Approximating a Behavioural Pseudometric without Discount for Probabilistic Systems

Franck van Breugel[1], Babita Sharma[1], and James Worrell[2] ⋆

[1] York University
4700 Keele Street, Toronto, M3J 1P3, Canada
[2] Oxford University Computing Laboratory
Parks Road, Oxford, OX1 3QD, England

Abstract. Desharnais, Gupta, Jagadeesan and Panangaden introduced a family of behavioural pseudometrics for probabilistic transition systems. These pseudometrics are a quantitative analogue of probabilistic bisimilarity. Distance zero captures probabilistic bisimilarity. Each pseudometric has a discount factor, a real number in the interval $(0, 1]$. The smaller the discount factor, the more the future is discounted. If the discount factor is one, then the future is not discounted at all. Desharnais et al. showed that the behavioural distances can be calculated upto any desired degree of accuracy if the discount factor is smaller than one. In this paper, we show that the distances can also be approximated if the future is not discounted. A key ingredient of our algorithm is Tarski's decision procedure for the first order theory over real closed fields. By exploiting the Kantorovich-Rubinstein duality theorem we can restrict to the existential fragment.

1 Introduction

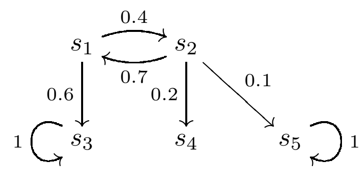
For systems that contain quantitative information, like, for example, probabilities, time and costs, several *behavioural pseudometrics* (and closely related notions) have been introduced (see, for example, [5, 7, 9, 12, 13, 16–19, 24, 29]). Such a pseudometric assigns a distance, a nonnegative real number, to every pair of states. This distance captures the behavioural similarity of the states. The smaller the distance, the more alike the states behave. Distance zero captures that the states are behaviourally equivalent.

In this paper, we focus on *probabilistic transition systems*, which are a variant of Markov chains. Desharnais, Gupta, Jagadeesan and Panangaden [16] introduced a family of behavioural pseudometrics for these systems. These pseudometrics assign a distance, a real number in the interval $[0, 1]$, to each pair of states of the probabilistic transition system. The distance captures the behavioural similarity of the states. The smaller the distance, the more alike the states behave. The distance is zero if and only if the states are *probabilistic bisimilar*, a behavioural equivalence introduced by Larsen and Skou [23].

⋆ The first and second author were supported by NSERC.

The definition of the behavioural pseudometrics of Desharnais et al. is parametrized by a *discount factor* δ, a real number in the interval $(0, 1]$. The smaller the discount factor, the more (behavioural differences in) the future are discounted. In the case that δ equals one, the future is not discounted. All differences in behaviour, whether in the near or far future, contribute alike to the distance. For systems that (in principle) run forever, we may be interested in all these differences and, hence, in the pseudometric that does not discount the future.

Consider the following probabilistic transition system.



The distance between the states s_1 and s_2 is 0.01 if $\delta = 0.25$, 0.06 if $\delta = 0.5$, 0.15 if $\delta = 0.75$ and 0.32 if $\delta = 1$. The less we discount the future, the more s_1 and s_2 behave alike.

In [14], Desharnais et al. presented an *algorithm* to *approximate* the behavioural distances for δ smaller than one. The first and third author [4] presented also an approximation algorithm for δ smaller than one.

There is a fundamental difference between pseudometrics that discount the future and the one that does not. This is, for example, reflected by the fact that all pseudometrics that discount the future give rise to the same topology, whereas the pseudometric that does not discount the future gives rise to a different topology (see, for example, [16, page 350]). As a consequence, it may not be surprising that neither approximation algorithm mentioned in the previous paragraph can be modified in an obvious way to handle the case that δ equals one.

The main contribution of this paper is an algorithm that approximates the behavioural distances in the case that δ equals one. Starting from the definition of the pseudometric by Desharnais et al., we first give a characterization of the pseudometric as the greatest (post-)fixed point of a function from a complete lattice to itself. Next, we dualize this characterization exploiting the Kantorovich-Rubinstein duality theorem [22]. Subsequently, we show, exploiting the dual characterization, that a pseudometric being a post-fixed point can be expressed in the existential fragment of the first order theory over real closed fields. Based on the fact that this first order theory is decidable, a result due to Tarski [27], we show how to approximate the behavioural distances. Finally, we discuss an implementation of our algorithm in Mathematica.

2 Systems and pseudometrics

Some basic notions that will play a role in the rest of this paper are presented below. First we introduce the systems of interest: probabilistic transition systems.

Definition 1. *A probabilistic transition system is a tuple $\langle S, \pi \rangle$ consisting of*

- *a finite set S of states and*
- *a function $\pi : S \times S \to [0,1] \cap \mathbb{Q}$ satisfying $\sum_{s' \in S} \pi(s, s') \in \{0, 1\}$.*

We write $s \to$ if $\sum_{s' \in S} \pi(s, s') = 1$ and $s \not\to$ if $\sum_{s' \in S} \pi(s, s') = 0$.

For states s and s', $\pi(s, s')$ is the probability of making a transition to state s' given that the system is in state s. Each state s either has no outgoing transitions ($s \not\to$) or a transition is taken with probability 1 ($s \to$). To simplify the presentation, we do not consider the case that a state s may refuse to make a transition with some probability, that is, $\sum_{s' \in S} \pi(s, s') \in (0, 1)$. However, all our results can easily be generalized to handle that case as well. We also do not consider transitions that are labelled with actions. All our results can also easily be modified to handle labelled transitions. In the labelled case, the definition of probabilistic transition system is a mild generalisation of the notion of Markov chain.

In the rest of this paper, we will use the following probabilistic transition system as our running example.

Example 1. We consider a probabilistic transition system with five states: s_1, s_2, s_3, s_4 and s_5. The following table contains the transition probabilities and, hence, captures π.

	s_1	s_2	s_3	s_4	s_5
s_1	0	0.4	0.6	0	0
s_2	0.7	0	0	0.2	0.1
s_3	0	0	1	0	0
s_4	0	0	0	0	0
s_5	0	0	0	0	1

This probabilistic transition system is depicted in the introductory section.

We consider states of a probabilistic transition system behaviourally equivalent if they are probabilistic bisimilar [23].

Definition 2. *Let $\langle S, \pi \rangle$ be a probabilistic transition system. An equivalence relation $\mathcal{R}$ on the set of states S is a probabilistic bisimulation if $s_1 \, \mathcal{R} \, s_2$ implies $\sum_{s \in E} \pi(s_1, s) = \sum_{s \in E} \pi(s_2, s)$ for all $\mathcal{R}$-equivalence classes E. States s_1 and s_2 are probabilistic bisimilar, denoted $s_1 \sim s_2$, if $s_1 \, \mathcal{R} \, s_2$ for some probabilistic bisimulation $\mathcal{R}$.*

Note that probabilistic bisimilar states s_1 and s_2 have the same probability of transitioning to an equivalence class E of probabilistic bisimilar states.

Example 2. Consider the probabilistic transition system of Example 1. The smallest equivalence relation containing (s_3, s_5) is a probabilistic bisimulation. Hence, the states s_3 and s_5 are probabilistic bisimilar.

The behavioural pseudometrics that we study in this paper yield pseudometric spaces on the state space of probabilistic transition systems.

Definition 3. *A 1-bounded pseudometric space is a pair* (X, d_X) *consisting of a set* X *and a distance function* $d_X : X \times X \to [0, 1]$ *satisfying*

1. *for all* $x \in X$, $d_X(x, x) = 0$,
2. *for all* x, $y \in X$, $d_X(x, y) = d_X(y, x)$, *and*
3. *for all* x, y, $z \in X$, $d_X(x, z) \leq d_X(x, y) + d_X(y, z)$.

Instead of (X, d_X) *we often write* X *and we denote the distance function of a metric space* X *by* d_X.

Example 3. Let X be a set. The discrete metric $d_X : X \times X \to [0, 1]$ is defined by

$$d_X(x_1, x_2) = \begin{cases} 0 \text{ if } x_1 = x_2 \\ 1 \text{ otherwise.} \end{cases}$$

A (1-bounded) pseudometric space differs from a (1-bounded) metric space in that different points may have distance zero in the former and not in the latter. Since different states of a system may behave the same, such states will have distance zero in our behavioural pseudometrics.

In the characterization of a behavioural pseudometric in Section 4 nonexpansive functions play a key role.

Definition 4. *Let* X *be a 1-bounded pseudometric space. A function* $f : X \to [0, 1]$ *is nonexpansive if for all* x_1, $x_2 \in X$,

$$|f(x_1) - f(x_2)| \leq d_X(x_1, x_2).$$

The set of nonexpansive functions from X *to* $[0, 1]$ *is denoted by* $X \nrightarrow [0, 1]$.

Example 4. If the set X is endowed with the discrete metric, then every function from X to $[0, 1]$ is nonexpansive.

3 Behavioural pseudometrics

Desharnais, Gupta, Jagadeesan and Panangaden [16] introduced a family of behavioural pseudometrics for probabilistic transitions systems. Below, we will briefly review the key ingredients of their definition.

To define their behavioural pseudometrics, Desharnais et al. defined a real-valued semantics of a variant of Larsen and Skou's logic [23]. We describe this variant, adapted to the case of unlabelled transition systems, in Definition 5.

Definition 5. *The logic* $\mathcal{L}$ *is defined by*

$$\varphi ::= \text{true} \mid \Diamond\varphi \mid \varphi \wedge \varphi \mid \neg\varphi \mid \varphi \ominus q$$

where q *is a rational in* $[0, 1]$.

The main difference between the above logic and the one of Larsen and Skou is that we have $\Diamond\varphi$ and $\varphi \ominus q$ whereas they combine the operators $\Diamond$ and $\ominus q$ into one. Since they consider labelled transitions, they use the notation $\langle a \rangle_q$ for this combined operator.

Desharnais et al. provided a family of real-valued interpretations of the logic. That is, given a probabilistic transition system and a discount factor δ, the interpretation gives a quantitative measure of the validity of a formula φ of the logic in a state s of the system. The interpretation $[\![\varphi]\!]_\delta(s)$ is a real number in the interval $[0,1]$. It measures the validity of the formula φ in the state s. This real number can roughly be thought of as the probability that φ is true in s.

Definition 6. *Given a probabilistic transition system $\langle S, \pi \rangle$ and a discount factor $\delta \in (0,1]$, for each $\varphi \in \mathcal{L}$, the function $[\![\varphi]\!]_\delta : S \to [0,1]$ is defined by*

$$
\begin{aligned}
[\![\mathrm{true}]\!]_\delta(s) &= 1\\
[\![\Diamond\varphi]\!]_\delta(s) &= \delta \textstyle\sum_{s' \in S} \pi(s,s')[\![\varphi]\!]_\delta(s')\\
[\![\varphi \wedge \psi]\!]_\delta(s) &= \min\{[\![\varphi]\!]_\delta(s), [\![\psi]\!]_\delta(s)\}\\
[\![\neg\varphi]\!]_\delta(s) &= 1 - [\![\varphi]\!]_\delta(s)\\
[\![\varphi \ominus q]\!]_\delta(s) &= \max\{[\![\varphi]\!]_\delta(s) - q, 0\}
\end{aligned}
$$

Example 5. Consider the probabilistic transition system of Example 1. For this system, $[\![\Diamond\mathrm{true}]\!]_\delta(s_3) = \delta$ and $[\![\Diamond\mathrm{true}]\!]_\delta(s_4) = 0$.

Given a discount factor $\delta \in (0,1]$, the behavioural pseudometric d_δ assigns a distance, a real number in the interval $[0,1]$, to every pair of states of a probabilistic transition system. The distance is defined in terms of the logical formulae and their interpretation. Roughly speaking, the distance is captured by the logical formula that distinguishes the states the most.

Definition 7. *Given a probabilistic transition system $\langle S, \pi \rangle$ and a discount factor $\delta \in (0,1]$, the distance function $d_\delta : S \times S \to [0,1]$ is defined by*

$$d_\delta(s_1, s_2) = \sup_{\varphi \in \mathcal{L}} [\![\varphi]\!]_\delta(s_1) - [\![\varphi]\!]_\delta(s_2).$$

Example 6. Consider the probabilistic transition system of Example 1. For example, the states s_3 and s_4 are δ apart. This distance is witnessed by the formula $\Diamond$true.

The distances are collected in the following table. Since a distance function is symmetric and the distance from a state to itself is zero, we do not give all the entries.

	s_1	s_2	s_3	s_4
s_2	$\frac{25\delta^2 - 2\delta^4}{125 - 25\delta - 35\delta^2 + 7\delta^3}$			
s_3	$\frac{2\delta^3}{25 - 7\delta^2}$	$\frac{5\delta^2}{25 - 7\delta^2}$		
s_4	δ	δ	δ	
s_5	$\frac{2\delta^3}{25 - 7\delta^2}$	$\frac{5\delta^2}{25 - 7\delta^2}$	0	δ

Proposition 1. *d_δ is a 1-bounded pseudometric space.*

Each behavioural pseudometric d_δ is a quantitative analogue of probabilistic bisimilarity. This behavioural equivalence is exactly captured by those states that have distance zero.

Proposition 2 ([14]). *Given a probabilistic transition system $\langle S, \pi \rangle$ and a discount factor $\delta \in (0, 1]$,*

$$d_\delta(s_1, s_2) = 0 \text{ if and only if } s_1 \sim s_2$$

for all s_1, $s_2 \in S$.

In [14], Desharnais et al. present a decision procedure for the behavioural pseudometric d_δ when δ is smaller than one. Let us briefly sketch their algorithm. They define the depth of a logical formula as follows.

$$\begin{aligned} \text{depth}(\text{true}) &= 0 \\ \text{depth}(\Diamond\varphi) &= \text{depth}(\varphi) + 1 \\ \text{depth}(\varphi \wedge \psi) &= \text{depth}(\varphi) \max \text{depth}(\psi) \\ \text{depth}(\neg\varphi) &= \text{depth}(\varphi) \\ \text{depth}(\varphi \ominus q) &= \text{depth}(\varphi) \end{aligned}$$

One can easily verify that

$$[\![\varphi]\!]_\delta(s_1) - [\![\varphi]\!]_\delta(s_2) \leq \delta^{\text{depth}(\varphi)}$$

for each $\varphi \in \mathcal{L}$. This suggests that one can compute d_δ to any desired degree of accuracy by restricting attention to formulae ϕ of a fixed modal depth. Clearly, there exist infinitely many formulae of each fixed modal depth. Nevertheless Desharnais et al. show how to construct a finite subset $\mathcal{F}_n$ of the logical formulae of at most depth n such that

$$d_\delta(s_1, s_2) - \sup_{\varphi \in \mathcal{F}_n} [\![\varphi]\!]_\delta(s_1) - [\![\varphi]\!]_\delta(s_2) \leq \delta^n.$$

In this way, $d_\delta(s_1, s_2)$ can be approximated upto arbitrary accuracy *provided* $\delta < 1$.

4 A fixed point characterization and its dual

For the rest of this paper, we focus on the behavioural pseudometric that does not discount the future. That is, we concentrate on the pseudometric d_1. Below, we present an alternative characterization of this pseudometric. In particular, we characterize d_1 as the greatest (post-)fixed point of a function from a complete lattice to itself. This characterization can be viewed as a quantitative analogue of the greatest fixed point characterization of bisimilarity [25].

We also dualize the definition of Δ exploiting the Kantorovich-Rubinstein duality theorem [22]. As we will see in Section 5, this dual characterization will allow us to define Δ as the solution to a minimization problem rather than a maximization problem, as above. In turn this will allow us to capture the fact that a pseudometric is a post-fixed point of Δ in the existential fragment of the first order theory over real closed fields.

For the rest of this paper, we fix a probabilistic transition system $\langle S, \pi \rangle$. We endow the set of pseudometrics on S with the following order.

Definition 8. *The relation $\sqsubseteq$ on 1-bounded pseudometrics on S is defined by*

$$d_1 \sqsubseteq d_2 \text{ if } d_1(s_1, s_2) \geq d_2(s_1, s_2) \text{ for all } s_1,\ s_2 \in S.$$

Note the reverse direction of $\sqsubseteq$ and $\geq$ in the above definition. We decided to make this reversal so that d_1 is a greatest fixed point, in analogy with the characterization of bisimilarity, rather than a least fixed point. This choice has no impact on any results in this paper.

Proposition 3. *The set of 1-bounded pseudometrics on S endowed with the order $\sqsubseteq$ forms a complete lattice.*

Next, we introduce a function from this complete lattice to itself of which the behavioural pseudometric d_1 is the greatest fixed point.

Definition 9. *Let d be a 1-bounded pseudometric on S. The distance function $\Delta(d) : S \times S \to [0, 1]$ is defined by*

$$\Delta(d)(s_1, s_2) = \max \left\{ \sum_{s \in S} f(s)(\pi(s_1, s) - \pi(s_2, s)) \;\middle|\; f \in (S, d) \xrightarrow{1} [0, 1] \right\}$$

if $s_1 \to$ and $s_2 \to$, and

$$\Delta(d)(s_1, s_2) = \begin{cases} 0 \text{ if } s_1 \not\to \text{ and } s_2 \not\to \\ 1 \text{ otherwise.} \end{cases}$$

Proposition 4. *$\Delta(d)$ is a 1-bounded pseudometric on S.*

To conclude that Δ has a greatest fixed point, it suffices to show that Δ is order-preserving.

Proposition 5. *Δ is order-preserving.*

According to Tarski's fixed point theorem [28, Theorem 1], the fixed points of an order-preserving function on a complete lattice form a complete lattice and, hence, the function has a greatest fixed point. We denote the greatest fixed point of Δ by $\mathrm{gfp}(\Delta)$. This greatest fixed point of Δ is also the greatest post-fixed point of Δ (see, for example, [11, Theorem 4.11][3]).

Theorem 1. $d_1 = \mathrm{gfp}(\Delta)$.

[3] d is a *post-fixed point* of Δ if $d \sqsubseteq \Delta(d)$. In [11, page 94], such a d is called a prefixpoint.

The greatest fixed point of an order-preserving function on a complete lattice can be obtained by iteration (see, for example, [11, Exercise 4.13]).

Definition 10. *For each ordinal α, the 1-bounded pseudometric d^α on S is defined by*

$$\begin{aligned} d^0 &= \top \\ d^{\alpha+1} &= \Delta(d^\alpha) \\ d^\beta &= \bigsqcap_{\alpha \in \beta} d^\alpha \text{ if } \beta \text{ is a limit ordinal} \end{aligned}$$

For Δ, we need to iterate (at most) ω times before reaching the greatest fixed point.

Proposition 6. $\mathrm{gfp}(\Delta) = d^\omega$.

Let us recall (a minor variation of) the Kantorovich-Rubinstein duality theorem. Let X be a 1-bounded compact pseudometric space. Let μ_1 and μ_2 be Borel probability measures on X. We denote the set of Borel probability measures on the product space with marginals μ_1 and μ_2, that is, the Borel probability measures μ on X^2 such that for all Borel subsets B of X,

$$\mu(B \times X) = \mu_1(B) \text{ and } \mu(X \times B) = \mu_2(B),$$

by $\mu_1 \otimes \mu_2$. The Kantorovich-Rubinstein duality theorem tells us

$$\max \left\{ \int_X f d\mu_1 - \int_X f d\mu_2 \;\middle|\; f \in X \nrightarrow [0,1] \right\} = \min \left\{ \int_{X^2} d_X d\mu \;\middle|\; \mu \in \mu_1 \otimes \mu_2 \right\}.$$

The following proposition, which is a consequence of the Kantorovich-Rubinstein duality theorem, defines $\Delta(d)$ as a minimum as opposed to the maximum in Definition 9.

Proposition 7. *Let d be a 1-bounded pseudometric on S. Let s_1, $s_2 \in S$ such that $s_1 \rightarrow$ and $s_2 \rightarrow$. Then*

$$\Delta(d)(s_1, s_2) = \min \left\{ \sum_{(s_i, s_j) \in S^2} d(s_i, s_j) \mu(s_i, s_j) \;\middle|\; \mu \in \pi(s_1, \cdot) \otimes \pi(s_2, \cdot) \right\}$$

where $\mu \in \pi(s_1, \cdot) \otimes \pi(s_2, \cdot)$ if

$$\forall s_j \in S \sum_{s_i \in S} \mu(s_i, s_j) = \pi(s_1, s_j) \wedge \forall s_i \in S \sum_{s_j \in S} \mu(s_i, s_j) = \pi(s_2, s_i).$$

5 The algorithm

Before we present our algorithm, we first show that the fact that a pseudometric is a post-fixed point of Δ can be expressed in (the existential fragment of) the first order theory over real closed fields. This will allow us to exploit Tarksi's decision procedure to approximate the behavioural pseudometric.

For the rest of this paper, we assume that the probabilistic transition system $\langle S, \pi \rangle$ has N states $s_1, s_2, \ldots, s_N$. Instead of $\pi(s_i, s_j)$ we will write π_{ij}. We represent a 1-bounded pseudometric on the set S of states of the probabilistic transition system, as (the values of) a collection of real valued variables d_{ij}.

The fact that d is a 1-bounded pseudometric can now be captured as follows.

Definition 11. *The predicate* pseudo(d) *is defined by*

$$\begin{aligned}\mathrm{pseudo}(d) \equiv & \bigwedge_{1 \leq i,j \leq N} d_{ij} \geq 0 \wedge d_{ij} \leq 1 \wedge \\ & \bigwedge_{1 \leq i \leq N} d_{ii} = 0 \wedge \bigwedge_{1 \leq i,j \leq N} d_{ij} = d_{ji} \wedge \bigwedge_{1 \leq h,i,j \leq N} d_{hj} \leq d_{hi} + d_{ij}\end{aligned}$$

Furthermore, the fact that d is a post-fixed point of Δ can be captured as follows.

Definition 12. *The predicate* post-fixed(d) *is defined by*

$$\begin{aligned}&\text{post-fixed}(d) \\ &\equiv \bigwedge_{1 \leq i_0,j_0 \leq N} \text{post-fixed}_1(d, i_0, j_0) \vee \text{post-fixed}_2(d, i_0, j_0) \vee \text{post-fixed}_3(d, i_0, j_0)\end{aligned}$$

where

$$\begin{aligned}\text{post-fixed}_1(d, i_0, j_0) \equiv & \sum_{1 \leq i \leq N} \pi_{i_0 i} > 0 \wedge \sum_{1 \leq j \leq N} \pi_{j_0 j} > 0 \wedge \\ & \exists (\mu_{ij})_{1 \leq i,j \leq N} \bigwedge_{1 \leq i,j \leq N} \mu_{ij} \geq 0 \wedge \mu_{ij} \leq 1 \\ & \qquad \bigwedge_{1 \leq j \leq N} \sum_{1 \leq i \leq N} \mu_{ij} = \pi_{i_0 j} \wedge \\ & \qquad \bigwedge_{1 \leq i \leq N} \sum_{1 \leq j \leq N} \mu_{ij} = \pi_{j_0 i} \wedge \\ & \qquad \sum_{1 \leq i,j \leq N} d_{ij} \mu_{ij} \leq d_{i_0 j_0} \\ \text{post-fixed}_2(d, i_0, j_0) \equiv & \sum_{1 \leq i \leq N} \pi_{i_0 i} = 0 \wedge \sum_{1 \leq j \leq N} \pi_{j_0 j} = 0 \wedge 0 \leq d_{i_0 j_0} \\ \text{post-fixed}_3(d, i_0, j_0) \equiv & \left(\left(\sum_{1 \leq i \leq N} \pi_{i_0 i} > 0 \wedge \sum_{1 \leq j \leq N} \pi_{j_0 j} = 0 \right) \vee \right. \\ & \left. \left(\sum_{1 \leq i \leq N} \pi_{i_0 i} = 0 \wedge \sum_{1 \leq j \leq N} \pi_{j_0 j} > 0 \right) \right) \wedge \\ & 1 \leq d_{i_0 j_0}\end{aligned}$$

Now we are ready to present our algorithm. Consider the states s_{i_0} and s_{j_0}. We restrict our attention to the case that $s_{i_0} \rightarrow$ and $s_{j_0} \rightarrow$. In the other cases the computation of the distance is trivial.

In our algorithm, we use the algorithm `tarski` that takes as input a sentence of the first order theory of real closed fields and decides the truth or falsity of the given sentence. The fact that there exists such an algorithm was first proved by Tarski [27].

Let ϵ be the desired accuracy. That is, we want to find an interval $[\ell_0, u_0] \subseteq [0, 1]$ such that $u_0 - \ell_0 \leq \epsilon$ and $d_1(s_{i_0}, s_{j_0}) \in [\ell_0, u_0]$. The algorithm `approximate` takes as input an interval $[\ell, u] \subseteq [0, 1]$ such that $d_1(s_{i_0}, s_{j_0}) \in [\ell, u]$ and returns the desired result. As a consequence, `approximate`(0, 1) returns an approximation of $d_1(s_{i_0}, s_{j_0})$ with accuracy ϵ.

```
approximate(ℓ, u):
   if u − ℓ ≤ ε
      return [ℓ, u]
   else
      m = (ℓ+u)/2
      if tarski(∃d pseudo(d) ∧ post-fixed(d) ∧ d_{i0 j0} ≤ m)
         return approximate(ℓ, m)
      else
         return approximate(m, u)
```

Note that the argument of `tarski` is a sentence that is part of the existential fragment of the first order theory over real closed fields. For this fragment there are more efficient decision procedures than for the general theory (see, for example, [1]).

Let us sketch a correctness proof of our algorithm. Assume that $d_1(s_{i_0}, s_{j_0}) \in [\ell, u]$. We distinguish the following three cases.

– If $u - \ell \leq \epsilon$, then the algorithm obviously returns the desired result.
– Assume that $u - \ell > \epsilon$ and suppose that `tarski` returns true. Then there exists a 1-bounded pseudometric d that is a post-fixed point of Δ and $d(s_{i_0}, s_{j_0}) \leq m$. Since d_1 is the greatest post-fixed point of Δ, we have that $d \sqsubseteq d_1$. Hence, $d_1(s_{i_0}, s_{j_0}) \leq d(s_{i_0}, s_{j_0}) \leq m$. Therefore, $d_1(s_{i_0}, s_{j_0}) \in [\ell, m]$.
– Assume that $u - \ell > \epsilon$ and suppose that `tarski` returns false. Then $d(s_{i_0}, s_{j_0}) > m$ for every 1-bounded pseudometric d that is a post-fixed point of Δ. Since d_1 is a post-fixed point of Δ, we have that $d_1(s_{i_0}, s_{j_0}) > m$. Therefore, $d_1(s_{i_0}, s_{j_0}) \in [m, u]$.

Obviously, the algorithm terminates.

6 An implementation in Mathematica

A decision procedure for the first order theory of real closed fields based on quantifier elimination was first given by Tarski [27]. A number of algorithms have been developed thereafter for the theory (see, for example, [1, 10, 21]). Collin's algorithm is implemented in the tool Mathematica and can be used for solving our formulae. However, it works for very small examples and therefore it is essential to simplify the formula and reduce its size to make it solvable. To simplify the formula, we first compute some of the distances using the following results.

Proposition 8.

- *If* $s_1 \not\to$ *and* $s_2 \not\to$ *then* $d_1(s_1, s_2) = 0$.
- *If* $s_1 \not\to$ *and* $s_2 \to$, *or* $s_1 \to$ *and* $s_2 \not\to$ *then* $d_1(s_1, s_2) = 1$.

Example 7. Consider the probabilistic transition system of Example 1. State s_4 has distance one to all other states.

Next, we present a simple characterization of the distance between a state that never terminates (that is, the probability of reaching a state with no outgoing transitions is zero) and another state.

Given a state s and $n \in \omega + 1$, $\tau_n(s)$ is the probability of terminating in less than n transitions when started in s.

Definition 13. *For each* $n \in \omega + 1$, *the function* $\tau_n : S \to [0, 1]$ *is defined by*

$$\begin{aligned} \tau_0(s) &= 0 \\ \tau_{n+1}(s) &= \begin{cases} 1 & \text{if } s \not\to \\ \sum_{s' \in S} \pi(s, s')\tau_n(s') & \text{otherwise} \end{cases} \\ \tau_\omega(s) &= \sup_{n \in \omega} \tau_n(s) \end{aligned}$$

Example 8. Consider the probabilistic transition system of Example 1. Then we have that $\tau_\omega(s_1) = \frac{1}{9}$, $\tau_\omega(s_2) = \frac{5}{18}$, $\tau_\omega(s_3) = 0$, $\tau_\omega(s_4) = 1$ and $\tau_\omega(s_5) = 0$.

Obviously, for a state s without outgoing transitions, we have that $\tau_\omega(s) = 1$. For a state s that cannot reach any state without outgoing transitions, we have that $\tau_\omega(s) = 0$. For the remaining states, we can compute the probability of termination using standard techniques as described in, for example, [20, Section 11.2].

Proposition 9. *If* $\tau_\omega(s_2) = 0$ *then* $d_1(s_1, s_2) = \tau_\omega(s_1)$.

Example 9. Consider the probabilistic transition system of Example 1. From Proposition 9 we can conclude that $d_1(s_1, s_3) = \frac{1}{9}$, $d_1(s_2, s_3) = \frac{5}{18}$, $d_1(s_4, s_3) = 1$ and $d_1(s_5, s_3) = 0$.

Given a probabilistic bisimulation $\mathcal{R}$, we can quotient the probabilistic transition system $\langle S, \pi \rangle$ as follows.

Definition 14. *Let* $\mathcal{R}$ *be a probabilistic bisimulation. The probabilistic transition system* $\langle S_\mathcal{R}, \pi_\mathcal{R} \rangle$ *consists of*

- *the set* $S_\mathcal{R} = \{ [s] \mid s \in S \}$ *of* $\mathcal{R}$*-equivalence classes and*

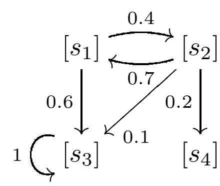
– *the function* $\pi_{\mathcal{R}} : S_{\mathcal{R}} \times S_{\mathcal{R}} \to [0,1]$ *defined by*

$$\pi_{\mathcal{R}}([s],[s']) = \sum_{s''\mathcal{R}s'} \pi(s,s').$$

Note that the function $\pi_{\mathcal{R}}$ is well-defined since $\mathcal{R}$ is a probabilistic bisimulation. We will apply the above quotient construction for the following bisimulation.

Proposition 10. *The smallest equivalence relation containing* $\{ \langle s_1, s_2 \rangle \mid s_1 \not\to \wedge s_2 \not\to \}$ *and* $\{ \langle s_1, s_2 \rangle \mid \tau_\omega(s_1) = 0 \wedge \tau_\omega(s_2) = 0 \}$ *is a probabilistic bisimulation.*

Example 10. Consider the probabilistic transition system of Example 1. According to Proposition 10, the smallest equivalence relation containing $\{\langle s_3, s_5\rangle\}$ is a bisimulation. The resulting quotient can be depicted as



By quotienting, the number of states that need to be considered and, hence, the number of variables in the formula may be reduced. However, we still have to check that the quotiented system gives rise to the same distances. Next we relate the behavioural pseudometric d_1 of the original system $\langle S, \pi\rangle$ with the behavioural pseudometric $d_{\mathcal{R}}$ of the quotiented system $\langle S_{\mathcal{R}}, \pi_{\mathcal{R}}\rangle$.

Proposition 11. *For all* s_1, $s_2 \in S$, $d_{\mathcal{R}}([s_1],[s_2]) = d_1(s_1, s_2)$.

To simplify the formula even further, we exploit the following three observations.

– Since d is a pseudometric, $d_1(s_i, s_i) = 0$ and $d_1(s_i, s_j) = d_1(s_j, s_i)$. Therefore, in pseudo($d$) $\wedge$ post-fixed(d) we can replace all d_{ii}'s with zero and all d_{ij}'s where $i > j$ with d_{ji}'s. As a consequence, we only need to consider d_{ij}'s with $i < j$. This reduces the number of variables in the formula considerably.
– Let C be the set of pairs of states for which the distances have already been computed. Then

$$\exists d \, \text{pseudo}(d) \wedge \text{post-fixed}(d) \wedge d_{i_0 j_0} \leq m$$

is equivalent to

$$\exists d \, \text{pseudo}(d) \wedge \text{post-fixed}(d) \wedge d_{i_0 j_0} \leq m \wedge \bigwedge_{(i,j)\in C} d_{ij} = d_1(s_i, s_j)$$

since d_1 is the greatest post-fixed point. As a consequence, we can replace all d_{ij}'s where $(i,j) \in C$ with their already computed distances $d_1(s_i, s_j)$. Again, the number of variables may be reduced.

– If $\pi_{i_0 j} = 0$, we can infer that $\mu_{ij} = 0$ for all $1 \leq i \leq N$. As a consequence, we can replace the occurrences of all those μ_{ij}'s with 0. Symmetrically, if $\pi_{j_0 i} = 0$ we can simplify the formula similarly. Also this simplification may reduce the number of variables.

We have implemented these simplifications in the form of a Java program that takes as input the probability matrix π and that produces as output the simplified formula in a format that can be fed to Mathematica.

Example 11. Consider the probabilistic transition system of Example 1. The simplified formula for this system is given below.

```
1  Reduce[
2    Exists[d12,
3      (0 <= d12 <= 1) &&
4      (0.11112 <= d12 + 0.27778) &&
5      (d12 <= 0.38889) &&
6      Exists[{u12,u13,u32,u42,u43,u33},
7        (0 <= u12 <= 1) && (0 <= u13 <= 1) && (0 <= u32 <= 1) &&
8        (0 <= u42 <= 1) && (0 <= u43 <= 1) &&
9        (u12 + u32 + u42 == 0.4) &&
10       (u13 + u43 + u33 == 0.6) &&
11       (u12 + u13 == 0.7) &&
12       (u32 + u33 == 0.1) &&
13       (u42 + u43 == 0.2) &&
14       (d12 * u12 + 0.11112 * u13 + 0.27778 * u32 + u42 + u43 <= d12)] &&
15     Exists[{u21,u23,u24,u31,u33, u34},
16       (0 <= u21 <= 1) && (0 <= u23 <= 1) && (0 <= u24 <= 1) &&
17       (0 <= u31 <= 1) && (0 <= u34 <= 1) &&
18       (u21 + u31 == 0.7) &&
19       (u23 + u33 == 0.1) &&
20       (u24 + u34 == 0.2) &&
21       (u21 + u23 + u24 == 0.4) &&
22       (u31 + u33 + u34 == 0.6) &&
23       (d12 * u21 + 0.27778 * u23 + u24 + 0.11112 * u31 + u34 <= d12)] &&
24     (0 <= d12 <= 0.5)]]
```

Line 3–5 correspond to pseudo(d), line 6–14 correspond to post-fixed$_1(d, 1, 2)$ and line 15–23 correspond to post-fixed$_1(d, 2, 1)$. The formula was reduced to true by Mathematica in 8.2 seconds on a 3GHz machine with 1GB RAM.

We also attempted to solve this example with a solver called QEPCAD B [8] but the performance of Mathematica on this example was better.

7 Conclusion

This paper combines a number of ingredients, known already for a long time, including the Kantorovich-Rubinstein duality theorem of the fifties, Tarski's fixed point theorem of the fourties and Tarski's decision procedure for the first order

theory of real closed fields of the thirties. We show that the behavioural pseudometric d_1, which does not discounts the future, can be approximated upto an arbitrary accuracy. While the combination of the above results into a decision procedure for the pseudometric is not technically difficult, we do solve a problem that has been open since 1999. Most of the results in Section 3 and 4 are (variations on) known results. As far as we know, Proposition 6 and the results in Section 5 and 6 are new. The techniques exploited in this paper can also be used to approximate other behavioural pseudometrics that do not discount the future like, for example, the ones presented in [2, 13]. Since the satisfiability problem for the existential fragment of the first order theory of the real closed fields is PSPACE-complete [1], it is not surprising that our algorithm can only handle small examples as we have show in Section 6. As a consequence, the quest for practical algorithms to approximate d_1 is still open. Since the closure ordinal of Δ is ω, as proved in Proposition 6, an iterative algorithm might be feasible.
Acknowledgements The authors would like to thank Christel Baier for providing some pointers to the literature.

References

1. S. Basu, R. Pollack, and M.-F. Roy. On the combinatorial and algebraic complexity of quantifier elimination. *Journal of the ACM*, 43(6):1002–1045, November 1996.
2. F. van Breugel. A behavioural pseudometric for metric labelled transition systems. In M. Abadi and L. de Alfaro, editors, *Proceedings of the 16th International Conference on Concurrency Theory*, volume 3653 of *Lecture Notes in Computer Science*, pages 141–155, San Francisco, August 2005. Springer-Verlag.
3. F. van Breugel, C. Hermida, M. Makkai, and J. Worrell. Recursively defined metric spaces without contraction. *Theoretical Computer Science*. To appear.
4. F. van Breugel and J. Worrell. An algorithm for quantitative verification of probabilistic transition systems. In K.G. Larsen and M. Nielsen, editors, *Proceedings of 12th International Conference on Concurrency Theory*, volume 2154 of *Lecture Notes in Computer Science*, pages 336–350, Aalborg, August 2001. Springer-Verlag.
5. F. van Breugel and J. Worrell. A behavioural pseudometric for probabilistic transition systems. *Theoretical Computer Science*, 331(1):115–142, February 2005.
6. F. van Breugel and J. Worrell. An algorithm for approximating behavioural pseudometrics for probabilistic transition systems. *Theoretical Computer Science*, 360(1/3):373–385, August 2006.
7. M. Broucke. Regularity of solutions and homotopic equivalence for hybrid systems. In *Proceedings of the 37th IEEE Conference on Decision and Control*, volume 4, pages 4283–4288, Tampa, December 1998. IEEE.
8. C.W. Brown. An overview of QEPCAD B: a tool for real quantifier elimination and formula simplification. *Journal of Japan Society for Symbolic and Algebraic Computation*, 10(1):13–22, 2003.
9. P. Caspi and A. Benveniste. Toward an approximation theory for computerised control. In A.L. Sangiovanni-Vincentelli and J. Sifakis, editors, *Proceedings of the 2nd International Conference on Embedded Software*, volume 2491 of *Lecture Notes in Computer Science*, pages 294–304, Grenoble, October 2002. Springer-Verlag.
10. G.E. Collins. Quantifier elimination for real closed fields by cylindrical algebraic decomposition. In H. Barkhage, editor, *Proceedings of the 2nd GI Conference on*

Automata Theory and Formal Languages, volume 33 of *Lecture Notes in Computer Science*, pages 134–183, Kaiserslautern, May 1975. Springer-Verlag.

11. B.A. Davey and H.A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, Cambridge, 1990.
12. L. de Alfaro, T.A. Henzinger, and R. Majumdar. Discounting the future in systems theory. In J.C.M. Baeten, J.K. Lenstra, J. Parrow, and G.J. Woeginger, editors, *Proceedings of 30th International Colloquium on Automata, Languages and Programming*, volume 2719 of *Lecture Notes in Computer Science*, pages 1022–1037, Eindhoven, June/July 2003. Springer-Verlag.
13. Y. Deng, T. Chothia, C. Palamidessi, and J. Pang. Metrics for action-labelled quantitative transition systems. In A. Cerone and H. Wiklicky, editors, *Proceedings of 3rd Workshop on Quantitative Aspects of Programming Languages*, volume 153(2) of *Electronic Notes in Theoretical Computer Science*, pages 79–96, Edinburgh, April 2005. Elsevier.
14. J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for labeled Markov systems. In J.C.M. Baeten and S. Mauw, editors, *Proceedings of 10th International Conference on Concurrency Theory*, volume 1664 of *Lecture Notes in Computer Science*, pages 258–273, Eindhoven, August 1999. Springer-Verlag.
15. J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. The metric analogue of weak bisimulation for probabilistic processes. In *Proceedings of 17th Annual IEEE Symposium on Logic in Computer Science*, pages 413–422, Copenhagen, July 2002. IEEE.
16. J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for labelled Markov processes. *Theoretical Computer Science*, 318(3):323–354, June 2004.
17. A. Di Pierro, C. Hankin, and H. Wiklicky. Quantitative relations and approximate process equivalences. In R. Amadio and D. Lugiez, editors, *Proceedings of the 14th International Conference on Concurrency Theory*, volume 2761 of *Lecture Notes in Computer Science*, pages 508–522, Marseille, September 2003. Springer-Verlag.
18. A. Giacalone, C.-C. Jou, and S.A. Smolka. Algebraic reasoning for probabilistic concurrent systems. In *Proceedings of the IFIP WG 2.2/2.3 Working Conference on Programming Concepts and Methods*, pages 443–458, Sea of Gallilee, April 1990. North-Holland.
19. A. Girard and G.J. Pappas. Approximate bisimulations for nonlinear dynamical systems. In *Proceedings of the 44th IEEE Conference on Decision and Control and the European Control Conference*, pages 684–689, Seville, December 2005. IEEE.
20. C.M. Grinstead and J.L. Snell. *Introduction to Probability*. AMS, Providence, second revised edition, 1997.
21. L. Hörmander. *The Analysis of Linear Partial Differential Operators II: Differential Operators with Constant Coefficients*. Classics in Mathematics. Springer-Verlag, Berlin, 2005.
22. L.V. Kantorovich and G.Sh. Rubinstein. On the space of completely additive functions. *Vestnik Leningradskogo Universiteta*, 3(2):52–59, 1958. In Russian.
23. K.G. Larsen and A. Skou. Bisimulation through probabilistic testing. *Information and Computation*, 94(1):1–28, September 1991.
24. J.C Mitchell, A. Ramanathan, A. Scedrov, and V. Teague. A probabilistic polynomial-time process calculus for the analysis of cryptographic protocols. *Theoretical Computer Science*, 353(1/3):118–164, March 2006.
25. D. Park. Concurrency and automata on infinite sequences. In P. Deussen, editor, *Proceedings of 5th GI-Conference on Theoretical Computer Science*, volume 104 of *Lecture Notes in Computer Science*, pages 167–183, Karlsruhe, March 1981. Spring-Verlag.

26. G. Strang. *Introduction to Applied Mathematics*. Wellesley-Cambridge Press, Wellesley, 1986.
27. A. Tarski. *A decision method for elementary algebra and geometry*. University of California Press, Berkeley, 1951.
28. A. Tarski. A lattice-theoretic fixed point theorem and its applications. *Pacific Journal of Mathematics*, 5(2):285–309, June 1955.
29. M. Ying. Bisimulation indexes and their applications. *Theoretical Computer Science*, 275(1/2):1–68, March 2002.

A Proof of Proposition 6

In this appendix we prove that the closure ordinal of Δ is ω, that is, we show that $\Delta(d^\omega) = d^\omega$. As a consequence, d^ω is the greatest fixed point of Δ (see, for example, [11, Example 4.13]). As we will see below, the fact that d^ω is a fixed point of Δ follows from the facts that Δ is order-preserving (Proposition 5) and Lipschitz (Proposition 15).

States having distance zero defines an equivalence relation. That is, the relation $\equiv_d$ on states defined by

$$s_1 \equiv_d s_2 \text{ if } d(s_1, s_2) = 0$$

is an equivalence relation. We denote the equivalence class that contains the state s by $[s]_d$, that is,

$$[s]_d = \{\, s' \in S \mid d(s, s') = 0 \,\}.$$

From each equivalence class $[s]_d$ we pick a designated state which we denote by $\langle s \rangle_d$. Hence, $\langle s \rangle_d \in [s]_d$ and also $d(s, \langle s \rangle_d) = 0$.

Proposition 12. *For all s_1, $s_2 \in S$,*

$$d(\langle s_1 \rangle_d, \langle s_2 \rangle_d) = d(s_1, s_2).$$

Proof.

$$\begin{aligned}
& d(\langle s_1 \rangle_d, \langle s_2 \rangle_d) \\
& \leq d(\langle s_1 \rangle_d, s_1) + d(s_1, s_2) + d(s_2, \langle s_2 \rangle_d) \\
& = d(s_1, s_2) \\
& \leq d(s_1, \langle s_1 \rangle_d) + d(\langle s_1 \rangle_d, \langle s_2 \rangle_d) + d(\langle s_2 \rangle_d, s_2) \\
& = d(\langle s_1 \rangle_d, \langle s_2 \rangle_d).
\end{aligned}$$

□

Let $d_1 \sqsubseteq d_2$. The ratio $\rho(d_1, d_2)$ of d_1 and d_2 is defined by

$$\rho(d_1, d_2) = \min \left\{ \frac{d_2(s_1, s_2)}{d_1(s_1, s_2)} \;\middle|\; d_2(s_1, s_2) > 0 \right\}$$

Note that we never divide by zero since $d_1 \sqsubseteq d_2$ and, hence, $d_1(s_1, s_2) \geq d_2(s_1, s_2)$.

In this appendix, we will use the convention that the minimum of the empty set is one and the maximum of the empty set is zero.

Given pseudometrics d_1 and d_2 such that $d_1 \sqsubseteq d_2$ and given an $f \in (S, d_1) \twoheadrightarrow [0, 1]$, we next show that there exists a $g_f \in (S, d_2) \to [0, 1]$ that is nonexpansive.

Proposition 13. *Let $d_1 \sqsubseteq d_2$ and $f \in (S, d_1) \twoheadrightarrow [0, 1]$. Let $g_f : S \to [0, 1]$ be defined by*

$$g_f(s) = \rho(d_1, d_2) f(\langle s \rangle_{d_2}).$$

Then $g_f \in (S, d_2) \twoheadrightarrow [0, 1]$.

Proof. Let s_1, $s_2 \in S$. We have to show that

$$|g_f(s_1) - g_f(s_2)| \leq d_2(s_1, s_2).$$

We distinguish two cases. If $d_2(s_1, s_2) = 0$ then $\langle s_1 \rangle_{d_2} = \langle s_2 \rangle_{d_2}$ and, hence, $f(\langle s_1 \rangle_{d_2}) = f(\langle s_2 \rangle_{d_2})$. Therefore $g_f(s_1) = g_f(s_2)$ and, hence, the property is vacuously true. Let $d_2(s_1, s_2) > 0$. According to Proposition 12, $d_2(\langle s_1 \rangle_{d_2}, \langle s_2 \rangle_{d_2}) = > 0$. Also $d_1(s_1, s_2) > 0$ since $d_1 \sqsubseteq d_2$, and

$$\begin{aligned}
&|g_f(s_1) - g_f(s_2)| \\
&= |\rho(d_1, d_2) f(\langle s_1 \rangle_{d_2}) - \rho(d_1, d_2) f(\langle s_2 \rangle_{d_2})| \\
&= \rho(d_1, d_2) |f(\langle s_1 \rangle_{d_2}) - f(\langle s_2 \rangle_{d_2})| \\
&\leq \rho(d_1, d_2) d_1(\langle s_1 \rangle_{d_2}, \langle s_2 \rangle_{d_2}) \quad [f \in (S, d_1) \nrightarrow [0,1]] \\
&\leq \frac{d_2(\langle s_1 \rangle_{d_2}, \langle s_2 \rangle_{d_2})}{d_1(\langle s_1 \rangle_{d_2}, \langle s_2 \rangle_{d_2})} d_1(\langle s_1 \rangle_{d_2}, \langle s_2 \rangle_{d_2}) \\
&= d_2(\langle s_1 \rangle_{d_2}, \langle s_2 \rangle_{d_2}) \\
&= d_2(s_1, s_2) \quad [\text{Proposition 12}]
\end{aligned}$$

□

Next, we bound $f - g_f$ from above.

Proposition 14. *Let* $\mu = \min\{\, d_1(s_1, s_2) \mid d_1(s_1, s_2) > 0 \,\}$. *Then*

$$f(s) - g_f(s) \leq \frac{\mu + 1}{\mu} \max_{s'_1, s'_2 \in S} d_1(s'_1, s'_2) - d_2(s'_1, s'_2)$$

for all $s \in S$.

Proof. Let $s \in S$. Then

$$\begin{aligned}
&f(s) - g_f(s) \\
&= f(s) - \rho(d_1, d_2) f(\langle s \rangle_{d_2}) \\
&= (f(s) - f(\langle s \rangle_{d_2})) + (f(\langle s \rangle_{d_2}) - \rho(d_1, d_2) f(\langle s \rangle_{d_2})).
\end{aligned}$$

Furthermore,

$$\begin{aligned}
&f(s) - f(\langle s \rangle_{d_2}) \\
&\leq d_1(s, \langle s \rangle_{d_2}) \quad [f \in (S, d_1) \nrightarrow [0,1]] \\
&= d_1(s, \langle s \rangle_{d_2}) - d_2(s, \langle s \rangle_{d_2}) \quad [d_2(s, \langle s \rangle_{d_2}) = 0] \\
&\leq \max_{s'_1, s'_2 \in S} d_1(s'_1, s'_2) - d_2(s'_1, s'_2)
\end{aligned}$$

and

$$\begin{aligned}
&f(\langle s \rangle_{d_2}) - \rho(d_1, d_2) f(\langle s \rangle_{d_2}) \\
&\leq 1 - \rho(d_1, d_2)
\end{aligned}$$

$$
\begin{aligned}
&= 1 - \min\left\{ \frac{d_2(s_1,s_2)}{d_1(s_1,s_2)} \;\middle|\; d_2(s_1,s_2) > 0 \right\} \\
&= \max\left\{ \frac{d_1(s_1,s_2) - d_2(s_1,s_2)}{d_1(s_1,s_2)} \;\middle|\; d_2(s_1,s_2) > 0 \right\} \\
&\leq \frac{1}{\mu} \max\{\, d_1(s_1,s_2) - d_2(s_1,s_2) \mid d_2(s_1,s_2) > 0 \,\} \\
&\leq \frac{1}{\mu} \max_{s_1',s_2' \in S} d_1(s_1',s_2') - d_2(s_1',s_2').
\end{aligned}
$$

□

Now we can prove that Δ is Lipschitz, that is,

$$
\max_{s_1,s_2 \in S} \Delta(d_1)(s_1,s_2) - \Delta(d_2)(s_1,s_2) \leq \lambda \max_{s_1',s_2' \in S} d_1(s_1',s_2') - d_2(s_1',s_2').
$$

for some constant λ.

Proposition 15. *For all s_1, $s_2 \in S$,*

$$
\Delta(d_1)(s_1,s_2) - \Delta(d_2)(s_1,s_2) \leq |S| \frac{\mu+1}{\mu} \max_{s_1',s_2' \in S} d_1(s_1',s_2') - d_2(s_1',s_2').
$$

Proof. Let s_1, $s_2 \in S$. Then

$$
\begin{aligned}
&\Delta(d_1)(s_1,s_2) - \Delta(d_2)(s_1,s_2) \\
&= \max\left\{ \sum_{s \in S} f(s)(\pi(s_1,s) - \pi(s_2,s)) \;\middle|\; f \in (S,d_1) \nrightarrow [0,1] \right\} - \\
&\quad \max\left\{ \sum_{s \in S} g(s)(\pi(s_1,s) - \pi(s_2,s)) \;\middle|\; g \in (S,d_2) \nrightarrow [0,1] \right\} \\
&= \max\left\{ \min\left\{ \sum_{s \in S} f(s)(\pi(s_1,s) - \pi(s_2,s)) - \sum_{s \in S} g(s)(\pi(s_1,s) - \pi(s_2,s)) \right.\right. \\
&\qquad\qquad \left.\left. \;\middle|\; g \in (S,d_2) \nrightarrow [0,1] \right\} \;\middle|\; f \in (S,d_1) \nrightarrow [0,1] \right\} \\
&= \max\left\{ \min\left\{ \sum_{s \in S} (f(s) - g(s))(\pi(s_1,s) - \pi(s_2,s)) \right.\right. \\
&\qquad\qquad \left.\left. \;\middle|\; g \in (S,d_2) \nrightarrow [0,1] \right\} \;\middle|\; f \in (S,d_1) \nrightarrow [0,1] \right\} \\
&\leq \max\left\{ \sum_{s \in S} (f(s) - g_f(s))(\pi(s_1,s) - \pi(s_2,s)) \;\middle|\; f \in (S,d_1) \nrightarrow [0,1] \right\} \\
&\qquad \text{[Proposition 13]}
\end{aligned}
$$

$$\leq \max\left\{ \sum_{s \in S} f(s) - g_f(s) \;\middle|\; f \in (S, d_1) \nrightarrow [0,1] \right\}$$
$$\leq |S| \frac{\mu + 1}{\mu} \max_{s'_1, s'_2 \in S} d_1(s'_1, s'_2) - d_2(s'_1, s'_2) \quad \text{[Proposition 14]}$$

□

Proposition 16. $\Delta(d^\omega) = d^\omega$.

Proof. First, we show that $\Delta(d^\omega) \sqsubseteq d^\omega$. By definition, $d^\omega = \bigsqcap_{n \in \omega} d^n \sqsubseteq d^n$ for all $n \in \omega$. Since Δ is order-preserving, $\Delta(d^\omega) \sqsubseteq \Delta(d^n) = d^{n+1}$ for all $n \in \omega$. Obviously, $\Delta(d^\omega) \sqsubseteq d^0$. Therefore, $\Delta(d^\omega)$ is a lowerbound of $\{ d^n \mid n \in \omega \}$. Since d^ω is the greatest lowerbound by definition, $\Delta(d^\omega) \sqsubseteq d^\omega$. Consequently, $d^\omega(s_1, s_2) \leq \Delta(d^\omega)(s_1, s_2)$ for all s_1, $s_2 \in S$.

We have left to show that $\Delta(d^\omega)(s_1, s_2) \leq d^\omega(s_1, s_2)$ for all s_1, $s_2 \in S$. Let s_1, $s_2 \in S$. Let $\epsilon > 0$. It suffices to show that there exists an n such that $\Delta(d^\omega)(s_1, s_2) - d^{n+1}(s_1, s_2) \leq \epsilon$. Let $\mu = \min\{ d^\omega(s_1, s_2) \mid d^\omega(s_1, s_2) > 0 \}$. Since the set S is finite, for every $\delta > 0$ there exists an n such that for all s'_1, $s'_2 \in S$,

$$d^\omega(s'_1, s'_2) - d^n(s'_1, s'_2) \leq \delta.$$

Here we pick δ to be $\frac{\mu\epsilon}{(\mu+1)|S|}$. From Proposition 15 we can conclude that

$$\begin{aligned} &\Delta(d^\omega)(s_1, s_2) - d^{n+1}(s_1, s_2) \\ &= \Delta(d^\omega)(s_1, s_2) - \Delta(d^n)(s_1, s_2) \\ &\leq \epsilon. \end{aligned}$$

□

B Proofs

Proposition 1. *d_δ is a 1-bounded pseudometric space.*

Proof. First, observe that

$$[\![\varphi]\!]_\delta(s_1) - [\![\varphi]\!]_\delta(s_2) = [\![\neg\varphi]\!]_\delta(s_2) - [\![\neg\varphi]\!]_\delta(s_1).$$

As a consequence, we can replace $[\![\varphi]\!]_\delta(s_1) - [\![\varphi]\!]_\delta(s_2)$ in the definition of d_δ with $|[\![\varphi]\!]_\delta(s_1) - [\![\varphi]\!]_\delta(s_2)|$. Checking now that d_δ satisfies the three conditions of Definition 3 is straightforward. □

A similar result is presented in [16, Theorem 5.2].

Proposition 2. *Given a probabilistic transition system $\langle S, \pi \rangle$ and a discount factor $\delta \in (0, 1]$,*

$$d_\delta(s_1, s_2) = 0 \textit{ if and only if } s_1 \sim s_2$$

for all s_1, $s_2 \in S$.

Proof. We split the proof in two parts.

– Assume that $s_1 \sim s_2$. It suffices to show that $[\![\varphi]\!]_\delta(s_1) = [\![\varphi]\!]_\delta(s_2)$ for all φ. We can prove this by structural induction on φ. We focus here on the only nontrivial case: $\Diamond\varphi$. Let $\{\, E_i \mid i \in I \,\}$ be the $\sim$-equivalence classes. Assume that e_i is an element of E_i. By induction, the function $[\![\varphi]\!]_\delta$ restricted to E_i is constant. Hence,

$$\begin{aligned}
[\![\Diamond\varphi]\!]_\delta(s_1) &= \delta \sum_{s \in S} \pi(s_1, s)[\![\varphi]\!]_\delta(s) \\
&= \delta \sum_{i \in I} \sum_{s \in E_i} \pi(s_1, s)[\![\varphi]\!]_\delta(s) \\
&= \delta \sum_{i \in I} [\![\varphi]\!]_\delta(e_i) \sum_{s \in E_i} \pi(s_1, s) \\
&= \delta \sum_{i \in I} [\![\varphi]\!]_\delta(e_i) \sum_{s \in E_i} \pi(s_2, s) \quad [s_1 \sim s_2] \\
&= [\![\Diamond\varphi]\!]_\delta(s_2).
\end{aligned}$$

– We show that the relation

$$\mathcal{R} = \{\, (s_1, s_2) \mid d_\delta(s_1, s_2) = 0 \,\}$$

is a probabilistic bisimulation. Obviously, $\mathcal{R}$ is an equivalence relation. Assume that $s_1 \,\mathcal{R}\, s_2$. That is, $d_\delta(s_1, s_2) = 0$. Let E be an $\mathcal{R}$-equivalence class. Without loss of any generality, we may assume that E is of the form $[s]_{d_\delta}$. All states in $[s]_{d_\delta}$ assign the same value to each formula. For each state $s' \notin [s]_{d_\delta}$ there exists a formula $\varphi_{s'}$ such that $[\![\varphi_{s'}]\!]_\delta(s) \neq [\![\varphi_{s'}]\!]_\delta(s')$. Without loss of any generality, we may assume that $[\![\varphi_{s'}]\!]_\delta(s) > [\![\varphi_{s'}]\!]_\delta(s')$. Hence, there exists a rational $q_{s'}$ in $[0, 1]$ such that $[\![\varphi_{s'} \ominus q_{s'}]\!]_\delta(s') = 0$ and $[\![\varphi_{s'} \ominus q_{s'}]\!]_\delta(s) > 0$. Now consider the formula

$$\varphi = \bigwedge_{s' \notin [s]_{d_\delta}} \varphi_{s'} \ominus q_{s'}.$$

Then $[\![\varphi]\!]_\delta(s'') > 0$ iff $s'' \in [s]_{d_\delta}$. As a consequence,

$$\begin{aligned}
&\delta[\![\varphi]\!]_\delta(s) \sum_{s' \in [s]_{d_\delta}} \pi(s_1, s') \\
&= \delta \sum_{s' \in [s]_{d_\delta}} \pi(s_1, s')[\![\varphi]\!]_\delta(s') \\
&= \delta \sum_{s'' \in S} \pi(s_1, s'')[\![\varphi]\!]_\delta(s'') \quad [[\![\varphi]\!]_\delta(s'') = 0 \text{ for all } s'' \notin [s]_{d_\delta}] \\
&= [\![\Diamond\varphi]\!]_\delta(s_1) \\
&= [\![\Diamond\varphi]\!]_\delta(s_2) \quad [d_\delta(s_1, s_2) = 0] \\
&= \delta[\![\varphi]\!]_\delta(s) \sum_{s' \in [s]_{d_\delta}} \pi(s_2, s').
\end{aligned}$$

Therefore, $\sum_{s' \in [s]_{d_\delta}} \pi(s_1, s') = \sum_{s' \in [s]_{d_\delta}} \pi(s_2, s')$ and, hence, $\mathcal{R}$ is a probabilistic bisimulation. □

This result has also been proved in [16, Theorem 4.10].

Proposition 3. *The set of 1-bounded pseudometrics on S endowed with the order $\sqsubseteq$ forms a complete lattice.*

Proof. Obviously, $\sqsubseteq$ is a partial order. The top element is the 1-bounded pseudometric $\top$ defined by

$$\top(s_1, s_2) = 0.$$

The bottom element is the 1-bounded pseudometric $\perp$ defined by

$$\perp(s_1, s_2) = \begin{cases} 0 \text{ if } s_1 = s_2 \\ 1 \text{ otherwise.} \end{cases}$$

Let D be a nonempty set of 1-bounded pseudometrics on S. The meet of D is the 1-bounded pseudometrics $\bigsqcap D$ defined by

$$(\bigsqcap D)(s_1, s_2) = \sup_{d \in D} d(s_1, s_2).$$

The join of D can be expressed in terms of the meet of D (see, for example, [11, Lemma 2.15]). □

A similar proof can be found in [15, Lemma 3.2] and [13, Lemma 2.3].

Proposition 4. *$\Delta(d)$ is a 1-bounded pseudometric on S.*

Proof. Note that $f \in (S, d) \nrightarrow [0, 1]$ implies $1 - f \in (S, d) \nrightarrow [0, 1]$. Furthermore, if $s_1 \rightarrow$ and $s_2 \rightarrow$ then

$$\begin{aligned}
&\sum_{s \in S} (1 - f)(s)(\pi(s_1, s) - \pi(s_2, s)) \\
&= \sum_{s \in S} \pi(s_1, s) - \sum_{s \in S} \pi(s_2, s) + \sum_{s \in S} f(s)(\pi(s_2, s) - \pi(s_1, s)) \\
&= \sum_{s \in S} f(s)(\pi(s_2, s) - \pi(s_1, s)) \\
&= \sum_{s \in S} f(s)\pi(s_2, s) - \sum_{s \in S} f(s)\pi(s_1, s).
\end{aligned}$$

As a consequence, if $s_1 \rightarrow$ and $s_2 \rightarrow$ then

$$\Delta(d)(s_1, s_2) = \max \left\{ \left| \sum_{s \in S} f(s)\pi(s_1, s) - \sum_{s \in S} f(s)\pi(s_2, s) \right| \;\middle|\; f \in (S, d) \nrightarrow [0, 1] \right\}.$$

Now that we have this alternative representation of $\Delta(d)$, checking that it satisfies the three conditions of Definition 3 is straightforward. □

Proposition 5. *Δ is order-preserving.*

Proof. Let d_1 and d_2 be 1-bounded pseudometrics on S with $d_1 \sqsubseteq d_2$.
Assume that $f \in (S, d_2) \nrightarrow [0,1]$. Then

$$\begin{aligned}
&|f(s_1) - f(s_2)| \\
&\leq d_2(s_1, s_2) \quad [f \in (S, d_2) \nrightarrow [0,1]] \\
&\leq d_1(s_1, s_2) \quad [d_1 \sqsubseteq d_2]
\end{aligned}$$

As a consequence,

$$(S, d_1) \nrightarrow [0,1] \supseteq (S, d_2) \nrightarrow [0,1]. \tag{1}$$

We have to show that $\Delta(d_1)(s_1, s_2) \geq \Delta(d_2)(s_1, s_2)$. We focus on the only nontrivial case: $s_1 \rightarrow$ and $s_2 \rightarrow$. In this case,

$$\begin{aligned}
&\Delta(d_1)(s_1, s_2) \\
&= \max \left\{ \sum_{s \in S} f(s)(\pi(s_1, s) - \pi(s_2, s)) \;\middle|\; f \in (S, d_1) \nrightarrow [0,1] \right\} \\
&\geq \max \left\{ \sum_{s \in S} f(s)(\pi(s_1, s) - \pi(s_2, s)) \;\middle|\; f \in (S, d_2) \nrightarrow [0,1] \right\} \quad [(1)] \\
&= \Delta(d_2)(s_1, s_2).
\end{aligned}$$

$\square$

A similar result is presented in [3, Proposition 38].

Theorem 1. $d_1 = \mathrm{gfp}(\Delta)$.

Proof. We first prove that d_1 is a post-fixed point of Δ. That is, we show that $\Delta(d_1)(s_1, s_2) \leq d_1(s_1, s_2)$. To prove this, we distinguish the following three cases.

- If $s_1 \not\rightarrow$ and $s_2 \not\rightarrow$ then the property is vacuously true.
- If $s_1 \not\rightarrow$ and $s_2 \rightarrow$, or $s_1 \rightarrow$ and $s_2 \not\rightarrow$, then the formula $\Diamond$true witnesses that the states s_1 and s_2 have distance one.
- Assume that $s_1 \rightarrow$ and $s_2 \rightarrow$. According to [5, Proposition 39], the set $\{ [\![\varphi]\!]_1 \mid \varphi \in \mathcal{L} \}$ is dense in $(S, d_1) \nrightarrow [0,1]$. As a consequence,

$$\begin{aligned}
&\max \left\{ \sum_{s \in S} f(s)(\pi(s_1, s) - \pi(s_2, s)) \;\middle|\; f \in (S, d_1) \nrightarrow [0,1] \right\} \\
&= \max \left\{ \sum_{s \in S} [\![\varphi]\!]_1(s)(\pi(s_1, s) - \pi(s_2, s)) \;\middle|\; \varphi \in \mathcal{L} \right\} \\
&= \max \left\{ \sum_{s \in S} \pi(s_1, s)[\![\varphi]\!]_1(s) - \sum_{s \in S} \pi(s_2, s)[\![\varphi]\!]_1(s) \;\middle|\; \varphi \in \mathcal{L} \right\} \\
&= \max \left\{ [\![\Diamond\varphi]\!]_1(s_1) - [\![\Diamond\varphi]\!]_1(s_2) \;\middle|\; \varphi \in \mathcal{L} \right\} \\
&\leq d_1(s_1, s_2).
\end{aligned}$$

Next we prove that d_1 is the greatest post-fixed point of Δ. Assume that d is a post-fixed point of Δ. We have to show that $d \sqsubseteq d_1$. That is, $d_1(s_1, s_2) \leq d(s_1, s_2)$. We restrict our attention to the case that $s_1 \rightarrow$ and $s_2 \rightarrow$. It suffices to show that

$$[\![\varphi]\!]_1(s_1) - [\![\varphi]\!]_1(s_2) \leq d(s_1, s_2)$$

for all φ. This can be proved by structural induction on φ. We consider only the nontrivial case: $\Diamond\varphi$.

$$\begin{aligned}
&[\![\Diamond\varphi]\!]_1(s_1) - [\![\Diamond\varphi]\!]_1(s_2)\\
&= \sum_{s \in S} \pi(s_1, s)[\![\varphi]\!]_1(s) - \sum_{s \in S} \pi(s_2, s)[\![\varphi]\!]_1(s)\\
&= \sum_{s \in S} [\![\varphi]\!]_1(s)(\pi(s_1, s) - \pi(s_2, s))\\
&\leq \max \left\{ \sum_{s \in S} f(s)(\pi(s_1, s) - \pi(s_2, s)) \;\middle|\; f \in (S, d) \nrightarrow [0, 1] \right\}\\
&\qquad [\text{by induction, } [\![\varphi]\!]_1 \in (S, d) \nrightarrow [0, 1]]\\
&= \Delta(d)(s_1, s_2)\\
&\leq d(s_1, s_2) \quad [d \text{ is a post-fixed point of } \Delta]
\end{aligned}$$

□

A similar result can be obtained by combining Theorem 40 and 44 of [3].

Proposition 7. *Let d be a 1-bounded pseudometric on S. Let s_1, $s_2 \in S$ such that $s_1 \rightarrow$ and $s_2 \rightarrow$. Then*

$$\Delta(d)(s_1, s_2) = \min \left\{ \sum_{(s_i, s_j) \in S^2} d(s_i, s_j)\mu(s_i, s_j) \;\middle|\; \mu \in \pi(s_1, \cdot) \otimes \pi(s_2, \cdot) \right\}$$

where $\mu \in \pi(s_1, \cdot) \otimes \pi(s_2, \cdot)$ if

$$\forall s_j \in S \sum_{s_i \in S} \mu(s_i, s_j) = \pi(s_1, s_j) \wedge \forall s_i \in S \sum_{s_j \in S} \mu(s_i, s_j) = \pi(s_2, s_i).$$

Proof. Since the set S is finite, the space (S, d) is compact. The probability distributions $\pi(s_1, \cdot)$ and $\pi(s_2, \cdot)$ define Borel probability measures on (S, d). Applying the Kantorovich-Rubinstein gives us the desired result. □

A similar result is presented in [6, Corollary 19].

Proposition 8.

- *If $s_1 \not\rightarrow$ and $s_2 \not\rightarrow$ then $d_1(s_1, s_2) = 0$.*
- *If $s_1 \not\rightarrow$ and $s_2 \rightarrow$, or $s_1 \rightarrow$ and $s_2 \not\rightarrow$ then $d_1(s_1, s_2) = 1$.*

Proof. We only consider the first case. The second one can be proved similarly. If $s_1 \not\rightarrow$ and $s_2 \not\rightarrow$ then $\delta(s_1, s_2) = \Delta(\delta)(s_1, s_2) = 0$. □

Proposition 9. *If* $\tau_\omega(s_2) = 0$ *then* $d_1(s_1, s_2) = \tau_\omega(s_1)$.

Proof. Assume that $\tau_\omega(s_2) = 0$. We prove that for all $n \in \omega + 1$,

$$d^n(s_1, s_2) = \tau_n(s_1)$$

by induction on n.

- Obviously, $d^0(s_1, s_2) = 0 = \tau_0(s_1)$.
- We have to prove that $d^{n+1}(s_1, s_2) = \tau_{n+1}(s_1)$. We distinguish the following two cases.
  - If $s_1 \not\rightarrow$ then $d^{n+1}(s_1, s_2) = 1 = \tau_{n+1}(s_1)$.
  - Now let us assume that $s_1 \rightarrow$. First we show that τ_n as a function from (S, d^n) to $[0, 1]$ is nonexpansive. For all s, s',

$$\begin{aligned} |\tau_n(s) - \tau_n(s')| &= |d^n(s, s_2) - d^n(s', s_2)| \quad \text{[induction]} \\ &\leq d^n(s, s') \quad \text{[triangle inequality]} \end{aligned}$$

Since

$$\begin{aligned} & d^{n+1}(s_1, s_2) \\ & = \Delta(d^n)(s_1, s_2) \\ & \geq \sum_{s \in S} \tau_n(s)(\pi(s_1, s) - \pi(s_2, s)) \quad [\tau_n \text{ is nonexpansive}] \\ & = \sum_{s \in S} \tau_n(s)\pi(s_1, s) - \sum_{s \in S} \tau_n(s)\pi(s_2, s) \\ & = \tau_{n+1}(s_1) - \tau_{n+1}(s_2) \\ & = \tau_{n+1}(s_1) \quad [\tau_\omega(s_2) = 0 \text{ and, hence, } \tau_{n+1}(s_2) = 0] \end{aligned}$$

Let $f \in (S, d^n) \rightarrow\!\!\!\!\!\!\nearrow [0, 1]$. For all s,

$$f(s) - f(s_2) \leq |f(s) - f(s_2)| \leq d^n(s, s_2) = \tau_n(s).$$

As a consequence,

$$\begin{aligned} & \sum_{s \in S} f(s)(\pi(s_1, s) - \pi(s_2, s)) \\ & = \sum_{s \in S} f(s)\pi(s_1, s) - \sum_{s \in S} f(s)\pi(s_2, s)) \\ & = \sum_{s \in S} (f(s) - f(s_2))\pi(s_1, s) - \sum_{s \in S} (f(s) - f(s_2))\pi(s_1, s) \quad [\textstyle\sum_{s \in S} f(s_2))\pi(s_1, s) = f(s_2)] \\ & = \sum_{s \in S} (f(s) - f(s_2))(\pi(s_1, s) - \pi(s_2, s)) \\ & \leq \sum_{s \in S} \tau_n(s)(\pi(s_1, s) - \pi(s_2, s)) \\ & = \tau_{n+1}(s_1). \end{aligned}$$

Since f was chosen arbitrarily, we can conclude that

$$d^{n+1}(s_1, s_2) \leq \tau_{n+1}(s_1).$$

- Finally,

$$\begin{aligned} d^{\omega}(s_1, s_2) &= \sup_n d^n(s_1, s_2) \\ &= \sup_n \tau_n(s_1) \quad \text{[by induction]} \\ &= \tau_\omega(s_1). \end{aligned}$$

From Theorem 1 and Proposition 6 we can conclude that $d_1(s_1, s_2) = d^{\omega}(s_1, s_2) = \tau_\omega(s_1)$.

□

Proposition 10. *The smallest equivalence relation containing* $\{ \langle s_1, s_2 \rangle \mid s_1 \not\rightarrow \wedge s_2 \not\rightarrow \}$ *and* $\{ \langle s_1, s_2 \rangle \mid \tau_\omega(s_1) = 0 \wedge \tau_\omega(s_2) = 0 \}$ *is a probabilistic bisimulation.*

Proof. Let us denote the relation by $\mathcal{R}$. Assume that $s_1 \mathcal{R} s_2$. It suffices to show that $\sum_{s' \mathcal{R} s} \pi(s_1, s') = \sum_{s' \mathcal{R} s} \pi(s_2, s')$ for each s. We distinguish the following three cases.

- If $s_1 = s_2$ then it is vacuously true.
- If $s_1 \not\rightarrow$ and $s_2 \not\rightarrow$ then $\sum_{s' \mathcal{R} s} \pi(s_1, s') = 0$ and $\sum_{s' \mathcal{R} s} \pi(s_2, s') = 0$.
- Assume $\tau_\omega(s_1) = 0$ and $\tau_\omega(s_2) = 0$. If $\tau_\omega(s) = 0$ then $\sum_{s' \mathcal{R} s} \pi(s_1, s') = 1$ and $\sum_{s' \mathcal{R} s} \pi(s_2, s') = 1$. Otherwise, $\sum_{s' \mathcal{R} s} \pi(s_1, s') = 0$ and $\sum_{s' \mathcal{R} s} \pi(s_2, s') = 0$.

□

Proposition 11. *For all* s_1*,* $s_2 \in S$*,* $d_{\mathcal{R}}([s_1], [s_2]) = d_1(s_1, s_2)$.

Proof. First all, note that

$$\sum_{s' \in S} \pi(s, s') = \sum_{[s'] \in S_{\mathcal{R}}} \sum_{s'' \mathcal{R} s'} \pi(s, s'') = \sum_{[s'] \in S_{\mathcal{R}}} \pi_{\mathcal{R}}([s], [s']).$$

As a consequence, we have left to consider the case $s_1 \rightarrow$ and $s_2 \rightarrow$. We prove that for all $n \in \omega{+}1$, $d^n_{\mathcal{R}}([s_1], [s_2]) = d^n_1(s_1, s_2)$ by induction on n. We distinguish the following three cases.

- If $n = 0$ then the property is vacuously true.
- Assume that $d^n_{\mathcal{R}}([s'_1], [s'_2]) = d^n_1(s'_1, s'_2)$ for all s'_1, $s'_2 \in S$. Let s_1, $s_2 \in S$. We have to prove that $d^{n+1}_{\mathcal{R}}([s_1], [s_2]) = d^{n+1}_1(s_1, s_2)$. In the proof of this case, we make use of the following two observations. For each $f \in (S_{\mathcal{R}}, d^n_{\mathcal{R}}) \rightarrow [0, 1]$, there exists a $g \in (S, d^{n+1}_1) \rightarrow [0, 1]$ such that $g(s) = f([s])$ for all $s \in S$, since

$$\begin{aligned} |g(s) - g(s')| &= |f([s]) - f([s'])| \\ &\leq d^n_{\mathcal{R}}(s, s') \quad [f \text{ is nonexpansive}] \\ &= d^{n+1}_1(s, s') \quad \text{[induction]}. \end{aligned}$$

Similarly, we can show that for each $g \in (S, d_1^{n+1}) \rightarrow [0,1]$, there exists $f \in (S_{\mathcal{R}}, d_{\mathcal{R}}^n) \rightarrow [0,1]$ such that $f([s]) = g(s)$ for all $s \in S$. Note that if states s and s' are probabilistic bisimilar then $d_1(s,s') = 0$ and, hence, $d_1^{n+1}(s,s') = 0$ and, therefore, $g(s) = g(s')$, since g is nonexpansive.

$$
\begin{aligned}
&d_{\mathcal{R}}^{n+1}([s_1],[s_2]) \\
&= \Delta(d_{\mathcal{R}}^n)([s_1],[s_2]) \\
&= \max\left\{ \sum_{[s]\in S_{\mathcal{R}}} f([s])(\pi_{\mathcal{R}}([s_1],[s]) - \pi_{\mathcal{R}}([s_2],[s])) \;\middle|\; f \in (S_{\mathcal{R}}, d_{\mathcal{R}}^n) \rightarrow [0,1] \right\} \\
&= \max\left\{ \sum_{[s]\in S_{\mathcal{R}}} f([s]) \sum_{s'\mathcal{R}s} (\pi(s_1,s') - \pi(s_2,s')) \;\middle|\; f \in (S_{\mathcal{R}}, d_{\mathcal{R}}^n) \rightarrow [0,1] \right\} \\
&= \max\left\{ \sum_{[s]\in S_{\mathcal{R}}} \sum_{s'\mathcal{R}s} f([s'])(\pi(s_1,s') - \pi(s_2,s')) \;\middle|\; f \in (S_{\mathcal{R}}, d_{\mathcal{R}}^n) \rightarrow [0,1] \right\} \\
&= \max\left\{ \sum_{s\in S} g(s)(\pi(s_1,s) - \pi(s_2,s)) \;\middle|\; g \in (S, d_1^n) \rightarrow [0,1] \right\} \\
&= \Delta(d_1^n)(s_1,s_2) \\
&= d_1^{n+1}(s_1,s_2).
\end{aligned}
$$

– Furthermore,

$$
\begin{aligned}
d_{\mathcal{R}}^{\omega}([s_1],[s_2]) &= \sup_n d_{\mathcal{R}}^n([s_1],[s_2]) \\
&= \sup_n d_1^n(s_1,s_2) \quad \text{[induction]} \\
&= d_1^{\omega}(s_1,s_2).
\end{aligned}
$$

□

C Calculation of Termination Probabilities

In this appendix we present the key ingredients that are used to compute the termination probabilities. We already defined the termination probability τ_ω in Definition 13. It is the probability of reaching a state with no outgoing transitions.

The set S^0 consists of those states that can reach a state without outgoing transition in zero transitions, that is, $S^0 = \{\, s \in S \mid s \not\rightarrow \,\}$. The set S^* consists of those states that can reach a state without outgoing transition. The set S^+ consists of those states that can reach a state without outgoing transition in at least one transition, that is, $S^+ = S^* \setminus S^0$. The set of states that cannot reach a state without outgoing transitions, $S \setminus S^*$, is denoted by $\overline{S^*}$. Obviously, S^0, S^+ and $\overline{S^*}$ form a partition of S.

Clearly, the probability of termination for states in S^0 and $\overline{S^*}$ is 1 and 0, respectively. For a state $s \in S^+$, the probability of termination $\tau_\omega(s)$ can be expressed as follows:

$$\tau_\omega(s) = \sum_{u \in S^+} \pi(s, s')\tau_\omega(s') + \sum_{s' \in S^0} \pi(s, s').$$

Let rename the states such that $S^+ = \{s_1, \ldots, s_M\}$ for some $M \geq 0$. Then the above equation can be expressed in matrix form as

$$T = P.T + R,$$

where $T[i] = \tau_\omega(s_i)$, $P[i,j] = \pi(s_i, s_j)$ and $R[i] = \sum_{s' \in S^0} \pi(s_i, s')$ for $1 \leq i, j \leq M$. We have that

$$\begin{aligned}
&T = P.T + R \\
&\Leftrightarrow (I - P).T = R \\
&\Leftrightarrow T = (I - P)^{-1}.R
\end{aligned}$$

Next, we prove that $(I-P)^{-1}$ exists and, therefore, the terminating probabilities can easily be computed using the above characterization.

Proposition 17. $\lim\limits_{n\to\infty} P^n = 0.$

Proof. This proof is a modification of the proof of [20, Theorem 11.3].

For each state $s_i \in S^+$, there exists a path from s_i to a state in $S^0 \cup \overline{S^*}$. Let m_i be the minimum length of such a path and let p_i be the probability of staying in S^+ in the first m_i transitions when started in s_i. Clearly, $p_i < 1$. Let $m = \max_{1 \leq i \leq M} m_i$ and $p = \max_{1 \leq i \leq M} p_i$. Then $p<1$. Obviously, the probability of staying in S^+ in the first m transitions when started in s_i is at most p. Hence, the probability of staying in S^+ in the first km transitions when started in s_i is at most p^k.

Note that $P^m[i,j]$ is the probability of reaching state s_j, starting from state s_i, in m transitions. To reach $s_j \in S^+$, one has to stay in S^+. From the above we can conclude that $P^{km}[i,j] \leq p^k$, which implies the desired result. □

Proposition 18. $(I - P)^{-1}$ *exists.*

Proof. This proof is a modification of the proof of [20, Theorem 11.4].

The matrix $I - P$ has an inverse iff 0 is not an eigenvalue of $I - P$ (see, for example, [26, page 47]). We prove the latter by contradiction. Assume that 0 is an eigenvalue of $I - P$. Then $(I - P).X = 0$ for some $X \neq 0$. Hence,

$$\begin{aligned}
&(I - P).X = 0 \\
&\Rightarrow I.X - P.X = 0 \\
&\Rightarrow I.X = P.X \\
&\Rightarrow X = P.X \\
&\Rightarrow X = P^n.X.
\end{aligned}$$

From Proposition 17 we can conclude that X has to be 0, which contradicts the assumption that $X \neq 0$. $\square$