Computing Distances between Probabilistic Automata

Mathieu Tracol

LRI, Université Paris-Sud, France

Josée Desharnais,*   Abir Zhioua*

Dép. d'informatique et de génie logiciel, Université Laval, Québec, Canada

We present relaxed notions of simulation and bisimulation on Probabilistic Automata (PA), that allow some error ε. When $\varepsilon = 0$ we retrieve the usual notions of bisimulation and simulation on PAs. We give logical characterisations of these notions by choosing suitable logics which differ from the elementary ones, $\mathscr{L}$ and $\mathscr{L}^{\neg}$, by the modal operator. Using flow networks, we show how to compute the relations in PTIME. This allows the definition of an efficiently computable non-discounted distance between the states of a PA. A natural modification of this distance is introduced, to obtain a discounted distance, which weakens the influence of long term transitions. We compare our notions of distance to others previously defined and illustrate our approach on various examples. We also show that our distance is not expansive with respect to process algebra operators.

Although $\mathscr{L}^{(\neg)}$ is a suitable logic to characterise ε-(bi)simulation on deterministic PAs, it is not for general PAs; interestingly, we prove that it does characterise weaker notions, called *a priori ε-(bi)simulation*, which we prove to be NP-difficult to decide.

Keywords: Metrics, Bisimulation, Logic, Probabilistic Automata

1 Introduction

Preorders and equivalence notions between processes are central to concurrency theory. One wants to compare terms of a process algebra for proving an axiomatisation sound, to compare processes to some abstractions of them, etc. For non-probabilistic processes, notions of bisimulation and simulation are widely acknowledged, with, of course, many variations. In the study of probabilistic systems it has been observed [17] that the comparison between processes should not be based on notions that rely strongly on exact numbers, as do the known notions of bisimulation and simulation for probabilistic systems. The most important reason is that the stochastic information in probabilistic processes often comes from *observations*, or from *theoretical estimations*. Hence a slight difference in the probabilities between two processes should be treated differently from important ones and certainly not be simply tagged as non equivalence. In this context, notions of approximate equivalence or distance are more useful. Distances have been defined for probabilistic processes [15, 4] and some have tried to estimate bisimulation with a certain degree of confidence [15]. Relaxing the definition of simulation and bisimulation is another avenue, which we follow.

We first extend previous work on deterministic processes [12] to their non deterministic version, Probabilistic Automata (PA) [22]. We present *relaxed* notions of simulation and bisimulation on them with respect to some accuracy ε. When $\varepsilon = 0$ we retrieve the usual notions of bisimulation and simulation on PAs. Our notions rely on a definition of *ε-lifting* of relations, which happens to be equivalent to the one presented in [23]. However, in this paper, the authors present different notions of ε-simulations which consider *distributions on the set executions*, whereas our relations are always between the *states* of the systems, and our purpose is different. We give logical characterisations of these notions: a state ε-simulates another state if and only if it ε-*satisfies* every formula that the other one (exactly) satisfies;

---

similarly for ε-bisimulation. The extension of previous work comprises also the definition of an efficiently computable *non-discounted* distance: two states are at distance less than or equal to ε if they are ε-bisimilar. Using flow networks, we show how to compute in PTIME our relaxed relations of (bi)simulation which helps to also compute efficiently the distance.

The nature of non determinism leads to new challenges and concepts. It is not suprising that the logics that we prove to characterise ε-bisimulation and ε-simulation differ from the elementary ones, $\mathscr{L}$ and $\mathscr{L}^{\neg}$. Although $\mathscr{L}^{(\neg)}$ is a suitable logic to characterise ε-(bi)simulation on deterministic PAs, it is not for general PAs; interestingly, we define weaker notions that it does characterise on PAs, called *a priori ε-(bi)simulation*. We also prove that a priori 0-simulation is NP-difficult to decide, contrarily to ε-bi/simulation.

We propose a natural modification of our basic distance in order to discount the influence of long term transitions. We illustrate the difference between the values of the two distances on various examples of two-dimensional grids. Both (pseudo-)distances are different from the ones defined in the past [11, 1, 15, 4], in that differences along paths are not accumulated, even in the discounted one. The other known distances all accumulate differences through paths, and most of them discount the future. Those that do not discount the future are intractable: it has recently been proven decidable [3], but with double exponential complexity. Our distance is determined with a polynomial algorithm.

Finally, we prove that our distances are not expansive with respect to process algebras operators, such as parallel composition and non-deterministic choice.

2   Probabilistic Automata and ε-relations

In this section we give the definitions of our models and the relaxed relations that we study. *Probabilistic Automata* are labelled transition systems where transitions are from states to distributions and that involve non determinism. We generalize slightly the standard model, allowing sub-distributions instead of distributions, to model non responsiveness of the system and to make simulation a richer notion. Given a countable set S, we write $\mathrm{Sub}(S)$ for the set of sub-distributions on S: the total probability out of a sub-distribution may be less than one. Given a relation R on $S \times S$ and $X \subseteq S$, $R(X) = \{y \in S | \exists x \in X \ s.t. \ xRy\}$. A set X is *R-closed* if $R(X) \subseteq X$.

Definition 1 (PA [22]) *A* probabilistic automaton*, or PA, is a tuple* $\mathscr{S} = (S, Act, \mathscr{D})$ *where S is a denumerable state space, Act is a finite set of actions, and* $\mathscr{D} \subseteq S \times Act \times Sub(S)$ *is the transition relation.* $\mathscr{S}$ *is* finitely branching *if for all* $s \in S$ *and* $a \in Act$, $\{\mu \in Sub(S) \mid (s, a, \mu) \in \mathscr{D})\}$ *is finite; if it is a singleton or empty, we say that* $\mathscr{S}$ *is* deterministic. *The disjoint union of PAs* $\mathscr{S}_1, ..., \mathscr{S}_k$ *is the PA* $\uplus_{i \in [1;k]} \mathscr{S}_i$ *whose states are the disjoint union of the* S_i *and transitions carry through.*

Closely related models restrict states to be either probabilistic or non deterministic [18]. Generalizations to uncountable state spaces have also been studied [7, 9]. We sometimes mark a state (or a distribution) as initial. We write $s \xrightarrow{a} \mu$ for a transition $(s, a, \mu) \in \mathscr{D}$.

An example of PA is given in Fig. 1: an arrow labelled with action l and value r represents an l-transition of probability r; in picture representations, we omit the distributions that are concentrated in one point, as can be seen for the transitions from s to s_0 and from t to itself. In contrast, state s has an a-transition to distribution μ_s giving s three possible successors for this a-transition.

In previous work [12], we relaxed the classical notion of simulation between deterministic PAs to ε-*simulation*. We now generalize this approach to the context of PAs.
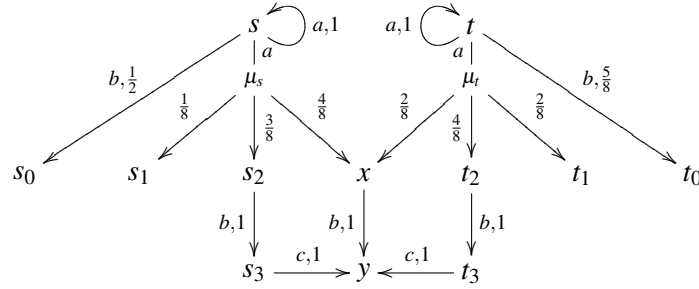
Figure 1: $s \prec_{\frac{1}{8}} t$ and $t \prec_{\frac{1}{8}} s$ but $s \not\sim_{\frac{1}{8}} t$

Definition 2 (ε-bi/simulation) *Let $\mathscr{S} = (S, Act, \mathscr{D})$ be a PA, and $\varepsilon \geq 0$. A relation $R \subseteq S \times S$ is an ε-simulation on $\mathscr{S}$ if whenever sRt, if $s \xrightarrow{a} \mu$, then there exists a transition $t \xrightarrow{a} v$ such that $\mu \mathscr{L}^{\varepsilon}(R) v$, where*

$$\mu \mathscr{L}^{\varepsilon}(R) v \quad \text{iff for all } E \subseteq S \text{ we have } \mu(E) \leq v(R(E)) + \varepsilon.$$

If R is symmetric, it is an ε-bisimulation. Two states s and s' of PAs $\mathscr{S}_1$ and $\mathscr{S}_2$ are ε-similar (resp. ε-bisimilar), written $s \prec_{\varepsilon} s'$ (resp. $s \sim_{\varepsilon} s'$), if there is some ε-simulation (resp. ε-bisimulation) that relates them in $\mathscr{S}_1 \uplus \mathscr{S}_2$.

We may omit ε in the notation when $\varepsilon = 0$, as it yields the classical notions.

Example 1 *In the PA of Fig. 1 $s \prec_{\frac{1}{8}} t$ and $t \prec_{\frac{1}{8}} s$. This is witnessed by the relations*

$$\{(s,t), (x,x), (x,t_2), (y,y), (y,t_3)\} \cup \{(s_i,t_i) \mid i = 0,1,2,3\},$$

$$\{(t,s), (x,x), (x,s_2), (y,y), (y,s_3)\} \cup \{(t_i,s_i) \mid i = 0,1,2,3\}.$$

However, s and t are not related for any $\varepsilon < \frac{1}{8}$. Notice that we have $s \not\sim_{\frac{1}{8}} t$. Indeed, x is $\frac{1}{8}$-bisimilar to no state but itself, but $\mu_s(\{x\}) = \frac{4}{8}$, which is strictly greater than $\mu_t(\{x\}) + \frac{1}{8} = \frac{3}{8}$.

This example shows that two-way ε-simulation is not ε-bisimulation, even for deterministic PAs. A deterministic example is obtained by removing a-loops in Fig. 1.

Proposition 1 *ε-bisimulation is different from two-way ε-simulation.*

As for the classical case, we define the largest relations as greatest fixed points: $F^{\varepsilon} : 2^{S \times S} \to 2^{S \times S}$ is defined as follows $\forall R \subseteq S \times S$, $\forall s, t \in S$, let $(s,t) \in F^{\varepsilon}(R)$ iff $\forall s \xrightarrow{a} \mu$, $\exists t \xrightarrow{a} v \mid \mu \mathscr{L}^{\varepsilon}(R) v$. Similarly, $G^{\varepsilon} : 2^{S \times S} \to 2^{S \times S}$ is defined as $(s,t) \in G^{\varepsilon}(R)$ iff $(s,t), (t,s) \in F^{\varepsilon}(R)$. We then define $\prec_{\varepsilon}^n$ iteratively as $\prec_{\varepsilon}^0 = S \times S$ and for all n, $\prec_{\varepsilon}^{n+1} = F^{\varepsilon}(\prec^n)$. As well, let $\sim_{\varepsilon}^0 = S \times S$ and for all n, $\sim_{\varepsilon}^{n+1} = G^{\varepsilon}(\sim^n)$.

Theorem 1 *$\mathscr{S}$ being finitely branching, $\prec_{\varepsilon}$ and $\sim_{\varepsilon}$ are the greatest fixpoints of F^{ε} and G^{ε}, respectively. In other words, $\prec_{\varepsilon} = \bigcap_{n \in \mathbb{N}} \prec_{\varepsilon}^n$, and $\sim_{\varepsilon} = \bigcap_{n \in \mathbb{N}} \sim_{\varepsilon}^n$.*

2.1   Lifting of relations and flow networks.

The lifting of a relation is a standard construction that transfers a relation R on states to a relation $\mathscr{L}(R)$ on sub-distributions over states. Contrarily to the way we formulate Def. 2, the usual definition of liftings is rather in terms of the existence of a weight function [22]. We show the equivalence of the definitions.

Definition 3 (ε-weight functions) *Let $\varepsilon \geq 0$, and $\mu, \nu \in Sub(S)$. An ε-weight function for (μ, ν) with respect to R is a function $\delta : S \times S \to [0;1]$ such that:*

- *If $\delta(s,t) > 0$ then $(s,t) \in R$.*

- *For all $s,t \in S$, $\sum_{s' \in S} \delta(s,s') \leq \mu(s)$ and $\sum_{s' \in S} \delta(s',t) \leq \nu(t)$.*

- *$\sum_{s,s' \in S} \delta(s,s') \geq \mu(S) - \varepsilon$.*

Before stating the equivalence between our formulation of $\mathscr{L}^\varepsilon(R)$ and the one with weight functions, we recall the notion of flow network, since it provides a convenient alternative definition for applications [2].

A network is a tuple $\mathscr{N} = (V, E, \perp, \top, c)$ where (V, E) is a finite directed graph in which every edge $(u,v) \in E$ has a non-negative, real-valued capacity $c(u,v)$. If $(u,v) \notin E$ we assume $c(u,v) = 0$. We distinguish two vertices: a source $\perp$ and a sink $\top$. For $v \in V$ let $in(v)$ be the set of incoming edges to node v, and $out(v)$ the set of outgoing edges from node v. A *flow function* is a real function $f : V \times V \to \mathbb{R}$ with the two following properties for all nodes u and v:

- Capacity constraints: $0 \leq f(u,v) \leq c(u,v)$. The flow along an edge cannot exceed its capacity.

- Flow conservation: for each node $v \in V - \{\perp, \top\}$, we have $\sum_{e \in in(v)} f(e) = \sum_{e \in out(v)} f(e)$.

The flow $\mathscr{F}(f)$ of f is given by $\mathscr{F}(f) = \sum_{e \in out(\perp)} f(e) - \sum_{e \in in(\top)} f(e)$.

Definition 4 (The network $\mathscr{N}(\mu, \nu, R)$) *Let S be a finite set, $R \subseteq S \times S$, and $\mu, \nu \in Sub(S)$. Let $S' = \{t' | t \in S\}$, where t' are pairwise distinct "new" states (i.e. $t' \notin S$). Let $\perp$ and $\top$ be two distinct new elements not contained in $S \cup S'$. The network $\mathscr{N}(\mu, \nu, R) = (V, E, \perp, \top, c)$ is defined as follows:*

- *$V = S \cup S' \cup \{\perp, \top\}$.*
- *$E = \{(s,t') | (s,t) \in R\} \cup \{(\perp, s) | s \in S\} \cup \{(t', \top) | t \in S\}$.*
- *The capacity function c is given by: $c(\perp, s) = \mu(s)$, $c(t, \top) = \nu(t)$, and $c(s,t) = 1$ for all $s,t \in S$.*

The following proposition gives various characterizations of the simulation relation.

Proposition 2 *Let S be a finite set, $R \subseteq S \times S$, and $\mu, \nu \in Sub(S)$. The following properties are equivalent:*

1. *$\mu \mathscr{L}^\varepsilon(R) \nu$.*

2. *The maximal flow in $\mathscr{N}(\mu, \nu, R)$ is greater than or equal to $\mu(S) - \varepsilon$.*

3. *There exists an ε-weight function for (μ, ν) with respect to R.*

4. *For all R-closed set $E \subseteq S$, we have $\mu(E) \leq \nu(E) + \varepsilon$.*

The equivalence with 4 applies only if the domain and image of R are considered disjoint.

Proof. $1 \Leftrightarrow 2$ is Theorem 7 [12]. $2 \Leftrightarrow 3$ is a slight generalization of a result of [2], in which $\varepsilon = 0$. $1 \Rightarrow 4$ is always true and $4 \Rightarrow 1$ is straightforward if the domain and image of R are disjoint. $\square$

The condition on the fourth statement may look restrictive, but it is quite natural. By taking two copies of the state space and relating a state to the copies of the states that R relates it to, one obtains a relation that satisfies the condition, yet representing the same relation as R. This allows to make a distinction between states that are *simulated* from states that are viewed as *simulating*.

3   Logic for ε-simulation and ε-bisimulation

3.1   The logic $\mathscr{L}$ and its corresponding notion of simulation

In the context of deterministic PAs, ε-bi/simulation are characterized [12] by the simple logic $\mathscr{L}^{\neg}$ [19], using a relaxed semantics "up to ε".

Definition 5 (The logics $\mathscr{L}$ and $\mathscr{L}^{\neg}$) *The syntax of $\mathscr{L}^{\neg}$ is as follows:*

$$\mathscr{L}^{\neg} \quad \phi ::= \top \mid \neg\phi \mid \phi_1 \wedge \phi_2 \mid \phi_1 \vee \phi_2 \mid \langle a \rangle_\delta \phi \quad \text{where } \delta \in \mathbb{Q} \cap [0;1].$$

We write $\mathscr{L}$ for the logic without negation. Given a PA with components $(S, Act, \mathscr{D})$, the relaxed ε-semantics $\models_\varepsilon$ is defined by structural induction on the formulas.

- $\bullet \forall s \in S, s \models_\varepsilon \top$ $\bullet\, s \models_\varepsilon \neg\theta$ *iff* $s \not\models_{-\varepsilon} \theta$ $\bullet\, s \models_\varepsilon \phi_1 \wedge \phi_2$ *iff* $s \models_\varepsilon \phi_1$ *and* $s \models_\varepsilon \phi_2$;

- $\bullet\, s \models_\varepsilon \langle a \rangle_\delta \phi$ *iff there exists a transition* $s \xrightarrow{a} \mu$ *s.t.* $\mu([\![\phi]\!]_\varepsilon) \geq \delta - \varepsilon$,

where $[\![\phi]\!]_\varepsilon = \{ s \in S \mid s \models_\varepsilon \phi \}$, *and the semantics of $\vee$ is similar to the one of $\wedge$. Given $s \in S$, we write $\mathscr{F}^\varepsilon(s)$ (resp. $\mathscr{F}^{\varepsilon,\neg}(s)$) for the set of formulas in $\mathscr{L}$ (resp. in $\mathscr{L}^{\neg}$) ε-satisfied by s.*

As for deterministic PAs, the logic is less expressive with the relaxed semantics [12] than with the standard one. Indeed, for each $\phi \in \mathscr{L}^{\neg}$, we can construct an associated formula $\phi_\varepsilon \in \mathscr{L}^{\neg}$ such that $[\![\phi]\!]_\varepsilon = [\![\phi_\varepsilon]\!]_0$. Here is how this is done: $\top_\varepsilon = \top$; $(\phi_1 \wedge \phi_2)_\varepsilon = (\phi_1)_\varepsilon \wedge (\phi_2)_\varepsilon$; $(\langle a \rangle_\delta \phi)_\varepsilon = \langle a \rangle_{\delta-\varepsilon} \phi_\varepsilon$; $(\neg\phi)_\varepsilon = \neg(\phi_{-\varepsilon})$. Here we use the fact that $\langle a \rangle_\lambda \phi$ is still a valid formula, even if $\lambda < 0$ or $\lambda > 1$, which gives in turn that $(\phi_\varepsilon)_{-\varepsilon} = \phi$. Clearly the transformation is additive, as $(\phi_\varepsilon)_{\varepsilon'} = \phi_{\varepsilon+\varepsilon'}$.

Example 2 *If $\phi = \langle a \rangle_{.5} (\neg \langle a \rangle_{.2} \top)$, then $\phi_\varepsilon = \langle a \rangle_{.5-\varepsilon} (\neg \langle a \rangle_{.2+\varepsilon} \top)$.*

The relaxed logic being less expressive is not an issue because we use the new semantics to simplify the formulations of the logical characterisations. It implies that model checking of formulas with the relaxed semantics can be done using the same technique as for the usual semantics.

The logics $\mathscr{L}$ and $\mathscr{L}^{\neg}$ induce ε-simulation and ε-bisimulation relations:

Definition 6 (Logical ε-simulation.) *Let $\varepsilon \in [0;1]$, $s,t \in S$. We say that t $\mathscr{L}_\varepsilon$-simulates s, written $s \prec^{\mathscr{L}}_\varepsilon t$, if for all formula $\phi \in \mathscr{L}$, $s \models \phi$ implies $t \models_\varepsilon \phi$. States t and s are said ε-logically equivalent, written $s \sim^{\mathscr{L}^{\neg}}_\varepsilon t$, if $s \prec^{\mathscr{L}^{\neg}}_\varepsilon t$ and $t \prec^{\mathscr{L}^{\neg}}_\varepsilon s$.*

The following theorem says that $\mathscr{L}$ characterizes ε-simulation on deterministic PAs. As Ex. 1 illustrates, ε-bisimulation is different from two-way ε-simulation when $\varepsilon > 0$, and hence we need negation to characterize ε-bisimulation.

Theorem 2 ([12]) *For deterministic PAs we have $\prec_\varepsilon = \prec^{\mathscr{L}}_\varepsilon$ and $\sim_\varepsilon = \sim^{\mathscr{L}^{\neg}}_\varepsilon$*

As can be expected, the logics $\mathscr{L}$ and $\mathscr{L}^{\neg}$ are not strong enough to characterize ε-bi/simulation for PAs. In the next section, we will present a stronger logic that will characterize these notions. Nevertheless, we can present the notions that correspond to $\prec^{\mathscr{L}}_\varepsilon$ and $\sim^{\mathscr{L}^{\neg}}_\varepsilon$. The name "a-priori" [1] comes from the order of the quantifiers in the definition: sets E in the following definition are chosen *before* the matching transition, which contrasts with Def. 2.

Definition 7 (A priori ε-simulation and bisimulation) *Let $\mathscr{S} = (S, Act, \mathscr{D})$ be a PA. A relation $R \subseteq S \times S$ is an a priori ε-simulation on $\mathscr{S}$ iff for all $s,t \in S$ such that sRt, for all $s \xrightarrow{a} \mu$ and for all $E \subseteq S$, there exists a transition $t \xrightarrow{a} \nu$ such that $\mu(E) \leq \nu(R(E)) + \varepsilon$. If R is symmetric, then it is an a priori ε-bisimulation. We write $\prec^{prio}_\varepsilon$ and $\sim^{prio}_\varepsilon$ for the largest relations of a priori ε-simulation and ε-bisimulation.*

Before proving that this relation is characterized by $\mathscr{L}$, we introduce some notation. We define $\prec_{\varepsilon}^{prio,n}$ iteratively in the same way as $\prec_{\varepsilon}^{n}$, using $F_{\varepsilon}^{prio} : 2^{S \times S} \to 2^{S \times S}$, which is defined as follows: $\forall R \subseteq S \times S$, $\forall s,t \in S$, let $(s,t) \in F_{\varepsilon}^{prio}(R)$ iff $\forall s \xrightarrow{a} \mu$, $\forall E \subseteq S$, $\exists t \xrightarrow{a} \nu$ s.t. $\mu(E) \leq \nu(R(E)) + \varepsilon$. As for $\prec_{\varepsilon} = \bigcap_{n \in \mathbb{N}} \prec_{\varepsilon}^{n}$, we can show that we have $\prec_{\varepsilon}^{prio} = \bigcap_{n \in \mathbb{N}} \prec_{\varepsilon}^{prio,n}$. The *depth* of a formula is the maximal number of imbrications of $\langle a \rangle_{\delta}$ operators. We write $\mathscr{F}^{n}$ for the set of formulas of $\mathscr{L}$ of depth at most n. Given $n \in \mathbb{N}$, given $s \in S$, $\mathscr{F}_{\varepsilon}^{n}(s)$ is the set of formulas $\phi \in \mathscr{L}$ of depth at most n such that $s \models_{\varepsilon} \phi$ and $\mathscr{F}_{\varepsilon}(s)$ is the set of formulas $\phi \in \mathscr{L}$ such that $s \models_{\varepsilon} \phi$. We define $\mathscr{F}^{n}(s) = \mathscr{F}_{0}^{n}(s)$ and $\mathscr{F}(s) = \mathscr{F}_{0}(s)$.

The next theorem proves the logical characterization of $\prec_{\varepsilon}^{prio}$ and $\sim_{\varepsilon}^{prio}$.
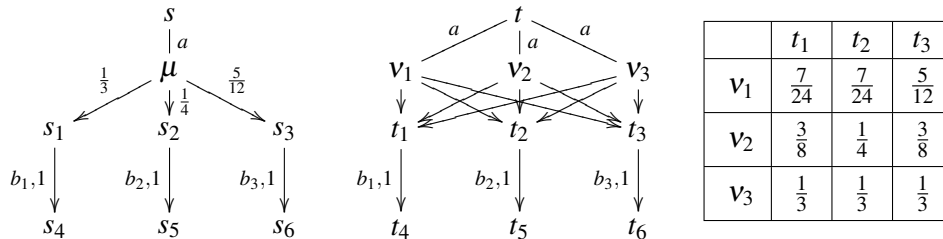
Theorem 3 *Let $\mathscr{S} = (S, Act, \mathscr{D})$ be a PA, and let $s,t \in S$. Then:*

1. *$s \prec_{\varepsilon}^{prio,n} t$ iff $\mathscr{F}^{n}(s) \subseteq \mathscr{F}_{\varepsilon}^{n}(t)$ for all $n \geq 0$*

2. *For all $n \geq 0$, for all $u,v \in S$, there exists $\phi_u \in \mathscr{F}_n$ such that $v \models_{\varepsilon} \phi_u$ iff $u \prec_{\varepsilon}^{prio,n} v$*

3. *$\prec_{\varepsilon} \subseteq \prec_{\varepsilon}^{prio} = \prec_{\varepsilon}^{\mathscr{L}}$ and the inclusion is strict.*

4. *$\sim_{\varepsilon} \subseteq \sim_{\varepsilon}^{prio} = \sim_{\varepsilon}^{\mathscr{L}^{\neg}}$ and the inclusion is strict.*

This proof generalizes to the context of countable state space systems, using a method close to the method used in [12] to extend logic characterizations to denumerable state spaces.

Proof. Inclusions are straightforward. The proof of the strictness of inclusion is given by the example following the proof. The structure of the proof is similar to the ones of [19] and [22] for the logical characterisation of simulation but we adapt them to systems with non determinism. The third point is a corollary of the first one. The fourth point is not more difficult and is the same kind of translation as the proof of [19]. We sketch the proof of the first two points, concentrating on the "$\Leftarrow$" direction. The two points are proven simultaneously by induction on n. The base case follows trivially from the definitions. Assume that the claims are true for n. We now prove 2 for $n + 1$. Fix $u \in S$. We define $\phi_u \in \mathscr{F}_{n+1}$ as follows. Let $v \in S$ such that $u \not\prec_{\varepsilon}^{prio,n} v$; by induction $\mathscr{F}^{n}(u) \not\subseteq \mathscr{F}_{\varepsilon}^{n}(v)$, that is, there exists a formula $\phi_{(u,v)} \in \mathscr{F}_n$, such that $u \models \phi_{(u,v)}$ and $v \not\models_{\varepsilon} \phi_{(u,v)}$. The formula $\phi_u := \bigwedge_{v \not\prec_{\varepsilon}^{prio,n} u} \phi_{(u,v)}$ is in $\mathscr{F}_{n+1}$ because S is finite. Now, $u \models \phi_u$, and any v such that $u \not\prec_{\varepsilon}^{prio,n+1} v$ verifies $u \not\prec_{\varepsilon}^{prio,n} v$ and hence $v \not\models_{\varepsilon} \phi_u$. Since $u \prec_{\varepsilon}^{prio,n} w$ implies $w \models \phi_u$ by induction, we get the result; hence 2 is proven for $n + 1$. As for 1, suppose $\mathscr{F}^{n+1}(s) \subseteq \mathscr{F}_{\varepsilon}^{n+1}(t)$. Let $s \xrightarrow{a} \mu$ be a transition from s, and let $E \subseteq S$. We are looking for a transition $t \xrightarrow{a} \nu$ such that $\nu(\prec_{\varepsilon}^{prio,n} E) \geq \mu(E) - \varepsilon$. Let $p = \mu(E)$. We construct a formula ϕ_E such that for all $u \in S$, $u \models \phi_E$ iff $u \in E$. By the second claim, we just have to consider $\phi_E = \bigvee_{u \in E} \phi_u$. Now, $s \models \langle a \rangle_p \phi_E$. Since $\langle a \rangle_p \phi_E \in \mathscr{F}^{n+1}$, we must have $t \models_{\varepsilon} \langle a \rangle_p \phi_E$. Since $\prec_{\varepsilon}^{prio,n} [\![\phi_E]\!] \subseteq [\![\phi]\!]_{E\varepsilon}$, we get that there exists a transition $t \xrightarrow{a} \nu$ from t such that $\nu(\prec_{\varepsilon}^{prio,n} E) \geq \mu(E) - \varepsilon$, which proves the result. $\square$

Example 3 *In the following PA (where the b_i's are different labels), we have $s \prec^{prio} t$ and $s \not\prec t$.*

	t_1	t_2	t_3
v_1	$\frac{7}{24}$	$\frac{7}{24}$	$\frac{5}{12}$
v_2	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{3}{8}$
v_3	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

The transitions has been chosen such that $v_3(t_1) = \mu(s_1)$, $v_2(t_2) = \mu(s_2)$, $v_1(t_3) = \mu(s_3)$, $v_1(\{t_1\} \cup \{t_2\}) = \mu(\{s_1\} \cup \{s_2\})$, $v_2(\{t_1\} \cup \{t_3\}) = \mu(\{s_1\} \cup \{s_3\})$, $v_3(\{t_2\} \cup \{t_3\}) = \mu(\{s_2\} \cup \{s_3\})$. Then it is

easy to see that s_i is simulated by t_i for all $i = 1, 2, 3$. Moreover, the last set of equalities shows that $s \prec_{prio} t$. However, we do not have $s \prec t$. Indeed for all transitions $t \xrightarrow{a} v$ (combined or not) from t, we can find a set $E \subseteq \cup_{i \in \{1,2,3\}} \{s_i, t_i\}$ containing s_i if and only if it contains t_i, and such that $\mu(E) > v(E)$.

Remark 1 *By Theorem 3 item 3, for the PA of Ex. 1, we have $s \prec_{\frac{1}{8}}^{prio} t$ and $t \prec_{\frac{1}{8}}^{prio} s$. Now, the only state $\frac{1}{8}$-a-priori bisimilar to x is, here again, x itself. Hence, for $s \xrightarrow{a} \mu_x$ and $E = \{x\}$, there exists no transition $t \xrightarrow{a} v$ such that $\mu(E) \leq v(\sim_{\frac{1}{8}}^{prio} (E)) + \frac{1}{8}$, hence s and t are not $\frac{1}{8}$-a-priori bisimilar. As a consequence, two-way a-priori simulation is different from a-priori bisimulation, and the negation is needed in the logical characterization of bisimilarity.*

    Decidability of A Priori Simulation. An interesting fact is that it is NP-hard to decide a-priori simulation and bisimulation, even when $\varepsilon = 0$. This contrasts with classical results on strong simulation and bisimulation whose decision procedures were proven to be in Poly-time (see [2, 24, 6]). The proof of the following theorem, not presented here due to lqck of space, is by reducing the *subset sum problem*, known to be NP-complete ([16]), to our problem.

Theorem 4 *The following problem is NP-complete:*
Input: *A PA $\mathscr{S}$, $s, t \in S$.* **Question:** *Do we have $s \prec^{prio} t$?*

3.2   The logic $\mathscr{L}^N$ for PAs

We saw in the previous subsection that $\mathscr{L}$ is not strong enough for PAs. We now give a logic characterizing our relaxed relations $\prec_\varepsilon$ and $\sim_\varepsilon$. The difference between this logic and $\mathscr{L}^\neg$ is the modal operator that permits to "isolate" a distribution out of a state, and write properties that it satisfies. This allows the semantics to be defined on states, as pointed out as well by D'Argenio et al. [9]. In contrast, Parma and Segala [20] used a semantics on distributions to prove the logical characterisation of bisimulation (with $\varepsilon = 0$).

Definition 8 (The logic $\mathscr{L}^{N,\neg}$) *The syntax differs by one operator from $\mathscr{L}^\neg$:*

$$\mathscr{L}^{N,\neg} \quad \phi := \top \,|\, \neg \phi \,|\, \phi_1 \wedge \phi_2 \,|\, \phi_1 \vee \phi_2 \,|\, \langle a \rangle \{(\phi_i, p_i)\}_{i \in I} \quad I \text{ finite, } p_i \in \mathbb{Q} \cap [0; 1].$$

We write $\mathscr{L}^N$ for the same logic without negation. Let $\varepsilon \in [0; 1]$. The relaxed semantics $\models_\varepsilon$ of $\mathscr{L}^{N,\neg}$ is defined by structural induction on the formulas, in the same way as for $\mathscr{L}^\neg$ except for the modal formula: $s \models_\varepsilon \langle a \rangle \{(\phi_i, p_i)\}_{i \in I}$ iff there exists a transition $s \xrightarrow{a} \mu$ from s such that for all $i \in I$, we have $\mu([\![\phi_i]\!]_\varepsilon) \geq p_i - \varepsilon$.

    As for the logic $\mathscr{L}^\neg$ we can, by structural induction on the formulas, construct for each $\phi \in \mathscr{L}^{N,\neg}$ an associated formula $\phi_\varepsilon \in \mathscr{L}^{N,\neg}$ such that $[\![\phi]\!]_\varepsilon = [\![\phi_\varepsilon]\!]$. Hence, here again, model checking of formulas with the relaxed semantics can be done using the same technique as for the usual semantics. The logics $\mathscr{L}^N$ and $\mathscr{L}^{N,\neg}$ induce the relations $\prec_\varepsilon^{\mathscr{L}^N}$ and $\sim_\varepsilon^{\mathscr{L}^N}$ as in Def. 6. The following example illustrates how this logic differs from $\mathscr{L}$. The key difference is that formula $\langle a \rangle \{(\phi_i, p_i)\}_{i \in I}$ is not equivalent to $\wedge_{i \in I} \langle a \rangle_{p_i} \phi_i$.

Example 4 *Consider the PAs of Ex. 3. As $s \prec^{prio} t$, we also have $s \prec_\varepsilon^{\mathscr{L}} t$ by Theorem 3, and hence every formula of $\mathscr{L}$ satisfied by s is also satisfied by t. However, $s \not\prec t$ and the formula*

$$\langle a \rangle \{((\langle b_1 \rangle_1 \top, \frac{1}{3}), ((\langle b_2 \rangle_1 \top, \frac{1}{4}), ((\langle b_3 \rangle_1 \top, \frac{5}{12})\}$$

of $\mathscr{L}^N$ is not satisfied by t. Note how the semantics forces the v_i's to commit before the three "b_i formulas" are checked.

The following theorem is a logical characterization of the ε-relations. Notice that we need negation once again, since two-way simulation is different from bisimulation.

Theorem 5 $\prec_\varepsilon = \prec_\varepsilon^{\mathscr{L}^N}$ *and* $\sim_\varepsilon = \sim_\varepsilon^{\mathscr{L}^{N,\neg}}$.

Proof. $[\prec_\varepsilon \subseteq \prec_\varepsilon^{\mathscr{L}^N}]$. Let R be an ε-simulation. We prove by structural induction that for all $\phi \in \mathscr{L}^N$, $R(\llbracket \phi \rrbracket) \subseteq \llbracket \phi \rrbracket_\varepsilon$. We prove the case where $\phi = \langle a \rangle \{(\phi_i, p_i)\}_{i \in I}$, since the other cases are trivial. Let $s \in \llbracket \phi \rrbracket$, $t \in R(\{s\})$ and let $s \xrightarrow{a} \mu$ be the associated transition such that for all $i \in I$, $\mu(\llbracket \phi_i \rrbracket) \geq p_i$. Since R is an ε-simulation, there exists a transition $t \xrightarrow{a} \nu$ such that $\mu \mathscr{L}^\varepsilon(R)\nu$. Thus, for all $E \subseteq S$, we have $\mu(E) \leq \nu(R(E)) + \varepsilon$. In particular, given $i \in I$, we get that $\mu(\llbracket \phi_i \rrbracket) \leq \nu(R(\llbracket \phi_i \rrbracket)) + \varepsilon$. But by induction hypothesis, we know that for all $i \in I$, $R(\llbracket \phi_i \rrbracket) \subseteq \llbracket \phi_i \rrbracket_\varepsilon$. This gives us that $\mu(\llbracket \phi_i \rrbracket) \leq \nu(\llbracket \phi_i \rrbracket_\varepsilon)) + \varepsilon$ hence the result since by hypothesis $\mu(\llbracket \phi_i \rrbracket) \geq p_i$.

$[\prec_\varepsilon \supseteq \prec_\varepsilon^{\mathscr{L}^N}]$. We prove that $\prec_\varepsilon^{\mathscr{L}}$ is an ε-simulation. Suppose that $s \prec_\varepsilon^{\mathscr{L}} t$, and let $s \xrightarrow{a} \mu$ be a transition from s. We need to find some $t \to \nu$ such that $\nu(\prec_\varepsilon^{\mathscr{L}}(X)) \geq \mu(X) - \varepsilon$ for all $X \subseteq S$. This will be constructed from a family of $t \to \nu_E$ for finite sets E. Let $n \in \mathbb{N}$ and let $E \subseteq S$ be a finite set such that $\mu(E) \geq \mu(S) - 1/n$.

The key idea is to define the formula $\phi_e^k = \wedge_{e \models \phi_j, \ j \leq k} \phi_j$ for every $e \in E$, and where $(\phi_j)_{j \in \mathbb{N}}$ is an enumeration of the formulas of $\mathscr{L}$. Then for every finite set $X \subseteq S$, we set $\phi_X^k = \vee_{e \in X} \phi_e^k$, and we let $p_X^k := \mu(\llbracket \vee_{e \in X} \phi_e^k \rrbracket)$. Since $e \models \phi_e^k$, we have $\llbracket \phi_e^k \rrbracket \supseteq \prec_\varepsilon^{\mathscr{L}}(\{e\})$,

$$p_X^k = \mu(\llbracket \vee_{e \in X} \phi_e^k \rrbracket) \geq \mu(\prec_\varepsilon^{\mathscr{L}}(X)) \geq \mu(X),$$

and $s \models \langle a \rangle \{(\vee_{e \in X} \phi_e^k, p_X^k)\}_{X \subseteq E}$ for all $k \geq 1$. By hypothesis, we get that $t \models_\varepsilon \langle a \rangle \{(\vee_{e \in X} \phi_e^k, p_X^k)\}_{X \subseteq E}$ for all $k \geq 1$. Let ν_E^k be the associated transition. Then for all $X \subseteq E$ and for all $k \in \mathbb{N}$:

$$\nu_E^k(\llbracket \vee_{e \in X} \phi_e^k \rrbracket) \geq p_X^k - \varepsilon \geq \mu(X) - \varepsilon.$$

Now, $\llbracket \vee_{e \in X} \phi_e^k \rrbracket_\varepsilon$ is decreasing to $\prec_\varepsilon^{\mathscr{L}}(X)$ as k goes to infinity. Since the systems we consider are finitely branching, we can define a transition $t \xrightarrow{a} \nu_E$ such that ν_E is the limit of a subsequence of $\{\nu_E^k\}_{k \in \mathbb{N}}$. That is, there exists an increasing function $\psi : \mathbb{N} \to \mathbb{N}$ such that for all set $Y \subseteq S$ we have $\lim_{k \to \infty} \nu_E^{\psi(k)}(Y) = \nu_E(Y)$. This implies that: $\nu_E(\prec_\varepsilon^{\mathscr{L}}(X)) \geq \mu(X) - \varepsilon$. We have proven the following: for all $s \xrightarrow{a} \mu$, for all $E \subseteq S$ finite, there exists $t \xrightarrow{a} \nu_E$ such that for any $X \subseteq E$ we have $\nu_E(\prec_\varepsilon^{\mathscr{L}}(X)) \geq \mu(X) - \varepsilon$. Let $E_k, k \in \mathbb{N}$ be a growing sequence of finite subsets of S such that $S = \cup_{k \in \mathbb{N}} E_k$. Again, since the system is finitely branching, let ν be the limit of a subsequence of $\{\nu_{E_k}\}_{k \in \mathbb{N}}$. As before, we get that for any $X \subseteq S$ finite, $\nu(\prec_\varepsilon^{\mathscr{L}}(X)) \geq \mu(X) - \varepsilon$.

$[\sim_\varepsilon = \sim_\varepsilon^{\mathscr{L}^N}]$ It can be proven that $\sim_\varepsilon^{\mathscr{L}^{N,\neg}}$ is an ε-bisimulation by following the proof above and using the fact that $\sim_\varepsilon^{\mathscr{L}^{N,\neg}}$ is a symmetric relation (which comes from the presence of negation in $\mathscr{L}_\neg^N$). $\square$

4 A Bisimulation Pseudo-Metric between PAs

4.1 The pseudo-metric d

The notion of ε-bisimulation induces a pseudo-metric on states of a PA, given by the smallest ε such that the states are ε-bisimilar.

Definition 9 (Bisimulation metric) *Given $s, t \in S$, let $d(s,t) = \inf\{\varepsilon \mid s \sim_\varepsilon t\}$.*

Using the finite branching of our systems, we can prove d is a bisimulation pseudo-metric, i.e., states at distance zero are bisimilar. We now discuss the computation of this distance between all states of a given PA. We propose three approaches, the first being exact, and the others approximate. The two first compute the distance iteratively, updating a function $d_i : S \times S \to [0;1]$, in the same way as it was done for deterministic PAs in [12]. This approach is close to the classical iterative algorithms for computing simulation and bisimulation on probabilistic systems, see [2, 6]. It makes use of a network flow computation. The algorithm for the first approach is the left one in Fig. 2. Given $d : S \times S \to [0;1]$, and $\varepsilon \geq 0$, let R^d_ε be the relation on $S \times S$ defined as: $R^d_\varepsilon(s,t)$ iff $d(s,t) \leq \varepsilon$.

Algorithm $\mathscr{A}$: exact computation
Input: A finite PA $\mathscr{S} = (S, Act, \mathscr{D})$.
Output: $d : S \times S \to [0;1]$.
Method:
Let $d_0(s,t) = 0 \ \forall s,t \in S \times S$. Let $j = 0$
Until $d_j = d_{j+1}$ do begin:
$\quad d_{j+1} = d_j$.
$\quad$ For all $(s,t) \in S \times S$ do begin:
$\quad\quad$ For all $a \in Act$ do begin:
$\quad\quad\quad$ For all $s \xrightarrow{a} \mu$ do begin:
$\quad\quad\quad\quad$ Let $d_{j+1}(s,t)$ be the smallest $\varepsilon \in [0;1]$ s.t. $\exists \, t \xrightarrow{a} \nu$ such that the maximum flow of network $\mathscr{N}(\mu, \nu, R^{d_j}_\varepsilon)$ is $\geq \mu(S) - \varepsilon$.
$\quad\quad$ end end end
$\quad j = j + 1$ end return d_{j-1}.

Algorithm $\mathscr{B}$: computation up to $1/n$
Input: A finite PA $\mathscr{S} = (S, Act, \mathscr{D})$, $n \in \mathbb{N}$.
Output: $d : S \times S \to [0;1]$.
Method:
Let $d_0(s,t) = 0 \ \forall s,t \in S \times S$. Let $j = 0$
For $m = n - 1$ to 0 do begin:
Until $d_j = d_{j+1}$ do begin:
$\quad d_{j+1} = d_j$.
$\quad$ For all (s,t) s.t. $d(s,t) = 0$ do begin:
$\quad\quad$ For all $a \in Act$ do begin:
$\quad\quad\quad$ For all $s \xrightarrow{a} \mu$ do begin:
$\quad\quad\quad\quad$ If $\exists \, t \xrightarrow{a} \nu$ s.t. the maximum flow of network $\mathscr{N}(\mu, \nu, R^{d_j}_0)$ is $< \mu(S) - m\varepsilon/n$ then let $d_{j+1}(s,t) = (m+1)\varepsilon/n$.
$\quad\quad$ end end end
$\quad j = j + 1$ end end return d_{j-1}.

Figure 2: Computations of metric d on $\mathscr{S}$

Proposition 3 *Algorithm $\mathscr{A}$ correctly outputs the distance d between all pairs of states in S. Moreover, algorithm $\mathscr{A}$ runs in time $O(|S|^9 \cdot |Act| \cdot l^2)$, where l is the maximal number of transitions with the same label issued from a single state.*

This algorithm is quite expensive, and hence we propose other approaches that approximate the distance. The first one is a variation of algorithm $\mathscr{A}$ and is the right-hand algorithm of Fig. 2. Let $1/n$ be the accuracy we are interested in, for $n \in \mathbb{N}$. The idea is to compute m/n-bisimulation iteratively, for m decreasing from $n-1$ to 0. These relations are decreasing as ε-bisimilarity implies ε'-bisimilarity for any $\varepsilon' > \varepsilon$. At each iteration of that loop, states whose distance have not been established yet have d-value 0. At step $m < n$, these states will be given distance $(m+1)/n$ if they are not m/n-bisimilar. The relation consisting of states at zero distance will decrease at every j step. For every pair of states, the worst number of flow networks to be computed will be n (this happens if the states are bisimilar). Hence the algorithm runs in $O(|S|^5 \cdot n \cdot |Act| \cdot l^2)$. Of course, some values of m can be ignored and we can save some time.

The last algorithm that we propose uses recent work of Zhang et al. [24], to update efficiently the flow computation in algorithm $\mathscr{A}$. The algorithm of Zhang et al. computes strong bisimularity on a probabilistic automaton in time $O(|S| \cdot m^2)$, where m is the total number of transitions. By a slight

generalization of this algorithm to our context, we can compute ε-bisimularity on $\mathscr{S}$ in time $O(|S| \cdot m^2)$, for any given ε. Using a dichotomic approach, given two states s and t, we can compute $d(s,t)$ up to an additive approximation factor δ in time $O(|S| \cdot m^2 \cdot log(\delta))$, and thus we can compute d up to an additive approximation factor δ in time $O(|S|^3 \cdot m^2 \cdot log(\delta))$.

4.2 The decayed distance d^λ

In the previous definitions, differences in the far future have as much importance as those in the near future. We can relax the impact of the future, or instead the impact of short term transitions, by using a decayed relaxation. Instead of a fixed relaxation of parameter ε, we can ask for a relaxation that changes as we get further from the starting state. As we get deeper through the transitions, the parameter could get bigger, hence diminishing the importance of further differences, or symmetrically, we could make the parameter smaller. In order to leave this flexible, we will use a function $\lambda : [0;1] \to [0;1]$. If $x \leq \lambda(x)$, the future will be neglected whereas $x \geq \lambda(x)$ will make the future more precise. If λ is the identity, we get the previous notions. We will describe below a natural choice for λ, but first, let us define the new semantics. We write λ^n for to the n-th self composition of function λ, $n \in \mathbb{N}$, and λ^0 is the identity function on $[0;1]$. Also, $\mathscr{L}_n^N$ will be the set of formulas of $\mathscr{L}^N$ of depth at most $n \in \mathbb{N}$.

Definition 10 (The (ε, λ)-semantics) *Let $\varepsilon \in [-1;1]$, and $\lambda : [0;1] \to [0;1]$ The syntax is the one of $\mathscr{L}^N$. The semantics $\models_\varepsilon^\lambda$ is defined similarly as for $\mathscr{L}^N$ except for the modal operator. Given $\phi \in \mathscr{L}^N$, let $[\![\phi]\!]_\varepsilon^\lambda = \{s \in S | s \models_\varepsilon^\lambda \phi\}$. Given $s \in S$, $s \models_\varepsilon^\lambda \langle a \rangle \{(\phi_i, p_i)\}_{i \in I}$ iff there exists a transition $s \xrightarrow{a} \mu$ such that for all $i \in I$, we have $\mu([\![\phi_i]\!]_{\lambda(\varepsilon)}^\lambda) \geq p_i - \varepsilon$. This semantics induces the relations $\prec_\varepsilon^{\mathscr{L}_n^N, \lambda}$ and $\sim_\varepsilon^{\mathscr{L}_n^{N,\neg}, \lambda}$ as in Def. 6.*

The relations $\prec_\varepsilon^{\mathscr{L}_n^N, \lambda}$ and $\sim_\varepsilon^{\mathscr{L}_n^{N,\neg}, \lambda}$ are defined for formulas of a given maximal depth n, because in order to compute $\models_\varepsilon^\lambda$ on $S \times S$ for a given $\varepsilon \in [0;1]$, we may have to compute the $\models_{\lambda^n(\varepsilon)}^\lambda$ for all $n \in \mathbb{N}$.

Most of the time, one wants to give less importance to the future. In these situations, the decay is called a discount and could be exponential, as in [10, 14]. In our case, this would correspond to asking that there is a constant $0 < c < 1$ such that $1 - \lambda(\varepsilon) = c(1 - \varepsilon)$, i.e. $\lambda(\varepsilon) = 1 - c \cdot (1 - \varepsilon)$.

The associated simulation and bisimulation are variations of Def. 2:

Definition 11 (Order n (ε, λ)-bi/simulation) *Given $n \in \mathbb{N}$, an order n (ε, λ) simulation on $\mathscr{S}$ is a decreasing sequence of relations $R_0, ..., R_n$ on S such that $R_0 = S \times S$, and for all $i \in [1;n]$, whenever sR_it, if $s \xrightarrow{a} \mu$, then there exists $t \xrightarrow{a} \nu$ such that $\mu \mathscr{L}^{\lambda^{n-i}(\varepsilon)}(R_{i-1})\nu$. We write $s \prec_\varepsilon^{\lambda,n} t$ if there exists an order n (ε, λ)-simulation R_n on $\mathscr{S}$ such that sR_nt, and we write $s \sim_\varepsilon^{\lambda,n} t$ if $s \prec_\varepsilon^{\lambda,n} t$ and $t \prec_\varepsilon^{\lambda,n} s$.*

Proposition 4 *Let $n \in \mathbb{N}$. Then $\prec_\varepsilon^{\mathscr{L}_n^N, \lambda} = \prec_\varepsilon^{\lambda,n}$ and $\sim_\varepsilon^{L_n^{\mathscr{N},\neg}, \lambda} = \sim_\varepsilon^{\lambda,n}$*

Given $s, t \in S$, we define $d^\lambda(s,t) = \min\{\lambda^{-n}(1) \mid s \sim_{\lambda^{-n}(1)}^{\lambda,n} t\}$. We can compute the distance d^λ using the algorithm of Fig. 3.

Proposition 5 *Algorithm $\mathscr{C}$ of Fig. 3 runs in time $O(|S|^5 \cdot |Act| \cdot l^2)$.*

Proof. Direct, using $O(|S|^3)$ flow network computations. $\square$

Algorithm $\mathscr{C}$: Computation of the discounted metric d_λ on $\mathscr{S}$
Input: A finite PA $\mathscr{S} = (S, Act, \mathscr{D})$ $N \in \mathbb{N}$.
Output: $d : S \times S \to [0; 1]$.
Method:
Let $d_0(s,t) = 1$ for all $s,t \in S \times S$. Let $\delta = 1/N$.
Let $R_0 = S \times S$.
For $n = 1$ to N do begin:
    For all $(s,t) \in S \times S$ do begin:
        For all $a \in Act$ do begin:
            For all $s \xrightarrow{a} \mu$ do begin:
                Let $R_{n+1} = R_n$.
                If there exists no transition $t \xrightarrow{a} \nu$ such that the maximum
                flow of the network $\mathscr{N}(\mu, \nu, R_n)$ is greater than or equal
                to $\mu(S) - (1 - n \cdot \delta)$, then: let $d(s,t) = 1 - n \cdot \delta$, and let
                $R_{n+1} = R_{n+1} - \{(s,t)\}$.
    end end end end
  return d.

Figure 3: Computation of metric d_λ on $\mathscr{S}$

4.3 Comparison to other metrics on probabilistic systems

During the past ten years, several metrics have been defined in the context of Probabilistic Automata or closely related models such as Labeled Markov Chains [11, 15, 13], reactive probabilistic transition systems [5], Markov Decision Processes [14, 13, 21], or more general game processes [1]. Most of these metrics are variations of the metric of [11]. In [5, Th. 4.6], an equivalent metric is defined as a terminal coalgebra, using category theory. In [4], the same authors give an algorithm to compute in polynomial time this metric, relying on linear programming computation for a transshipment optimization problem. This approach is for deterministic models and it is applied in [14] and related papers to compute metrics between Markov Decision Processes. Most of these algorithms introduce a decay factor to make the computation tractable. In [1] the authors consider metrics between systems which allow non determinism, but the complexity of the algorithms presented in [8] to compute the metrics is at best PSPACE.

The main difference between our metric and those is that differences along paths are not accumulated in ours, even in the discounted metric: other metrics all involve comparing (among others) the probability of paths, and this makes these metrics straightforwardly different from ours, as we never multiply probability values. In [11, 14], the metric can be computed using a familly of functional expressions $\mathscr{F}^c$ from states to $[0;1]$ that play the same role as the quantitative formulas of [1]. Given s,t states of the system, the distance $d^c(s,t)$ is then defined as $d^c(s,t) = \sup_{f \in \mathscr{F}^c} |f^c(s) - f^c(t)|$. This distance is incomparable with ours, as shows the following example.

Example 5 *Let $0 < \varepsilon < 1$, and consider the systems of Fig. 4. The distance between s and t is always greater with our distance, as $d^c(s,t) = c^2 \cdot \varepsilon^2 \leq \varepsilon = d(s,t)$. On the contrary, for $\varepsilon < 1/2$, there is some c for which the distance between u and v is smaller with our distance. Indeed, $d^c(u,v) >= c^2(\varepsilon/2 - \varepsilon^2)$, whereas $d(u,v) = \varepsilon$. Hence we obtain $d(u,v) \leq d^c(u,v)$ by taking $c \in [0;1]$ such that $c^2(3\varepsilon/2 - \varepsilon^2) > \varepsilon$; for example, with $\varepsilon = 1/4$ and $c^2 > 4/5$. The example can be adapted for $\varepsilon \geq 1/2$.*
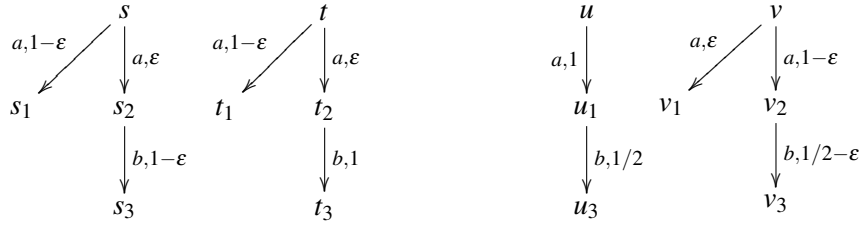
Figure 4: Our metric differs from those based on paths.

4.4 The Metric d on a Process Algebra

We consider a process algebra on an extension of the model of PAs: the extension is that we distinguish between *Input* and *Output* action labels. As in [15], we assume a set of underlying labels Σ, and suppose that the labels of the PAs belong to a set $L = L! \cup L?$, where $L? = \{a? | a \in \Sigma\}$ and $L! = \{a! | a \in \Sigma\}$ are the sets of Input and Ouput labels respectivelly. Two PAs $\mathscr{S}_1 = (S_1, Act_1, \mathscr{D}_1, s_1^0)$ and $\mathscr{S}_2 = (S_2, Act_2, \mathscr{D}_2, s_2^0)$ will synchronize on labels in $Act_1 \cap Act_2$.

We only present the operators for non-deterministic choice and parallel composition, as the other operators can be taken as in [15].

Non Deterministic Choice. Let $\mathscr{S}_1, ..., \mathscr{S}_k$ be PAs with respective state spaces $S_1, ..., S_k$ and initial states $s_1^0, ..., s_k^0$. Let $\{a_1, ..., a_l\} \subseteq L$, and for each a_i let $\{\mu_{i,j} | j \in [1; n_i]\}$ be a finite family of distributions on $\{s_1^0, ..., s_k^0\}$. We define: $\mathscr{S}' = \langle + \rangle \, _{i=1}^{l} \{s \xrightarrow{a_i} \mu_{i,j}, j \in [1; n_i]\} \{\mathscr{S}_1, ..., \mathscr{S}_k\}$, a PA whose state space is $S' = \{s\} \uplus_{i \in [1;k]} S_i$, and initial state s. Transitions from s are all the $s \xrightarrow{a_i} \mu_{i,j}$. There may be several transitions from s with the same label a_i. $\mathscr{S}'$ accepts the input label a or outputs the label a, depending on $a \in L?$ or $a \in L!$.

Parallel Composition. Given PAs $\mathscr{S}_i = (S_1, Act_i, \mathscr{D}_i, s_i^0)$, $i = 1, 2$, we define the parallel composition $\mathscr{S}' = \mathscr{S}_1 || \mathscr{S}_2$. The synchronisation is on labels in $Act_1 \cap Act_2$. The state space of $\mathscr{S}'$ is $S' = S_1 \times S_2$, with initial state $s'^0 = (s_1^0, s_2^0)$. The set of labels of $\mathscr{S}'$ is $Act' = Act_1 \cup Act_2$. Given μ and ν two distributions on disjoint sets S_1 and S_2, given $X \subseteq S_1 \times S_2$, let $\mu \otimes \nu(X) = \sum_{(s,t) \in X} \mu(s) \cdot \nu(t)$. Given states $s \in S_1$ and $t \in S_2$, we expect the following synchronized transitions:

- *Synchronization on input labels in $Act_1 \cap Act_2$*: $\forall a \in Act_1 \cap Act_2$, if $s \xrightarrow{a?} \mu$ and $t \xrightarrow{a?} \nu$, then there is a transition $(s,t) \xrightarrow{a?} \mu \otimes \nu$ on $\mathscr{S}$.

- *Synchronization in the Output/Input*: if $s \xrightarrow{a!} \mu$ and $t \xrightarrow{a?} \nu$, then there is a transition $(s,t) \xrightarrow{a!} \mu \otimes \nu$ (and symmetrically)

- *Asynchronous evolution on labels in $Act_1 \setminus Act_2$*: given $a \in Act_1 \setminus Act_2$ if $s \xrightarrow{a} \mu$, then $(s,t) \xrightarrow{a} \mu \otimes \delta_t$, where δ_t is the Dirac distribution on t (symmetrically on S_2).

We prove that the distance d is non expansive with respect to the parallel operator: when composing two processes with a third one, the distance does not increase. Non expansiveness with respect to other operators is more common.

Theorem 6 $d(\mathscr{S}_1 || \mathscr{S}, \mathscr{S}_2 || \mathscr{S}) \leq d(\mathscr{S}_1, \mathscr{S}_2)$.

Proof. Let s, s_1, s_2 be three states of $\mathscr{S}, \mathscr{S}_1$ and $\mathscr{S}_2$ respectivelly. Let $\varepsilon \geq 0$, and suppose $s_1 \sim_\varepsilon s_2$. We want to prove that $s_1 \times s$, which is a state of $\mathscr{S}_1 || \mathscr{S}$, is ε-bisimilar to $s_2 \times s$, a state of $\mathscr{S}_2 || \mathscr{S}$. We prove by induction on $n \in \mathbb{N}$ that if $s_1 \sim_\varepsilon^n s_2$, then $s_1 \times s \sim_\varepsilon^n s_2 \times s$. We will use the following notations: Given $\phi \in \mathscr{L}^N$, and $j \in \{1, 2\}$:

$$\llbracket \phi \rrbracket_{\mathscr{S}_j \| \mathscr{S}} = \{s_j \times s \mid s_j \in S_j, \ s \in S, \text{ and } s_j \times s \models \phi\}$$

Where the semantics is taken on the PA $\mathscr{S}_j \| \mathscr{S}$. Let

$$\text{left}_j(\llbracket \phi \rrbracket_{\mathscr{S}_j \| \mathscr{S}}) = \{v \in S_j \mid \exists u \in S \text{ s.t. } u \times v \models \phi\}$$

Given $v \in S$, let

$$\text{right}_j(v, \llbracket \phi \rrbracket_{\mathscr{S}_j \| \mathscr{S}}) = \{u \in S_1 \mid u \times v \models \phi\}$$

The key case is when $\phi = \langle a \rangle \{(\phi_i, p_i)\}_{i \in I} \in \mathscr{L}^N$, with depth n. Suppose $s_1 \times s \models \phi$. Then there exists a transition $s_1 \times s \xrightarrow{a} \mu_1 \otimes v$ on $\mathscr{S}'$ such that for all $i \in I$, $(\mu_1 \otimes v)(\llbracket \phi_i \rrbracket) \geq p_i$.

By hypothesis, $s_1 \sim^n_\varepsilon s_2$. Hence, there exists a transition $s_2 \xrightarrow{a} \mu_2$ such that $\mu_1 \mathscr{L}^\varepsilon(\sim_\varepsilon) \mu_2$.

We know that for all $i \in I$,

$$\mu_1 \otimes v(\llbracket \phi_i \rrbracket) = \sum_{v \in \text{left}_1(\llbracket \phi_i \rrbracket)} v(v) \cdot \sum_{u \in \text{right}_1(v, \llbracket \phi_i \rrbracket)} \mu_1(v).$$

Given $v \in \text{left}(\llbracket \phi_i \rrbracket)$ (hence $v \in S$), we know that $\mu_2(\sim^n_\varepsilon \text{right}_1(v, \llbracket \phi_i \rrbracket)) \geq \mu_1(\text{right}_1(v, \llbracket \phi_i \rrbracket)) - \varepsilon$.

Moreover, by induction hypothesis, $\sim^n_\varepsilon \text{right}_1(v, \llbracket \phi_i \rrbracket) \subseteq \text{right}_2(v, \llbracket \phi_i \rrbracket_\varepsilon)$. Indeed,

$$\sim^n_\varepsilon \text{right}_1(v, \llbracket \phi_i \rrbracket) \cap S_2 = \sim^n_\varepsilon \{u \in S_2 \mid \exists u' \in S_1 \text{ s.t. } u \sim^n_\varepsilon u' \text{ and } u' \models \phi_i\}.$$

We get the result since by induction hypothesis, if $u \sim^n_\varepsilon u'$ and $s' \in S$, we have $u \times s' \sim^n_\varepsilon u' \times s'$.

This implies that $\mu_2(\text{right}_2(v, \llbracket \phi_i \rrbracket_\varepsilon)) \geq \mu_1(\text{right}_1(v, \llbracket \phi_i \rrbracket)) - \varepsilon$. Finally,

$$\sum_{v \in \text{left}_2(\llbracket \phi_i \rrbracket_\varepsilon)} v(v) \cdot \sum_{u \in \text{right}_2(v, \llbracket \phi_i \rrbracket_\varepsilon)} \mu_2(v) \geq \sum_{v \in \text{left}_1(\llbracket \phi_i \rrbracket)} v(v) \cdot \sum_{u \in \text{right}_1(v, \llbracket \phi_i \rrbracket)} \mu_1(v) - \varepsilon.$$

Hence $\mu_2 \otimes v(\llbracket \phi_i \rrbracket_\varepsilon) \geq \mu_1 \otimes v(\llbracket \phi_i \rrbracket) - \varepsilon$. This proves that $s_2 \times s \models_\varepsilon \phi$. $\square$

5 Examples

We build a benchmark set of deterministic PAs to compare distances. The processes are variations of a basic one from which we delete some transitions. The state space of the basic PA is a square grid of $n \times n$. The set of actions is $\{a\}$. All the computations on the state indices are done modulo n: the grid is a torus. The a-transitions from state (i, j) are as follows:

from (i, j) to	$(i, j-1) : 0.1$	$(i, j+1) : 0.5$	$(i-1, j) : 0.25$	$(i+1, j) : 0.15$

This basic PA is compared to variations of it obtained by deleting in some states the transition of label a (to all successors). Note that the basic process is bisimilar to the one-state process that can do a with probability 1. We consider the distances between states with same indices of the different systems. Fig. 5 illustrates some PAs and the impact of the deletion of transitions for the two distances that we defined. The distance d^λ is illustrated in the bottom grids of the figure. The linear function for λ is the following, with $N = 20$. We take $\delta = 1/N$, and let $\lambda(x) = x + \delta$ if $x \in [0; 1 - \delta]$, and $\lambda(x) = 1$ if $x \in [1 - \delta; 1]$. One can observe that the decay distances fade out when further from the difference, whereas d is more constant.

It would be nice to compare these grids with others obtained from other known metrics. We leave that for future work, as we have no implementation of other metrics that can handle more than 25 states.
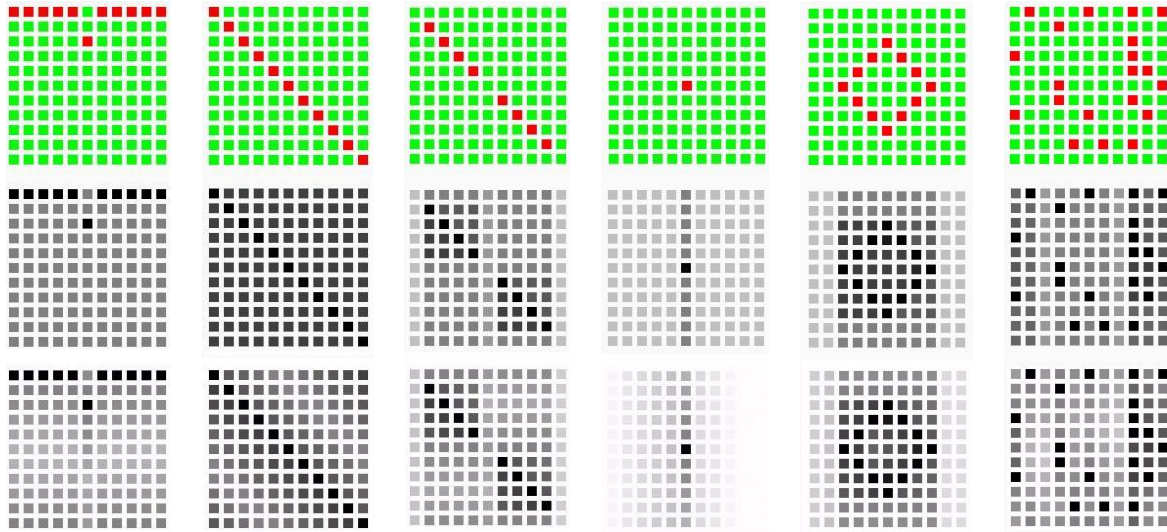
Figure 5: The red entries of the top grid are states that have lost their *a*-transition. The two other grids have at entry (i, j) the distance between state (i, j) of top grid to state (i, j) in the basic PA. The darker is the entry, the bigger the distance. The middle grid illustrates the basic distance, the other the decayed one.

6   Conclusion

We presented relaxed notions of ε-simulation and ε-bisimulation. When $\varepsilon = 0$ we retrieve the usual notions of bisimulation and simulation on PAs. We gave logical characterisations of these notions and algorithms to compute in PTIME two corresponding pseudo-metrics, one that discounts the future, and one that does not. We showed that our distance is not expansive with respect to process algebra operators. We also showed that the basic logic $\mathscr{L}^{(\neg)}$ characterises a notion weaker than ε-(bi)simulation, called *a priori ε-(bi)simulation*. Interestingly, we have proven this notion NP-difficult to decide. Further work includes relaxing what is called probabilistic bisimulation and studying the associated distances; implementing our third proposal of algorithm to compute d, using a modification of the algorithm of [24]; investigating further the weaknesses and strengths of the different metrics defined so far.

References

[1] Luca de Alfaro, Rupak Majumdar, Vishwanath Raman & Marielle Stoelinga (2007): *Game Relations and Metrics*. In: *LICS '07*. IEEE Computer Society, pp. 99–108, doi:10.1109/LICS.2007.22.

[2] Christel Baier (1996): *Polynomial Time Algorithms for Testing Probabilistic Bisimulation and Simulation*. In: *CAV'96. LNCS* 1102, pp. 38–49, doi:10.1007/3-540-61474-5_57.

[3] Franck van Breugel, Babita Sharma & James Worrell (2007): *Approximating a Behavioural Pseudometric Without Discount for Probabilistic Systems*. In: *FOSSACS. LNCS* 4423, Springer, pp. 123–137, doi:10.1007/978-3-540-71389-0_10.

[4] Franck van Breugel & James Worrell (2001): *An Algorithm for Quantitative Verification of Probabilistic Transition Systems*. In: *CONCUR '01*. Springer-Verlag, pp. 336–350, doi:10.1007/3-540-44685-0_23.

[5] Franck van Breugel & James Worrell (2004): *A behavioural pseudometric for probabilistic transition systems*. *Theor. Comput. Sci.* 331(1), pp. 115–142, doi:10.1007/3-540-48224-5_35.

[6] Stefano Cattani & Roberto Segala (2002): *Decision Algorithms for Probabilistic Bisimulation*. In: *CONCUR '02*. Springer-Verlag, pp. 371–385, doi:10.1007/3-540-45694-5_25.

[7] Stefano Cattani, Roberto Segala, Marta Kwiatkowska & Gethin Norman (2005): *Stochastic transition systems for continuous state spaces and non-determinism*. In: *FOSSACS'05*. *LNCS* 3441, Springer Verlag, pp. 125–139, doi:10.1007/978-3-540-31982-5_8.

[8] K. Chatterjee, Luca de Alfaro, Rupak Majumdar & Vishwanath Raman (2010, to appear.): *Algorithms for game metrics*. *LMCS: Logical Methods in Computer Science* doi:10.2168/LMCS-6(3:13)2010.

[9] Pedro R. D'Argenio, Nicolás Wolovick, Pedro Sánchez Terraf & Pablo Celayes (2009): *Nondeterministic labeled Markov processes: bisimulations and logical characterization*. *QEST'09* , pp. 11–20doi:10.1109/QEST.2009.17.

[10] Josée Desharnais, Abbas Edalat & Prakash Panangaden (2002): *Bisimulation for Labeled Markov Processes*. *Information and Computation* 179(2), pp. 163–193, doi:10.1.1.16.5653.

[11] Josée Desharnais, Vineet Gupta, R. Jagadeesan & P. Panangaden (1999): *Metrics for Labeled Markov Processes*. In: *CONCUR'99*. LNCS, Springer-Verlag, pp. 258–273, doi:10.1007/3-540-48320-9_19.

[12] Josée Desharnais, François Laviolette & Mathieu Tracol (2008): *Approximate analysis of probabilistic processes: logic, simulation and games*. In: *QEST'08*. IEEE Computer Society, pp. 264–273, doi:10.1109/QEST.2008.42.

[13] Josée Desharnais, François Laviolette & Sami Zhioua (2006): *Testing Probabilistic Equivalence Through Reinforcement Learning*. In: *FSTTCS'06*. *LNCS* 4337, Springer, pp. 236–247, doi:10.1007/11944836_23.

[14] Norm Ferns, Prakash Panangaden & Doina Precup (2004): *Metrics for finite Markov decision processes*. In: *UAI'04*. AUAI Press Arlington, Virginia, United States, pp. 162–169, doi:10.1.1.87.9485.

[15] Norman Ferns, Prakash Panangaden & Doina Precup (2005): *Metrics for Markov Decision Processes with Infinite State Spaces*. In: *UAI'05*. AUAI Press, p. 201, doi:10.1007/BF01908587.

[16] Michael R. Garey & David S. Johnson (1979): *Computers and Intractability: A Guide to the Theory of NP-completeness*. W. H. Freeman & Co., New York, NY, USA.

[17] Alessandro Giacalone, Chi-Chang Jou & Scott A. Smolka (1990): *Algebraic Reasoning for Probabilistic Concurrent Systems*. In: *Proc. IFIP TC2 Working Conference on Programming Concepts and Methods*. North-Holland, pp. 443–458, doi:10.1.1.56.3664.

[18] Holger Hermanns (2002): *Interactive Markov chains, and the quest for quantified quality*. Springer-Verlag, Berlin, Heidelberg, doi:10.1007/3-540-45804-2.

[19] Kim G. Larsen & Arne Skou (1991): *Bisimulation through Probablistic Testing*. Information and Computation 94, pp. 1–28, doi:10.1.1.158.9316.

[20] Augusto Parma & Roberto Segala (2007): *Logical characterisation of Bisimulations for Discrete Probabilistic Systems*. In: *FOSSACS'07*. *LNCS* 4423, Springer-Verlag, pp. 287–301, doi:10.1007/978-3-540-71389-0_21.

[21] Martin L. Puterman (1994): *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley.

[22] Roberto Segala & Nancy Lynch (1994): *Probabilistic Simulations for Probabilistic Processes*. In: *CONCUR'94*. *LNCS* 836, Springer-Verlag, pp. 481–496, doi:10.1007/BFb0015027.

[23] Roberto Segala & Andrea Turrini (2007): *Approximated Computationally Bounded Simulation Relations for Probabilistic Automata*. In: *CSF'07*. pp. 140–156, doi:10.1007/3-540-45694-5_25.

[24] Lijun Zhang, Holger Hermanns, Friedrich. Eisenbrand & David N. Jansen (2007): *Flow faster: efficient decision algorithms for probabilistic simulations*. In: *TACAS'07*. *LNCS* 4424, Springer-Verlag, pp. 155–169, doi:10.1007/978-3-540-71209-1_14.