Precise Approximations of the Probability Distribution of a Markov Process in Time: An Application to Probabilistic Invariance

Sadegh Esmaeil Zadeh Soudjani[1] and Alessandro Abate[2,1]

[1] Delft Center for Systems & Control, TU Delft, The Netherlands
[2] Department of Computer Science, University of Oxford, UK
{S.EsmaeilZadehSoudjani,A.Abate}@tudelft.nl

Abstract. The goal of this work is to formally abstract a Markov process evolving over a general state space as a finite-state Markov chain, with the objective of precisely approximating the state probability distribution of the Markov process in time. The approach uses a partition of the state space and is based on the computation of the average transition probability between partition sets. In the case of unbounded state spaces, a procedure for precisely truncating the state space within a compact set is provided, together with an error bound that depends on the asymptotic properties of the transition kernel of the Markov process. In the case of compact state spaces, the work provides error bounds that depend on the diameters of the partitions, and as such the errors can be tuned. The method is applied to the problem of computing probabilistic invariance of the model under study, and the result is compared to an alternative approach in the literature.

1 Introduction

Verification techniques and tools for deterministic, discrete time, finite-state systems have been available for many years [9]. Formal methods in the stochastic context is typically limited to discrete state structures, either in continuous or in discrete time [3, 12]. Stochastic processes evolving over continuous (uncountable) spaces are often related to undecidable problems (the exception being when they admit analytical solutions). It is thus of interest to resort to formal approximation techniques that allow solving corresponding problems over finite discretizations of the original models. In order to relate the approximate solutions to the original problems, it is of interest to come up with precise bounds on the error introduced by the approximations. The use of formal approximations techniques for such complex models can be looked at from the perspective of the research on abstraction techniques, which are of wide use in formal verification.

Successful numerical schemes based on Markov chain approximations of stochastic systems in continuous time have been introduced in the literature, e.g. [10]. However, the finite abstractions are only related to the original models asymptotically (at the limit), with no explicit error bounds. This approach has

E. Ábrahám and K. Havelund (Eds.): TACAS 2014, LNCS 8413, pp. 547–561, 2014.

been applied to the approximate study of probabilistic reachability or safety of stochastic hybrid models in [8, 15]. In [1] a technique has been introduced to instead provide formal abstractions of discrete-time, continuous-space Markov models [2], with the objective of investigating their probabilistic invariance (safety) by employing probabilistic model checking over a finite Markov chain. In view of scalability, the approach has been improved and optimized in [7].

In this work we show that the approach in [1, 7] can be successfully employed to approximately compute the statistics in time of a stochastic process over a continuous state space. This additionally leads to an alternative method for probabilistic safety analysis of the process. We first provide a forward recursion for the approximate computation of the state distribution of a Markov process in time. The computation of the state distribution is based on a state-space partitioning procedure, and on the abstraction of the Markov process as a finite-state Markov chain. An upper bound on the error related to the approximation is formally derived. Based on the information from the state distribution, we show how the method can be used to approximately compute probabilistic invariance (safety) for discrete-time stochastic systems over general state spaces.

Probabilistic safety is the dual problem to probabilistic reachability. Over deterministic models reachability and safety have been vastly studied in the literature, and computational algorithms and tools have been developed based on both forward and backward reachability for these systems. Similarly, for the probabilistic models under study, we compare the presented approach (based on forward computations) with the existing approaches in the literature [1, 5–7] (which hinge on backward computations), particularly in terms of the introduced error.

The article is structured as follows. Section 2 introduces the model under study and discusses some structural assumptions needed for the abstraction procedure. The procedure comprises two separate parts: Section 3 describes the truncation of the dynamics of the model, whereas Section 4 details the abstraction of the dynamics (approximation of the transition kernel) – both parts formally assess the associated approximation error. Section 5 discusses the application of the procedure to the computation of probabilistic invariance, and compares it against an alternative approach in the literature.

2 Model, Preliminaries, and Goal of This Work

We consider a discrete time Markov process $\mathcal{M}$ defined over a general state space, which is characterized by a pair $(\mathcal{S}, T_s)$, where $\mathcal{S}$ is the continuous state space that we assume endowed with a metric and Borel measurable. We denote by $(\mathcal{S}, \mathcal{B}(\mathcal{S}), \mathcal{P})$ the probability structure on $\mathcal{S}$, with $\mathcal{B}(\mathcal{S})$ being the associated sigma algebra and $\mathcal{P}$ a probability measure to be characterized shortly. T_s is a conditional stochastic kernel that assigns to each point $s \in \mathcal{S}$ a probability measure $T_s(\cdot|s)$, so that for any measurable set $A \in \mathcal{B}(\mathcal{S})$, $\mathcal{P}(s(1) \in A|s(0) = s) = \int_A T_s(d\bar{s}|s)$. We assume that the stochastic kernel T_s admits a density function t_s, namely $T_s(d\bar{s}|s) = t_s(\bar{s}|s)d\bar{s}$.

Suppose that the initial state of the Markov process $\mathcal{M}$ is random and distributed according to the density function $\pi_0 : \mathcal{S} \to \mathbb{R}^{\geq 0}$. The state distribution of $\mathcal{M}$ at time $t \in \mathbb{N} \doteq \{1, 2, 3, \ldots\}$ is characterized by a density function $\pi_t : \mathcal{S} \to \mathbb{R}^{\geq 0}$, which fully describes the statistics of the process at t and is in particular such that, for all $A \in \mathcal{B}(\mathcal{S})$,

$$\mathbb{P}(s(t) \in A) = \int_A \pi_t(s) ds,$$

where the symbol $\mathbb{P}$ is loosely used to indicate the probability associated to events over the product space $\mathcal{S}^{t+1}$ with elements $\mathbf{s} = [s(0), s(1), \ldots, s(t)]$, whereas the bold typeset is constantly used in the sequel to indicate vectors.

The state density functions $\pi_t(\cdot)$ can be computed recursively, as follows:

$$\pi_{t+1}(\bar{s}) = \int_{\mathcal{S}} t_s(\bar{s}|s) \pi_t(s) ds \quad \forall \bar{s} \in \mathcal{S}. \tag{1}$$

In practice the forward recursion in (1) rarely yields a closed form for the density function $\pi_{t+1}(\cdot)$. A special instance where this is the case is represented by a linear dynamical system perturbed by Gaussian process noise: due to the closure property of the Gaussian distribution with respect to addition and multiplication by a constant, it is possible to explicitly write recursive formulas for the mean and the variance of the distribution, and thus express in a closed form the distribution in time of the solution of the model. In more general cases, it is necessary to numerically (hence, approximately) compute the density function of the model in time.

This article provides a numerical approximation of the density function of $\mathcal{M}$ as the probability mass function (pmf) of a finite-state Markov chain $\mathcal{M}_f$ in time. The Markov chain $\mathcal{M}_f$ is obtained as an abstraction of the concrete Markov process $\mathcal{M}$. The abstraction is associated with a guaranteed and tunable error bound, and algorithmically it leverages a state-space partitioning procedure. The procedure is comprised of two steps:

1. since the state space $\mathcal{S}$ is generally unbounded, it is first properly truncated;
2. subsequently, a partition of the truncated dynamics is introduced.

Section 3 discusses the error generated by the state-space truncation, whereas Section 4 describes the construction of the Markov chain by state-space partitioning. We employ the following example throughout the article as a running case study.

Example 1. Consider the one-dimensional stochastic dynamical system

$$s(t+1) = as(t) + b + \sigma w(t),$$

where the parameters $a, \sigma > 0$, whereas $b \in \mathbb{R}$, and such that $w(\cdot)$ is a process comprised of independent, identically distributed random variables with a standard normal distribution. The initial state of the process is selected uniformly in

the bounded interval $[\beta_0, \gamma_0] \subset \mathbb{R}$. The solution of the model is a Markov process, evolving over the state space $\mathcal{S} = \mathbb{R}$, and fully characterized by the conditional density function

$$t_s(\bar{s}|s) = \phi_\sigma(\bar{s} - as - b), \quad \phi_\sigma(u) = \frac{1}{\sigma\sqrt{2\pi}} e^{-u^2/2\sigma^2}. \qquad \square$$

We raise the following assumptions in order to be able to later relate the state density function of $\mathcal{M}$ to the probability mass function of $\mathcal{M}_f$.

Assumption 1. *For given sets $\Gamma \subset \mathcal{S}^2$ and $\Lambda_0 \subset \mathcal{S}$, there exist positive constants ϵ and ε_0, such that $t_s(\bar{s}|s)$ and $\pi_0(s)$ satisfy the following conditions:*

$$t_s(\bar{s}|s) \leq \epsilon \quad \forall (s, \bar{s}) \in \mathcal{S}^2 \backslash \Gamma, \text{ and } \pi_0(s) \leq \varepsilon_0 \quad \forall s \in \mathcal{S} \backslash \Lambda_0. \tag{2}$$

Assumption 2. *The density functions $\pi_0(s)$ and $t_s(\bar{s}|s)$ are (globally) Lipschitz continuous, namely there exist finite constants λ_0, λ_f, such that the following Lipschitz continuity conditions hold:*

$$|\pi_0(s) - \pi_0(s')| \leq \lambda_0 \|s - s'\| \quad \forall s, s' \in \Lambda_0, \tag{3}$$

$$|t_s(\bar{s}|s) - t_s(\bar{s}'|s)| \leq \lambda_f \|\bar{s} - \bar{s}'\| \quad \forall s, \bar{s}, \bar{s}' \in \mathcal{S}. \tag{4}$$

Moreover, there exists a finite constant M_f such that

$$M_f = \sup \left\{ \int_{\mathcal{S}} t_s(\bar{s}|s) ds \middle| \bar{s} \in \mathcal{S} \right\}. \tag{5}$$

The Lipschitz constants λ_0, λ_f are effectively computed by taking partial derivatives of the density functions $\pi_0(\cdot), t_s(\cdot|s)$ and maximizing its norm. The sets Λ_0 and Γ will be used to truncate the support of density functions $\pi_0(\cdot)$ and $t_s(\cdot|\cdot)$, respectively. Assumption 1 enables the precise study of the behavior of density functions $\pi_t(\cdot)$ over the truncated part of the state space. Further, the Lipschitz continuity conditions in Assumption 2 are essential to derive error bounds related to the abstraction of the Markov process over the truncated state space. In order to compute these error bounds, we assign the infinity norm to the space of bounded measurable functions over the state space $\mathcal{S}$, namely

$$\|f(s)\|_\infty = \sup_{s \in \mathcal{S}} |f(s)| \quad \forall f : \mathcal{S} \to \mathbb{R}.$$

In the sequel the function $\mathbb{I}_A(\cdot)$ denotes the indicator function of a set $A \subseteq \mathcal{S}$, namely $\mathbb{I}_A(s) = 1$, if $s \in A$; else $\mathbb{I}_A(s) = 0$.

Example 1 (Continued). Select the interval $\Lambda_0 = [\beta_0, \gamma_0]$ and define the set Γ by the linear inequality

$$\Gamma = \{(s, \bar{s}) \in \mathbb{R}^2 \big| |\bar{s} - as - b| \leq \alpha\sigma\}.$$

The initial density function π_0 of the process can be represented by the function

$$\psi_0(s) = \mathbb{I}_{[\beta_0, \gamma_0]}(s)/(\gamma_0 - \beta_0).$$

Then Assumption 1 holds with $\epsilon = \phi_1(\alpha)/\sigma$ and $\varepsilon_0 = 0$. The constant M_f in Assumption 2 is equal to $1/a$. Lipschitz continuity, as per (3) and (4), holds for constants $\lambda_0 = 0$ and $\lambda_f = 1/\left(\sigma^2\sqrt{2\pi e}\right)$. $\square$

3 State-Space Truncation Procedure, with Error Quantification

We truncate the support of the density functions π_0, t_s to the sets Λ_0, Γ respectively, and recursively compute support sets Λ_t, as in (7) that are associated to the density functions π_t. Then we employ the quantities ϵ, ε_0 in Assumption 1 to compute error bounds ε_t, as in (6), on the value of the density functions π_t outside the sets Λ_t. Finally we truncate the unbounded state space to $\Upsilon = \cup_{t=0}^{N} \Lambda_t$.

As intuitive, the error related to the spatial truncation depends on the behavior of the conditional density function t_s over the eliminated regions of the state space. Suppose that sets Γ, Λ_0 are selected such that Assumption 1 is satisfied with constants ϵ, ε_0: then Theorem 2 provides an upper bound on the error obtained from evaluating the density functions in time $\pi_t(\cdot)$ over the truncated regions of the state space.

Theorem 1. *Under Assumption 1 the functions π_t satisfy the bound*

$$0 \leq \pi_t(s) \leq \varepsilon_t \quad \forall s \in \mathcal{S} \backslash \Lambda_t,$$

where the quantities $\{\varepsilon_t\}_{t=0}^{N}$ are defined recursively by

$$\varepsilon_{t+1} = \epsilon + M_f \varepsilon_t, \tag{6}$$

whereas the support sets $\{\Lambda_t\}_{t=0}^{N}$ are computed as

$$\Lambda_{t+1} = \Pi_{\bar{s}} \left(\Gamma \cap (\Lambda_t \times \mathcal{S}) \right), \tag{7}$$

where $\Pi_{\bar{s}}$ denotes the projection map along the second set of coordinates[1].

Remark 1. Notice that if the shape of the sets Γ and Λ_0 is computationally manageable (e.g., polytopes) then it is possible to implement the computation of the recursion in (7) by available software tools, such as the MPT toolbox [11].

Further, notice that if for some t_0, $\Lambda_{t_0+1} \supset \Lambda_{t_0}$, then for all $t \geq t_0$, $\Lambda_{t+1} \supset \Lambda_t$. Similarly, we have that

- if for some t_0, $\Lambda_{t_0+1} \subset \Lambda_{t_0}$, then for all $t \geq t_0$, $\Lambda_{t+1} \subset \Lambda_t$.
- if for some t_0, $\Lambda_{t_0+1} = \Lambda_{t_0}$, then for all $t \geq t_0$, $\Lambda_t = \Lambda_{t_0}$.

To clarify the role of Γ in the computation of Λ_t, we emphasize that $\Lambda_{t+1} = \cup_{s \in \Lambda_t} \Xi(s)$, where Ξ depends only on Γ and is defined by the set-valued map

$$\Xi : \mathcal{S} \to 2^{\mathcal{S}}, \quad \Xi(s) = \{\bar{s} \in \mathcal{S} | (s, \bar{s}) \in \Gamma\}.$$

Figure 1 provides a visual illustration of the recursion in (7). □

Let us introduce a quantity $\kappa(t, M_f)$, which plays a role in the solution of (6) and will be frequently used shortly:

$$\kappa(t, M_f) = \begin{cases} \frac{1 - M_f^t}{1 - M_f}, & M_f \neq 1 \\ t, & M_f = 1. \end{cases} \tag{8}$$

[1] Recall that both Γ and $\Lambda \times \mathcal{S}$ are defined over $\mathcal{S}^2 = \mathcal{S} \times \mathcal{S}$.

The following theorem provides a truncation procedure, valid over a finite time horizon $\{0, 1, \ldots, N\}$, which reduces the state space $\mathcal{S}$ to the set $\Upsilon = \bigcup_{t=0}^{N} \Lambda_t$. The theorem also formally quantifies the associated truncation error.

Theorem 2. *Suppose that the state space of the process $\mathcal{M}$ has been truncated to the set $\Upsilon = \bigcup_{t=0}^{N} \Lambda_t$. Let us introduce the following recursion to compute functions $\mu_t : \mathcal{S} \to \mathbb{R}^{\geq 0}$ as an approximation of the density functions π_t:*

$$\mu_{t+1}(\bar{s}) = \mathbb{I}_{\Upsilon}(\bar{s}) \int_{\mathcal{S}} t_s(\bar{s}|s)\mu_t(s)ds, \quad \mu_0(s) = \mathbb{I}_{\Lambda_0}(s)\pi_0(s) \quad \forall \bar{s} \in \mathcal{S}. \tag{9}$$

Then the introduced approximation error is $\|\pi_t - \mu_t\|_\infty \leq \varepsilon_t$, for all $t \leq N$.

To recapitulate, Theorem 2 leads to the following procedure to approximate the density functions π_t of $\mathcal{M}$ over an unbounded state space $\mathcal{S}$:

1. truncate π_0 in such a way that μ_0 has a bounded support Λ_0;
2. truncate the conditional density function $t_s(\cdot|s)$ over a bounded set for all $s \in \mathcal{S}$, then quantify $\Gamma \subset \mathcal{S}^2$ as the support of the truncated density function;
3. leverage the recursion in (7) to compute the support sets Λ_t;
4. use the recursion in (9) to compute the approximate density functions μ_t over the set $\Upsilon = \cup_{t=0}^{N} \Lambda_t$. Note that the recursion in (9) is effectively computed over the set Υ, since $\mu_t(s) = 0$ for all $s \in \mathcal{S} \backslash \Upsilon$.

Note that we could as well deal with the support of $\mu_t(\cdot)$ over the time-varying sets Λ_t by adapting recursion (9) with $\mathbb{I}_{\Lambda_{t+1}}$ instead of $\mathbb{I}_{\Upsilon}$. While employing the (larger) set Υ may lead to a memory increase at each stage, it will considerably simplify the computations of the state-space partitioning and the abstraction as a Markov chain: indeed, employing time-varying sets Λ_t would render the partitioning procedure also time-dependent, and the obtained Markov chain would be time-inhomogeneous. We opt to work directly with Υ to avoid these difficulties.

Example 1 (Continued). We can easily obtain a closed form for the sets $\Lambda_t = [\beta_t, \gamma_t]$, via

$$\beta_{t+1} = a\beta_t + b - \alpha\sigma, \quad \gamma_{t+1} = a\gamma_t + b + \alpha\sigma.$$

The set Υ is the union of intervals $[\beta_t, \gamma_t]$. The error of the state-space truncation over set Υ is

$$\|\pi_t - \mu_t\|_\infty \leq \varepsilon_t = \kappa(t, M_f)\frac{\phi_1(\alpha)}{\sigma}, \quad M_f = \frac{1}{a}.$$

□

4 State-Space Partitioning Procedure, with Error Quantification

In this section we assume that the sets Γ, Λ_0 have been properly selected so that $\Upsilon = \cup_{i=0}^{N} \Lambda_t$ is bounded. In order to formally abstract process $\mathcal{M}$ as a finite

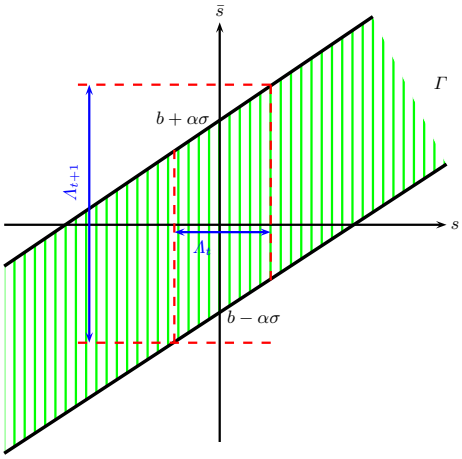

Fig. 1. Graphical representation of the recursion in (7) for Λ_t

Markov chain $\mathcal{M}_f$ and to approximate its state density functions, we select a finite partition of the bounded set Υ as $\Upsilon = \cup_{i=1}^{n}\mathcal{A}_i$, where the sets $\mathcal{A}_i$ have non-trivial measure. We then complete the partition over the whole state space $\mathcal{S} = \cup_{i=1}^{n+1}\mathcal{A}_i$ by considering the set $\mathcal{A}_{n+1} = \mathcal{S}\backslash\Upsilon$. This results in a finite Markov chain $\mathcal{M}_f$ with $n+1$ discrete abstract states in the set $\mathbb{N}_{n+1}\doteq\{1, 2, \cdots, n, n+1\}$, and characterized by the transition probability matrix $P = [P_{ij}] \in \mathbb{R}^{(n+1)^2}$, where the probability of jumping from any pair of states i to j (P_{ij}) is computed as

$$\begin{aligned} &P_{ij} = \tfrac{1}{\mathcal{L}(\mathcal{A}_i)}\int_{\mathcal{A}_j}\int_{\mathcal{A}_i} t_s(\bar{s}|s)dsd\bar{s} \quad \forall i \in \mathbb{N}_n, \\ &P_{(n+1)j} = \delta_{(n+1)j}, \end{aligned} \tag{10}$$

for all $j \in \mathbb{N}_{n+1}$, and where $\delta_{(n+1)j}$ is the Kronecker delta function (the abstract state $n+1$ of $\mathcal{M}_f$ is absorbing), and $\mathcal{L}(\cdot)$ denotes the Lebesgue measure of a set. The quantities in (10) are well-defined since the set Υ is bounded and the measures $\mathcal{L}(\mathcal{A}_i), i \in \mathbb{N}_n$, are finite and non-trivial.

Notice that matrix P for the Markov chain $\mathcal{M}_f$ is stochastic, namely

$$\begin{aligned} \sum_{j=1}^{n+1} P_{ij} &= \sum_{j=1}^{n+1}\frac{1}{\mathcal{L}(\mathcal{A}_i)}\int_{\mathcal{A}_j}\int_{\mathcal{A}_i} t_s(\bar{s}|s)dsd\bar{s} = \frac{1}{\mathcal{L}(\mathcal{A}_i)}\int_{\mathcal{A}_i}\left(\sum_{j=1}^{n+1}\int_{\mathcal{A}_j} t_s(\bar{s}|s)d\bar{s}\right)ds \\ &= \frac{1}{\mathcal{L}(\mathcal{A}_i)}\int_{\mathcal{A}_i}\int_{\mathcal{S}} t_s(\bar{s}|s)d\bar{s}ds = \frac{1}{\mathcal{L}(\mathcal{A}_i)}\int_{\mathcal{A}_i} ds = 1. \end{aligned}$$

The initial distribution of $\mathcal{M}_f$ is the pmf $\mathbf{p_0} = [p_0(1), p_0(2), \ldots, p_0(n+1)]$, and it is obtained from π_0 as $p_0(i) = \int_{\mathcal{A}_i}\pi_0(s)ds, \forall i \in \mathbb{N}_{n+1}$. Then the pmf associated to the state distribution of $\mathcal{M}_f$ at time t can be computed as $\mathbf{p_t} = \mathbf{p_0}P^t$.

It is intuitive that the discrete pmf $\mathbf{p_t}$ of the Markov chain $\mathcal{M}_f$ approximates the continuous density function π_t of the Markov process $\mathcal{M}$. In the rest of the section we show how to formalize this relationship: $\mathbf{p_t}$ is used to construct an

approximation function, denoted by ψ_t, of the density function π_t. Theorem 3 shows that ψ_t is a piece-wise constant approximation (with values that are the entries of the pmf $\mathbf{p_t}$ normalized by the Lebesgue measure of the associated partition set) of the original density function π_t. Moreover, under the continuity assumption in (4) (ref. Lemma 1) we can establish the Lipschitz continuity of π_t, which enables the quantification in Theorem 3 of the error of its piece-wise constant approximation.

Lemma 1. *Suppose that the inequality in* (4) *holds. Then the state density functions* $\pi_t(\cdot)$ *are globally Lipschitz continuous with constant* λ_f *for all* $t \in \mathbb{N}$*:*

$$|\pi_t(s) - \pi_t(s')| \leq \lambda_f \|s - s'\| \quad \forall s, s' \in \mathcal{S}.$$

Theorem 3. *Under Assumptions 1 and 2, the functions* $\pi_t(\cdot)$ *can be approximated by piece-wise constant functions* $\psi_t(\cdot)$*, defined as*

$$\psi_t(s) = \sum_{i=1}^{n} \frac{p_t(i)}{\mathcal{L}(\mathcal{A}_i)} \mathbb{I}_{\mathcal{A}_i}(s) \quad \forall t \in \mathbb{N}, \tag{11}$$

where $\mathbb{I}_B(\cdot)$ *is the indicator function of a set* $B \subset \mathcal{S}$*. The approximation error is upper-bounded by the quantity*

$$\|\pi_t - \psi_t\|_\infty \leq \varepsilon_t + E_t \quad \forall t \in \mathbb{N}, \tag{12}$$

where E_t *can be recursively computed as*

$$E_{t+1} = M_f E_t + \lambda_f \delta, \quad E_0 = \lambda_0 \delta, \tag{13}$$

and δ *is an upper bound on the diameters of the partition sets* $\{\mathcal{A}_i\}_{i=1}^n$*, namely* $\delta = \sup\{\|s - s'\|,\ s, s' \in \mathcal{A}_i,\ i \in \mathbb{N}_n\}$.

Note that the functions ψ_t are defined over the whole state space $\mathcal{S}$, but (11) implies that they are equal to zero outside the set Υ.

Corollary 1. *The recursion in* (13) *admits the explicit solution*

$$E_t = \left[\kappa(t, M_f)\lambda_f + M_f^t \lambda_0\right] \delta,$$

where $\kappa(t, M_f)$ *is introduced in* (8).

Underlying Theorem 3 is the fact that $\psi_t(\cdot)$ are in general sub-stochastic density functions:

$$\begin{aligned}\int_{\mathcal{S}} \psi_t(s)ds &= \int_{\mathcal{S}} \sum_{i=1}^{n} \frac{p_t(i)}{\mathcal{L}(\mathcal{A}_i)} \mathbb{I}_{\mathcal{A}_i}(s)ds = \sum_{i=1}^{n} \frac{p_t(i)}{\mathcal{L}(\mathcal{A}_i)} \int_{\mathcal{S}} \mathbb{I}_{\mathcal{A}_i}(s)ds \\ &= \sum_{i=1}^{n} \frac{p_t(i)}{\mathcal{L}(\mathcal{A}_i)} \mathcal{L}(\mathcal{A}_i) = \sum_{i=1}^{n} p_t(i) = 1 - p_t(n+1) \leq 1.\end{aligned}$$

This is clearly due to the fact that we are operating on the dynamics of $\mathcal{M}$ truncated over the set Υ. It is thus intuitive that the approximation procedure and the derived error bounds are also valid for the case of sub-stochastic density functions, namely

$$\int_{\mathcal{S}} t_s(\bar{s}|s)d\bar{s} \leq 1 \quad \forall s \in \mathcal{S}, \quad \int_{\mathcal{S}} \pi_0(s)ds \leq 1,$$

the only difference being that the obtained Markov chain $\mathcal{M}_f$ is as well sub-stochastic.

Further, whenever the Lipschitz continuity requirement on the initial density function, as per (3) in Assumption 2, does not hold, (for instance, this is the case when the initial state of the process is deterministic) we can relax this continuity assumption on the initial distribution of the process by starting the discrete computation from the time step $t = 1$. In this case we define the pmf $\mathbf{p_1} = [p_1(1), p_1(2), \ldots, p_1(n+1)]$, where

$$p_1(i) = \int_{\mathcal{A}_i} \int_{\mathcal{S}} t_s(\bar{s}|s)\pi_0(s)dsd\bar{s} \quad \forall i \in \mathbb{N}_{n+1},$$

and derive $\mathbf{p_t} = \mathbf{p_1}P^{t-1}$ for all $t \in \mathbb{N}$. Theorem 3 follows along similar lines, except for eqn. (13), where the initial error is set to $E_0 = 0$ and the time-dependent terms E_t can be derived as $E_t = \kappa(t, M_f)\lambda_f\delta$.

It is important to emphasize the computability of the derived errors and the fact that they can be tuned. Further, in order to attain abstractions that are practically useful, it imperative to seek improvements on the derived error bounds: in particular, the approximation errors can be computed locally (under corresponding local Lipschitz continuity assumptions), following the procedures discussed in [7].

Example 1 (Continued). The error of proposed Markov chain abstraction can be expressed as

$$\|\pi_t - \psi_t\|_\infty \leq \kappa(t, M_f)\left[\frac{\delta}{\sigma^2\sqrt{2\pi e}} + \frac{\phi_1(\alpha)}{\sigma}\right], \quad M_f = \frac{1}{a}.$$

The error can be tuned in two distinct ways:

1. by selecting larger values for α, which on the one hand leads to a less conservative truncation, but on the other requires the partition of a larger interval;
2. by reducing partitions diameter δ, which of course results in a larger cardinality of the partition sets.

Let us select values $b = 0, \beta_0 = 0, \gamma_0 = 1, \sigma = 0.1$, and time horizon $N = 5$. For $a = 1.2$ we need to partition the interval $\Upsilon = [-0.75\alpha, 2.49 + 0.75\alpha]$, which results in the error $\|\pi_t - \psi_t\|_\infty \leq 86.8\delta + 35.9\phi_1(\alpha)$ for all $t \leq N$. For $a = 0.8$ we need to partition the smaller interval $\Upsilon = [-0.34\alpha, 0.33 + 0.34\alpha]$, which results in the error $\|\pi_t - \psi_t\|_\infty \leq 198.6\delta + 82.1\phi_1(\alpha)$ for all $t \leq N$. In the case of $a = 1.2$, we partition a larger interval and obtain a smaller error, while for $a = 0.8$ we

partition a smaller interval with correspondingly a larger error. It is obvious that the parameters δ, α can be chosen properly to ensure that a certain error precision is met. This simple model admits a solution in closed form, and its state density functions can be obtained as the convolution of a uniform distribution (the contribution of initial state) and a zero-mean Gaussian distribution with time-dependent variance (the contributions of the process noise). Figure 2 displays the original and the approximated state density functions for the set of parameters $\alpha = 2.4, \delta = 0.05$. □

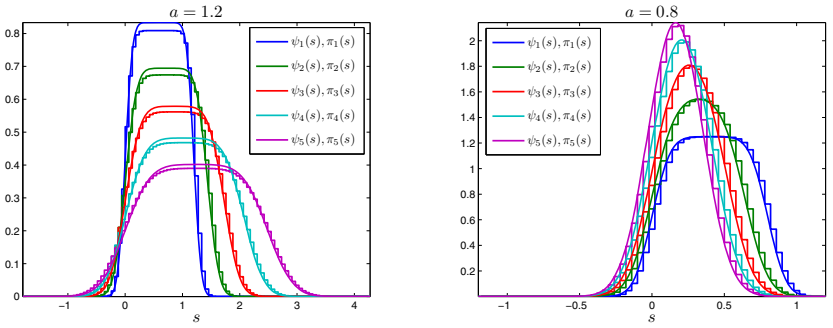

Fig. 2. Piece-wise constant approximation of the state density function $\psi_t(\cdot)$, compared to the actual function $\pi_t(\cdot)$ (derived analytically) for $a = 1.2$ (left) and $a = 0.8$ (right)

5 Application of the Formal Approximation Procedure to the Probabilistic Invariance Problem

The problem of probabilistic invariance (safety) for general Markov processes has been theoretically characterized in [2] and further investigated computationally in [1, 4–6]. With reference to a discrete-time Markov process $\mathcal{M}$ over a continuous state space $\mathcal{S}$, and a safe set $\mathcal{A} \in \mathcal{B}(\mathcal{S})$, the goal is to quantify the probability

$$p_s^N(\mathcal{A}) = \mathbb{P}\{s(t) \in \mathcal{A}, \text{ for all } t \in [0, N] | s(0) = s\}.$$

More generally, it is of interest to quantify the probability $p_{\pi_0}^N(\mathcal{A})$, where the initial condition of the process $s(0)$ is a random variable characterized by the density function $\pi_0(\cdot)$. We present a forward computation of probabilistic invariance by application of the approximation procedure above in Section 5.1, then review results on backward computation [1, 4–6] in Section 5.2. Section 5.3 compares the two approaches.

5.1 Forward Computation of Probabilistic Invariance

The approach for approximating the density function of a process in time can be easily employed for the approximate computation of probabilistic invariance. Define sub-density functions $W_t : \mathcal{S} \to \mathbb{R}^{\geq 0}$, characterized by

$$W_{t+1}(\bar{s}) = \mathbb{I}_{\mathcal{A}}(\bar{s}) \int_{\mathcal{S}} W_t(s) t_s(\bar{s}|s) ds, \quad W_0(\bar{s}) = \mathbb{I}_{\mathcal{A}}(\bar{s}) \pi_0(\bar{s}) \quad \forall \bar{s} \in \mathcal{S}. \tag{14}$$

Then the solution of the problem is obtained as $p^N_{\pi_0}(\mathcal{A}) = \int_{\mathcal{S}} W_N(s) ds$. A comparison of the recursions in (14) and in (9) reveals how probabilistic invariance can be computed as a special case of the approximation procedure. In applying the procedure, the only difference consists in replacing set Υ by $\mathcal{A}$, and in restricting Assumption 2 to hold over the safe set (the solution over the complement of this set is trivially known) – in this case the error related to the truncation of the state space can be disregarded. The procedure consists in partitioning the safe set, in constructing the Markov chain $\mathcal{M}_f$ as per (10), and in computing $\psi_t(\cdot)$ as an approximation of $W_t(\cdot)$ based on (11). The error of this approximation is $\|W_t - \psi_t\|_\infty \leq E_t$, which results in the following:

$$\left| p^N_{\pi_0}(\mathcal{A}) - \int_{\mathcal{A}} \psi_t(s) ds \right| \leq E_N \mathcal{L}(\mathcal{A}) = \kappa(N, M_f) \lambda_f \delta \mathcal{L}(\mathcal{A}) \doteq E_f.$$

Note that these sub-density functions satisfy the inequalities

$$1 \geq \int_{\mathcal{S}} W_0(s) ds \geq \int_{\mathcal{S}} W_1(s) ds \geq \cdots \geq \int_{\mathcal{S}} W_N(s) ds \geq 0.$$

5.2 Backward Computation of Probabilistic Invariance

The contributions in [1, 4–6] have characterized specifications in PCTL with a formulation based on backward recursions. In particular, the computation of probabilistic invariance is obtained via the value functions $V_t : \mathcal{S} \to [0, 1]$, which are characterized as

$$V_t(s) = \mathbb{I}_{\mathcal{A}}(s) \int_{\mathcal{S}} V_{t+1}(\bar{s}) t_s(\bar{s}|s) d\bar{s}, \quad V_N(s) = \mathbb{I}_{\mathcal{A}}(s) \quad \forall s \in \mathcal{S}. \tag{15}$$

The desired probabilistic invariance is expressed as $p^N_{\pi_0}(\mathcal{A}) = \int_{\mathcal{S}} V_0(s) \pi_0(s) ds$. The value functions always map the state space to the interval $[0, 1]$ and they are non-increasing, $V_t(s) \leq V_{t+1}(s)$ for any fixed $s \in \mathcal{S}$. [1, 4–6] discuss efficient algorithms for the approximate computation of the quantity $p^N_{\pi_0}(\mathcal{A})$, relying on different assumptions on the model under study. The easiest and most straightforward procedure is based on the following assumption [1].

Assumption 3. *The conditional density function of the process is globally Lipschitz continuous within the safe set with respect to the conditional state. Namely, there exists a finite constant λ_b, such that*

$$|t(\bar{s}|s) - t(\bar{s}|s')| \leq \lambda_b \|s - s'\| \quad \forall s, s', \bar{s} \in \mathcal{A}.$$

A finite constant M_b is introduced as $M_b = \sup_{s \in \mathcal{A}} \int_{\mathcal{A}} t_s(\bar{s}|s) d\bar{s} \leq 1$.

The procedure introduces a partition of the safe set $\mathcal{A} = \cup_{i=1}^{n} \mathcal{A}_i$ and extends it to $\mathcal{S} = \cup_{i=1}^{n+1} \mathcal{A}_i$, with $\mathcal{A}_{n+1} = \mathcal{S} \backslash \mathcal{A}$. Then it selects arbitrary representative

points $s_i \in \mathcal{A}_i$ and constructs a finite-state Markov chain $\mathcal{M}_b$ over the finite state space $\{s_1, s_2, \ldots, s_{n+1}\}$, endowed with transition probabilities

$$P(s_i, s_j) = \int_{\mathcal{A}_j} t_s(\bar{s}|s_i)d\bar{s}, \quad P(s_{n+1}, s_j) = \delta_{(n+1)j}, \tag{16}$$

for all $i \in \mathbb{N}_n, j \in \mathbb{N}_{n+1}$. The error of such an approximation is [5]:

$$E_b = \kappa(N, M_b)\lambda_b \delta \mathcal{L}(\mathcal{A}),$$

where δ is the max partitions diameter, $\mathcal{L}(\mathcal{A})$ is the Lebesgue measure of set $\mathcal{A}$.

5.3 Comparison of the Two Approaches

We first compare the two constructed Markov chains. The Markov chain in the forward approach $\mathcal{M}_f$ is a special case of Markov chain of backward approach $\mathcal{M}_b$, where the representative points can be selected intelligently to determine the average probability of jumping from one partition set to another. More specifically, the quantities (10) are a special case of those in (16) (based on mean value theorem for integration). We claim that this leads to a less conservative (smaller) error bound for the approximation.

The forward computation is in general more informative than the backward computation since it provides not only the solution of the safety problem in time, but also the state distribution over the safe set. Further the forward approach may provide some insight to the solution of the infinite-horizon safety problem [16, 17], for a given initial distribution. Finally, the forward approach can be used to approximate the solution of safety problem over unbounded safe sets, while boundedness of the safe set is required in all the results in the literature that are based on backward computations.

Next, we compare errors and related assumptions. The error computations rely on different assumptions: the Lipschitz continuity of the conditional density function with respect to the current or to the next states, respectively. Further, the constants M_f and M_b are generally different and play an important role in the error. M_b represents the maximum probability of remaining inside the safe set, while M_f is an indication of the maximum concentration at one point over a single time-step of the process evolution. M_b is always less than or equal to one, while M_f could be any finite positive number.

Example 1 (Continued) The constants λ_f, M_f and λ_b, M_b for the one dimensional dynamical system of Example 1 are

$$\lambda_f = \frac{1}{\sigma^2\sqrt{2\pi e}}, \quad \lambda_b = a\lambda_f, \quad M_f = \frac{1}{a}, \quad M_b \leq 1.$$

If $0 < a < 1$ the system trajectories converge to an equilibrium point (in expected value). In this case the system state gets higher chances of being in the neighborhood of the equilibrium in time and the backward recursion provides a better error bound. If $a > 1$ the system trajectories tend to diverge with time. In this case the forward recursion provides a much better error bound, compared to the backward recursion.

For the numerical simulation we select a safety set $\mathcal{A} = [0, 1]$, a noise level $\sigma = 0.1$, and a time horizon $N = 10$. The solution of the safety problem for the two cases $a = 1.2$ and $a = 0.8$ is plotted in Figure 3. We have computed constants $\lambda_f = 24.20, M_b = 1$ in both cases, while $\lambda_b = 29.03, M_f = 0.83$ for the first case, and $\lambda_b = 19.36, M_f = 1.25$ for the second case. We have selected the center of the partition sets (distributed uniformly over the set $\mathcal{A}$) as representative points for Markov chain $\mathcal{M}_b$. In order to compare the two approaches, we have assumed the same computational effort (related to the same partition size of $\delta = 0.7 \times 10^{-4}$), and have obtained an error $E_f = 0.008, E_b = 0.020$ for $a = 1.2$ and $E_f = 0.056, E_b = 0.014$ for $a = 0.8$. The simulations show that the forward approach works better for $a = 1.2$, while the backward approach is better suitable for $a = 0.8$. Note that the approximate solutions provided by the two approaches are very close: the difference of the transition probabilities computed via the Markov chains $\mathcal{M}_f, \mathcal{M}_b$ are in the order of 10^{-8}, and the difference in the approximate solutions (black curve in Figure 3) is in the order of 10^{-6}. This has been due to the selection of very fine partition sets that have resulted in small abstraction errors. □

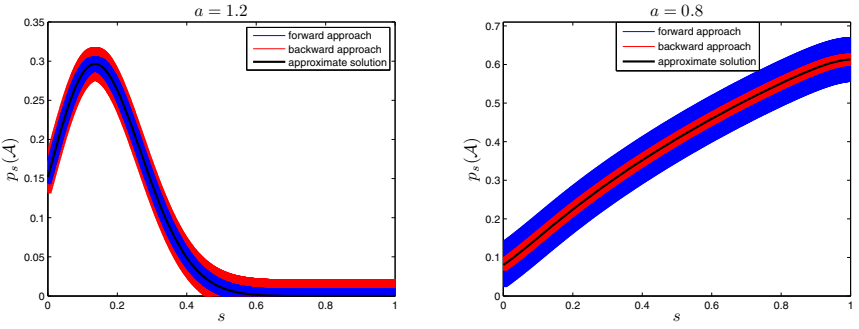

Fig. 3. Approximate solution of the probabilistic invariance problem (black line), together with error intervals of forward (red band) and backward (blue band) approaches, for $a = 1.2$ (left) and $a = 0.8$ (right)

Remark 2. Over deterministic models, [13] compares forward and backward reachability analysis and provides insights on their differences: the claim is that for systems with significant contraction, forward reachability is more effective than backward reachability because of numerical stability issues. On the other hand, for the probabilistic models under study, the result indicates that under Lipschitz continuity of the transition kernel the backward approach is more effective in systems with convergence in the state distribution. If we treat deterministic systems as special (limiting) instances of stochastic systems, our result is not contradicting with [13] since the Lipschitz continuity assumption on the transition kernels of probabilistic models does not hold over deterministic ones. □

Motivated by the previous case study we study how convergence properties of a Markov process are related to the constant M_f.

Theorem 4. *Assume that the initial density function $\pi_0(s)$ is bounded and that the constant M_f is finite and $M_f < 1$. If the state space is unbounded, the sequence of density functions $\{\pi_t(s)|t \geq 0\}$ uniformly exponentially converges to zero. The sequence of probabilities $\mathbb{P}\{s(t) \in \mathcal{A}\}$ and the corresponding solution of the safety problem for any compact safe set $\mathcal{A}$ exponentially converge to zero.*

Theorem 4 indicates that under the invoked assumptions the probability "spreads out" over the unbounded state space as time progresses. Moreover, the theorem ensures the absence of absorbing sets [16, 17], which are indeed known to characterize the solution of infinite-horizon properties. Example 2 studies the relationship between constant M_f and the stability of linear stochastic difference equations.

Example 2. Consider the stochastic linear difference equations

$$x(t+1) = Ax(t) + w(t), \quad x(\cdot), w(\cdot) \in \mathbb{R}^n,$$

where $w(\cdot)$ are i.i.d. random vectors with known distributions. For such systems $M_f = 1/|\det A|$, then the condition $M_f < 1$ implies instability of the system in expected value. Equivalently, mean-stability of the system implies $M_f \geq 1$. Note that for this class of systems $M_f > 1$ does not generally imply stability, since $\det A$ is only the product of the eigenvalues of the system. □

The Lipschitz constants λ_f and λ_b have a different nature. Example 3 clarifies this point.

Example 3. Consider the dynamical system

$$s(t+1) = f(s(t), w(t)),$$

where $w(\cdot)$ are i.i.d. with known distribution $t_w(\cdot)$. Suppose that the vector field $f : \mathbb{R}^{\mathfrak{n}} \times \mathbb{R}^{\mathfrak{n}} \to \mathbb{R}^{\mathfrak{n}}$ is continuously differentiable and that the matrix $\frac{\partial f}{\partial w}$ is invertible. Then the *implicit function theorem* guarantees the existence and uniqueness of a function $g : \mathbb{R}^{\mathfrak{n}} \times \mathbb{R}^{\mathfrak{n}} \to \mathbb{R}^{\mathfrak{n}}$ such that $w(t) = g(s(t+1), s(t))$. The conditional density function of the system in this case is [14]:

$$t_s(\bar{s}|s) = \left|\det\left[\frac{\partial g}{\partial \bar{s}}(\bar{s}, s)\right]\right| t_w(g(\bar{s}, s)).$$

The Lipschitz constants λ_f, λ_b are specified by the dependence of function $g(\bar{s}, s)$ from the variables $\bar{s}, s$, respectively. As a special case the invertability of $\frac{\partial f}{\partial w}$ is guaranteed for systems with additive process noise, namely $f(s, w) = f_a(s) + w$. Then $g(\bar{s}, s) = \bar{s} - f_a(s)$, λ_f is the Lipschitz constant of $t_w(\cdot)$, while λ_b is the multiplication of Lipschitz constant of $t_w(\cdot)$ and of $f_a(\cdot)$. □

References

1. Abate, A., Katoen, J.-P., Lygeros, J., Prandini, M.: Approximate model checking of stochastic hybrid systems. European Journal of Control 6, 624–641 (2010)
2. Abate, A., Prandini, M., Lygeros, J., Sastry, S.: Probabilistic reachability and safety for controlled discrete time stochastic hybrid systems. Automatica 44(11), 2724–2734 (2008)

3. Baier, C., Katoen, J.-P., Hermanns, H.: Approximate symbolic model checking of continuous-time Markov chains (Extended abstract). In: Baeten, J.C.M., Mauw, S. (eds.) CONCUR 1999. LNCS, vol. 1664, pp. 146–162. Springer, Heidelberg (1999)
4. Esmaeil Zadeh Soudjani, S., Abate, A.: Adaptive gridding for abstraction and verification of stochastic hybrid systems. In: Proceedings of the 8th International Conference on Quantitative Evaluation of Systems, Aachen, DE, pp. 59–69 (September 2011)
5. Esmaeil Zadeh Soudjani, S., Abate, A.: Higher-Order Approximations for Verification of Stochastic Hybrid Systems. In: Chakraborty, S., Mukund, M. (eds.) ATVA 2012. LNCS, vol. 7561, pp. 416–434. Springer, Heidelberg (2012)
6. Esmaeil Zadeh Soudjani, S., Abate, A.: Probabilistic invariance of mixed deterministic-stochastic dynamical systems. In: ACM Proceedings of the 15th International Conference on Hybrid Systems: Computation and Control, Beijing, PRC, pp. 207–216 (April 2012)
7. Esmaeil Zadeh Soudjani, S., Abate, A.: Adaptive and sequential gridding procedures for the abstraction and verification of stochastic processes. SIAM Journal on Applied Dynamical Systems 12(2), 921–956 (2013)
8. Koutsoukos, X., Riley, D.: Computational methods for reachability analysis of stochastic hybrid systems. In: Hespanha, J.P., Tiwari, A. (eds.) HSCC 2006. LNCS, vol. 3927, pp. 377–391. Springer, Heidelberg (2006)
9. Kurshan, R.P.: Computer-Aided Verification of Coordinating Processes: The Automata-Theoretic Approach. Princeton Series in Computer Science. Princeton University Press (1994)
10. Kushner, H.J., Dupuis, P.G.: Numerical Methods for Stochastic Control Problems in Continuous Time. Springer, New York (2001)
11. Kvasnica, M., Grieder, P., Baotić, M.: Multi-parametric toolbox, MPT (2004)
12. Kwiatkowska, M., Norman, G., Segala, R., Sproston, J.: Verifying quantitative properties of continuous probabilistic timed automata. In: Palamidessi, C. (ed.) CONCUR 2000. LNCS, vol. 1877, pp. 123–137. Springer, Heidelberg (2000)
13. Mitchell, I.M.: Comparing forward and backward reachability as tools for safety analysis. In: Bemporad, A., Bicchi, A., Buttazzo, G. (eds.) HSCC 2007. LNCS, vol. 4416, pp. 428–443. Springer, Heidelberg (2007)
14. Papoulis, A.: Probability, Random Variables, and Stochastic Processes, 3rd edn. Mcgraw-hill (1991)
15. Prandini, M., Hu, J.: Stochastic reachability: Theory and numerical approximation. In: Cassandras, C.G., Lygeros, J. (eds.) Stochastic Hybrid Systems. Automation and Control Engineering Series, vol. 24, pp. 107–138. Taylor & Francis Group/CRC Press (2006)
16. Tkachev, I., Abate, A.: On infinite-horizon probabilistic properties and stochastic bisimulation functions. In: Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference, Orlando, FL, pp. 526–531 (December 2011)
17. Tkachev, I., Abate, A.: Characterization and computation of infinite-horizon specifications over markov processes. Theoretical Computer Science 515, 1–18 (2014)