Approximation Metrics Based on Probabilistic Bisimulations for General State-Space Markov Processes: A Survey

Alessandro Abate [*],[1]

Delft Center for Systems & Control
Technische Universiteit Delft
Mekelweg 2, 2628 CD Delft
The Netherlands

Abstract

This article provides a survey of approximation metrics for stochastic processes. We deal with Markovian processes in discrete time evolving on general state spaces, namely on domains with infinite cardinality and endowed with proper measurability and metric structures. The focus of this work is to discuss approximation metrics between two such processes, based on the notion of probabilistic bisimulation: in particular we investigate metrics characterized by an approximate variant of this notion. We suggest that metrics between two processes can be introduced essentially in two distinct ways: the first employs the probabilistic conditional kernels underlying the two stochastic processes under study, and leverages notions derived from algebra, logic, or category theory; whereas the second looks at distances between trajectories of the two processes, and is based on the dynamical properties of the two processes (either their syntax, via the notion of bisimulation function; or their semantics, via sampling techniques). The survey moreover covers the problem of constructing formal approximations of stochastic processes according to the introduced metrics.

Keywords: Markov Processes, Formal Approximations, Probabilistic Bisimulations, Metrics over Probability Measures, Lyapunov Theory, Stochastic Contractivity, Probabilistic Reachability, Randomized Methods.

1 Motivations and Objective

In order to cope with the increasing complexity of real-world engineering systems and with the intractability of their corresponding mathematical models, a number of studies have explored the development of techniques aimed at quantitatively

putting in relationship two models, a concrete one and an abstract one. The latter is generally obtained via a simplification of the first (e.g., by means of model reduction, low-order approximation, state-space clustering, or lumping and factorization) and represents a tractable version of it. Ideally, the abstract model should be, in a certain sense, equivalent to the concrete one. Equivalence between models is usually expressed with the notion of language correspondence, trajectory or trace equality, or with that of bisimulation relation between pairs of states of the two models. Unidirectional and thus less stringent versions of the notion of bisimulation are also used to express the idea of inclusion between the abstract model dynamics and that of the concrete one: in this sense, the abstract model represents a simulation of the concrete one.

From a different perspective, since frequently the exact notion of bisimulation translates into rather conservative requirements on the models under study, and because it is a notion that lacks robustness (against model parameter perturbations, for instance), the concept of approximate bisimulation has been introduced as a relaxed version of that of strict bisimulation. This approximate concept leads to the use of proper metrics (or pseudo-metrics) over the dynamics of the models. The use of approximate relations between models appears to be quite pertinent for models that are dynamically rich, such as models with continuous (uncountable) or even hybrid state spaces, as well as stochastic processes. For the latter in particular, the use of approximate notions allows the development of metrics that are robust to small perturbations of the models parameters, and accommodates quantitative correspondences between the realization likelihood of models trajectories.

This contribution focuses on probabilistic processes with general (continuous) state spaces and aims at surveying and discussing approximation metrics between pairs of such processes. In particular, we provide an overview of results in the literature that are based on the notion of approximate bisimulation. We decide to concentrate on approximate version of the notion of strong (rather than weak) bisimulation, and we only touch upon the concepts of (probabilistic) simulations.

This work suggests that metrics between processes, based on the concept of approximate bisimulation, can be introduced essentially in two separate ways. The first approach employs the probabilistic conditional kernels underlying the stochastic processes under study – in this sense, the approximation comes from metrics between (marginals of) probability measures related to the two processes. The second procedure looks at distance metrics between trajectories of the two processes and utilizes the dynamical properties of the two processes to define such metrics: this can be done either by analyzing the models syntax, or by directly employing their semantics in order to compare realizations of the two models.

The two approaches are depicted in Figure 1. The blue and red dots represent realizations of two "similar" stochastic processes, which evolve over the Euclidean plane and are both initialized on the black dot (top right). The first approach refers to the magnified square at the bottom, which portrays the two corresponding conditional kernels. The second approach is pictorially represented by the top square, where the (point-wise) distance between the realization of the trajectories of the
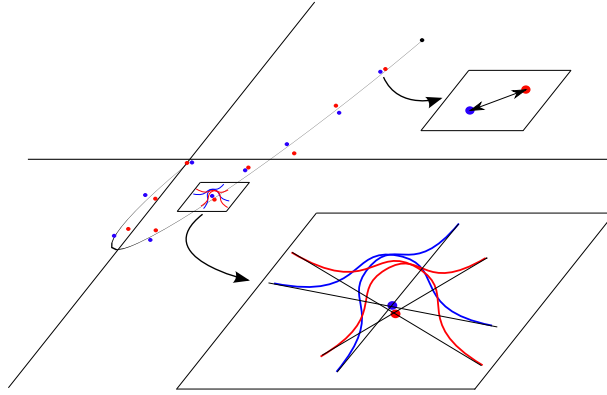
Fig. 1. Approximation metrics for general state-space Markov processes: two approaches.

two processes is highlighted.

This work is structured as follows.

- **Section 2** provides a comprehensive coverage of the work on simulations, bisimulations and approximate versions thereof, with focus on probabilistic models living on general state spaces.
- **Section 3** introduces and discusses the models under study.
- **Section 4** presents the concept of exact and approximate (strong) probabilistic bisimulation, and provides related characterizations based on algebra, logic, and category theory.
- **Section 5** puts forward the definition and the characterization of the notion of probabilistic bisimulation function, which leads to the introduction of approximate metrics between the trajectories of two processes.
- **Section 6** finally looks at semantic-based computations of distance metrics between comparable probabilistic processes.
- **Section 7** discusses the surveyed techniques and looks forward at future research directions.

2   Review of Literature Background

The concept of strong probabilistic bisimulation over a discrete-time, finite-state Markov chain has been introduced in [46], based on earlier notions for non-probabilistic models [49,51]. The work in [36] uses similar notions for Markov decision processes with finite state spaces, and puts forward procedures for finding factored bisimilar models. The notion of weak bisimulation is discussed in [9,38,54] for a number of (finite-state) probabilistic processes. The contributions in [39,59] cover the notion of probabilistic simulation relations for classes of probabilistic automata. [10,11] provide a recapitulation and draw relationships between these notions. These concepts are of applicative interest and build on earlier work on approximation techniques, such as that of lumpability for Markov chains [13]

and in queuing theory [41].

The interesting work in [24] discusses approximate notions of bisimulations for finite state labeled Markov chains, and elaborates on this notions by using a logical approach as well as one based on games. The use of approximate notions is advocated in [32] and motivated by robustness issues related to the verification of specifications over probabilistic models. Furthermore, approximate notions appear much less restrictive than the exact one, particularly when applied over models with continuous state spaces – this is precisely what has been observed also for deterministic models, where notions of exact bisimulation have been developed only for limited classes of models, e.g. timed automata [7] (via the region graph construction), linear hybrid automata [37], o-minimal hybrid systems [45] and certain classes of linear (control) systems [50,61]. The introduction of approximate versions [34] based on distance between trajectories of deterministic models has lead to the study of approximate abstractions for nonlinear [55] and switched systems [35].

For continuous space processes (namely, discrete-time labeled Markov processes as in Section 3), [21] provides a relational and logical characterization of bisimulation (see Section 4). Alternatively, probabilistic bisimulations relations can be introduced via coalgebraic [20] or categorical arguments [64]. Building on these results, the material in [22] is relevant in that metrics for labeled Markov processes are discussed (see Section 4), whereas [23] proposes metrics via weak bisimulations, and the contributions in [29,30] discusses metrics for respectively finite- and infinite-state Markov decision processes.

Related to the notions above, [60] introduces exact bisimulations for communicating piecewise-deterministic Markov processes (which are models related to [38]), [25] discusses bisimulation of continuous-time processes, [8] elaborates abstraction notions based on bisimulations for probabilistic process algebras, whereas [14] attempts definitions of bisimulations for stochastic hybrid models [12,17]. None of these works proposes approximate variants of the respective exact notions.

With focus on probabilistic models and on the development of metrics over systems trajectories (realizations) [34], the concept of probabilistic bisimulation functions (see Section 5) is introduced in [40] and elaborated in [1]. The recent work in [63] puts forward a reachability problem to find metrics between discrete-time stochastic processes.

From a different perspective, [5] puts forward an approach based on randomization techniques to characterize approximation distances between processes over finite time horizons, with no assumptions on their dynamics (see Section 6). This approach also promises to provide model reduction or approximation techniques for classes of stochastic processes. Along this line of research, [26] introduces an approximation for such processes. This approximation can be related to the work in [3,4] (which works with discrete-time stochastic hybrid systems), as well as to that in [62] (which uses Wasserstein Pseudometrics over continuous space processes) and to the classical reference in [43,44], which discusses weak approximations of stochastic processes, which has been applied on hybrid models in [42,56], but which offers no explicit approximation bound. Related to this works, [58] has proposed explicit

error bounds on a time and space discretization of a Markov process with certain ergodic properties.

3　Markov Processes over General State Spaces

We consider probabilistic processes defined over continuous spaces with Polish structure [48]. We assume to be working on a topological space that is homeomorphic to a Borel subset of a complete (i.e., a metric space where every Cauchy sequence converges) and separable (i.e., which contains a countable dense subset) metric space. Furthermore, we assume that the space is endowed with a Borel σ-algebra, which is characterized by sets that are Borel measurable. The reference metric can be reduced to the usual Euclidean metric (more on this in the following).

The continuous state-space is denoted by $\mathcal{S}$, whereas $\mathcal{B}(\mathcal{S})$ is the associated σ-algebra. Processes will be evolving in discrete time over the interval $[0, N]$ on a sample space $\Omega_{N+1} = \mathcal{S}^{N+1}$, equipped with the canonical product σ-algebra $\mathcal{B}(\Omega_{N+1})$. P is a probability measure defined on this event space.

We also introduce a control space $\mathcal{U}$, which we assume to be Borel measurable and in general continuous.

The following definition first appeared in [6], which focused on a rather rich state-space structure, namely a hybrid state space [12,17] – it can be shown [19] that a metric that is equivalent to the Euclidean one can be defined over this space. This model can be equivalently regarded as a Markov decision process [57] over a general state space with no rewards.

Definition 3.1 [Controlled Markov Process] Consider a discrete time controlled Markov processes (CMP) $\mathscr{S} = (\mathcal{S}, \mathcal{T}, \mathcal{U})$ defined over the state space $\mathcal{S}$, and characterized by $\mathcal{T}$, a conditional stochastic kernel that assigns to each point $s \in \mathcal{S}$ and control $u \in \mathcal{U}$ a probability measure $\mathcal{T}(\cdot|s, u)$. For any set $A \in \mathcal{B}(\mathcal{S})$, $\mathsf{P}_{s,u}(X(1) \in A) = \int_A \mathcal{T}(ds|X(0) = s, u)$, where $\mathsf{P}_{s,u}$ denotes the conditional probability $\mathsf{P}(\cdot|s, u)$. Process $\mathscr{S} = (\mathcal{S}, \mathcal{T}, \mathcal{U})$ is initialized according to a probability distribution $\pi : \mathcal{B}(\mathcal{S}) \to [0, 1]$.　　□

The syntax in Definition 3.1 leads to the following semantics for a trajectory $X(k)$ over the time horizon $[0, N]$. Let us fix a control string $\{u_0, u_1, \ldots, u_{N-1}; u_i \in \mathcal{U}\}$. Given an initial condition $x \in \mathcal{S}$ sampled from the probability distribution π, and given the control input $u_0 \in \mathcal{U}$, the value of the process at time $k = 1$ point is described by a probability law characterized by the conditional kernel $\mathcal{T}(\cdot|x, u_0)$. Likewise, for any $k = \{1, \ldots, N-1\}, X(k+1) \sim \mathcal{T}(\cdot|X(k), u_k)$.

Example 3.2 Consider the process $\mathscr{S}$ characterized as the solution $X(k), k \in \mathbb{N}$, of the following stochastic difference equation:

$$X(k+1) = X(k) + a(X(k))u(k) + b(X(k))w(k),$$

where $X(\cdot) \in \mathbb{R}^n$, $u(\cdot)$ takes values in a bounded set $\mathcal{U}$, the functions $a(\cdot), b(\cdot)$ are Lipschitz continuous and with linearly bounded growth, and $w(k) \sim \mathcal{N}(0, 1)$ is

a standard normal random variable for any $k \in \mathbb{N}$. Then, process $\mathscr{S}$ exists, is uniquely defined, and its dynamics can be characterized by the following kernel: $\mathcal{T}(\cdot|x, u) = \mathcal{N}(x + a(x)u, \sqrt{b(x)})$.

An instance of a realization of this process can be either of the two traces depicted in Figure 1, which have a deterministic initialization (black dot) and evolve in discrete time according to (Gaussian) conditional kernels. □

Let us now assume that $\mathcal{U}$ a *finite* set of labels (that is, elements taken from a finite alphabet). The following definition is derived from the work in [21,22,26] [2].

Definition 3.3 [Labeled Markov Process] A labeled Markov process (LMP) $\mathscr{S}$ is a structure

$$(\mathcal{S}, s_0, \mathcal{B}(\mathcal{S}), \{\tau_u | u \in \mathcal{U}\}),$$

where $\mathcal{S}$ is the state space, $s_0 \in \mathcal{S}$ is the initial state, $\mathcal{B}(\mathcal{S})$ is the Borel σ-field on $\mathcal{S}$, $\mathcal{U}$ is the set of labels, and

$$\forall u \in \mathcal{U}, \tau_u : \mathcal{S} \times \mathcal{B}(\mathcal{S}) \longrightarrow [0, 1]$$

is a transition probability function, namely a set-valued function $\tau_u(s, \cdot)$ that is a probability measure on $\mathcal{B}(\mathcal{S})$ for each $s \in \mathcal{S}$, and such that for each $S \in \mathcal{B}(\mathcal{S})$ the function $\tau_u(\cdot, S)$ is measurable. □

Definition 3.3 is generalized in [21,22,26] by allowing sub-probability measures τ_u. In this work, for the sake of consistency, we will only refer to complete probability measures. Notice that in general, unlike LMP, CMP do not specify the initial condition within the model definition, instead allowing for any choice of it within the state space.

If we semantically equate the labels of the LMP with the controls of the CMP, it is clear that labeled Markov processes in Definition 3.3 are a subclass of discrete time controlled Markov processes as in Definition 3.1, since the latter allows for a richer control structure. In fact the measure $\tau_u(s, \cdot)$ corresponds to the conditional kernel $\mathcal{T}(\cdot|s, u)$. We thus proceed by utilizing the model with syntax in Definition 3.1. In the next sections we will come back to the differences in the semantical characterization of the control structure between LMP and CMP.

4 Exact and Approximate Probabilistic Bisimulations: Relations, Logics, and Categories

In the following we introduce exact and approximate notions of bisimulations for CMP. We emphasize that both concepts are to be regarded as *strong* notions, as opposed to weak versions as in [9,38,54] [3]. The definitions can be looked at from three different aspects: via relations, via logics, and via categories.

---

[2] Some of these contributions work with analytic spaces, which are generalization of the Borel measurable ones that this work focuses on. However, since the properties of analytic spaces are not needed in this work and since Borel measurable sets are also analytically measurable, we have decided to focus on the latter.

[3] Weak notions are introduced in order to abstract from "internal" moves that do not influence the future behaviour of a process.

4.1 Exact Characterization via Relations, Logics, and Categories

Recall that a relation over a given set is an equivalence relation if it is reflexive, symmetric, and transitive. (In this work we assume to be working with relations with finite or countable cardinality.)

Definition 4.1 [(Exact) Probabilistic Bisimulation] Consider two CMP $\mathscr{S}_1 = (\mathcal{S}_1, \mathcal{T}_1, \mathcal{U})$ and $\mathscr{S}_2 = (\mathcal{S}_2, \mathcal{T}_2, \mathcal{U})$. An equivalence relation $R \subseteq \mathcal{S}_1 \times \mathcal{S}_2$ is a bisimulation relation on $\mathcal{S}_1 \times \mathcal{S}_2$ if, whenever $s_1 R s_2$ for any $s_1 \in \mathcal{S}_1, s_2 \in \mathcal{S}_2$, for given $u \in \mathcal{U}$ and set $\tilde{S}_1 \times \tilde{S}_2 \in (\mathcal{S}_1 \times \mathcal{S}_2)/R$ (which is Borel measurable), it holds that

$$\mathcal{T}_1(\tilde{S}_1|s_1, u) = \mathcal{T}_2(\tilde{S}_2|s_2, u).$$

A pair of states $s_1 \in \mathcal{S}_1, s_2 \in \mathcal{S}_2$ is said to be (probabilistically) bisimilar if $\exists R$, a bisimulation, such that $s R t$, whereas two CMP $\mathscr{S}_1, \mathscr{S}_2$ are said to be (probabilistically) bisimilar (denoted $\mathscr{S}_1 R \mathscr{S}_2$) if there exists a bisimulation relation R that is total over respectively $\mathscr{S}_1$ and $\mathscr{S}_2$ [4] . □

Notice that the autonomous case, characterized by a labels set with a single element, can be obtained as a special case of the above definition.

Example 4.2 Consider two processes $\mathscr{S}_i, i = 1, 2$, characterized by the models (as per Example 3.2)

$$X_i(k+1) = X_i(k) + a_i(X_i(k))u_i(k) + b_i(X_i(k))w_i(k),$$

where we assume $a_i(X_i) = -X_i, u_i = 1 \, \forall k \geq 0 \, (\mathcal{U} = \{1\}), b_1(X_1) = 0.4X_1, b_2(X_2) = 0.3X_2$, and where $X_i \in \mathbb{R} = \mathcal{S}_i$. The dynamics are quite trivial, since a trajectory is reset at any point in time to a neighborhood of the origin according to a Gaussian kernel with state-dependent variance. The following simple relation

$$R = \begin{cases} (-\infty, 0] \times (-\infty, 0], \\ (-\infty, 0] \times (0, +\infty), \\ (0, +\infty) \times (-\infty, 0], \\ (0, +\infty) \times (0, +\infty) \end{cases}$$

induces an (exact) bisimulation relation between the two processes, since conditional on any pair of states in $\mathcal{S}_1 \times \mathcal{S}_2 = \mathbb{R}^2$, the probability that either process X_i transitions to the respective projection of an element of R is equal to 0.5. Notice that R induces a partition of the composed state space $\mathbb{R}^2$. □

A logic $\mathcal{L}$ can be defined, which allows to show that two states are bisimilar if and only if they satisfy the same formulas ϕ of the logic $\mathcal{L}$ [21]. This approach emphasizes the fact that bisimulation is an equivalence relation.

---

[4] As a special case, if the initial conditions of $\mathscr{S}_1, \mathscr{S}_2$ are characterized by two given sets, then the relation should hold over pairs of states extracted from each of these sets respectively.

The work in [21] further characterizes probabilistic bisimulations via categorical notions (based on zigzag morphisms). Related to this approach, [20] employ coalgebraic notions to precisely relate probabilistic models. Similarly, probabilistic bisimulations relations over continuous-space processes can be introduced via categorical arguments, as discussed in [64].

4.2 *Approximate Characterization via Metrics based on Functions*

The exact relational and logical characterizations are formal, but have to be relaxed in order to accommodate for computational robustness [65] and for real-world engineering applications.

Probabilistic bisimulation can be sufficiently characterized by a family of functional expressions [22]. More specifically, given a process $\mathscr{S}$, consider a family $\mathcal{F}^c$ of real-valued functions $f_{\mathscr{S}} : \mathcal{S} \to [0,1]$, which are defined by a grammar (a set of operations). The operations induced by the grammar can be related to the rules of the logic $\mathcal{L}$. The parameter $c \in (0,1]$ is utilized in the definition of an operator in order to rescale the computation of the expected value at successive times (notice that this operation also depends on the labels of the process), and is in practice put forward to discount the future. The introduction of the family $\mathcal{F}^c$ of functions further allows to define a metric on processes.

Definition 4.3 [Metric between processes] Consider two CMP $\mathscr{S}_i = (\mathcal{S}_i, \mathcal{T}_i, \mathcal{U}), i = 1, 2$. A family $\mathcal{F}^c$ of functional expressions on $\mathscr{S}_i$ induces a distance as follows:

$$d^c(\mathscr{S}_1, \mathscr{S}_2) = \sup_{f \in \mathcal{F}^c} |f_{\mathscr{S}_1} - f_{\mathscr{S}_2}|,$$

where $f_{\mathscr{S}_i}$ are functions in $\mathcal{F}^c$ evaluated over the respective spaces $\mathcal{S}_i$. □

It can be shown that, for any $c \in (0,1]$, d^c is a pseudo-metric [5]. As a special instance, d^0 characterizes bisimilar processes [22]. Quite interestingly, for $c < 1$ it can be shown that, given an approximation parameter $\epsilon > 0$, the problem of checking $d^c(\mathscr{S}_1, \mathscr{S}_2) < \epsilon$ is decidable.

This discussion leads to the notion of approximate bisimulation with level ϵ, or simply of ϵ-bisimulation [24]. Let R be a relation on a set A. A set $\tilde{A} \subseteq A$ is said to be R-closed if $R(\tilde{A}) = \{t | sR t, s \in \tilde{A}\} \subseteq \tilde{A}$.

Definition 4.4 [Approximate Probabilistic Bisimulation] Consider two CMP $\mathscr{S}_1 = (\mathcal{S}_1, \mathcal{T}_1, \mathcal{U})$ and $\mathscr{S}_2 = (\mathcal{S}_2, \mathcal{T}_2, \mathcal{U})$. A relation $R_\epsilon \subseteq \mathcal{S}_1 \times \mathcal{S}_2$ is an ϵ-bisimulation relation if, for any $s_1 \in \mathcal{S}_1$ there is a $s_2 \in \mathcal{S}_2$ such that $s_1 R_\epsilon s_2$, and for any $u \in \mathcal{U}$ and R_ϵ-closed set $\tilde{S}_1 \times \tilde{S}_2 \subseteq \mathcal{S}_1 \times \mathcal{S}_2$, it holds that

$$\left| \mathcal{T}_1(\tilde{S}_1 | s_1, u) - \mathcal{T}_2(\tilde{S}_2 | s_2, u) \right| \leq \epsilon.$$

---

[5] In conformity with the discussions in [22,64], in the following we will not be formally distinguishing between pseudo- (or semi-)metrics and actual metrics, since we are simply interested in (pseudo-)metrics that are sufficient for characterizing bisimilarity (or trajectory equivalence, see Section 5) – the necessity is not fundamentally important.

In this case we say that the two CMP are ϵ-bisimilar (denoted $\mathscr{S}_1 R_\epsilon \mathscr{S}_2$). $\quad\square$

In general R_ϵ does not satisfy the transitive property, and as such is not an equivalence relation [24]. Hence, it induces a cover of $\mathcal{S}_1 \times \mathcal{S}_2$ but in general not a partition.

Example 4.5 An example of approximate probabilistic bisimulation relation R_ϵ is obtained by adapting the models $\mathscr{S}_i, i = 1, 2$ from Example 4.2 allow for a noise term with a small drift η_i, namely $w_i \sim \mathcal{N}(\eta_i, 1)$ and $\eta_1 \neq \eta_2$. Based on the corresponding probabilistic kernels, this allows computing an upper bound ϵ on the (absolute value of the) difference in the marginalization of the corresponding kernels over the sets in $\mathbb{R}^2$ induced by $R = R_\epsilon$. In this specific instance R_ϵ induces a partition of the composed state space $\mathbb{R}^2$ — this fact does not old in general. $\quad\square$

The use of a metric between processes, as in Definition 4.3, allows to relate the distance in time between processes that are "similar." We then employ a result from [22] and use it in the case where similarity between processes is precisely characterized by an approximate bisimulation relation.

Theorem 4.6 *Consider two CMP $\mathscr{S}_1 = (\mathcal{S}_1, \mathcal{T}_1, \mathcal{U})$ and $\mathscr{S}_2 = (\mathcal{S}_2, \mathcal{T}_2, \mathcal{U})$ that are ϵ-bisimilar, namely $\mathscr{S}_1 R_\epsilon \mathscr{S}_2$. Then $d^c(\mathscr{S}_1, \mathscr{S}_2) < k\epsilon$, with $c < 1$ and where $k = \sup_{n \in \mathbb{R}^+} nc^n$.*

Proof. The claim follows from [22, Prop. 7.5], where an ϵ-approximate bisimulation relation between $\mathscr{S}_1$ and $\mathscr{S}_2$ is used in place of the ϵ-perturbation notion of [22, Def. 7.4]. $\quad\square$

The idea to define discounted metrics over probability measures that admit bisimulation as a fixed point is taken up in [30], which uses the Kantorovic distance between probability measures to approximate MDP over infinite state spaces. This distance is related to that discussed in Section 6. The cited work in [64] introduces a discounted metric that is both closely related to that presented in this survey and which is also based on the Kantorovic distance.

Remark 4.7 While for processes over discrete, finite state spaces there exist algorithmic procedures to compute exact [11] and approximate [24] probabilistic bisimulations, the computational aspects related to these notions for processes over continuous state spaces require further research. Presently, based on these notions a few results [3,18,26] have put forward techniques to approximate these processes with finite state ones — however their scalability properties ought to be more thoroughly assessed. Next section proposes an alternative approach to synthesize approximate probabilistic bisimulations, which hinges on the computation of a function relating the two processes. $\quad\square$

5 Approximate Bisimulations via Probabilistic Bisimulation Functions

Consider a CMP $\mathscr{S}_1 = (\mathcal{S}_1, \mathcal{T}_1, \mathcal{U}_1)$ with associated realizations $X_1(k), k \in \mathbb{N}$, and a second model $\mathscr{S}_2$. The quantification of similarity between $\mathscr{S}_1$ and $\mathscr{S}_2$ can be assessed by comparing trajectories of the two models. A formal comparison can be set up by seeking a function $g : \mathcal{S}_1 \times \mathcal{S}_2 \to \mathbb{R}_0^+$ that induces a metric over the distance between the trajectories [40], for instance

$$g(X_1, X_2) = \|X_1 - X_2\|^2, \quad \text{if } \mathcal{S}_1 = \mathcal{S}_2.$$

If $\mathcal{S}_1 \neq \mathcal{S}_2$, in order to effectively relate the two processes, we need to assume the existence of proper output maps $\mathcal{Y}_i : \mathcal{S}_i \to \mathcal{S}^o$ taking values over the same observation space $\mathcal{S}^o$. In this instance, we would then consider a function

$$g(X_1, X_2) = \|\mathcal{Y}_1(X_1) - \mathcal{Y}_2(X_2)\|^2.$$

In general, given such a measurable, non-negative function g evaluated over $(X_1(k), X_2(k)), k \in \mathbb{N}$ (the Markov process related to the joint system $(\mathscr{S}^1, \mathscr{S}^2)$), the quality of the approximation between $\mathscr{S}_1$ and $\mathscr{S}_2$ is then characterized over a finite and an infinite time horizon, by the following two quantities:

$$V_\delta^N(x) = \mathsf{P}_x \left\{ \sup_{0 \leq k \leq N} g(X_1(k), X_2(k)) \geq \delta \right\} \tag{1}$$

and

$$V_\delta(x) = \mathsf{P}_x \left\{ \sup_{k \geq 0} g(X_1(k), X_2(k)) \geq \delta \right\}. \tag{2}$$

Here $x \in \mathcal{S}_1 \times \mathcal{S}_2$ represents a pair of initial conditions, and δ is a non-negative real number denoting the approximation quality.

It is of interest to provide meaningful and possibly tight bounds for the probabilistic quantities in (1)-(2). In order to do so, let us start by recalling the following classical notion [27]:

Definition 5.1 [(Super-) Martingale] Consider an autonomous stochastic process $X(k), k \geq 0$, taking values in $\mathcal{S}$. A function $\chi : \mathcal{S} \to \mathbb{R}$ is called a martingale for the process $X(k), k \geq 0$, if for any $x = X(0) \in \mathcal{S}, k \geq 0, \mathsf{E}_x[\chi(X(k))] = \chi(x)$. The function χ is called a supermartingale if $\mathsf{E}_x[\chi(X(k))] \leq \chi(x)$. □

In words, a (super-)martingale is a function of the process which, conditional on any initial condition, has an expected value that remains equal (does not increase) in time. Notice that the controls have not been introduced so far – they do play a role in the following. Let us introduce the notion of stochastic bisimulation function (SBF), as presented for continuous-time models in [40].

Definition 5.2 [Stochastic Bisimulation Function] Let the measurable function $\varphi : \mathcal{S}_1 \times \mathcal{S}_2 \to \mathbb{R}_0^+$ satisfy the following conditions:

(i) $\varphi(x) \geq g(x)$ for all $x \in \mathcal{S}_1 \times \mathcal{S}_2$;

(ii) for any $u_1 \in \mathcal{U}_1$, there exists $u_2 \in \mathcal{U}_2$ such that the function $(\varphi(X_1(k), X_2(k)))_{k \geq 0}$ is a P_x-supermartingale for any fixed $x \in \mathcal{S}_1 \times \mathcal{S}_2$.

Then φ is a stochastic simulation function of $\mathscr{S}_1$ by $\mathscr{S}_2$. If φ is also a stochastic simulation function of $\mathscr{S}_2$ by $\mathscr{S}_1$, then it is an SBF for the function g with respect to the joint process $(\mathscr{S}_1, \mathscr{S}_2)$. If two processes admit an SBF, they are said to be probabilistically bisimilar with precision $\varphi(x)$. □

Notice how in this definition the controls of the two CMP are treated quite differently than those in Definitions 4.1 and 4.4, which were originally stated for LMP and associated labels. We will comment on these semantical differences in Remark 5.6.

The existence of an SBF can be directly used to compute an upper bound for the quantities in (1) and (2). More precisely, selecting a parameter $\delta > 0$, any two initial conditions $x_i \in \mathcal{S}_i, i = 1, 2$, and by resorting to the properties of the SBF (as described in Definition 5.2) and to the Markov inequality [27], the following holds:

$$
\begin{aligned}
&\mathsf{P}_{(x_1, x_2)} \left(\sup_{0 \leq k < \infty} \|\mathcal{Y}_1(X_1(k)) - \mathcal{Y}_2(X_2(k))\|^2 \geq \delta \right) \\
&= \mathsf{P}_{(x_1, x_2)} \left(\sup_{0 \leq k < \infty} g\left(X_1(k), X_2(k)\right) \geq \delta \right) \\
&\leq \mathsf{P}_{(x_1, x_2)} \left(\sup_{0 \leq k < \infty} \varphi\left(X_1(k), X_2(k)\right) \geq \delta \right) \\
&\leq \frac{\varphi(x_1, x_2)}{\delta}.
\end{aligned}
\tag{3}
$$

We have shown that the knowledge of an SBF allows deriving bounds on the approximation quality between two processes. Next, we survey three conceptually different approaches to find such an SBF.

5.1 *Characterization of Stochastic Bisimulation Function based on Stochastic Stability*

The contribution in [40] puts forward conditions to construct an SBF for certain classes of continuous-time stochastic processes, namely models that are linear in the drift, in the diffusion coefficient, and in the observation map. The setup allows for spontaneous jumps (under homogeneous arrivals) with related (linear) resets, thus resulting in a model with hybrid structure [12,17]. The reader is referred to [40] for practical examples of computation of stochastic bisimulation functions.

In the present work we re-derive a condition for the existence of an SBF similar to [40] for discrete time models $\mathscr{S}_i, i = 1, 2$, and for the sake of clarity we focus on the following simpler (non hybrid) dynamics:

$$
\begin{cases}
X_i(k + 1) = A_i X_i(k) + B_i u_i(k) + F_i X_i(k) w_i(k), \\
Y_i(k) = C_i X_i(k).
\end{cases}
\tag{4}
$$

Here w_i are independent standard normal random variables, and we assume that the

outputs of the two processes take values over the same space. Consider a candidate SBF with the following quadratic form:

$$\varphi(x) = x^T M x, \tag{5}$$

where $x = [X_1 \, X_2]^T$ and M is a properly sized, symmetric, non-negative definite, constant matrix.

Theorem 5.3 *Assume that $\mathscr{S}_1$ and $\mathscr{S}_2$ are autonomous (namely, disregard the terms B_i). Consider the joint process $(\mathscr{S}_1, \mathscr{S}_2)$. A function φ as in (5) is a stochastic bisimulation function for $\mathscr{S}_1$ and $\mathscr{S}_2$ if and only if*

$$M - C^T C \succeq 0,$$
$$A^T M A + F^T M F - M \preceq 0,$$

where $C = [C_1 \; -C_2]$, $A = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}$, and $F = \begin{bmatrix} F_1 & 0 \\ 0 & F_2 \end{bmatrix}$.

Proof. The two conditions are directly derived from the corresponding requirements in Definition 5.2. In particular, notice that

$$g(X_1, X_2) = \|C_1 X_1 - C_2 X_2\|^2 = x^T C^T C x$$

and compare with (i). Furthermore, with reference to (ii) and Def. 5.1 and recalling the independence of the sample realizations of the noise process,

$$\begin{aligned}
&\mathsf{E}_x \left[\varphi(X_1(k+1), X_2(k+1)) \right] \\
&= \mathsf{E}_x \left[[X_1(k+1) \, X_2(k+1)] M [X_1(k+1) \, X_2(k+1)]^T \right] \\
&= x^T \left(A^T M A + F^T M F \right) x.
\end{aligned}$$

$\square$

Example 5.4 Consider the models $\mathscr{S}_i, i = 1, 2$ from Example 4.2, where $A_i X_i = X_i + a_i(X_i)$, and where $F_i X_i = b_i(X_i)$. The processes are (semantically) autonomous, since $u_i = 1$ is fixed. An SBF for $\mathscr{S}_i, i = 1, 2$ is obtained considering

$$M = C^T C = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix},$$

which is clearly positive semi-definite, and since matrix A is degenerate, M is such that

$$M - F^T M F = C^T C - (CF)^T CF = (C(I - F))^T C(I - F) \succeq 0.$$

$\square$

Let us generalize the previous result to the non-autonomous case.

Theorem 5.5 *Consider two non-autonomous processes $\mathscr{S}_1$ and $\mathscr{S}_2$. A function φ as in (5) is a stochastic bisimulation function for $\mathscr{S}_1$ and $\mathscr{S}_2$ if and only if, $\forall x \in \mathcal{S}_1 \times \mathcal{S}_2$*

$$M - C^T C \succeq 0,$$
$$\max_{u_1 \in \mathcal{U}_1} \min_{u_2 \in \mathcal{U}_2} \left(x^T \left(A^T M A + F^T M F - M \right) x + 2 x^T A^T M B u \right) \leq 0,$$
$$\max_{u_2 \in \mathcal{U}_2} \min_{u_1 \in \mathcal{U}_1} \left(x^T \left(A^T M A + F^T M F - M \right) x + 2 x^T A^T M B u \right) \leq 0,$$

where $B = \begin{bmatrix} B_1 & 0 \\ 0 & B_2 \end{bmatrix}$ *and* $u = [u_1 \, u_2]^T$.

Proof. It follows similarly to that of Theorem 5.3. □

Let us now elaborate on the different role that controls in CMP play as opposed to labels in LMP.

Remark 5.6 [Labels vs Controls] Notice that the condition in Theorem 5.5 is set up as a dynamical game between the two models. This is in accord with the role that control inputs play in Definition 5.2, and is in contrast with Definition 4.1 or 4.4 (originally stated for LMP), which fixed the same control input for both models. This difference highlights two distinct ways to conceive the role of labels for LMP on the one hand, and that of control inputs for CMP on the other. In LMP labels are intended as predefined schedules or actions taken by the environment (or adversary), which the system reacts to. This adheres to the role that nondeteminism classically plays in LMP. For the second models (CMP, as well as MDP in the systems and control literature) control inputs are actions to be synthesized based on an objective function, or policies ("strings" of control actions over a time span) that are chosen for the model. Notice how this difference reflects in the game-theoretical definition of approximate bisimulation given in [24] (as a game between a "prover," i.e. a model, and an "adversary"), in contrast to the formulation in Theorem 5.5. □

Theorem 5.5 can be re-stated according to the LMP interpretation as follows:

Corollary 5.7 *Consider two non-autonomous processes $\mathscr{S}_1$ and $\mathscr{S}_2$ with the same input space $\mathcal{U}$. A function φ as in (5) is a stochastic bisimulation function for $\mathscr{S}_1$ and $\mathscr{S}_2$ if and only if for any $\bar{u} = [u \, u]^T \in \mathcal{U} \times \mathcal{U}$ the following holds, $\forall x \in \mathcal{S}_1 \times \mathcal{S}_2$:*

$$M - C^T C \succeq 0,$$
$$x^T \left(A^T M A + F^T M F - M \right) x + 2 x^T A^T M B \bar{u} \leq 0.$$

Conversely, for the sake of completeness, we provide a statement of Definition 4.1 according to the CMP interpretation as follows (notice that now $\mathcal{U}_1 \neq \mathcal{U}_2$):

Definition 5.8 Consider two CMP $\mathscr{S}_1 = (\mathcal{S}_1, \mathcal{T}_1, \mathcal{U}_1)$ and $\mathscr{S}_2 = (\mathcal{S}_2, \mathcal{T}_2, \mathcal{U}_2)$. A binary relation $R \subseteq \mathcal{S}_1 \times \mathcal{S}_2$ is a simulation of $\mathscr{S}_1$ by $\mathscr{S}_2$ if, for any $s_1 \in \mathcal{S}_1$ there is a $s_2 \in \mathcal{S}_2$ such that $s_1 R s_2$, and for any $u_1 \in \mathcal{U}_1$ and R-closed set $\tilde{S}_1 \times \tilde{S}_2 \in \mathcal{S}_1 \times \mathcal{S}_2$, there exists a $u_2 \in \mathcal{U}_2$ such that

$$\mathcal{T}(\tilde{S}_1 | s_1, u_1) = \mathcal{T}(\tilde{S}_2 | s_2, u_2).$$

If R is also a simulation of $\mathscr{S}_2$ by $\mathscr{S}_1$, then it is a bisimulation of $\mathscr{S}_1$ and $\mathscr{S}_2$ and corresponds to an equivalence relation between pairs of states. □

A similar restatement can be introduced for the approximate version of probabilistic bisimulation of Definition 4.4.

In this section we have raised structural assumptions on the joint process under study to derive sufficient conditions for the existence of an SBF with the shape of equation (5). These conditions can be shown to lead to certain stochastic stability properties of the models under study, and equation (5) to be related to a Lyapunov function for the process [31].

5.2 Characterization of Stochastic Bisimulation Function based on Stochastic Contractivity

The contribution in [1] introduces sufficient conditions for the existence of an SBF, based on the use of contractivity analysis [47] for probabilistic systems. (Please refer to this contribution for practical examples of computation of stochastic bisimulation functions.) Furthermore, it shows that the notion of stochastic contractivity is related to a probabilistic version of the concept of incremental stability. Interestingly, the results presented in the previous paragraph and based on [40] leveraged assumptions on model stability that are analogous to the deterministic equivalents in [50,61], and similarly the contractivity assumptions, related to incremental stability, parallel similar results in corresponding deterministic literature [35,55]. The work in [2] extends the characterization of SBF to more general hybrid models [12,17].

In the present work we tailor the conditions in [1] to discrete time processes $\mathscr{S}_i$ of the following kind:

$$\begin{cases} X_i(k+1) = a_i(X_i(k), u_i(k)) + f_i(X_i(k)) w_i(k), \\ Y_i(k) = c_i(X_i(k)), \qquad k \in \mathbb{N}. \end{cases} \tag{6}$$

In general, the functions a_i, f_i and c_i can be nonlinear. As usual, the process $X_i \in \mathcal{S}_i$ (e.g. $\mathbb{R}^n$) and $w_i(k)$ are independent standard normal random variables. We assume that a solution is well defined, which solely requires boundedness assumptions on the quantities at the right-hand side of the state equation. Moreover, we assume that the observation functions c_i vanish at the origin and that they are Lipschitz continuous with constant $0 \leq \nu_i < \infty$.

Let us focus on autonomous models (i.e., let us disregard the effect of u_i). The following definition is inspired by [52,53], which extends earlier studies for deter-

ministic models [47].

Definition 5.9 [Stochastic Contractivity] Consider the process $\mathscr{S}_i$ in (6) characterized by its state equation with no control input. Assume that the following conditions are valid:

(i) $a_i(\cdot)$ is such that, for all $\tilde{x} \in \mathcal{S}_i, \exists \Lambda_i < \infty : \lambda_{\max} \left(\frac{\partial a_i}{\partial x}(\tilde{x})^T \frac{\partial a_i}{\partial x}(\tilde{x}) \right) \leq \Lambda_i$,
   where $\partial a_i/\partial x(\tilde{x})$ is the Jacobian of a_i evaluated at $\tilde{x}$, and $\lambda_{\max}(\cdot)$ is a function computing the maximum among the real parts of the eigenvalues of a matrix;

(ii) $f_i(\cdot)$ is Lipschitz continuous, with finite and positive constant K_i.

Then the system $\mathscr{S}_i$ in (6) is said to be stochastically contractive (in the identity metric) if $\Lambda_i + 2K_i < 1$.   □

Properties of stochastically contractive processes (such as extensions to non-identity, weighted metrics, as well as the relationship to probabilistic incremental stability) are further discussed in [1].

Next, given two processes $\mathscr{S}_i, i = 1, 2$, we show that the property of stochastic contractivity of the joint process $(\mathscr{S}_1, \mathscr{S}_2)$ entails a condition of probabilistic bisimilarity between the processes. The contractivity is intended to hold for the same metric (as discussed above, we consider here the identity metric) for both processes. Consider a parallel composition of the two processes,

$$
a = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \qquad f = \begin{pmatrix} f_1 & 0 \\ 0 & f_2 \end{pmatrix}, \qquad c = [I - I] \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = [c_1 - c_2],
$$

where the new output map computes the difference between the two original ones. Let us again start considering autonomous models.

Theorem 5.10 *Consider two autonomous processes, solutions of systems $\mathscr{S}_1, \mathscr{S}_2$ as in (6). If the composition of $\mathscr{S}_1, \mathscr{S}_2$ is stochastically contractive, then $\mathscr{S}_1, \mathscr{S}_2$ are probabilistically bisimilar.*
When existing, a probabilistic bisimulation function has the form $\psi(X_1, X_2) = 2\nu \|[X_1, X_2]^T\|^2$, where $\nu = \max\{\nu_1, \nu_2\}$.   □

Example 5.11 Consider the models $\mathscr{S}_i, i = 1, 2$ from Example 4.2, where the vector field $a_i(X_i)$ takes the value $X_i + a_i(X_i)$ as in Example 4.2, whereas $f_i(X_i) = b_i(X_i)$. The processes are again (semantically) autonomous, since $u_i = 1$ is fixed. Notice that $\Lambda_i = 0$, whereas $K_i < 1/2$, so that $\Lambda_i + 2K_i < 1$. Since $c_i(X_i) = X_i$, we have that $\nu_i = 1$. Given that both processes are contractive, an SBF for $\mathscr{S}_i, i = 1, 2$ is obtained considering $\psi(X_1, X_2) = 2\|[X_1, X_2]^T\|^2$.   □

The extension to the non-autonomous case follows.

Corollary 5.12 *Consider two processes, solutions of systems $\mathscr{S}_1, \mathscr{S}_2$ as in (6). $\mathscr{S}_1, \mathscr{S}_2$ are probabilistically bisimilar if (1.) for any $u_1 \in \mathcal{U}_1$ there exists a $\tilde{u}_2 \in \mathcal{U}_2$ and if (2.) for any $u_2 \in \mathcal{U}_2$ there exists a $\tilde{u}_1 \in \mathcal{U}_1$ such that the composition of $\mathscr{S}_1, \mathscr{S}_2$*

is stochastically contractive in the following two instances, $\forall (x_1, x_2) \in \mathcal{S}_1 \times \mathcal{S}_2$:

$$a_{1\cdot} = \begin{bmatrix} a_1(\cdot, u_1) \\ a_2(\cdot, \tilde{u}_2) \end{bmatrix}, \quad and \; a_{2\cdot} = \begin{bmatrix} a_1(\cdot, \tilde{u}_1) \\ a_2(\cdot, u_2) \end{bmatrix},$$

with associated parameters $\Lambda_{1\cdot}$ and $\Lambda_{2\cdot}$ respectively. Here $f_{1\cdot} = f_{2\cdot}$ and we assume it is Lipschitz with parameter K. The above condition can be expressed as:

$$\max_{u_1 \in \mathcal{U}_1} \min_{\tilde{u}_2 \in \mathcal{U}_2} \Lambda_{1\cdot} + 2K < 1, \qquad \max_{u_2 \in \mathcal{U}_2} \min_{\tilde{u}_1 \in \mathcal{U}_1} \Lambda_{2\cdot} + 2K < 1.$$

When existing, a probabilistic bisimulation function has the form $\psi(X_1, X_2) = 2\nu \| [X_1, X_2]^T \|^2$.   □

In contrast to the approach in the previous section and based on matrix inequalities, the contractivity conditions are directly computable on the system dynamics (abstractly, it is possible to characterize the portion of the state space where such conditions are valid, and this region is – in a certain sense – invariant); also, the probabilistic bisimulation function is directly obtained; finally, the conditions are applicable to nonlinear dynamics; however, at present the former results are applicable to models with richer dynamics. Both approaches can potentially yield bounds that are conservative.

5.3   Characterization of Stochastic Bisimulation Function as solution of a Probabilistic Reachability Problem

For a measurable function $g : \mathcal{S}_1 \times \mathcal{S}_2 \to \mathbb{R}_0^+$ and a parameter $\delta \in \mathbb{R}_0^+$ define the superlevel set

$$\mathbb{S}_g(\delta) = \{ x \in \mathcal{S}_1 \times \mathcal{S}_2 : g(x) > \delta \}.$$

Consider the event set corresponding to the sample space Ω over respectively a finite and infinite time horizon $N \in \mathbb{N} \cup \{\infty\}$, $\Omega_{N+1} = (\mathcal{S}_1 \times \mathcal{S}_2)^{N+1}$, and equipped with the canonical product topology. Let us define the following events over Ω_{N+1}, for any $N \geq 0$ and $A \in \mathcal{B}(\mathcal{S}_1 \times \mathcal{S}_2)$:

$$r_N(A) = \{ \omega \in \Omega_{N+1} | \exists n \in [0, N] : X(n, \omega) \in A \},$$
$$r(A) = \{ \omega \in \Omega_\infty | \exists n \geq 0 : X(n, \omega) \in A \}.$$

The quantity $r_N(A)$ expresses the event that the joint process $X = (X_1, X_2)$ enters set A within the time horizon $[0, N]$, whereas $r(A)$ extends this quantity to the infinite horizon.

For a finite horizon N, it can be noticed that

$$V_\delta^N(x) = \mathsf{P}_x[r_N(\mathbb{S}_g(\delta))],$$

thus the metrics of interest $V_\delta^N(x)$ can be calculated by solving a probabilistic reachability problem in discrete time [6]. A number of contributions have put forward techniques to approximately compute this quantity [4,28].

Next, we focus on the infinite horizon case, where

$$V_\delta(x) = \mathsf{P}_x[r(\mathbb{S}_g(\delta))].$$

Notice this latter quantity is in general difficult to precisely quantify. Recent work [63] has provided methods for bounding this quantity or raised conditions for computing it with finite-step procedures. These conditions critically hinge on deciding and computing the presence of "absorbing" sets for the dynamics [63]: it is interesting to notice that these sets are related to the notions of stability and contractivity presented above.

The extension to the controlled case involves again setting up a game over the control inputs of the two models. The reader is referred to [63] for an example of computation of such a stochastic bisimulation function.

6   Approximation Metrics via Sampling Techniques and Randomization Algorithms

The collection of definitions and concepts described above allow establishing metrics for the a-priori quantification of the similarity between (the distributions or trajectories of) two processes when considered over a finite or infinite time horizon. As such, they relied on structural assumptions over the models under study.

Next we present an approach, first described in [5], which has the advantage to be valid for general models, with no specific structural assumptions raised on them. It examines sample trajectories of the two processes over finite horizons. In other words, while the approaches above focused on the syntax of the models, this technique directly exploits the process semantics. The material focuses on the autonomous case.

Consider two autonomous processes $\mathscr{S}_1, \mathscr{S}_2$, for which equation (3) can be interpreted as follows:

$$\mathsf{P}_x\big(d_T(\mathscr{S}_1, \mathscr{S}_2) > \delta\big) \leq \epsilon \quad \Leftrightarrow \quad \mathsf{P}_x\big(d_T(\mathscr{S}_1, \mathscr{S}_2) \leq \delta\big) \geq 1 - \epsilon,$$

where $d_T(\cdot, \cdot)$ represents a metric between trajectories evaluated over the finite time horizon $[0, T]$ and started at $x \in \mathcal{S}_1 \times \mathcal{S}_2$, whereas δ is a given desired parameter quantifying the approximation precision, and ϵ is an a-priori quantity (probabilistic confidence on the approximation) depending on the models. The choice of the metric $d_T(\cdot, \cdot)$ is unrestricted; [5] employs either a distance in time between the trajectories, or a Hausdorff distance between the traces of the two processes.

From a different perspective, the inequalities above can be interpreted as a quantification of the approximation (δ), given a certain certainty level ($1-\epsilon$) on the similarity of the two processes. The quality δ of the approximation up to level $1 - \epsilon$ can be assessed as the solution of the following chance-constrained optimization

problem [15]:

$$\min_{\delta \in \Re} \delta, \quad \text{subject to:} \tag{7}$$

$$\mathsf{P}_x\big(d_T(\mathscr{S}_1, \mathscr{S}_2) \le \delta\big) \ge 1 - \epsilon.$$

Notice that (7) is a semi-infinite optimization program, since the number of probabilistic constraints is in general infinite while the number of optimization variables (δ) is finite. Denote with δ_ϵ the solution of (7): while the computation of this solution is in general hard, it can be mitigated by using a randomized approach, which provides an estimate of δ_ϵ with approximation guarantees.

The randomized algorithm executes N trajectories of the two processes $\mathscr{S}_1, \mathscr{S}_2$ over $[0, T]$, for random extractions of the initial condition x and of the driving uncertainty. It then computes their distance $d_T(\mathscr{S}_1, \mathscr{S}_2)$ and discards the $k < N$ obtained largest values, thus finding an approximate solution $\hat{\delta}_\epsilon$. Based on arguments developed in [16], the work in [5] shows that a proper choice of the parameters k, N allows ensuring the feasibility of the solution $\hat{\delta}_\epsilon$ (namely, the verification of the probabilistic constraints), and provides bounds on its performance degradation. This result holds up to a second confidence parameter that can be properly tuned. Intuitively, by extracting at random N executions of the processes and discarding a-posteriori a fraction k/N of them that corresponds to the largest discrepancies between the processes, one can improve the quality bound δ while guaranteeing that the violation set has size smaller than or equal to the prescribed ϵ value.

Additionally, the approach also enables an additional feature: the *design* of an approximation. Recall that, thus far, we have assumed that both models $\mathscr{S}_1, \mathscr{S}_2$ have been given. Here instead, we postulate that $\mathscr{S}_2(\theta), \theta \in \Theta$, is a parameterized approximation of $\mathscr{S}_1$, where Θ is either a finite or an infinite but compact set. The synthesis problem can be stated as:

$$\min_{\delta \in \Re, \theta \in \Theta} \delta, \quad \text{subject to:}$$

$$\mathsf{P}_x\big(d_T(\mathscr{S}_1, \mathscr{S}_2(\theta)) \le \delta\big) \ge 1 - \epsilon.$$

The problem can be solved similarly as that in (7). In the instance of a continuous Θ, the argument within the probabilistic constraints needs to be convex in the optimization variable θ [5]. In the above problem we are actually selecting an optimal abstraction $\mathscr{S}_2$ of a given model $\mathscr{S}_1$, while quantifying its approximation level.

A few comments are due. The advantages of this approach over those based on the synthesis of a probabilistic bisimulation function are

(i) the absence of assumptions on the dynamics of the two processes,

(ii) and the possibility of approaching the problem of synthesis of an approximation.

On the other hand, the limitations are

 (i) the presence of a confidence level on the obtained bounds – this level can nevertheless be finely tuned,

 (ii) the validity of the outcomes over finite horizons of time,

(iii) as of yet, the absence of an approach for non-autonomous models.

As discussed in Section 4 with reference to [26,64], the distance between two (comparable) stochastic processes can also be studied by setting up metrics on the corresponding probability distributions over their sample spaces. There is a vast literature on the use of metrics between probability measures [33]. Related to the work presented in this Section, the approach in [62] leverages Wasserstein pseudo-metrics between two processes, and approximates them by empirical quantities, obtained by taking samples of the trajectories of the two processes. While the empirical quantities are proven to converge to the actual distance with the number of samples taken, this approach does not provide explicit bounds based on finite samples for the distance between the two processes.

7 Discussion and Conclusions

The use of metrics to quantify distances between processes has a long history [33]. This survey has focused on distances based on the approximate notion of probabilistic bisimulation, which has seen a recent increased interest both from the dependability and formal verification community, as well as within the systems and control field. The two communities clearly differ in the respective approach to the problem: in a quest for categorization, it superficially looks like the first method opts for employing the underlying conditional kernels of the processes under study, whereas the second favors a trajectory-based approach to the problem. Furthermore, the two techniques are grounded on different mathematics: algebra, logics, and category the first, versus dynamical systems (Lyapunov theory, contractivity and invariance analysis) the second. As an alternative, sampling approaches and randomized methods look at the approximation problem from a totally different perspective. Here the focus is on the semantics of the processes and on the possibility to extract trajectories over a finite time horizon. The latter approach appears to yield results that are perhaps less formal (they hold with given confidence bounds, though extremely high), yet with outcomes that are less conservative and not stymied by assumptions on the model syntax (such as model stability, contractivity, etc.).

Looking forward, this survey would like to draw the attention to and place some emphasis on two topics:

 (i) practical computation of approximation levels between two given processes, and

 (ii) synthesis of approximations of a given general state-space stochastic process.

Indeed, while the majority of the examined approaches assume to be given two similar processes to compare (one of which may be regarded as an approximation or an abstraction of the other), only a few put forward procedures for model ap-

proximation or abstraction with quantified quality. In particular:

- The work in [26], followed by [18], puts forward an approximation based on time unfolding and relates it to the distance metrics developed by the same authors.
- The work in [1] puts forward a procedure that constructs a discrete approximation of a diffusion process. The procedure is based on the discretization of space and time. Given a diffusion process, sufficient conditions for the existence of such an approximation are raised. It shows that the abstraction is probabilistically bisimilar to the original process, up to a certain approximation precision.
- The work in [5] allows the design of an approximation of a continuous-space process. Similarly, the work in [62] allows for a synthesis of approximating processes.
- The contributions in [3,58] have proposed two separate techniques to abstract a stochastic model into a finite-state Markov chain with probabilistic bounds on the distance in time between the trajectories of the two processes.

An emphasis on the computability aspects as well as on the problem of synthesis of abstractions will lead to practically relevant procedures helping with the analysis, verification, and control of general state-space Markov processes.

Acknowledgments

References

[1] Abate, A., *A contractivity approach for probabilistic bisimulations of diffusion processes*, in: *Proceedings of the 48th IEEE Conference on Decision and Control and the 28th Chinese Control Conference*, Peking, PRC, 2009, pp. 2230–2235.

[2] Abate, A., *Probabilistic bisimulations of switching and resetting diffusions*, in: *Proceedings of the 49th IEEE Conference on Decision and Control*, Atlanta, GA, 2010, pp. 5918–5923.

[3] Abate, A., A. D'Innocenzo and M. Di Benedetto, *Approximate abstractions of stochastic hybrid systems*, IEEE Transactions on Automatic Control **56** (2011), pp. 2688–2694.

[4] Abate, A., J.-P. Katoen, J. Lygeros and M. Prandini, *Approximate model checking of stochastic hybrid systems*, European Journal of Control **16** (2010), pp. 1–18.

[5] Abate, A. and M. Prandini, *Approximate abstractions of stochastic systems: a randomized method*, in: *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference*, Orlando, FL, 2011, pp. 4861–4866.

[6] Abate, A., M. Prandini, J. Lygeros and S. Sastry, *Probabilistic reachability and safety for controlled discrete time stochastic hybrid systems*, Automatica **44** (2008), pp. 2724–2734.

[7] Alur, R. and D. L. Dill, *A theory of timed automata*, Theoretical Computer Science **126** (1994), pp. 183–235.

[8] Andova, S. and J. Baeten, *Abstraction in probabilistic process algebra*, in: *Tools and Algorithms for the Construction and Analysis of Systems*, LNCS **2031**, 2001, pp. 204–219.

[9] Baier, C. and H. Hermanns, *Weak bisimulation for fully probabilistic processes*, in: *Computer-Aided Verification*, LNCS **1254**, 1997, pp. 119–130.

[10] Baier, C., J. Katoen, H. Hermanns and V. Wolf, *Comparative branching-time semantics for Markov chains*, Information and Computation **200** (2003), pp. 492–508.

[11] Baier, C. and J.-P. Katoen, "Principles of Model Checking," The MIT Press, Cambridge, Massachusetts, 2008.

[12] Blom, H. and J. Lygeros (Eds.), "Stochastic Hybrid Systems: Theory and Safety Critical Applications," Number 337 in Lecture Notes in Control and Information Sciences, Springer-Verlag, Berlin, 2006.

[13] Buchholz, P., *Exact and ordinary lumpability in finite Markov chains*, Journal of Applied Probability **31** (1994), pp. 59–74.

[14] Bujorianu, M., J. Lygeros and M. Bujorianu, *Bisimulation for general stochastic hybrid systems*, in: *Hybrid Systems: Computation and Control (HSCC 05)* (2005), pp. 198–214.

[15] Calafiore, G. and M. Campi, *Uncertain convex programs: randomized solutions and confidence levels*, Mathematical Programming **102** (2005), pp. 25–46.

[16] Campi, M., S. Garatti and M. Prandini, *The scenario approach for systems and control design*, Annual Reviews in Control **33** (2009), pp. 149–157.

[17] Cassandras, C. and J. Lygeros (Eds.), "Stochastic Hybrid Systems," Number 24 in Control Engineering, CRC Press, Boca Raton, 2006.

[18] Danos, V., J. Desharnais and P. Panangaden, *Labeled Markov processes: stronger and faster approximations*, in: *In Proceedings of the 18th Symposium on Logic in Computer Science (LICS 03)*, 2003, pp. 341–350.

[19] Davis, M. H. A., "Markov Models and Optimization," Chapman & Hall, 1993.

[20] de Vink, E. and J. Rutten, *Bisimulation for probabilistic transition systems: A coalgebraic approach*, in: *Proceedings of the 24th International Colloquium on Automata Languages and Programming (ICALP 98)*, 1998.

[21] Desharnais, J., A. Edalat and P. Panangaden, *Bisimulation for labeled Markov processes*, Information and Computation **179** (2002), pp. 163–193.

[22] Desharnais, J., V. Gupta, R. Jagadeesan and P. Panangaden, *Metrics for labelled Markov processes*, Theoretical Computer Science **318** (2004), pp. 323–354.

[23] Desharnais, J., R. Jagadeesan, V. Gupta and P. Panangaden, *The metric analogue of weak bisimulation for probabilistic processes*, in: *Proceedings of the 17th Annual IEEE Symposium on Logic in Computer Science (LICS 02)*, 2002, pp. 413 – 422.

[24] Desharnais, J., F. Laviolette and M. Tracol, *Approximate analysis of probabilistic processes: logic, simulation and games*, in: *Proceedings of the International Conference on Quantitative Evaluation of SysTems (QEST 08)*, 2008, pp. 264–273.

[25] Desharnais, J. and P. Panangaden, *Continuous stochastic logic characterizes bisimulation of continuous-time Markov processes*, The Journal of Logic and Algebraic Programming **56** (2003), pp. 99–115.

[26] Desharnais, J., P. Panangaden, R. Jagadeesan and V. Gupta, *Approximating labeled Markov processes*, in: *Proceedings of the 15th Annual IEEE Symposium on Logic in Computer Science (LICS 00)*, 2000, pp. 95–105.

[27] Durrett, R., "Probability: Theory and Examples - Third Edition," Duxbury Press, 2004.

[28] Esmaeil Zadeh Soudjani, S. and A. Abate, *Adaptive gridding for abstraction and verification of stochastic hybrid systems*, in: *Proceedings of the 8th International Conference on Quantitative Evaluation of SysTems (QEST 11)*, Aachen, DE, 2011, pp. 59–69.

[29] Ferns, N., P. Panangaden and D. Precup, *Metrics for finite Markov decision processes*, in: *Proceedings of the 20th conference on Uncertainty in Artificial Intelligence (UAI 04)*, 2004, pp. 162–169.

[30] Ferns, N., P. Panangaden and D. Precup, *Metrics for Markov decision processes with infinite state spaces*, in: *Proceedings of the 21st Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI 05)*, Arlington, VA, 2005, pp. 201–208.

[31] Florchinger, P., *Lyapunov-like techniques for stochastic stability*, SIAM Journal on Control and Optimization **33** (1995), pp. 1151–1169.

[32] Giacalone, A., C. Jou and S. Smolka, *Algebraic reasoning for probabilistic concurrent systems*, in: *Proceedings of the IFIP TC2 Working Conference on Programming Concepts and Methods*, 1990, pp. 443–458.

[33] Gibbs, A. and F. Su, *On choosing and bounding probability metrics*, International Statistical Review (2001), pp. 419–435.

[34] Girard, A. and G. Pappas, *Approximation metrics for discrete and continuous systems*, IEEE Trans. on Automatic Control **52** (2007), pp. 782–798.

[35] Girard, A., G. Pola and P. Tabuada, *Approximately bisimilar symbolic models for incrementally stable switched systems*, IEEE Transactions on Automatic Control **55** (2010), pp. 116–126.

[36] Givan, R., T. Dean and M. Greig, *Equivalence notions and model minimization in Markov decision processes*, Artificial Intelligence **147** (2003), pp. 163–223.

[37] Henzinger, T., P. Ho and H. Wong-Toi, *HyTech: A model checker for hybrid systems*, Software Tools for Technology Transfer **1** (1997), pp. 110–122.

[38] Hermanns, H., "Interactive Markov Chains, and the Quest for Quantified Quality," Lecture Notes in Computer Science **2428**, Springer Verlag, 2002.

[39] Jonsson, B. and K. Larsen, *Specification and refinement of probabilistic processes*, in: *Proceedings of Sixth Annual IEEE Symposium on Logic in Computer Science (LICS 91)*, 1991, pp. 266 –277.

[40] Julius, A. and G. Pappas, *Approximations of stochastic hybrid systems*, IEEE Transactions on Automatic Control **54** (2009), pp. 1193–1203.

[41] Kemeny, J. and L. Snell, "Finite Markov Chains," D. Van Nostrand Company, Princeton, NJ, 1963.

[42] Koutsoukos, X., *Optimal control of stochastic hybrid systems based on locally consistent Markov decision processes*, International Journal of Hybrid Systems **4** (2004), pp. 301–318.

[43] Kushner, H. J., "Approximation and Weak Convergence Methods for Random Processes with Applications to Stochastic Systems Theory," MIT Press, Cambridge, Massachussets, 1984.

[44] Kushner, H. J. and P. Dupuis, "Numerical Methods for Stochastic Control Problems in Continuous Time," Springer-Verlag, New York, 2001.

[45] Lafferriere, G., G. Pappas and S. Sastry, *O-minimal hybrid systems*, Mathematics of Control, Signals, and Systems **13** (2000), pp. 1–21.

[46] Larsen, K. and A. Skou, *Bisimulation through probabilistic testing*, Information and Computation **94** (1991), pp. 1–28.

[47] Lohmiller, W. and J. Slotine, *On contraction analysis for nonlinear systems*, Automatica **34** (1998), pp. 671–682.

[48] Meyn, S. and R. Tweedie, "Markov Chains and Stochastic Stability," Springer Verlag, 1993.

[49] Milner, R., "A Calculus of Communicating Systems," Springer-Verlag New York, Inc., Secaucus, NJ, USA, 1982.

[50] Pappas, G., *Bisimilar linear systems*, Automatica **39** (2003), pp. 2035–2047.

[51] Park, D., *Concurrency and automata on infinite sequences*, in: *Proceedings of the 5th GI-Conference on Theoretical Computer Science*, 1981, pp. 167–183.

[52] Pham, Q., *Analysis of discrete and hybrid stochastic systems by nonlinear contraction theory*, in: *Proceedings of the 10th International Conference on Control, Automation, Robotics and Vision (ICARCV 08)*, Hanoi, Vietnam, 2008, pp. 1054–1059.

[53] Pham, Q., N. Tabareau and J. Slotine, *A contraction theory approach to stochastic incremental stability*, IEEE Transactions on Automatic Control **54** (2009), pp. 816–820.

[54] Philippou, A., I. Lee and O. Sokolsky, *Weak bisimulation for probabilistic systems*, in: *CONCUR*, LNCS **1877** (2000), pp. 334–349.

[55] Pola, G., A. Girard and P. Tabuada, *Approximately bisimilar symbolic models for nonlinear control systems*, Automatica **44** (2008), pp. 2508–2516.

[56] Prandini, M. and J. Hu, *A numerical approximation scheme for reachability analysis of stochastic hybrid systems with state-dependent switchings*, in: *In Proceedings of the 46th IEEE International Conference on Decision and Control*, New Orleans, LA, 2007, pp. 4662–4667.

[57] Puterman, M., "Markov decision processes: discrete stochastic dynamic programming," John Wiley & Sons, Inc., 1994.

[58] Sarich, M., F. Noe and C. Schütte, *On the approximation quality of Markov state models*, Multiscale Modeling and Simulation **8** (2010), pp. 1154–1177.

[59] Segala, R. and N. Lynch, *Probabilistic simulations for probabilistic processes*, Nordic Journal of Computing **2** (1995), pp. 250–273.

[60] Strubbe, S. and A. van der Schaft, *Bisimulation for communicating piecewise deterministic Markov processes (CPDPs)*, in: *Hybrid Systems: Computation and Control (HSCC 05)* (2005), pp. 623–639.

[61] Tabuada, P. and G. Pappas, *Bisimilar control affine systems*, Systems & Control Letters **52** (2004), pp. 49–58.

[62] Thorsley, D. and E. Klavins, *Approximating stochastic biochemical processes with Wasserstein pseudometrics*, Systems Biology, IET **4** (2010), pp. 193 –211.

[63] Tkachev, I. and A. Abate, *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, 2011, pp. 526–531.

[64] van Breugel, F. and J. Worrell, *Towards quantitative verification of probabilistic transition systems*, in: *Proceedings of the 28th International Colloquium on Automata, Languages and Programming (ICALP 01)*, 2001, pp. 421–432.

[65] Wimmer, R. and B. Becker, *Correctness issues of symbolic bisimulation computation for Markov chains*, in: *Proceedings of MMB-DFT*, 2010, pp. 287–301.