions and \( t \)-complete behaviours. We introduce the notion of probabilistic behavioural inclusion, which is instrumental in describing the relation between a deterministic (but randomly sampled) model and a transition system constructed upon the collected system’s behaviours. Leveraging the non-convex scenario theory, we establish PAC guarantees for the inclusion of the concrete system’s (in)finite behaviours in those of the abstractions. This enables the verification of infinite horizon properties while retaining the PAC guarantees. We show throughout this work that the infinite horizon verification problem is equivalent to providing PAC guarantees on having observed the full support of an unknown probability mass function (with finite support) from a set of i.i.d. realizations. In other words, the PAC procedure bounds the probability of witnessing a new, unseen system behaviour.

Related Work. The notion of system abstraction is generally related to a model-based technique, where an expert or an automated technique exploits the knowledge of a concrete system to extract some prominent features [4, 25] in order to verify some system’s properties or to synthesise a policy (or controller) that ensures a user-defined requirement. When the model is unknown, one may resort to model identification (e.g. [15]) to then verify the identified model via traditional formal techniques, in a two-step procedure. Recently however, a few works propose a one-step method, i.e. the use of PAC guarantees to directly synthesise an abstraction from data, with guarantees of correctness, without the need to identify an underlying model. In [3, 8, 18] a sampled-based interval MDP is provided, employing the scenario approach to bound the transition probabilities of a stochastic dynamical model. In [12, 17], the authors define a PAC alternating simulation relationship between a symbolic abstraction and an underlying deterministic system, using one-step transitions.

In [20], PAC over-approximations of monotone systems are computed, which are then used to build models for unknown monotone systems. Finally, [16] computes the growth rate of a system from...
data, which is then used to construct a model abstraction and synthesise a controller. The use of data-driven f-complete models is briefly presented in [21] for linear PETC models. In [2, 23] the authors synthesise barrier certificates for unknown systems using template-based candidates, providing PAC bounds for their correctness, for stochastic and deterministic systems, respectively. However, when PAC bounds are related to a system’s transitions, the extension of behavioural guarantees from finite steps to infinite horizon is somewhat cumbersome, as we discuss in Section 3.

The scenario theory can be used also for verification purposes, as in [1], where the authors translate the probabilistic specifications into a Value-at-Risk determination problem for a scalar random variable. In [26, 27], a scenario-based set invariance verification approach is proposed, which relies on the observation of trajectories of an unknown deterministic system.

The works [10, 11, 13] discuss two approaches for the data-driven construction of forward reachability set of unknown autonomous systems, in the form of n-dimensional spheres and using Christoffel functions.

2 PRELIMINARIES

2.1 Notation

The set of real numbers is denoted by \( \mathbb{R} \). We use a string notation for sequences, e.g. \( r = abc \) means \( r(1) = a, \ r(2) = b, \ r(3) = c. \) When the length of a string \( r \) is not clear from the context, we use the subscript \( r \in \mathbb{N}_0 \) to denote its length, i.e. \( r. \) Given two sequences \( r_m \) and \( p_n \) with \( \infty \geq m > n \), we say that \( r \) exhibits \( p \) if there exists \( k \geq 0 \) such that \( r(k+i) = p(i) \) for \( i = \{1, \ldots, n\} \), denoted \( r \models p_{\omega} \). Given a set of sequences \( S \) and a sequence \( p \) we say that \( S \) exhibits \( p \) if there exists \( s \in S \) such that \( s \models p_{\omega} \), denoted \( S \models p_{\omega} \). We denote the uniform distribution supported on a domain \( \mathcal{D} \subset \mathbb{R}^n \) by \( \mathcal{U}_{\mathcal{D}} \).

2.2 Scenario Theory Background

Let us offer an overview on the scenario theory, as outlined in [6, 14], that we will use throughout this work. Let \( (\mathcal{D}, \mathcal{F}, \mathbb{P}) \) be a probability space, where \( \mathcal{D} \) is the sample space, endowed with a \( \sigma \)-algebra \( \mathcal{F} \) and a probability measure \( \mathbb{P} \); further, we denote by \( \mathcal{D}^\mathcal{N} \) the N-Cartesian product of the sample space and with \( \mathbb{P}^\mathcal{N} \) its product measure. A point in \( (\mathcal{D}^\mathcal{N}, \mathcal{F}^\mathcal{N}, \mathbb{P}^\mathcal{N}) \) is thus a sample \( (\delta_1, \ldots, \delta_N) \) of \( N \) elements drawn independently from \( \mathcal{D} \) according to the same probability \( \mathbb{P} \). Each \( \delta_i \) is regarded as an observation, or scenario. A set \( \Theta \), called the decision space, contains the decisions, i.e. the optimization space – no particular structure is assumed for this set. To every \( \delta \in \mathcal{D} \) there is associated a constraint set \( \Theta_\delta \subseteq \Theta \) which identifies the decisions that are admissible for the situation represented by \( \delta \).

Typically, the scenario theory refers to an optimisation program, which computes \( \theta_\mathcal{N}^* \), i.e. the solution of a optimisation program based on \( N \) samples. Once \( \theta_\mathcal{N}^* \) is computed, we are interested in assessing how it generalises to unseen scenarios \( \delta \in \mathcal{D} \), or, rather, the probability of extracting a sample that violates the constraints defined by \( \theta_\mathcal{N}^* \). To this end, we define the violation probability:

**Definition 1 (Violation [6]).** The violation probability of a given \( \theta \in \Theta \) is defined as

\[
V(\theta) = \mathbb{P}[\delta \in \mathcal{D} | \theta \not\in \Theta_\delta].
\]

V(\theta) quantifies the probability with which a new randomly selected constraint \( \Theta_\delta \) is violated by \( \theta \). If \( V(\theta) \leq \epsilon \), we say that \( \theta \) is \( \epsilon \)-robust against constraint violation. \( \square \)

Notice that in general \( V(\theta) \) is not directly computable since \( \mathbb{P} \) is not known. In [14], under mild assumptions, the authors show that a confidence bound can be derived as follows:

**Theorem 1 ([14]).** Given a confidence parameter \( \beta \in (0, 1) \) and the solution \( \theta_\mathcal{N}^* \), it holds that

\[
\mathbb{P}^\mathcal{N}[V(\theta_\mathcal{N}^*) \leq \epsilon(s_\mathcal{N}^*, \beta, N)] \geq 1 - \beta,
\]

where \( \epsilon(\cdot) \) is the solution of a polynomial equation (omitted here for brevity) and \( s_\mathcal{N}^* \) is the so-called complexity of the solution – it represents the minimum number of constraints \( m \leq N \) that yield the same solution \( \theta_\mathcal{N}^* \) computed with \( N \) scenarios.

**Remark 1.** The canonical scenario theory assumes continuous spaces \( \mathcal{D} \) and \( \Theta \). We instead refer to non-convex scenario theory results, as the sample space \( \mathcal{D} \) is discrete: this, in the scenario literature, is called a degenerate problem. For a detailed discussion on the derivation of these results, the interested reader may refer to [6, 14].

In this work, we solve the following problem.

**Problem Statement.** Given an unknown dynamical system, build an abstraction such that, with high confidence, the probability of witnessing a behaviour of the concrete system that is not exhibited by the abstraction’s behaviours is below a threshold value.

3 SAMPLING AND ABSTRACTIONS

3.1 Modeling Framework

Consider a time-invariant dynamical system described by

\[
\Sigma(x) := \begin{cases} 
  x_{k+1} = f(x_k), \\
  y_k = h(x_k), \\
  x_0 = x,
\end{cases}
\]

where \( x_k \in \mathcal{D} \subseteq \mathbb{R}^{n_x} \) is the plant’s state at time \( k \in \mathbb{N}_+ \), with initial value \( x_0 \in \mathcal{D}, y_k \in \mathcal{Y} \) is the system output with \( |\mathcal{Y}| < \infty \), and \( n_x \) is the state-space dimension. If the trajectory \( x_k \) exits \( \mathcal{D} \) at time \( k \), the output map returns a special symbol \( \mathcal{Y}^\top \) for all \( t \geq k \). We may think of the map \( h(\cdot) \) as a partitioning map, that returns the partition label (or index) corresponding to any state \( x_k \). The flow \( f(\cdot) \) and output map \( h(\cdot) \) are unknown, but we assume that given an initial condition \( x_0 \) we can observe the output sequence \( y_0, y_1, \ldots \) generated by \( \Sigma(x_0) \). We denote \( \mathcal{B}(\Sigma) \) the set of infinite horizon behaviours (i.e. output trajectories) exhibited by the system \( \Sigma \). With \( \mathcal{B}_H(\Sigma) \), we denote the set of behaviours for the time interval \( k = [0, H - 1] \). When the system of interest is evident, we remove the notation \( \Sigma \). Let us now introduce the notion of **equivalence class** [25]:

\[
[y] = \{x \in \mathcal{D} | y = h(x)\},
\]

and similarly, we define the equivalence class for an output sequence \( y_i = y_{i_1}, y_{i_2}, \ldots y_{i_j} \in \mathcal{Y}^j \) as

\[
[y_{i_j}] = \{x \in \mathcal{D} | y_{i_j} = h(f^{j-1}(x)) \text{ for } j = 1, \ldots, r\},
\]

with \( f^0(x) = x \). Equation (4) states that for \( i = 1, \ldots, |\mathcal{Y}|^j \), i.e. for every \( t \)-sequence \( y_t \in \mathcal{Y}^j \) the output equivalence class \([y_{i_j}]\) is
the set of points \( x \) such that if the dynamical system is initialized at \( x \), then the output sequence over the time interval \([0, t - 1]\)

conforms to \( y_t \). Note that an equivalence class might be empty if no initial condition gives an output behaviour corresponding to a particular \( t \)-sequence. Further, for all \( t \geq 1 \), the set of all \( [y_t] \) forms a partition of the domain \( D \).

To compute the abstraction of an unknown dynamical system we assume that \( \Sigma \) is randomly initialised by drawing initial conditions according to a probability measure space \((D, \mathcal{F}(D), \mathcal{P})\). Formally, we consider

\[ 
\Sigma_x := \{ \Sigma(x_0), \quad x_0 \sim \mathcal{P}(D) \}.
\]

The addition of a probability measure over the initial states affects the behaviours of \( \Sigma_x \). Let us refer to the probability of drawing an initial condition \( x_0 \) within the equivalence class \( [y_t] \) as

\[ 
\mathbb{P}[x_0 \in [y_t]] = \int_{[y_t]} p(v)dv,
\]

where \( p(\cdot) \) represents the probability density function of \( \mathcal{P} \). Throughout this work, abusing slightly notation, we consider \( \mathcal{B}_H(\Sigma_x) \)

\[ 
\mathcal{B}_H(\Sigma_x) := \{ \Sigma \in \mathcal{B}_H(\Sigma) \mid \mathbb{P}[x_0 \in [\Sigma]] > 0 \},
\]

that is the set of behaviours associated to equivalence classes with strictly positive probability measure. Obviously, it holds that if a behaviour belongs to \( \mathcal{B}_H(\Sigma_x) \) then it also belongs to \( \mathcal{B}_H(\Sigma) \), but the contrary is not necessarily true. Some behaviours present on \( \mathcal{B}_H(\Sigma) \) may have a zero probability measure associated in \( \Sigma_x \), therefore we say that \( \Sigma_x \) almost surely does not exhibit them – and we almost surely do not sample them. For instance, an unstable equilibrium with an output associated exclusively to \( y_{eq} = h(x_{eq}) \). In that case, while \( \mathcal{B}_H(\Sigma) \) contains the behaviour \( (y_{eq})^H \), this behaviour is only exhibited if the singleton \( \{x_{eq}\} \) is sampled, which almost surely does not happen. Note that, for the purpose of leveraging the scenario theory in this context, the presence of \( \mathcal{P} \) is unavoidable. Abusing notation, we denote \( \mathcal{B}^o(\Sigma_x) \) and \( \mathcal{B}_H(\Sigma_x) \) the behaviours, with strictly positive probability measure, of infinite and finite time horizon \( H \), respectively. Notice that the PAC guarantees are derived for the behaviours of \( \Sigma_x \), rather than for \( \Sigma \).

### 3.2 Abstractions via Transition Systems

The abstraction of a system is a tool that facilitates the analysis of large (even infinite) models. In this work, we assume that the concrete system is deterministic and non-blocking, and we employ a finite-state abstraction in the form of a transition system (TS).

**Definition 2 (Transition System [25]).** A transition system \( S \) is a tuple \((X, X_0, E, Y, H)\) where:

- \( X \) is the (possibly infinite) set of states,
- \( X_0 \subseteq X \) is the set of initial states,
- \( E \subseteq X \times X \) is the set of edges, or transitions,
- \( Y \) is the set of outputs, and
- \( H : X \rightarrow Y \) is the output map.

We consider non-blocking transition systems, i.e. systems where every state is equipped with at least one outgoing transition.

Let us define \( r = x_0, x_1, x_2, \ldots \) an infinite internal behaviour, or run of \( S \) if \( x_0 \in X_0 \) and \((x_i, x_{i+1}) \in E \) for all \( i \in \mathbb{N} \), and, with slight abuse of notation, \( B(r) = y_0, y_1, \ldots \) its corresponding external behaviour, or trace, if \( H(x_i) = y_i \) for all \( i \in \mathbb{N} \). With \( B^o(S(x_0)) \) and \( B_H(S(x_0)) \) we denote the sets of all infinite and of all finite external behaviours of length \( H \) of \( S \) starting from state \( x \), respectively; when the system of interest is clear from the context we use the shorthand notation \( B^o(x) \) and \( B_H(x) \) respectively. Models (3) and (5) can be described as a transition system, as follows.

**Definition 3 (Embedding of a dynamical system).** The embedding of a system \( \Sigma \) is the transition system \( \Sigma_S \) defined by \( \Sigma_S = (D, D_0, E, Y, H) \) where:

- \( E \subseteq D \times D \) is the set of all pairs \( x_i, x_j \) s.t. \( E = \{(x_i, x_j) \mid x_i \neq x_j \} \).
- \( H : D \rightarrow Y \) is the map defined by \( h(x) \).

Analogously if a system \( \Sigma_x \) is endowed with a probability distribution over the initial states as in (5), then the initial state is a random variable, and for every \( x_0 \in D_0 \), it holds that \( x_0 \sim \mathcal{P}(D_0) \). \( \square \)

Obviously, the embedding of a system does not change the set of all infinite (or finite) behaviours. Hence, the sets \( B^o(S_{\Sigma_S}) \) and \( B^o(S_{\Sigma_S}) \) differ only by zero probability measure behaviours. When no ambiguity arises, we simply denote the embedding of \( \Sigma \) as \( S \), omitting the subscript.

In order to construct the embedding of (5), we need full knowledge of the flow \( f(\cdot) \) to compute the transitions \( E \). As we only have access to the output \( y_k \), we recall the notion of **behavioural inclusion**: the abstraction is oblivious to the internal behaviour of a system, but we require that all behaviours that the concrete system can exhibit are within the set of behaviours of the abstraction.

**Definition 4 (Behavioural inclusion [25]).** Consider two systems \( S_a \) and \( S_b \) with \( Y_a = Y_b \). We say that \( S_b \) behaviourally includes \( S_a \), denoted by \( S_a \preceq S_b \), if \( B^o(S_a) \subseteq B^o(S_b) \). We say that \( S_a \) is behaviourally included in \( S_b \) until horizon \( H \) if this holds until horizon \( H \), i.e. \( B_H(S_a) \subseteq B_H(S_b) \), denoted \( S_a \preceq_{B_H} S_b \). \( \square \)

As we deal with systems where probability is embedded in the initial states, we define the probabilistic version of the behavioural inclusion relation.

**Definition 5 (Probabilistic Behavioural Inclusion).** Consider two systems \( S_a \) and \( S_b \) with \( Y_a = Y_b \). We say that \( S_a \) is behaviourally included in \( S_b \) with probability greater or equal than \( 1 - \epsilon \), denoted by

\[ 
\mathbb{P}[S_a \preceq_{B_H} S_b] \geq 1 - \epsilon,
\]

if for \( x_0 \sim \mathcal{P} \) it holds that:

\[ 
\mathbb{P}[B^o(S_a(x_0)) \subseteq B^o(S_b) \mid x_0 \sim \mathcal{P}] \geq 1 - \epsilon,
\]

where \( S_a(x_0) \) denotes the internal behaviour of system \( S_a \) starting from \( x_0 \). We say that \( S_a \) is almost surely behaviourally included in \( S_b \) if \( \epsilon = 0 \), written \( S_a \preceq_{B_H} S_b \) a.s. If the probabilistic behavioural inclusion holds until horizon \( H \), analogously, we write

\[ 
\mathbb{P}[S_a \preceq_{B_H} S_b] \geq 1 - \epsilon.
\]

A natural way of building a behavioural inclusion abstraction is by mapping each possible output sequence \( y_t \) to an abstract state. We may elaborate this intuition through a so-called \( t \)-complete model:
3.3 Trajectory Sampling
Several approaches [3, 8, 12, 17, 18, 20] provide PAC guarantees based on the concrete system’s transitions: in particular, samples are used to compute either transition probabilities or the probability of missing a transition amongst abstract states. As probabilities are related to transitions, the analysis of infinite-horizon properties becomes cumbersome: at every step, the transition probabilities accumulate until the abstraction becomes uninformative.

Further, we need to carefully address the extension of guarantees stemming from one-step transitions into multiple steps. The samples belonging to a trajectory are inherently correlated and they belong to different probability distributions. In practice, we sample the initial state $x_0$ according to a probability distribution $\mathcal{P}$, i.e.

$$x_0 \sim \mathcal{P},$$

and we then compute $x_1$, which may arise from a deterministic or stochastic system. Recall that the scenario theory can be used to provide guarantees about the behaviours arising from samples extracted from the same distribution. In general, however, the value $x_1 = f(x_0) + \eta_0 + \mathcal{P}$, belongs to a different probability distribution, as $f(x_0)$ belongs to a $\mathcal{P}$ distribution transformed by $f(\cdot)$ and the noise $\eta_0$ – if present – may belong to any other distribution; overall, $x_0$ and $x_1$ belong to different distributions. Hence, although one may use one-step transitions to solve a scenario program, i.e.

$$\{\delta_i\}_{i=1}^N = \{(x_0 \rightarrow x_1)_{i=1}^N\},$$

the scenario theory provides guarantees for a new one-step transition $(x_0 \rightarrow x_1)_{N+1}$. Nothing can be said on $(x_1 \rightarrow x_2)_{N+1}$, without further assumptions, as $x_1 \sim \mathcal{P}$.

The following example shows the shortcomings of using one-step transitions to infer (in)finite-horizon properties.

**Example 1.** Let us assume to partition the state space via a gridding procedure, as depicted in Fig. 2, and let us sample initial conditions from $P_0$, the partition in the bottom left corner. All trajectories starting from $P_0$ reach the blue region of $P_4$, where the blue region is strictly smaller than $P_0$. Let us further assume that every trajectory starting from the blue region reaches an unsafe region ($P_6$) in one step, whereas the trajectories starting from the white portion reach a safe set ($P_5$) in one step. If we sample trajectories from the partition $P_4$, we see that some of these reach the safe set (the ones starting from the white portion) and others reach the unsafe set (starting from the blue portion). Whilst the probability of reaching the unsafe set from $P_0$ is 1, using the sampled one-step transitions yield a probability of reaching the unsafe set strictly smaller than 1 and actually suggests the safe set can be reached.

Our approach is crucially different from the related literature: we employ the scenario approach to evaluate the probability of (existence of) states, whilst we use a deterministic rule (i.e. the domino rule) for transitions.

3.4 Data-driven Abstraction: Finite-time Guarantees
To overcome the limitations stemming from one-step transitions, we collect trajectories up to time $H$ and reason about the system behaviour over such horizon. Each $H$-long trajectory can be divided into several $\ell$-sequences (or one, if $\ell = H$). For instance, consider
\[ t = 2, \ H = 4, \ \text{and assume we sample the trajectory } y_1 y_2 y_3 y_1; \] the trajectory is split into \((H - t + 1)\) 2-sequences: \(y_1 y_2, y_2 y_3, y_3 y_4, \) and each 2-seques corresponds to an abstract state of the St-CA.

The data-driven procedure works as follows. We collect \(N\) \(H\)-long trajectories and construct \(X_N^N\), the set of witnessed \(t\)-sequences \(y_t\), that acts as the state set of the data-driven St-CA.

**Definition 7 (Data-driven St-CA).** The \(t\)-complete abstraction \(S_N^t = (X_N^N, \Delta^t, \Theta, \gamma, \mathbb{H})\) is called the data-driven \(t\)-complete abstraction (St-CA) of \(S\), where

- \(X_N^N\) is the state space built from the \(t\)-sequences collected from \(N\) trajectories of an underlying concrete system.
- The transitions, output space, and output map follow Definition 6. \(\square\)

**Remark 2 (Domino Completion).** The structure of the domino transitions, coupled with the sample-based states, may give rise to a blocking transition system. To illustrate this issue, consider \(t = 3\) and a dataset that exhibits solely the sequences \(\{y_1 y_1 y_1, y_1 y_2 y_2, y_2 y_2 y_1\}\). As the dataset does not include any sequence starting with \(y_2 y_1\), the state corresponding to \(y_2 y_1\) has no outgoing transitions, as depicted in Fig. 3. On the other hand, the existence of the sequence \(y_1 y_2 y_1\) implies the existence of at least one sequence starting with \(y_2 y_1\). To overcome this impediment, we artificially add all states corresponding to sequences \(y_2 y_1\). We repeat the procedure until we obtain a non-blocking transition system. \(\square\)

From here on we assume that the data-driven St-CA is non-blocking. Let us recall that we sample trajectories from \(\Sigma_x\) in (5): the set \(X_N^N\) almost surely does not contain \(t\)-sequences related to zero-measure sets. Once we collect \(N\) trajectories from the concrete system and construct the corresponding data-driven St-CA, we leverage the scenario theory to provide bounds on the probabilistic behavioural inclusion between the data-driven abstraction \(S_N^N\) and the embedding of the concrete system \(S\). In practice, we upper bound the probability of drawing an initial condition \(x_0\) in the concrete system that produces an \(H\)-long behaviour which cannot be generated by the abstraction.

As a first step towards the definition of the scenario problem, let us outline a chance-constrained program, where we assume to know both the probability distribution \(P\) and the concrete system \(\Sigma\). We denote by \(y_t\) the \(i\)-th \(t\)-sequence of the set \(\gamma\), for \(i \in \{1, \ldots, |\gamma|\}\). For the sake of simplicity, let us formulate the program for \(t = H\) (if \(H > t\) the same reasoning applies), and \(H \geq 1\). Let us define the random vector \(\delta \in \Delta\) as follows:

\[
\delta = [X_1, X_2, \ldots, X_{|\gamma|}]^T, \quad p_i = P[x_0 : B_H(x_0) \models \Theta y_t]. \tag{7}
\]

for \(i \in \{1, \ldots, |\gamma|\}\) and \(x_0 \sim P\). Note that, since the flow \(f(\cdot)\) is deterministic, the \(\delta\) depends entirely on \(x_0\), hence the \(X_i\)'s are mutually dependent. The vector \(\delta\) has one unique entry equal to 1 when \(t = H\), hence \(\delta\) follows a categorical distribution; otherwise it may have multiple ones.

**Example 2.** Let us assume a model has output space \(\gamma = \{a, b, c, d, e\}\), and let us consider \(t = H = 2\), thus giving \(|\gamma| = 25\). Let us assume that the underlying system produces solely the 2-sequences \(aa, bc, cd, de, eb\), whilst it does not admit any other 2-sequence. The concrete system partitions the domain into 5 regions, as depicted in Fig. 4, which correspond to the equivalence classes \([aa], [bc], [cd], [de], [eb]\). The vector \(\delta\) is composed of 25 elements; and \(p_i, i\) the probability of sampling \(aa\) is equal to 0.04. The probabilities of \(bc, \cdots, eb\) are respectively \(p_2, \ldots, p_5\), which are equal to 0.24. The remaining \(p_i\) for \(i > 5\) are zero. At each initial condition \(x_0\), randomly sampled within the domain, corresponds one \(t\)-sequence and one vector \(\delta\). For instance, sampling the 2-sequence \(aa\) has probability 0.04 and is encoded as the vector \(\delta = [1, 0, \ldots, 0]\). Notice that knowing that there are only 5 possible sequences reduces the support of \(\delta\); however, when this knowledge is unavailable, we use \(|\gamma|\).

**Remark 3.** For the purpose of obtaining an exact St-CA of \(S\) such that \(S \preceq S_t\), we solely require the knowledge about the possible (i.e. realisable) \(t\)-sequences; in practice it would be sufficient to know for which \(i \in \{1, \ldots, |\gamma|\}\) it holds that \(p_i > 0\). In that case, the relation \(S \preceq S_t\) holds almost surely (see Definition 5). \(\square\)

As \(\Delta\) is a discrete set with cardinality not greater than \(|\gamma|\), we can rewrite the \(i\)-th \(t\)-sequence \(y_t\) as a one-hot vector \(\delta_i\) of length \(|\gamma|\), as outlined in (7), and let \(\Theta = [\mathbb{R}^{|\gamma|}]\).

We are now ready to state the chance-constrained problem:

\[
\begin{align*}
\min_{\theta \in \Theta} & \quad 1^T [\gamma]^T \theta, \\
\text{s.t.} & \quad \sum_{i=1}^{|\gamma|} p_i \cdot \mathbb{1}_{\Delta_e} (\delta_i) \geq 1 - \epsilon, \\
& \quad (\theta - \delta_i) \geq 0, \text{for } \delta_i \in \Delta_e
\end{align*} \tag{8}
\]

where \(\Delta_e \subseteq \Delta\), with \(P(\Delta_e) \geq 1 - \epsilon\), \(1\)\(|\gamma|\) is a column vector of ones with length \(|\gamma|\), \(\mathbb{1}_{\Delta_e} (\cdot)\) is the indicator function. The program can
be interpreted as follows. Given a threshold $\epsilon$, find the maximum number of $\delta_i$ such that the sum of their probabilities is smaller than $\epsilon$. Recalling Example 2, given $\epsilon = 0.05$, we shall find a set of events $\Lambda_\epsilon$ such that its probability mass is not smaller than $1 - \epsilon$. From the event space $\Lambda = \{aa, bc, \ldots eb\}$ we then discard $aa$, since its probability is 0.04. We can then build an abstraction composed of the four remaining $t$-sequences, which accounts for a cumulative probability $P(\Lambda_\epsilon) \geq 1 - \epsilon$.

The solution $\theta^*$ represents a vector encoding the $t$-sequences that satisfy the probabilistic constraint $P(\Lambda_\epsilon) \geq 1 - \epsilon$. We can thus construct an approximate $S^t$-CA, where the states derive from $\theta^*$, and the transitions are governed by the domino rule. In fact, if the $i$-th row of $\theta^*$ is non-zero ($\theta^* \in \{0,1\}^{|Y|}$), then $\delta_i > 0$, and $y_i$ is part of the state set of $S^t_\ell$. Moreover, the domino completion might be necessary if the approximate $S^t$-CA has blocking states. Note that there could be more than one optimal solution, since $\Lambda$ is a discrete (and finite) set — in that case, we choose the smallest $\theta^*$ in lexicographic order. The following result follows trivially:

**Proposition 1.** Given $\epsilon$, for any optimal solution $(\theta^*, \Lambda^*_\epsilon)$ to (8) and the corresponding approximate $S^t$-CA, $S^t_\ell$, for a new initial condition $x_0$ sampled from $P$ it holds that

$$P \left[ B_H(S(x_0)) \in B_H(S^t_\ell) \right] \geq 1 - \epsilon,$$

where $B_H(S(x_0))$ denotes the $H$-long behaviour exhibited by $S$, i.e. the embedding of $\Sigma_s$, starting from $x_0$, and $B_H(S^t_\ell)$ denotes the set of all $H$-long behaviours of $S^t_\ell$.

Proposition 1 states that the set of $H$-long behaviours which belong to $S$ but do not belong to $S^t_\ell$ have a probability measure not bigger than $\epsilon$. As outlined in Definition 5, we can rewrite (9) as

$$P \left[ S \leq B_H(S^t_\ell) \right] \geq 1 - \epsilon.$$
Example 3. The scenario theory bounds the probability of sampling a new, unseen $\ell$-sequence within horizon $H$. Let us consider a one dimensional system,

$$x_{k+1} = \begin{cases} \frac{1}{2} x_k, & \text{if } x_k \in (\lambda, 1] \\ \frac{1}{2} x_k + \frac{1}{2}, & \text{if } x_k \in [0, \lambda] \end{cases}$$

where $x_0 \sim U([0,1])$, and $0 < \lambda < 2^{-4}$.

The state space $[0,1]$ is partitioned into five regions according to $P_i = (2^{-i}, 2^{-i+1})$ for $i = 1, ..., 4$, and $P_5 = [0, 2^{-4}]$, as shown in Fig. 5. Let us denote by $y_i$ the output of the system if the state belongs to $P_i$. It is easy to see that this system visits infinitely many times all 5 partitions, no matter what the initial condition is. For instance, if $x_0 \in P_1$, the system generates the repeating sequence $(y_1 y_2 y_3 y_4 y_5)^{10}$, where $\ell$ denotes the number of repetitions of the output $y$ and depends on $\lambda$, i.e. the width of the window that makes the system jump back to $P_1$.

Let us now assume we sample uniformly $[0,1]$ and take $\ell = H = 2$: this is equivalent to considering one-step transitions. It is easy to see that the probability of witnessing any $y_1 y_2 y_3 y_4 y_5$ is $2^{-4}$ for $i = 1, ..., 4$. The probability of sampling $y_5 y_1 y_2 y_3 y_4 y_5$ is $2^{-4} - \lambda$, and we witness the sequence $y_5 y_1 y_2$ only if $x_0 \in [0, \lambda] \subset P_5$, an event that occurs with probability $\lambda$, as summarized below:

$$P[\mathcal{B}_2(x_0) \models \diamond y_1 y_2] = 2^{-1}, \quad P[\mathcal{B}_2(x_0) \models \diamond y_2 y_3] = 2^{-2},$$
$$P[\mathcal{B}_2(x_0) \models \diamond y_3 y_4] = 2^{-3}, \quad P[\mathcal{B}_2(x_0) \models \diamond y_4 y_5] = 2^{-4},$$
$$P[\mathcal{B}_2(x_0) \models \diamond y_5 y_1] = 2^{-4} - \lambda, \quad P[\mathcal{B}_2(x_0) \models \diamond y_1 y_2] = \lambda,$$
$$P[\mathcal{B}_2(x_0) \models \diamond y_2 y_5] = 0, \quad P[\mathcal{B}_2(x_0) \models \diamond y_1 y_2] = 0.$$ (14)

Note that $\lambda$ is a parameter of the system and can be arbitrarily small.

Let us now consider a longer horizon $H' > 2$ and focus on the sequence $y_5 y_1 y_2$. It is easy to see that the subsequence $y_5 y_1 y_2$ will eventually be generated by every trajectory, if the horizon $H'$ is long enough\(^1\). Formally, for $H' \geq 1 - \lceil \log_2 (\lambda) \rceil$ it holds that

$$P[\mathcal{B}_{H'}(x_0) \models \diamond y_1 y_2] = 1.$$

This represents a challenge for our approach: whilst the probability of witnessing $y_5 y_1 y_2$ as an initial sequence is arbitrarily small, the probability of seeing $y_5 y_1 y_2$ over a sufficiently long horizon is actually 1. In general, we cannot use the scenario bounds generated from samples over time horizon $H$ to infer properties over longer time horizons. Finally, observe that the system exhibits the behaviour $y_5 y_2$ if and only if it is initialized exactly at $x_0 = 0$. Such event has a zero probability measure since $x_0 \sim U([0,1])$, hence, any data-driven St-CA will almost surely not include such behaviour.

3.5 Trajectory Horizon and Choice of $\ell$

Let us offer some intuition on the role of $\ell$ and $H$. On the one hand, we are interested in considering a large $H$ as we aim to extend the PAC guarantees over a long time horizon. In addition, if $H \gg \ell$, we expect to obtain a data-driven St-CA containing a large portion of all the possible $\ell$-sequences, as every trajectory contains several $y_i$'s. Further, a small $\ell$ limits the number of possible sequences (upper bounded by $|\mathcal{V}|^k$), which results in more compact abstractions. On the other hand, $\ell$ can be considered a refinement parameter: a larger $\ell$ provides a finer and more precise abstraction. Increasing the state space typically reduces the nondeterminism of the abstraction. Notice that this refinement applies also in case of a black-box model, where we do not have access to the partitioning map. The “classical” state discretization, i.e. gridding procedure, needs access to the actual state variables, not only to the output ones, to refine the partitioning.

We highlight that neither $\ell$ nor $H$ appear explicitly in the computation of $\epsilon$, however, they affect the complexity $s_N^\epsilon$. Recall the definition of $s_N^\epsilon$ from Theorem 1: for a given set of $N$ scenarios $\{\delta_i\}_{i=1}^N$ and the corresponding optimal solution $\theta_N^\epsilon$, the complexity $s_N^\epsilon$ is equal to the cardinality of the smallest subset $\{\delta_{i_1}, ..., \delta_{i_{\epsilon N}}\} \subseteq \{\delta_i\}_{i=1}^N$ returning the same scenario program solution, that is, $\theta_N^\epsilon = \theta_{m^*}^\epsilon$ and $s_N^\epsilon = m$. The choice of $H$ is directly linked to the duration allowed for exploration of the state space of the system. In fact, every $H$-trajectory contains between 1 and $\min(|\mathcal{Y}|^\ell, H - \ell + 1)$ different $\ell$-sequences. It follows that, as we increase the gap between $H$ and $\ell$, it is likely that the complexity will decrease, since one $H$-sequence can include multiple $\ell$-sequences. For a fixed $N$, the lower the complexity $s_N^\epsilon$, the lower the $\epsilon(s_N^\epsilon, N, \beta)$, and therefore, the higher is the probability of behavioural inclusion.

Note that finding the complexity of (10) is equivalent to solving a set cover problem [7], where the universe is given by all the witnessed $\ell$-sequences, and every covering set is obtained by considering the unique $\ell$-subsequences present in every sampled $H$-sequence. In practice, we use a greedy algorithm for the (unweighted) set cover problem to find an upper bound of the complexity (and hence an upper bound for $\epsilon$), see [29]. We construct the minimum cardinality of $H$-sequences $\{\delta_{i_1}, ..., \delta_{i_{\epsilon N}}\}$ holding the same solution $\theta_N^\epsilon$, by adding iteratively the $H$-sequence containing the largest number of the witnessed $\ell$-sequences.

4 INFINITE BEHAVIOURS

Proposition 2 states that we can construct an abstraction that behaviourally includes the concrete system, with PAC guarantees, up to the horizon $H$. Let us now discuss how to extend the guarantees to infinite horizon properties.

First, we notice that the dynamics of the concrete system define the probability of sampling a finite trace $y_\ell$ at any given time: let us recall Example 3. The probability of sampling a trajectory exhibiting the sub-sequence $y_\ell = y_5 y_1 y_2$ as an initial sub-sequence is rather small (\(\lambda\)) but it increases with $H$, until it becomes 1. Let us define the probability of encountering a trace $y_\ell$ at any given time $t$, as

$$P[ x_0 : \mathcal{B}^{\text{Pre}\ell}(x_0) \models \diamond y_\ell ] = P[ \bigcup_{i=0}^{\infty} \text{Pre}_{\mathcal{D}}^{i}(\{y_\ell\}) ] \geq P[ \bigcup_{i=0}^{k} \text{Pre}_{\mathcal{D}}^{i}(\{y_\ell\}) ],$$

where

$$\text{Pre}_{\mathcal{D}}^{i}(S) := \{ x' \in \mathcal{D} | f^{i}(x') \in S, f^{j}(x') \in \mathcal{D} \text{ for all } 0 \leq j \leq i \},$$

i.e. the points in the domain whose trajectory remains within $\mathcal{D}$ for all steps $j \leq i$, and the $i$-th step is within $S$. For simplicity we...
denote, for an arbitrary set $S$:
\[ \mu_0^k(S) := \mathbb{P}\left( \bigcup_{i=0}^{k} \text{Pre}^k_D(S) \right) . \]  
(15)

In words, the probability of sampling an initial state $x_0$ that eventually leads to witness the trace $y_t$ corresponds to the probability of the equivalence class of $y_t$ together with all the sets eventually leading to it, i.e. the $\text{Pre}^k_D(y_t)$. This quantity grows with $k$: whilst for a small $k$ the probability of sampling $y_t$ may be negligible, for $k \to \infty$, it may reach a significantly large value. To improve the probability of sampling $y_t$, we might increase the horizon $k$; whilst this increases the chance of sampling $y_t$, it also increases the computational needs of our procedure. We use $\mu(S)$ in place of $\mu_0^k(S)$ to denote the probability measure of a set $S$. We then introduce the following assumption.

**Assumption 1.** Given a system (5), assume that a monotonically non-decreasing function $\phi$ is known for some $k \in \mathbb{N}$, such that for all sets $S$ corresponding to arbitrary unions of equivalence classes, i.e. $S = \bigcup_{j \in J} [y_{t_j}]$ with $J \subseteq \{1, 2, \ldots, |\mathcal{Y}|\}$, the following holds
\[ \mu_0^k(S) \geq \phi(k) \cdot \mu_0^\infty(S). \]  
(16)

This trivially implies that
\[ \mu_0^k(S) < \epsilon \implies \mu_0^\infty(S) < \frac{1}{\phi(k)} \cdot \epsilon . \]

In practice Assumption 1 allows to link the probability measure of visiting any $f$-sequence’s equivalence class $S$ in $k$ steps with the probability of visiting it in an infinite number of steps. Importantly, the function $\phi(k)$ describes how the proportion of the measure assigned to the set $\bigcup_{i=0}^{k} \text{Pre}^k_D(S)$ with respect to the measure of $\bigcup_{i=0}^{\infty} \text{Pre}^k_D(S)$ changes over time, due to the dynamics of the system. Note that the function $\phi$ necessarily needs to be monotonically non-decreasing. Being able to obtain such a bounding function $\phi$ requires some a-priori knowledge of properties of the dynamics and the equivalence classes defined by the output map of $\Sigma$.

Assumption 1 allows us to extend the previous results to behaviours of arbitrary length.

**Proposition 3.** Consider $N$ trajectories of length $H$ collected from (5), the data-driven $\text{St}-\text{CA} S_N^\infty$ based on the corresponding $f$-sequences in the sampled trajectories, and, for a given $\beta$, the bound $\epsilon(s^*, N, \beta) = \overline{\epsilon}$ resulting from solving program (10). Let Assumption 1 hold with $k = H - t$ and $\overline{\epsilon}$ as defined therein. Then, for a new initial state $x_0$ sampled from $\mathcal{P}(D)$,
\[ \mathbb{P}[S_N \preceq_\beta S_N^\infty] \geq 1 - \overline{\epsilon} \geq 1 - \beta . \]  
(17)

i.e. with confidence $1 - \beta$, $S_N^\infty$ probably behaviourally includes the model (5) with probability not smaller than $1 - \overline{\epsilon}$.

**Proof.** Given $N$ i.i.d. scenarios, let us denote $\theta_N^m$ as the optimal solution of the corresponding scenario program and $V(\theta_N^m)$ as the violation probability. $V(\theta_N^m)$ represents the probability of drawing a new initial condition $x_0$ which results in a $H$-long behaviour exhibiting one (or more) previously unseen $f$-sequence, i.e.
\[ V(\theta_N^m) := \mathbb{P}[x_0 : B_H(S_N^\infty(x_0)) \models \bigoplus y_t \land \gamma_t \notin X_N^\infty] , \]

where $y_t$ represents any sub-sequence of length $t \in B_H(S_N^\infty(x_0))$, the $H$-long behaviour of $S_N^\infty$ starting from $x_0$. The scenario theory assures that the violation is upper-bounded by $\epsilon$, i.e. $\epsilon > V(\theta^m)$, with confidence not smaller than $1 - \beta$. Denote by $S$ the set of unseen $f$-sequences $y_t$, such that $B^\infty_S(y_t)$ exhibits $y_t$ but $y_t \notin X_N^\infty$. The scenario theory ensures that
\[ V(\theta_N^m) = \mathbb{P}[x_0 : B_H(x_0) \models \bigoplus y_t, y_t \in S] = \mu^H_{H - t} \left( \bigcup_{y_t \in S} [y_t] \right) , \]

where $V_{H - t}(\cdot)$ is the probability measure of all initial conditions $x_0 \in D$ which exhibit any $y_t \in S$ in at most $H$ steps. Equivalently, $x_0$ must visit any $[y_t]$ in at most $H - t$ steps. Let us apply Assumption 1, with $k = H - t$. It holds that
\[ \mu^\infty_0 \left( \bigcup_{y_t \in S} [y_t] \right) \leq \frac{1}{\phi(k)} \cdot \mu^k_0 \left( \bigcup_{y_t \in S} [y_t] \right) \leq \frac{1}{\phi(k)} \cdot \overline{\epsilon} = \overline{\epsilon} . \]

It follows that with confidence $1 - \beta$, all unseen sequences $y_t$ have a total probability measure of being exhibited by an infinite behaviour upper-bounded by $\overline{\epsilon}$, hence the behavioural relationship between the concrete model and abstraction holds with PAC bound $\overline{\epsilon}$. □

We can think of the scenario bound $\overline{\epsilon}$ as a bound on the probability measure of encountering an unseen $f$-sequence (with confidence $1 - \beta$) until time $H$. Assumption 1 allows us to extend the scenario guarantees from the finite to the infinite horizon.

4.1 Systems Accepting a Bisimulation Relation

A trivial consequence of Assumption 1 is the following corollary.

**Corollary 1.** Assume there exists $\overline{k}$ such that $\phi(\overline{k}) = 1$, then for the data-driven $\text{St}-\text{CA} S_N^\infty$ constructed from $N$ trajectories of length $H = \overline{k} + t$, Proposition 3 holds with $\overline{\epsilon} = \overline{\epsilon}$.

Corollary 1 gives a direct solution to extend finite time scenario guarantees to infinite time scenario guarantees since it provides the simulation horizon necessary to capture the transient behaviours. This follows from the observation that if $\phi(\overline{k}) = 1$ then $\mu^\infty_0([y_t]) = \mu^k_0([y_t])$ for every $y_t$. Corollary 1 is useful to provide guarantees on a broad family of models, i.e. dynamical systems accepting a bisimulation relation.

Let us assume the existence of a deterministic $\text{St}-\text{CA} S_T$ for the embedding of the concrete system $S_N$: the equivalence relation (4) for $t = \overline{T}$ defined on $S_N$ is a bisimulation relation if and only if $S_T$ is deterministic, see [5, Ch. 1]. We exploit this result to construct a data-driven $\text{St}-\text{CA}$ that probably behaviourally includes the concrete system for infinite horizon, as the next proposition shows.

**Proposition 4.** Suppose there exists a deterministic $\text{St}-\text{CA} S_T$ for the embedding of system (3) $S_N$ with $t = \overline{T}$, and consider the data-driven $\text{St}-\text{CA} S_N^\infty$ constructed from $N$ trajectories. Proposition 3 holds with $\overline{\epsilon} = \overline{\epsilon}$ if either
1. $H = |\mathcal{Y}|^{\overline{T} - 1} + \overline{T} - 1$,
2. $\overline{T} \leq H < |\mathcal{Y}|^{\overline{T} - 1} + \overline{T} - 1$ and $S_N^\infty$ is non-blocking.
4.2 Uncertain Affine Stable Systems

Let us consider the class of affine stable systems, that can be written without loss of generality as

\[ \Sigma(x):= \begin{cases} x_{k+1} = f(x_k) = Ax_k + b, \\ y_k = h(x_k), \\ x_0 = x, \end{cases} \tag{18} \]

where both \( A \) and the equilibrium \( x_{eq} \) of the map \( f \) are unknown. Note that \( b \) can be computed from the relation \( b = (I - A)x_{eq} \). We assume \( A \) to be full-rank and to know a bound on its eigenvalues, i.e. \( ||A||_2 \leq \alpha < 1 \) (where \( || \cdot ||_2 \) denotes the induced 2-norm).

For the affine stable system (18), (where \( x_{eq} \) is known to be full-rank and to know a bound on its eigenvalues), let us consider the class of linear systems.

Proposition 5. For the affine stable system (18), where two scalars \( \rho, \alpha \in \mathbb{R} \) such that \( 0 < ||A||_2 \leq \alpha < 1, \rho \geq |\det(A^{-1})| > 1 \) are known, and \( x_0 \sim U \), if there exists an equivalence class \( [y^*] \) such that the equilibrium \( x_{eq} \) belongs to its interior, Assumption 1 holds for every \( k \in \mathbb{N} \), with

\[ \varphi(k) = \begin{cases} \min \left( \psi(k), \rho^{k-\overline{k}} \right) & \text{for } k \leq \overline{k} - 1, \\ 1 & \text{for } k \geq \overline{k}, \end{cases} \tag{19} \]

where

\[ \psi(k) = \left( 1 + \rho^{\overline{k} - 1} - k \sum_{i=0}^{k} \rho^{-i(k+1)} \right)^{-1} , \tag{20} \]

and where \( \overline{k} = \left[ \log_{d_{\text{max}} / d_{\text{min}}} \right] z(k) = \overline{k}/(k+1) \), \( d_{\text{min}} \) and \( d_{\text{max}} \) are respectively the radius of the largest ball entirely contained in \([y^*]\) and the radius of the smallest ball entirely containing \( D \) centered at \( x_{eq} \).

Proof. See Appendix A.2. \( \square \)

Let us offer some intuition on the above proposition. Denote by \( S_0 \) the partition containing \( x_{eq} \) and define:

\[ \overline{k} = \min_k x(T) \in S_0 \text{ for all } t \geq T, \text{ for all } x \in D, \]

i.e. the time instant for which all trajectories reach and remain within \( S_0 \). Proposition 5 captures the intuitive fact that \( \overline{k} \) must depend on the ratio between the volume of \( S_0 \) and the volume of \( D \), as well as on the contractivity rate, i.e. the eigenvalues, of (18). Notice that \( \overline{k} \) exists and it is finite, as \( D \) is finite and the system is asymptotically stable. Hence, the set \( \text{Pre}^k(S_0) \) at time \( k = \overline{k} \) covers the whole domain, i.e. \( D \subseteq \text{Pre}^{\overline{k}}(S_0) \).

A similar reasoning applies to unstable affine systems: defining again \( S_0 \) as the partition containing the equilibrium point, with output symbol \( y^* \). All trajectories starting at any point in \( D \setminus S_0 \) exit the domain in a finite number of steps, denoted \( \overline{k} \). On the other hand, all trajectories starting in \( S_0 \) generate \( t \)-sequences that start with \( y^* \); we denote them as \( y^*\sigma_{t-1} \), where \( \sigma_{t-1} \in Y^{t-1} \) represent any possible output symbol sequence of length \( t - 1 \). This allows us to easily establish that \( \varphi(k) = 1, \forall k \geq \overline{k} \). One can envision a similar generalisation to classes of non-linear dynamics, based on contractivity properties, which we leave as future work.

Remark 5. Note that the assumed property enabling these extensions to infinite horizons, in the previous classes of systems, amount to bounding the possible length of transients of the systems (without exact knowledge of the system itself).

5 EXPERIMENTAL EVALUATION

5.1 Linear Stable System

We first consider a linear stable model to show our approach. The system evolves according to

\[ x_{k+1} = Ax_k, \text{ where } A = \begin{bmatrix} 1/3 & 1/2 \\ -1 & 1 \end{bmatrix}. \]

The state space \( D = [-1,1]^2 \) is partitioned into 81 regions by a uniform grid. Note that in order to include the equilibrium with a squared partition we need to assess that \( ||A||_\infty \leq 1 \). We sample \( N = 10^4 \) initial states from the uniform distribution \( U_D \) and let the trajectories run until horizon \( H = 4 \). Let us consider \( t \)-sequences with \( t = 2 \). We collect 189 \( t \)-sequences and construct the corresponding abstraction. We set the confidence value to \( \beta = 10^{-12} \) and compute the scenario bounds for (11), which result in

\[ \epsilon(s^*, \beta, N) = \overline{T} = 7.45 \cdot 10^{-9} \].

In order to extend the guarantees from horizon \( H = 4 \) to the infinite horizon we employ Assumption 1. We compute \( \rho = |\det(A)^{-1}| \) and \( \alpha = ||A||_2 \), which give \( \overline{T} = 9 \). Since \( H = 4 \), we evaluate the function \( \varphi \) for \( k = H - \ell = 2 \) and, using (19), we get \( \varphi(k) = 4.57 \cdot 10^{-4} \). We can thus guarantee that the abstraction holds for an infinite horizon. Hence, the measure of \( \rho_0^\infty(S) \) is upper bounded by

\[ \rho_0^\infty(S) = \frac{1}{\varphi(k)} \overline{T} = 1.62 \cdot 10^{-5}. \]

Our abstraction holds for infinite horizon properties with PAC guarantees

\[ \mathbb{P}[S \gtrless_{\uparrow} S_T^N] \geq 1 - 1.62 \cdot 10^{-5} \geq 1 - 10^{-12}. \]

5.2 One-dimensional Hybrid System

Let us consider the system (13) from Example 3 with \( \lambda = 10^{-2} \) and let us collect \( N \) trajectories from it. The output of each partition \( P_i \) is denoted \( y_i \) for \( i \in [1, 5] \).

Trivially, for \( \ell = 1 \), the possible \( \ell \)-sequences are \( \{y_i\}_{i=1}^5 \). The probability of sampling each \( y_i \) is simply

\[ \mathbb{P}[B_1(x_0) = y_i] = 2^{-1}, \quad i \in [1, 5]. \]

Let us consider \( \ell = 2 \); the possible \( \ell \)-sequences and the probability of being sampled are reported in (14) where the smallest non zero probability value \( \lambda \) is attained by sequence \( y_2 y_3 \).

We collect \( N = 10^4 \) trajectories of horizon \( H = 2 \), which provide all the \( 2 \)-sequences: hence, we know that the resulting \( SF-CA \) is a
We consider a path planning problem, as depicted in Fig. 7, where an agent lies within a grid state space $X^N = \{y_1 y_2, y_2 y_3, y_3 y_4, y_4 y_5, y_5 y_6, y_6 y_7\}$, shown in Fig. 6. Let us set a confidence value $\beta = 10^{-12}$; the scenario program has a complexity $s_N^\beta = 3$, and by inserting this value in (17) we get

$$
\epsilon(s_N^\beta, \beta, N) = 4.80 \cdot 10^{-3},
$$

thus the data-driven abstraction holds for horizon $H = 2$ with PAC bound $\epsilon = 4.8 \cdot 10^{-3}$. To prove that the bounds hold we compute the exact probability of violation, assuming full knowledge of the system: by taking into account the probability of each sequence, it is possible to show that

$$
\mathbb{P}[V(\partial_N^\epsilon)] \leq \epsilon = \mathbb{P}[\mathcal{S}^{\leq \ell} \subseteq \mathcal{S}_N^{\ell}] \geq 1 - \epsilon \approx 1 - 2.25 \cdot 10^{-44},
$$

hence, we show that the scenario bounds we obtain are conservative.

Let us now consider the infinite behaviours of system (13).

Analysing the Pre sets of the equivalence classes, we find $\bar{k} = 7$: this value represents the time horizon $k$ such that $\mu_k^S(S) = \mu_k^\mu(S) = 1$ for all sets $S$ corresponding to a (union of) equivalence classes. As discussed in Corollary 1, by sampling trajectories of length $H = \bar{k} + \ell = 9$ we can provide guarantees for the infinite horizon behaviours. We repeat the experiment considering $N = 10^4$ trajectories with $\ell = 2$ and $H$ as above. Again, we collect the $6 \ell$-sequences shown above, which can be obtained from $s_N^\epsilon = 1$ trajectory; we get a new scenario bound

$$
\epsilon(s_N^\epsilon, \beta, N) = 3.47 \cdot 10^{-3}.
$$

### 5.3 Path Planning

We consider a path planning problem, as depicted in Fig. 7, where an agent lies within a $10 \times 10$ grid state space. It is tasked to reach the green target area – with coordinates $[7, 8] \times [7, 9]$ – whilst avoiding the obstacles (shown in red) and remaining within the borders of the state space. The agent’s initial state is chosen uniformly at random within the white area of the state space, and it can choose among four actions (up, down, left, right) at every time step, in order to reach the target area. First, we run a standard Q-learning algorithm [28] to train the agent, with time horizon $H = 40$.

After the training, we use the newly synthesised control policy for a continuous-space experiment, where the agent can take positions over the continuous $[0, 10]^2$ domain, and its actions are obtained as a weighted average of the actions corresponding to the closest grid points. Formally, the action $a(x)$ results $a(x) = \sum_{d(x, g) < 1} w_d \cdot a(g)$, where $x$, $g$ are the locations in the continuous and grid space, respectively, $d(x, g)$ is the distance between points $x$ and $g$, and $w_d$ is a coefficient depending on the $d(x, g)$ – the weights $w_d$ sum up to 1. We sample the system and collect the agent’s position in terms of $W, R, G$ labels (white, red, and green, respectively). We obtain $N = 10^4$ trajectories with horizon $H = 40$, and consider subsequences of length $\ell = 27$. We collect 27 different $\ell$-sequences, which can be obtained from a total of $s_N^\epsilon = 3$ trajectories. By setting the confidence to $\beta = 10^{-12}$, the scenario bound evaluates at

$$
\epsilon(s_N^\epsilon, \beta, N) = 4.06 \cdot 10^{-3}.
$$

We construct the data-driven St-CA with the 27 $\ell$-sequences, and we verify a safety property: the system always reaches (and remains within) the target set avoiding the obstacles. Hence, with confidence $1 - \beta$, the concrete model always reaches the target set with probability greater or equal to $1 - \epsilon$ (for a horizon $H = 40$).

## 6 CONCLUSIONS AND FUTURE WORK

We have presented a method to construct a finite, data-driven abstraction of a deterministic system with unknown dynamics. We introduce the notion of probabilistic behavioural inclusion, and use it to bound the probability of unseen behaviours of the concrete system. We then build an $\ell$-complete automaton that generates behaviours of the concrete system, based on trajectories up to time $H$ – which can be useful in machine learning approaches as RL. We further prove that this construction can be used to simulate the behaviour of the system for infinite time, if the concrete system admits a bisimulation relation.

Our approach can be used to verify any logic specification: the PAC guarantees provided by our approach rely upon the collected behaviours, and whether these satisfy the desired property. As such, the concrete model complexity (e.g. nonlinearities) is irrelevant.

The current method finds applications in the verification of unknown systems, whilst future work includes the extension to control synthesis and stochastic systems.

## ACKNOWLEDGMENTS

This work was supported by the European Research Council through the SENTIENT project (ERC-2017-STG #755953).

We would like to thank Gabriel de Albuquerque Gleizer and Giannis Delimpaltadakis for their helpful discussions throughout the development of this work.
A PROOFS

A.1 Proof of Proposition 4

Proof of (1). Consider the deterministic $S^*$-CA for $S_\ell$, denoted $S_\ell^*$. Since we assume $S_\ell$ to be non-blocking, necessarily $S_\ell^*$ is also non-blocking. We begin by bounding the cardinality of the state set of $S_\ell^*$. Given a state $y_\ell^*$, consider the set of reachable states

$$\mathcal{R}(y_\ell^*) = \{ y \in S_\ell^* \mid (y_\ell^*, y) \in E_\ell \}. $$

By the domino rule, the first $\ell - 1$ elements of the sequence $y_\ell$ are identical to the last $\ell - 1$ elements of $y_\ell^*$, and by determinism, $|\mathcal{R}(y_\ell^*)| = 1$. Thus, there cannot exist two states $y_\ell^*, y_\ell^*$ sharing the same first $\ell - 1$ elements. We conclude that the maximum cardinality of the set $S_\ell^*$ is equivalent to the number of possible prefixes of length $(\ell - 1)$ given the output map $\mathcal{Y}$, which amounts to $|\mathcal{Y}|^{\ell-1}$. One example of transition system with the maximum number of states is represented by a 'chain' of states, i.e. when each state has one outgoing and one incoming transition. Given $S_\ell$, we denote with $\bar{\ell}$ the diameter of the underlying graph, i.e. the longest path connecting any two states. In general $\bar{\ell} \leq |\mathcal{Y}|^{\ell-1} - 1$, and this is a strict bound. The diameter of the abstraction affects $\mu_\ell^k(y_\ell^*)$ and allows us to invoke Corollary 1. Indeed, from (4)

$$\text{Pre}_\ell^D(y_\ell^*) = \left\{ \left( y_{\ell-1} \right) \mid \exists y \in S \right\} \subseteq \text{Pre}_\ell^D(y_\ell^*) \subseteq \left\{ \left( y_{\ell-1} \right) \mid \exists y \in S \right\}$$

where $S = \{ y \in \mathcal{Y} \mid \text{Pre}_\ell^D(y_\ell) \cap \{ y \neq \emptyset \} \}$ and $(S_\ell^*)$ implies that for any $y \in S$

$$\{ y_{\ell-1} \} \subseteq \left( y_{\ell-1} \right) \subseteq \left( y_{\ell-1} \right).$$

To see this, suppose there exists $\tilde{y}_\ell = y_{\ell-1} \ldots y_0$, such that $[\tilde{y}_\ell] \neq [y_{\ell-1} \ldots y_0]$. Trivially, from (4), it follows that $\{ y_{\ell-1} \} \subseteq \{ \tilde{y}_\ell \} \subseteq \{ y_{\ell-1} \}$. Since $S_\ell^*$ is non-blocking and $[\tilde{y}_\ell] \neq \emptyset$ it follows that $|\mathcal{R}(\tilde{y}_\ell^*)| > 1$ which is a contradiction. Thus, we have shown that

$$\text{Pre}_\ell^D(y_\ell^*) = \left\{ \left( y_{\ell-1} \right) \mid \exists y \in S \right\} \subseteq \text{Pre}_\ell^D(y_\ell^*) \subseteq \left\{ \left( y_{\ell-1} \right) \mid \exists y \in S \right\}$$

or, in words, that the image of the operator $\text{Pre}_\ell^D$ on any equivalence class $[y_{\ell-1}]$ consists of a union of equivalence classes of other $\bar{\ell}$-sequences. Hence, for any $y_{\ell-1}$

$$\mu_\ell^k(y_{\ell-1}) = \mathbb{P} \left[ \bigcup_{t = 0}^{k} \text{Pre}_\ell^D(y_{\ell-1}) \right] = \mu_\ell^{\infty}(y_{\ell-1}), \text{ for } k \geq \bar{\ell}. $$

By inspection of the $S_\ell$, we can track the Pre’s of any $[y_{\ell-1}]$, simply by following the abstraction’s edges in the opposite direction, until we reach either an empty set or a state previously visited; the value $\bar{\ell}$ denotes the maximum number of steps it takes for this to happen. Let us now consider the data-driven perspective. It is easy to see that any trajectory of length

$$H = \bar{\ell} + \bar{\ell} \leq |\mathcal{Y}|^{\ell-1} + \bar{\ell} - 1,$$

ensures to collect all possible $\bar{\ell}$-sequences that can be generated from a given initial condition $x_0$. Thus, there exists a function $\varphi$, such that $\varphi(\bar{\ell}) = 1$ holds. By Corollary 1, for the data-driven $S^*$-CA
Let us consider a set $A$. Consider an over-approximation of the system’s behaviours stemming from the collected states $S$. It can either be disjoint from $S^N$ i.e. the internal behaviour never reaches any $y_T \in X^N_T$ or, in finite time, it reaches one of the collected states $y_T \in X^N_T$. In either situations, the only way to witness $y_T$ is by sampling them as initial portion of a trajectory. This event falls under the guarantees offered by the scenario theory; formally

$$V(\theta_T^N) = \mathbb{P} \left[ x_0 : B_{fT}(S(x_0)) \models \varphi_T \wedge y_T \not\in X_T^N \right] = \mathbb{P} \left[ x_0 : B_{\omega T}(S(x_0)) \models \varphi_T \wedge y_T \not\in X_T^N \right] \leq \varepsilon,$$

hence the data-driven abstraction probably behaviourally contains the concrete model with bound $\varepsilon = \varepsilon$, satisfying Proposition 3.

Note that if $S_T^N$ is blocking after collecting $N$ trajectories of length $T = H < |Y|^{-1} + \bar{F} - 1$ it can be rendered non-blocking by domino completion. In this case $S_T^N$ includes all the possible behaviours stemming from the collected $\bar{F}$-sequences, and, in this sense, provides an over-approximation of the system’s behaviours.

### A.2 Proof of Proposition 5

**Proof.** Consider $[y^i]$ the equivalence class containing the equilibrium, defined by the output map $h(x) = y^i$. By assumption $x_{eq}$ is an interior point of $[y^i]$. Define the following quantities

$$d_{min} = \max \{r > 0 \mid |x - x_{eq}|_2 \leq r \implies x \in [y^i]\}$$
$$d_{max} = \min \{r > 0 \mid |x - x_{eq}|_2 \geq r \implies x \not\in \mathcal{D}\}$$

Since $|A|_2 \leq \alpha < 1$, the set $S_0 = \{x \in [y^i] \mid |x - x_{eq}|_2 \leq d_{min}\}$ is such that 

$$S_0 \subseteq f^{-1}(S_0) \cup \mathcal{D}, \tag{21}$$

or equivalently, since $A$ is invertible, $f(S_0) \subseteq S_0$. Moreover, for all $x \in \mathcal{D}$,

$$\|f^{-1}(x - x_{eq})\|_2 \geq \alpha^{-1} \|x - x_{eq}\|_2.$$

Let us consider a set $Q$ such that $Q \cap S_0 = \emptyset$. For any point in $Q$, there exists a finite number of steps such that $\text{Pre}_{f}(Q) \cap \mathcal{D} = \emptyset$. Indeed, for all $x \in Q$

$$\|f^{-k}(x - x_{eq})\|_2 > \alpha^{-k} \min d_{min} \geq d_{max},$$

for all $k \geq \bar{k} = \left\lceil \frac{\log(d_{min})}{\log(d_{max})} \right\rceil$.

We conclude that

$$\text{Pre}_{f}^{-k}(Q) \cap \mathcal{D} = \emptyset,$$

i.e. $\text{Pre}_{f}(Q)$ is entirely outside $\mathcal{D}$. Therefore, recalling (15), it holds that

$$\mu_0^{-1}(Q) = \mu_{\bar{k}}^u(Q) = \mu_{\bar{k}}^u(Q). \tag{22}$$

In other words, there exists a finite number of times that the $\text{Pre}(\cdot)$ operator can be applied to any set in $\mathcal{D} \setminus S_0$ before the image does not intersect with the domain $\mathcal{D}$. Note that if only a lower bound on $d_{min}$ and/or an upper bound on $d_{max}$ are known we obtain an upper bound for $\bar{k}$, hence the same reasoning still applies.

Let us consider the partition $S_0$ containing the equilibrium. Combining (21) with the argument above, it is easy to see that

$$\text{Pre}_{f}^{-1}(S_0) \subseteq \text{Pre}_{f}(S_0) \cup \mathcal{D},$$

$$\text{Pre}_{f}(S_0) \cap \mathcal{D} = \text{Pre}_{f}^{-k}(S_0) \cap \mathcal{D} = \emptyset \tag{23}$$

We have thus proved that $\text{Pre}_{f}(S_0) = \mathcal{D}$, which gives

$$\mu_0^{-1}(S_0) = \mu_{\bar{k}}^u(Q) \tag{24}.$$
To conclude, we provide a bound for the union of $S_o$ with any other set $Q \in \mathcal{D}$. We have
\[ \mu_0^F(S_o \cup Q) = \mu_0^F(S_o) = \mu(\mathcal{D}) = 1. \]
Therefore
\[ \mu_k^0(S_o \cup Q) \geq \mu_k^0(S_o) \geq \mu_k^0(S_o) \rho^k \kappa = \mu_k^0(S_o \cup Q) \rho^k \kappa. \] (25)

Thus, by combining (23)-(24)-(25),
\[ \mu_k^0(W) \geq \mu_k^0(W) \cdot \min \left( 1 + \rho^{k-1} \sum_{i=0}^{z(k)-1} \rho^{-2(i+1)} \right)^{-1} \cdot \rho^{k-\kappa}, \]
for any $W \in \mathcal{D}$ as claimed. \hfill \qed