The Behavior of Probabilistic Systems: From Equivalences to Behavioral Distances

Giorgio Bacci Giovanni Bacci Kim G. Larsen Radu Mardare
Department of Computer Science, Aalborg University, Denmark

Abstract

In this paper we synthesize our recent work on behavioral distances for probabilistic systems and present an overview of the current state of the art in the field. We mainly focus on behavioral distances for Markov chains, Markov decision processes, and Segala systems. We illustrate three different methods used for the definition of such metrics: logical, order theoretic, and measure-testing; and we discuss the relationships between them and provide the main arguments in support of each of them.

We also overview the problem of computing such distances, both from a theoretical and a practical view point, including the exact and the approximated methods.

1 Introduction

Probabilistic bisimulation of Larsen and Skou (1989) and probabilistic trace equivalence are acknowledged to be the basic equivalences for equating probabilistic systems from the point of view of their behaviors.

An example of probabilistic system is the labelled Markov chain depicted in Figure 1 (left). Here states s_1 and s_2 goes to state s_4 with probability $\frac{2}{3}$ and $\frac{1}{3}$, respectively. Although they move with different probabilities to s_4, states s_1 and s_2 are bisimilar because they reach any bisimilarity class with the same probability (clearly, also s_4 and s_5 are bisimilar).

When the probabilistic models are obtained as approximations of others, e.g., as simplified model abstractions or inferred from empirical data, then an equivalence is too strong a concept. This issue is illustrated in Figure 1 (right), where the states t_1 and t_2 (i.e., the counterpart of s_1 and s_2, respectively, after a perturbation of the transition probabilities) are not bisimilar. This motivated the quest for a robust notion of behavioral equivalence and the development of a theory of behavioral "nearness".

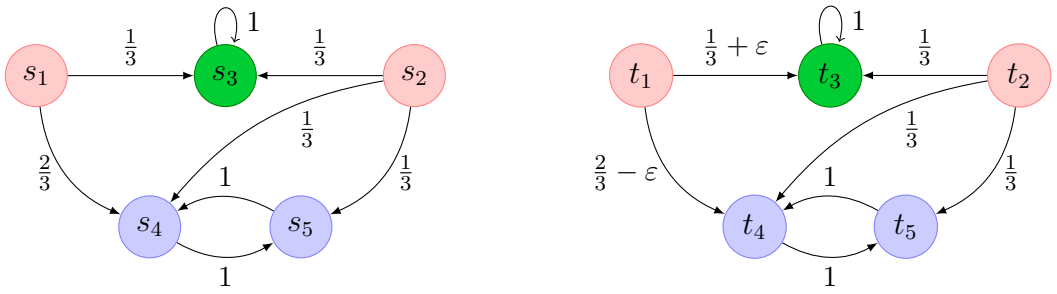

Figure 1: A labelled Markov chain (on the left) and an ε-perturbation of it (on the right), for some $0 < \epsilon < \frac{2}{3}$. Labels are represented by different colors.

To this end, Giacalone et al. (1990) proposed to use *pseudometrics* in place of equivalences aiming at measuring the behavioral similarities between states. Differently from an ordinary metric, a pseudometric allows different states to have distance zero, hence it can be thought of as a quantitative generalization of the notion of equivalence. In this respect, a pseudometric is said *behavioral* if states are at distance zero if and only if they are equivalent w.r.t. some behavioral semantics (e.g., bisimilarity, trace equivalence, etc.).

Behavioral distances do not come only as a robust notion of semantical equivalence, but they can also be used to address some important problems that are challenging computer science nowadays. One of these comes from systems biology and consists in providing analytical tools to help biologists understand the general mechanisms that characterize biological systems.

A considerable step forward in this direction is due to Luca Cardelli with his work on process algebra (Regev et al. 2004; Cardelli 2005, 2013) and on stochastic/ODE semantics (Cardelli 2008; Cardelli and Mardare 2010, 2013; Mardare et al. 2012), etc. Recently, Cardelli and Csikász-Nagy (2012) pointed out how the *Cell Cycle switch* (CC) —a fundamental biomolecular network that regulates the mitosis in eukaryotes— is surprisingly related, both in the structure and in the dynamics, to the *Approximate Majority* (AM) algorithm from distributed computing. The AM algorithm decides which of two populations is in majority by switching the majority into the totality, and it does so in a way that is *fast*, *reliable*, *robust*, and *asymptotically optimal* in the number of reactions required to obtain the result. The comparison between AM and CC is carried out by successive transformations that turn the network structure of the former into that of the latter. The difference is dictated by biological constraints, that are gradually introduced in the structure of AM during its transformation into CC while preserving both the computational and the dynamical properties.

Due to these observations, one may argue that CC, even though constrained

by biological limitations, tends to behave as similar as possible to AM, the theoretical "optimal" one. In this respect, behavioral distances seem the appropriate analytical tool to measure the quality of a candidate network model for CC: the closer the network model is to AM the better it is.

In this paper we overview our recent work on behavioral distances in comparison with the current state of the art in the field, focusing on four types of probabilistic systems: discrete-time Markov Chains (MCs), Markov Decision Processes with rewards (MDPs), continuous-time Markov Chains (CTMC), and Segala Systems (SS). We examine the main techniques that have been used in the literature to characterize behavioral distances for probabilistic systems, namely, logical, fixed point, and coupling characterizations. For each technique, we show how they have been applied on the different types of probabilistic systems we are considering, we point out their differences and similarities, and provide practical and theoretical arguments supporting each one.

Finally, we consider the problem of computing such distances, including both the exact and the approximated methods, for which we overview the most recent theoretical complexity results that are known in the literature.

2 Preliminaries

A *probability distribution* over a finite set S is a function $\mu\colon S \to [0,1]$ such that $\sum_{s\in S} \mu(s) = 1$. We denote by $\mathcal{D}(S)$ the set of probability distributions over S.

Let S and L be nonempty finite sets of *states* and *labels*, respectively. A *discrete-time Markov chain* (MC) is a tuple $\mathcal{M} = (S, L, \tau, \ell)$ where $\tau\colon S \to \mathcal{D}(S)$ is a transition probability function, and $\ell\colon S \to L$ a labeling function. Intuitively, labels represent properties that hold at a given state, $\tau(s)(s')$ is the probability to move from state s to a successor state s'. A *continuous-time Markov Chain* (CTMC) $\mathcal{M} = (S, L, \tau, \rho, \ell)$ extends an MC with a rate function $\rho\colon S \to \mathbb{R}_+$ associating with each state s the rate of an exponential distribution representing the residence-time distribution at s. A *Segala system* (SS) extends an MC by adding nondeterminism: is a tuple $\mathcal{M} = (S, L, \theta, \ell)$ where $\theta\colon S \to 2^{\mathcal{D}(S)}$ assigns with each state a set of possible probabilistic outcomes. Finally, a *Markov decision process with rewards* (MDP) is a tuple $\mathcal{M} = (S, A, \vartheta, r)$ consisting a finite nonempty set A of *action labels*, a labeled transition probability function $\vartheta\colon S \times A \to \mathcal{D}(S)$, and a reward function $r\colon S \times A \to \mathbb{R}_+$ assigning to each state the reward associated with the chosen action.

A (1-bounded) *pseudometric* on a set S is a map $d\colon S \times S \to [0,1]$ such that for all $s, t, u \in S$, $d(s,s) = 0$, $d(s,t) = d(t,s)$ and $d(s,u) \leq d(s,t) + d(t,u)$. For a set S, the *indiscrete pseudometric* is defined as $\mathcal{I}_S(s,s') = 1$ if $s \neq s'$, otherwise 0.

For a pseudometric $d\colon S \times S \to [0,1]$ on S, we recall two pseudometrics on $\mathcal{D}(S)$

$$\|\mu - \nu\|_{\mathrm{TV}} = \sup_{E \subseteq S} |\mu(E) - \nu(E)|\,, \qquad \text{(Total Variation)}$$
$$\mathcal{K}(d)(\mu,\nu) = \sup\left\{ \left| \int f \,\mathrm{d}\mu - \int f \,\mathrm{d}\nu \right| \mid |f(x) - f(y)| \leq d(x,y) \right\}\,, \qquad \text{(Kantorovich)}$$

and one pseudometric on 2^S defined, for $A, B \subseteq S$, as follows

$$\mathcal{H}(d)(A,B) = \max\{\sup_{a\in A} \inf_{b\in B} d(a,b), \sup_{b\in B} \inf_{a\in A} d(a,b)\}\,. \qquad \text{(Hausdorff)}$$

The set of 1-bounded pseudometrics on S, endowed with the pointwise pre-order $d_1 \sqsubseteq d_2$ iff $d_1(s,s') \leq d_2(s,s')$ for all $s, s' \in S$, forms a complete lattice, with bottom the constant 0 pseudometrics and top the indiscrete pseudometric $\mathcal{I}_S$.

Let $\lambda \in [0,1]$ and $a, b \in \mathbb{R}$, $a \oplus_\lambda b$ denotes the convex combination $\lambda a + (1-\lambda)b$.

3 Behavioral Distances

We overview the three main techniques that have been used in the literature for the characterization of behavioral pseudometrics over probabilistic systems.

3.1 Logical Characterizations

Real-valued logics. The first authentic behavioral pseudometric on probabilistic systems, due to Desharnais et al. (2004), is defined in terms of a family of functional expressions to be interpreted as real-valued modal formulas. Given a probabilistic model $\mathcal{M}$ over the set of states S, a functional $f \in \mathcal{F}$ is interpreted as a function $f_{\mathcal{M}}\colon S \to [0,1]$, and the pseudometric $\delta^{\mathcal{M}}\colon S \times S \to [0,1]$ assigns a distance to any given pair of states of $\mathcal{M}$ according to the following definition:

$$\delta^{\mathcal{M}}(s,s') = \sup_{f\in\mathcal{F}} |f_{\mathcal{M}}(s) - f_{\mathcal{M}}(s')|\,.$$

Their work builds on an idea of Kozen (1985) to generalize logic to handle probabilistic phenomena, and was first developed on MCs. Later, this approach has been adapted to MDPs, by de Alfaro et al. (2007) and Ferns et al. (2014), and extended to Segala systems by Mio (2013). Figure 2 shows the set of functional expressions for the case of MCs. One can think of those as logical formulas: a represents an atomic proposition, $\mathbf{1} - f$ negation, $\min(f_1, f_2)$ conjunction, $\Diamond f$ the modal operator, and $f \ominus q$ the "greater then q" test. The key result for such a metric is that two states are at distance 0 iff they are probabilistic bisimilar.

$$
\begin{array}{rl}
\mathcal{F} \ni f ::= a & a_{\mathcal{M}}(s) = \mathcal{I}_L(a, \ell(s)) \\
\mid \mathbf{1} - f & (\mathbf{1} - f)_{\mathcal{M}}(s) = 1 - f_{\mathcal{M}}(s) \\
\mid \min(f, f) & (\min(f_1, f_2))_{\mathcal{M}}(s) = \min\{(f_1)_{\mathcal{M}}(s), (f_2)_{\mathcal{M}}(s)\} \\
\mid f \ominus q & (f \ominus q)_{\mathcal{M}}(s) = \max\{f_{\mathcal{M}}(s) - q, 0\} \\
\mid \Diamond f & (\Diamond f)_{\mathcal{M}}(s) = \int f_{\mathcal{M}} \, \mathrm{d}\tau(s)
\end{array}
$$

Figure 2: Real-valued logic: syntax (on the left) and interpretation (on the right), where $\mathcal{M} = (S, L, \tau, \ell)$ is an MC, $a \in L$ a label, and $q \in \mathbb{Q} \cap [0, 1]$.

Linear time logics. If the models can be observed by only testing single execution runs, then bisimulation is to stringent as an equivalence and trace equivalence is preferred instead. Similar arguments justify the introduction of behavioral distances that focus on linear time properties only.

In (Bacci et al. 2014) we compared CTMCs against linear real-time specifications expressed as Metric Temporal Logic (MTL) formulas (Alur and Henzinger 1993). MTL is a continuous-time extension of LTL, where the *next* and *until* operators are annotated with a closed time interval $I = [t, t']$, for $t, t' \in \mathbb{Q}_+$.

$$\varphi ::= p \mid \bot \mid \varphi \to \varphi \mid \mathsf{X}^I \varphi \mid \varphi \, \mathsf{U}^I \, \varphi \,,$$

The satisfiability relation $\pi \models \varphi$ is defined over timed paths $\pi = s_0, t_0, s_1, t_1 \ldots$, where $t_i \in \mathbb{R}_+$ represents the *residence time* in s_i before moving to s_{i+1}. Modalities are interpreted as in LTL, with the additional requirement that in the next operator the step is taken at a time $t \in I$, and the until is satisfied with total accumulated time in I. We denote by $[\![\varphi]\!]$ the set of timed paths that satisfy φ.

The quantitative model checking of an CTMC $\mathcal{M}$ against an MTL formula φ consists in computing the probability $\mathbb{P}_s^{\mathcal{M}}([\![\varphi]\!])$ that $\mathcal{M}$, starting from the state s, generates a timed path that satisfies φ. Then, the following pseudometric

$$\delta^{\mathcal{M}}_{\mathrm{MTL}}(s, s') = \sup_{\varphi \in \mathrm{MTL}} |\mathbb{P}_s^{\mathcal{M}}([\![\varphi]\!]) - \mathbb{P}_{s'}^{\mathcal{M}}([\![\varphi]\!])| \tag{1}$$

guarantees that any result obtained by testing one state against an MTL formula can be reflected to the other with absolute error bounded by their distance[1].

Interestingly, we proved that the measurable sets generated by MTL formulas coincide with those generated by Deterministic Timed Automata (DTAs) specifications. Moreover, we singled out a *dense* subclass of specifications, namely that of resetting single-clock DTAs (1-RDTA), which implies that (i) the probability of satisfying any real-time specification can be approximated arbitrarily close by

[1]Clearly, the *discrete-time* case can be treated analogously by considering LTL formulas.

an 1-RDTA; (ii) the pseudometric (1) can be alternatively characterized letting range the supremum over 1-RDTA specifications only. This has practical applications in the quantitative model checking of CTMCs, since this allows one to exploit algorithms designed by Chen et al. (2011) for single-clock DTAs.

3.2 Fixed Point Characterizations

Often, behavioral distances are defined as fixed points of functional operators on pseudometrics. The first to use this technique were van Breugel and Worrell (2001), who showed the pseudometric of Desharnais et al. (2004) (see §3.1) can be defined as the least fixed point of an operator based on the Kantorovich metric.

This technique is very flexible and adapts easily in different contexts. The key observation is that functional operators can be composed to capture the different characteristics of the system. Next we show some examples from the literature.

Markov Chains. Let $\mathcal{M} = (S, L, \tau, \ell)$ be an MC. The functional operator defined by van Breugel and Worrell (2001) is as follows:

$$\mathcal{F}^{\mathcal{M}}_{\mathrm{MC}}(d)(s, s') = \max\left\{\mathcal{I}_L(\ell(s), \ell(s')), \mathcal{K}(d)(\tau(s), \tau(s'))\right\}. \tag{2}$$

Intuitively, $\mathcal{I}_L$ handles the difference in the labels, whereas the Kantorovich distance $\mathcal{K}(d)$ deals with the probabilistic choices by lifting the underlying pseudometric d over states to distributions over states. The two are combined by taking the maximum between them.

Markov Decision Processes. Let $\mathcal{M} = (S, A, \vartheta, r)$ be an MDP. Ferns et al. (2004) defined a pseudometric using the following operator, for $\lambda \in (0, 1)$:

$$\mathcal{F}^{\mathcal{M}}_{\mathrm{MDP}}(d)(s, s') = \max_{a \in A}\left\{|r(s, a) - r(s', a)| \oplus_\lambda \mathcal{K}(d)(\vartheta(s, a), \vartheta(s', a))\right\}. \tag{3}$$

The functional mixes in a convex combination the maximal differences w.r.t. the rewards associated with the choice of an action label in the current state and the difference in the probabilities of the succeeding transitions.

Segala Systems. Let $\mathcal{M} = (S, L, \theta, \ell)$ be an SS. van Breugel and Worrell (2014) extended the pseudometric on MCs using the following operator:

$$\mathcal{F}^{\mathcal{M}}_{\mathrm{SS}}(d)(s, s') = \max\left\{\mathcal{I}_L(\ell(s), \ell(s')), \mathcal{H}(\mathcal{K}(d))(\theta(s), \theta(s'))\right\}. \tag{4}$$

This functional extends (2) by handling the additional nonderminism with the Hausdorff metric on sets. In (de Alfaro et al. 2007; Mio 2013) a different functional operator is considered; this is obtained from (4) by replacing the sets $\theta(s)$ and $\theta(s')$ with their convex closures.

Continuous-time Markov Chains. Let $\mathcal{M} = (S, L, \tau, \rho, \ell)$ be a CTMC. In (Bacci et al. 2014), we proposed the following operator:

$$\mathcal{F}^{\mathcal{M}}_{\mathrm{CTMC}}(d)(s, s') = \max\left\{\mathcal{I}_L(\ell(s), \ell(s')), 1 \oplus_\alpha \mathcal{K}(d)(\tau(s), \tau(s'))\right\}, \tag{5}$$

where $\alpha = \|Exp(\rho(s)) - Exp(\rho(s'))\|_{\text{TV}}$ is the total variation distance between the exponential residence time distributions associated with the current states. Note that α equals the probability that the residence time in the two states differs. Therefore, the operator above can be seen as an extension of (2) that takes into consideration both the probability that the steps occur at different time moments or that the distinction can be seen later on in the probabilistic choices.

3.3 Coupling Characterizations

Given two probability distrubutions $\mu, \nu \in \mathcal{D}(X)$, a *coupling* for (μ, ν) is a joint probability distribution $\omega \in \mathcal{D}(X \times X)$ s.t. for all $E \subseteq X$, $\omega(E \times X) = \mu(E)$ and $\omega(X \times E) = \nu(E)$. Hereafter, $\Omega(\mu, \nu)$ will denote the set of couplings for (μ, ν).

Couplings have come to be used primarily for estimating the total variation distances between measures, since the following equality holds

$$\|\mu - \nu\|_{\text{TV}} = \inf \{\omega(\neq) \mid \omega \in \Omega(\mu, \nu)\} \ , \tag{6}$$

but they also work well for comparing probability distributions in general. Another notable equality is the following, a.k.a. *Kantorovich duality*

$$\mathcal{K}(d)(\mu, \nu) = \inf \left\{ \int d \, \mathrm{d}\omega \mid \omega \in \Omega(\mu, \nu) \right\} . \tag{7}$$

Based on these equalities, behavioral pseudometrics have been given alternative characterizations in terms of couplings.

Couplings & linear time logics. In (Bacci et al. 2014), we provided an alternative characterization of (1) that works as follows

$$\delta^{\mathcal{M}}_{\text{MTL}}(s, s') = \sup_{E \in \sigma(\text{MTL})} |\mathbb{P}^{\mathcal{M}}_s(E) - \mathbb{P}^{\mathcal{M}}_{s'}(E)| \tag{8}$$

$$= \inf \left\{ \omega(\not\equiv_{\text{MTL}}) \mid \omega \in \Omega(\mathbb{P}^{\mathcal{M}}_s, \mathbb{P}^{\mathcal{M}}_{s'}) \right\} \ , \tag{9}$$

where $\sigma(\text{MTL})$ denotes the σ-algebra generated by the sets $[\![\varphi]\!]$, for $\varphi \in \text{MTL}$, and $\equiv_{\text{MTL}}$ is the logical equivalence on timed paths. Equation (8) follows by showing that the generator is dense in $\sigma(\text{MTL})$, whereas (9) is proven generalizing (6).

Couplings & fixed points. Due to the Kantorovich duality, to the behavioral distances seen in §3.2 it can be given an alternative characterization based on a notion of *coupling structure* (varying w.r.t. the model) as the following minimum

$$\min \left\{ \gamma^{\mathcal{C}} \mid \mathcal{C} \text{ coupling structure for } \mathcal{M} \right\} . \tag{10}$$

where $\gamma^{\mathcal{C}}$ is the least fixed point of certain operators $\Gamma^{\mathcal{C}}$, that we describe below. A coupling structure $\mathcal{C}$ is said *optimal* if it achieves the minimum in (10).

Markov Chains. A coupling structure for an MC $\mathcal{M} = (S, L, \tau, \ell)$ is a tuple $\mathcal{C} = (\tau_{\mathcal{C}}, \ell)$ where $\tau_{\mathcal{C}} \colon (S \times S) \to \mathcal{D}(S \times S)$ is a probability transition function over pair of states such that, for all $s, s' \in S$, $\tau_{\mathcal{C}}(s, s') \in \Omega(\tau(s), \tau(s'))$, to be though of as a probabilistic pairing of two copies of τ.

Chen et al. (2012) showed that the distance of Desharnais et al. (2004) can be described as in (10) by means of the following operator

$$\Gamma^{\mathcal{C}}_{\mathrm{MC}}(d)(s, s') = \max\left\{\mathcal{I}_L(\ell(s), \ell(s')), \kappa^{\mathcal{C}}(d)(s, s')\right\}, \tag{11}$$

where $\kappa^{\mathcal{C}}(d)(s, s') = \sum_{u,v \in S} d(u, v) \cdot \tau_{\mathcal{C}}(s, s')(u, v)$. From (7), one may think of $\kappa^{\mathcal{C}}$ as the specialization of $\mathcal{K}$ on a fixed coupling structure $\mathcal{C}$.

Markov Decision Processes. A coupling structure for an MDP $\mathcal{M} = (S, A, \vartheta, r)$ is a tuple $\mathcal{C} = (\vartheta_{\mathcal{C}}, r)$ where $\vartheta_{\mathcal{C}} \colon (A \times S \times S) \to \mathcal{D}(S \times S)$ is a labelled probability transition function over pair of states such that, for all $a \in A$ and $s, s' \in S$, $\vartheta_{\mathcal{C}}(a, s, s') \in \Omega(\vartheta(a, s), \vartheta(a, s'))$. In (Bacci et al. 2013a) we showed that the distance of Desharnais et al. (2004), can be described as in (10) using the following operator

$$\Gamma^{\mathcal{C}}_{\mathrm{MDP}}(d)(s, s') = \max_{a \in A}\left\{|r(s, a) - r(s', a)| \oplus_{\lambda} \kappa^{\mathcal{C}}_a(d)(s, s')\right\}. \tag{12}$$

where $\kappa^{\mathcal{C}}_a(d)(s, s') = \sum_{u,v \in S} d(u, v) \cdot \vartheta_{\mathcal{C}}(a, s, s')(u, v)$. Once again, (12) may be seen the specialization of (3) w.r.t the coupling structure $\mathcal{C}$.

Continuous-time Markov Chains. In (Bacci et al. 2014) we extended the case of MCs in a continuous-time setting, by considering also the couplings over residence time distributions. A coupling structure for an CTMC $\mathcal{M} = (S, L, \tau, \rho, \ell)$ is a tuple $\mathcal{C} = (\tau_{\mathcal{C}}, \rho_{\mathcal{C}}, \ell)$ where $\tau_{\mathcal{C}}$ is defined as above, and $\rho_{\mathcal{C}} \colon S \times S \to \mathcal{D}(\mathbb{R}_+ \times \mathbb{R}_+)$ is such that $\rho_{\mathcal{C}}(s, s') \in \Omega(\mathit{Exp}(\rho(s)), \mathit{Exp}(\rho(s')))$, for all $s, s' \in S$.

In this case the functional operator is defined as follows

$$\Gamma^{\mathcal{C}}_{\mathrm{CTMC}}(d)(s, s') = \max\left\{\mathcal{I}_L(\ell(s), \ell(s')), 1 \oplus_{\beta} \kappa^{\mathcal{C}}(d)(s, s')\right\}, \tag{13}$$

where $\kappa^{\mathcal{C}}$ is defined as above and $\beta = \rho_{\mathcal{C}}(s, s')(\neq)$. Equation (6) justifies the value chosen for β, making (13) a specialization of (5) w.r.t. a coupling structure $\mathcal{C}$.

Notably, due to the characterization above and (9), in (Bacci et al. 2014) we have been able to prove that the behavioral distance defined as the least fixed point of $\mathcal{F}^{\mathcal{M}}_{\mathrm{CTMC}}$ is an upper bound of $\delta^{\mathcal{M}}_{\mathrm{MTL}}$ in §3.1(1). In fact, this generalizes from a quantitative point of view the inclusion of probabilistic trace equivalence into probabilistic bisimilarity.

4 Computational Aspects

Historically, behavioral distances have been defined in logical terms, but effective methods for computing distances arose only after the introduction of fixed point

characterizations. Indeed, one can easily *approximate* the distance from below by iteratively applying the fixed point operator, and improving the accuracy with the increased number of iterations. To this end, the operator needs to be computed efficiently. For instance, this is the case when the fixed point operator is based on the Kantorovich metric (see §3.2). Indeed, for $\mu, \nu \in \mathcal{D}(S)$ and S *finite*, the value $\mathcal{K}(d)(\mu, \nu)$ is achieved by the optimal solution of the following linear program (a.k.a. *transportation problem*)

$$TP(d)(\mu, \nu) = \arg\min_{\omega \in \Omega(\mu,\nu)} \sum_{u,v \in S} \omega(u, v) \cdot d(u, v) \tag{14}$$

where $\Omega(\mu, \nu)$ describes a transportation polytope. The above problem is in **P** and comes with efficient algorithms (Dantzig 1951; Ford and Fulkerson 1956).

In (Bacci et al. 2013b), we proposed an alternative *exact* method that computes the distance of Desharnais et al. (2004) over MCs efficiently, that adopts an on-the-fly strategy to avoid an exhaustive exploration of the state space. Our technique is based on the coupling characterization seen in §3.3. Given an MC $\mathcal{M}$ and an initial coupling structure $\mathcal{C}_0$ for it, we adopt a *greedy search strategy* that moves toward an optimal coupling by updating the current one, say $\mathcal{C}_i$, as $\mathcal{C}_{i+1} = \mathcal{C}_i[(s, s')/\omega]$ by locally replacing, at some pair of states (s, s'), a coupling, which is not optimal, with the optimal solution $\omega = TP(\gamma^{\mathcal{C}_i})(\tau(s), \tau(s'))$. Each update strictly improves the current coupling (i.e., $\gamma^{\mathcal{C}_{i+1}} \sqsubset \gamma^{\mathcal{C}_i}$) and ensures a fast convergence to an optimal one.

The method is sound independently from the initial starting coupling. Moreover, since the update is local, when the goal is to compute the distance only between certain pairs, the construction of the coupling structures can be done *on-the-fly*, delimiting the exploration only on those states that are demanded during the computation. Experimental results show that our method outperforms the iterative one by orders of magnitude even when one computes the distance on all pairs of states.

In (Bacci et al. 2013a) we further improved this technique in the case the input model is given as a composition of others. In summary, we identified a well behaved class of operators, called *safe*, for which is it possible to exploit the compositional structure of the system to obtain a heuristic for constructing a good initial coupling to start with the on-the-fly algorithm described above. It is worth noting that this is the first method that exploits the compositionality of the system to compute behavioral distances.

Complexity Results. Most of the theoretical complexity results about the problem of computing behavioral distances relies on fixed point characterizations.

Based on the fixed point characterization of van Breugel and Worrell (2001) (see §3.2(2)), Chen et al. (2012) showed that the bisimilarity distance of Desharnais et al. (2004) over MCs can be computed in polynomial time as the solution of

a linear program that can be solved using the ellipsoid method (Schrijver 1986); this result has been later extended in (Bacci et al. 2014) to CTMCs (see §3.2(5)).

Not surprisingly, the integration of non determinism on top of the probabilistic behavior, as in MDPs and Segala systems, has consequences also from a complexity perspective. Fu (2012) proved that the fixed point characterization of the bisimilarity pseudometric of de Alfaro et al. (2007) is in $\mathbf{NP} \cap \text{co-}\mathbf{NP}$. This is done by showing that the problem of deciding if a (rational) fixed point is the least one is in $\mathbf{P}$, then he provided a nondeterministic procedure to guess (rational) fixed points. Recently, van Breugel and Worrell (2014) proved that the problem of computing the distance on Segala systems (see §3.2(4)) belongs to $\mathbf{PPAD}$[2] by using a result by Etessami and Yannakakis (2010) that states that computing fixed points of *polynomial piecewise linear* functionals is in $\mathbf{PPAD}$.

In the case of linear (real-)time behavioral distances (see §3.1(1)), the decidability problem is still open. However, in a recent work, we showed that the MTL (resp. LTL) distance on CTMCs (resp. MCs) is $\mathbf{NP}$-hard to compute (Bacci et al. 2014, Corollary 23). We proved this result following arguments from Lyngsø and Pedersen (2002), who proved the $\mathbf{NP}$-hardness of comparing hidden Markov models with respect to the L_1 norm.

References

R. Alur and T. A. Henzinger. Real-Time Logics: Complexity and Expressiveness. *Information and Computation*, 104(1):35–77, 1993.

G. Bacci, G. Bacci, K. G. Larsen, and R. Mardare. Computing Behavioral Distances, Compositionally. In *MFCS*, volume 8087 of *Lecture Notes in Computer Science*, pages 74–85, 2013a.

G. Bacci, G. Bacci, K. G. Larsen, and R. Mardare. On-the-Fly Exact Computation of Bisimilarity Distances. In *TACAS*, volume 7795 of *Lecture Notes in Computer Science*, pages 1–15, 2013b.

G. Bacci, G. Bacci, K. G. Larsen, and R. Mardare. Topologies of Stochastic Markov Models: Computational Aspects. *ArXiv e-prints*, Mar. 2014.

L. Cardelli. Brane Calculi. In *CMSB*, volume 3082 of *LNCS*, pages 257–278. Springer, 2005. ISBN 978-3-540-25375-4.

[2]The complexity class $\mathbf{PPAD}$, which stands for *polynomial parity argument in a directed graph*, was introduced by Papadimitriou (1994) and it has received increased attention after it has been shown that finding Nash equilibria of two player games is a $\mathbf{PPAD}$-complete problem.

L. Cardelli. On process rate semantics. *Theoretical Computer Science*, 391(3): 190–215, 2008. ISSN 0304-3975.

L. Cardelli. Two-domain DNA strand displacement. *MSCS*, 23:247–271, 4 2013. ISSN 1469-8072.

L. Cardelli and A. Csikász-Nagy. The Cell Cycle Switch Computes Approximate Majority. *Scientific Reports*, 2, 2012. doi: 10.1038/srep00656.

L. Cardelli and R. Mardare. The Measurable Space of Stochastic Processes. In *QEST*, pages 171–180, 2010.

L. Cardelli and R. Mardare. Stochastic Pi-calculus Revisited. In *ICTAC*, volume 8049 of *LNCS*, pages 1–21. Springer, 2013. ISBN 978-3-642-39717-2.

D. Chen, F. van Breugel, and J. Worrell. On the Complexity of Computing Probabilistic Bisimilarity. In *FoSSaCS*, volume 7213 of *Lecture Notes in Computer Science*, pages 437–451. Springer, 2012.

T. Chen, T. Han, J.-P. Katoen, and A. Mereacre. Model Checking of Continuous-Time Markov Chains Against Timed Automata Specifications. *Logical Methods in Computer Science*, 7(1), 2011.

G. B. Dantzig. Application of the Simplex method to a transportation problem. In *Activity analysis of production and allocation*, pages 359–373. Wiley, 1951.

L. de Alfaro, R. Majumdar, V. Raman, and M. Stoelinga. Game Relations and Metrics. In *LICS*, pages 99–108, July 2007.

J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for labelled Markov processes. *Theoretical Computer Science*, 318(3):323–354, 2004.

K. Etessami and M. Yannakakis. On the Complexity of Nash Equilibria and Other Fixed Points. *SIAM Journal on Computing*, 39(6):2531–2597, 2010.

N. Ferns, P. Panangaden, and D. Precup. Metrics for finite Markov Decision Processes. In *UAI*, pages 162–169. AUAI Press, 2004. ISBN 0-9749039-0-6.

N. Ferns, D. Precup, and S. Knight. Bisimulation for Markov Decision Processes through Families of Functional Expressions. In *Horizons of the Mind. A Tribute to Prakash Panangaden*, volume 8464 of *LNCS*, pages 319–342, 2014.

L. R. Ford and D. R. Fulkerson. Solving the Transportation Problem. *Management Science*, 3(1):24–32, 1956.

H. Fu. Computing Game Metrics on Markov Decision Processes. In A. Czumaj, K. Mehlhorn, A. Pitts, and R. Wattenhofer, editors, *Automata, Languages, and Programming*, volume 7392 of *Lecture Notes in Computer Science*, pages 227–238. Springer Berlin Heidelberg, 2012. ISBN 978-3-642-31584-8.

A. Giacalone, C.-C. Jou, and S. A. Smolka. Algebraic Reasoning for Probabilistic Concurrent Games. In *IFIP WG 2.2/2.3*, pages 443–458, 1990.

D. Kozen. A Probabilistic PDL. *J. Comput. Syst. Sci.*, 30(2):162–178, 1985.

K. G. Larsen and A. Skou. Bisimulation Through Probabilistic Testing. In *POPL*, pages 344–352, 1989.

R. B. Lyngsø and C. N. Pedersen. The consensus string problem and the complexity of comparing hidden Markov models. *Journal of Computer and System Sciences*, 65(3):545–569, 2002.

R. Mardare, L. Cardelli, and K. G. Larsen. Continuous Markovian Logics - axiomatization and quantified metatheory. *LMCS*, 8(19):247–271, 2012.

M. Mio. Upper-Expectation Bisimilarity and Real-valued Modal Logics. *CoRR*, abs/1310.0768, 2013.

C. H. Papadimitriou. On the Complexity of the Parity Argument and Other Inefficient Proofs of Existence. *J. Comput. Syst. Sci.*, 48(3):498–532, 1994.

A. Regev, E. M. Panina, W. Silverman, L. Cardelli, and E. Shapiro. BioAmbients: an abstraction for biological compartments. *TCS*, 325(1):141–167, 2004.

A. Schrijver. *Theory of Linear and Integer Programming*. John Wiley & Sons, Inc., New York, NY, USA, 1986. ISBN 0-471-90854-1.

F. van Breugel and J. Worrell. Towards Quantitative Verification of Probabilistic Transition Systems. In *ICALP*, volume 2076 of *LNCS*, pages 421–432, 2001.

F. van Breugel and J. Worrell. The Complexity of Computing a Bisimilarity Pseudometric on Probabilistic Automata. In *Horizons of the Mind. A Tribute to Prakash Panangaden*, volume 8464 of *LNCS*, pages 191–213, 2014.