On the Continuity of the Probabilistic Bisimilarity Distance

Syyeda Zainab Fatmi
University of Oxford, UK

Stefan Kiefer
University of Oxford, UK

David Parker
University of Oxford, UK

Franck van Breugel
York University, Toronto, Canada

Abstract

The probabilistic bisimilarity distance provides a quantitative measure of behavioural difference for labelled Markov chains, but it may be discontinuous under perturbations of the transition probabilities. This lack of continuity undermines its applicability to empirically derived models, where transition probabilities are often approximations. Recently, we introduced robust probabilistic bisimilarity as a sufficient condition for continuity at distance zero. In this paper, we show that it is also a necessary condition, that is, two states are robustly probabilistic bisimilar if and only if their probabilistic bisimilarity distance is small for any small enough perturbation of the transition probabilities. We further extend robustness to non-bisimilar state pairs to establish a complete characterization for continuity of the probabilistic bisimilarity distance. Based on this characterization, we develop a polynomial time algorithm to decide continuity. Finally, we complement our theoretical contributions with an experimental evaluation demonstrating the proposed approach in practice. Our results show that the extra step of deciding continuity requires minimal additional cost when compared to computing the probabilistic bisimilarity distance.

2012 ACM Subject Classification Theory of computation → Logic and verification; Theory of computation → Verification by model checking; Theory of computation → Probabilistic computation

Keywords and phrases probabilistic model checking, labelled Markov chain, probabilistic bisimilarity distance

Digital Object Identifier 10.4230/LIPIcs.CONCUR.2026.34

Related Version *Full Version*: `https://doi.org/10.48550/arXiv.2606.27209` [16]

Supplementary Material *Software*: `https://github.com/zainabfatmi/prism/tree/concur` [14]
archived at `swh:1:dir:4efb0e4307476e883191835aed030dce0af9782c`

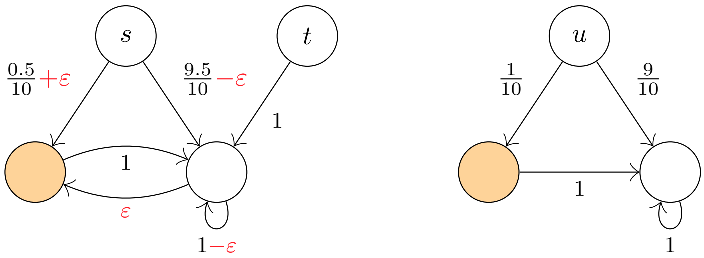
Funding *Syyeda Zainab Fatmi*: Clarendon Fund
Stefan Kiefer: EPSRC grant EP/Z536179/1
Franck van Breugel: Natural Sciences and Engineering Research Council of Canada

1 Introduction

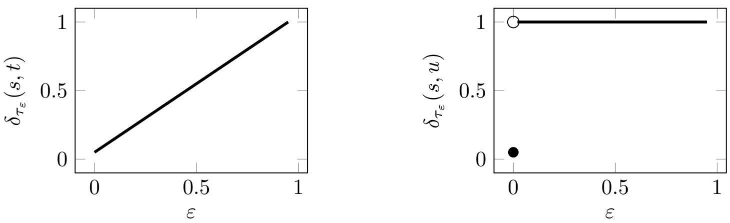
Probabilistic bisimilarity (henceforth just *bisimilarity*) is a concept of behavioural equivalence for probabilistic systems, such as labelled Markov chains. Kemeny and Snell [24] introduced the notion of lumpability for Markov chains and it was adapted to the setting of labelled Markov chains by Larsen and Skou [28]. Intuitively, two states of a labelled Markov chain are considered bisimilar if they have the same labels and they transition into the same equivalence classes with the same probabilities. Bisimilarity can be used to minimize the state space of a system by identifying and merging states that are behaviourally indistinguishable, which often reduces the time required to verify properties of the system [23].

However, since behavioural equivalences are sensitive to the exact values of the transition probabilities, Giacalone et al. [18] proposed using behavioural pseudometrics, a quantitative generalisation of behavioural equivalences, to capture the behavioural similarity of states in a probabilistic system. Rather than classifying states as either equivalent or inequivalent, the pseudometric maps each pair of states to a real value in the unit interval, thus also quantifying the behavioural difference between non-equivalent states.

In probabilistic verification, the most widely studied example of such a behavioural pseudometric is the *probabilistic bisimilarity distance* (henceforth just *bisimilarity distance*), introduced by Desharnais et al. [8], which quantitatively generalises bisimilarity. The lower the distance between two states, the less difference in their behaviour, and two states have distance zero if and only if they are bisimilar.



(a) A family of labelled Markov chains for $\varepsilon \in [0, \frac{9.5}{10}]$. The state labels are represented by colours.



(b) An illustration of the distances between the state pairs (s, t) and (s, u), respectively. When $\varepsilon = 0$, then the distance between these pairs is 0.05. As ε increases, the distance between s and t increases proportionally, whereas the distance between s and u jumps up to 1.

Figure 1 A motivating example.

▶ **Example 1.** The bisimilarity distance is based on a variation of Larsen and Skou's probabilistic modal logic [6, 26, 28, 29]. A formula φ of this logic assigns to each state s of a labelled Markov chain a numerical value $[\![\varphi]\!](s) \in [0, 1]$. For example, consider the labelled Markov chain in Figure 1a when $\varepsilon = 0$. The states s and t have a bisimilarity distance of 0.05. Then it follows [3, Equation 2.3] that $|[\![\varphi]\!](s) - [\![\varphi]\!](t)| \leq 0.05$, where φ ranges over all formulas. Thus, the bisimilarity distance bounds the maximal difference in logical observations between states. This is relevant for model checking. Suppose we are interested in computing the probability of eventually reaching a state labelled with *orange*. This property can be expressed as the quantitative mu-calculus formula $\varphi = \mu V.\, \mathit{orange} \vee \mathrm{next}(V)$. For efficiency one might want to approximate these values $[\![\varphi]\!]$, up to some acceptable error threshold, say 0.1. In this case, for states s and t, it suffices to compute $[\![\varphi]\!](s)$ only. The same holds for *every* formula φ of the logic. Similarly, since states s and u also have a bisimilarity distance of 0.05, it follows that $|[\![\varphi]\!](s) - [\![\varphi]\!](u)| \leq 0.05 < 0.1$, for all φ. Hence, for states s and u, it suffices to compute $[\![\varphi]\!](s)$ only as well, speeding up model checking. ⌟

However, as recognised by Jaeger et al. [21], the bisimilarity distance is sometimes not continuous. Small changes in the transition probabilities can lead to unexpected and abrupt changes in behaviour between two states. This is especially concerning when relying on bisimilarity as a measure of system equivalence for empirically derived models, where transition probabilities are often approximated from experimental data [11, 30, 31, 33, 36]. It may be dangerous to merge bisimilar states if the transition probabilities are not known precisely, as the distance may jump up. In such settings, continuity analysis provides a way to assess the sensitivity of the bisimilarity distance to perturbations. If the distance is continuous, then sufficiently small errors in the transition probabilities result in only small changes in the distance; if it is discontinuous, arbitrarily small perturbations may cause a non-negligible increase in the distance, indicating that the apparent similarity of two states may not be robust to uncertainty in the model.

More recently, in our prior work [15], we introduced a new notion of probabilistic equivalence, namely *robust probabilistic bisimilarity* (henceforth just *robust bisimilarity*). Our main result showed that this definition ensures the continuity of the bisimilarity distance function under perturbations of the transition probabilities. However, several fundamental questions remain unanswered. First, although we conjectured in [15] that robust bisimilarity is not only a sufficient but also a necessary condition for continuity, this was left open. Second, our analysis in [15] is restricted to bisimilar pairs of states and does not address how one could characterize and decide the continuity of the bisimilarity distance for non-bisimilar pairs of states.

In this paper, we close these gaps. We prove that robust bisimilarity exactly characterizes continuity for bisimilar state pairs, thereby settling the conjecture of [15]. As a consequence, we obtain a polynomial time algorithm to decide continuity for bisimilar pairs of states. Moreover, we go beyond the setting of bisimilar states and provide a complete characterization of continuity for arbitrary pairs of states. We also develop a polynomial-time algorithm to decide continuity for all pairs of states. Finally, we present experimental results on an implementation of this algorithm, which demonstrate that deciding continuity incurs only a modest overhead compared to computing the bisimilarity distance, and is often significantly faster in practice.

▶ **Example 2.** Recall the model checking scenario in Example 1 involving the states s, t, and u. The argument to compute $[\![\varphi]\!](s)$ only is still valid when the transition probabilities are not known precisely, *provided* that the distance is continuous with respect to the transition probabilities. As shown in Figure 1b, the distance between s and t is continuous. Thus, as long as the perturbation is small enough (in particular $\varepsilon < 0.05$), the distance between s and t is less than 0.1 and therefore $|[\![\varphi]\!](s) - [\![\varphi]\!](t)| < 0.1$. In contrast, the distance between s and u is discontinuous. Indeed, when $\varepsilon > 0$, for the formula $\varphi = \mu V.\, \mathit{orange} \vee \mathrm{next}(V)$, we have $|[\![\varphi]\!](s) - [\![\varphi]\!](u)| > 0.1$. Therefore, in this case, we cannot compute $[\![\varphi]\!](s)$ only. ⌟

The rest of the paper is structured as follows. Section 2 provides necessary preliminaries, such as the definitions of labelled Markov chains and traditional bisimulation. Section 3 recalls the formal definition and several properties of the bisimilarity distance. In Section 4 we build on the key result from [15] by proving the conjecture posed therein, establishing that robust bisimilarity is not only sufficient but also necessary for the continuity of the bisimilarity distance for bisimilar state pairs. Our central technical contribution appears in Section 5, where we complete the theoretical characterisation of the continuity of the bisimilarity distance. In Section 6 we provide a polynomial time algorithm to decide continuity and in Section 7 we report experimental results on the algorithm's implementation. Finally, Section 8 summarises our results and discusses directions for future research. The full version of this paper, found in [16], includes omitted proofs and further details.

2 Preliminaries

In this section, we present some fundamental concepts that underpin this paper. We begin with basic notions on probability distributions and Markov chains.

Let X be a nonempty finite set. A function $\mu : X \to [0,1]$ is a *subprobability distribution* on X if $\sum_{x \in X} \mu(x) \leq 1$. We denote the set of subprobability distributions on X by $\mathcal{S}(X)$. For $\mu \in \mathcal{S}(X)$ and $A \subseteq X$, we often write $\mu(A)$ instead of $\sum_{x \in A} \mu(x)$. For a distribution $\mu \in \mathcal{S}(X)$ we define the *support* of μ by $\text{support}(\mu) = \{ x \in X \mid \mu(x) > 0 \}$. A subprobability distribution μ on X is a *probability distribution* if $\mu(X) = 1$. We denote the set of probability distributions on X by $\mathcal{D}(X)$.

A *Markov chain* is a pair $\langle S, \tau \rangle$ consisting of a finite set S of states and a transition probability function $\tau : S \to \mathcal{D}(S)$. A *labelled Markov chain* is a tuple $\langle S, L, \tau, \ell \rangle$ where $\langle S, \tau \rangle$ is a Markov chain, L is a finite set of labels and $\ell : S \to L$ is a labelling function. A *path* in a Markov chain $\langle S, \tau \rangle$ is a sequence of states s_0, s_1, $s_2 \ldots$ such that $s_i \in S$ and $\tau(s_i)(s_{i+1}) > 0$ for all $i \geq 0$.

We say that states of a Markov chain *communicate* with each other if each is reachable from the other via a (possibly empty) path. This is an equivalence relation which yields a set of *communication classes*. These communication classes coincide with the strongly connected components (SCCs) of the directed graph underlying the labelled Markov chain. A communication class is *closed* if the probability of leaving the class is zero, that is, no transition with positive probability leads from a state in the class to a state outside the class. Equivalently, a closed communication class is a bottom SCC.

For the remainder, we fix a labelled Markov chain $\langle S, L, \tau, \ell \rangle$, and we will study perturbations of the transition probability function τ.

We now introduce couplings, which play a central role throughout the rest of this paper. Couplings provide a way to relate two probability distributions by describing how their probability mass can be "matched." Intuitively, a coupling specifies a joint distribution whose marginals coincide with the original distributions. Formally, for all μ, $\nu \in \mathcal{D}(X)$, the set $\Omega(\mu, \nu)$ of *couplings* of μ and ν is defined by

$$\Omega(\mu, \nu) = \{ \omega \in \mathcal{D}(X \times X) \mid \forall x \in X : \omega(x, X) = \mu(x) \text{ and } \omega(X, x) = \nu(x) \}.$$

We write $\omega(x, X)$ for $\sum_{y \in X} \omega(x, y)$.

▶ **Definition 3.** *An equivalence relation $R \subseteq S \times S$ is a* bisimulation *if for all $(s,t) \in R$, $\ell(s) = \ell(t)$ and there exists $\omega \in \Omega(\tau(s), \tau(t))$ such that* $\text{support}(\omega) \subseteq R$. *States s and t are* bisimilar*, denoted $s \sim t$, if $(s,t) \in R$ for some bisimulation R.*

Definition 3 [22, Definition 4.3] differs from the standard definition [28, Definition 6.3] which defines a bisimulation as an equivalence relation $R \subseteq S \times S$ such that for all $(s,t) \in R$, $\ell(s) = \ell(t)$ and for all R-equivalence classes C, $\tau(s)(C) = \tau(t)(C)$. Nevertheless, an equivalence relation R is a bisimulation by Definition 3 if and only if it is a bisimulation as per the standard definition (see [22, Theorem 4.6]). We use the coupling-based formulation of bisimulation in Definition 3 as it aligns naturally with the definition of the bisimilarity distance introduced in the next section.

▶ **Example 4.** Consider the states b_1 and b_2 in the labelled Markov chain of Figure 2a with $\varepsilon = 0$. These states are bisimilar, as shown in Figure 2b. By the standard definition of bisimilarity, it suffices to observe that b_1 and b_2 have the same label and assign the same probability mass to each $\sim$-equivalence class: both states transition with probability $\frac{1}{2}$ to the class $\{ c_1, c_2, c_3, c_4 \}$ and with probability $\frac{1}{2}$ to the class $\{ d_1, d_2 \}$.

Equivalently, using Definition 3, there exists a coupling $\omega \in \Omega(\tau(b_1), \tau(b_2))$ whose support is contained in the bisimulation relation. Concretely, one such coupling ω matches the successors of b_1 and b_2 by assigning probability $\frac{1}{2}$ each to the pairs (c_2, c_4) and (d_1, d_2). Since $c_2 \sim c_4$ and $d_1 \sim d_2$, it follows that $\text{support}(\omega) \subseteq \sim$. ⌟

We write $\ell(X)$ for $\{ \ell(x) \mid x \in X \}$. If $|\ell(S)| = 1$ then $\sim = S \times S$. In the remainder, we assume that the labelled Markov chain contains states with different labels, that is, $|\ell(S)| \geq 2$. Hence, we also have that $|S| \geq 2$.

3 Probabilistic Bisimilarity Distance

The bisimilarity distance [8] measures the behavioural difference between two states. A smaller distance indicates more similar behaviour, with a distance of zero corresponding precisely to behavioural equivalence.

Before presenting the formal definition of the bisimilarity distance, which is phrased as a fixed point, we first outline the underlying idea. If a pair of states, say s and t, have different labels, the states are maximally different and, therefore, have distance 1. Otherwise, one compares their transition distributions. This is done by selecting a coupling that matches successors of s with successors of t in an optimal way, and then measuring the behavioural difference of these successor pairs. As we will see in Proposition 11 below, the distance between two states corresponds to the minimal probability that, when exploring the two states jointly, one eventually encounters a pair of states with different labels.

▶ **Definition 5.** *The* bisimilarity distance, $\delta_\tau : S \times S \to [0, 1]$*, is the least fixed point of the function* $\Delta_\tau : (S \times S \to [0, 1]) \to (S \times S \to [0, 1])$ *defined by*

$$\Delta_\tau(d)(s, t) = \begin{cases} 1 & \text{if } \ell(s) \neq \ell(t) \\ \inf\limits_{\omega \in \Omega(\tau(s), \tau(t))} \sum\limits_{u,v \in S} \omega(u, v)\, d(u, v) & \text{otherwise.} \end{cases}$$

▶ **Theorem 6** ([34, Theorem 2.1.30])**.** δ_τ *is a pseudometric.*

▶ **Theorem 7** ([9, Theorem 4.10])**.** *For all* $s, t \in S$*,* $s \sim t$ *if and only if* $\delta_\tau(s, t) = 0$*.*

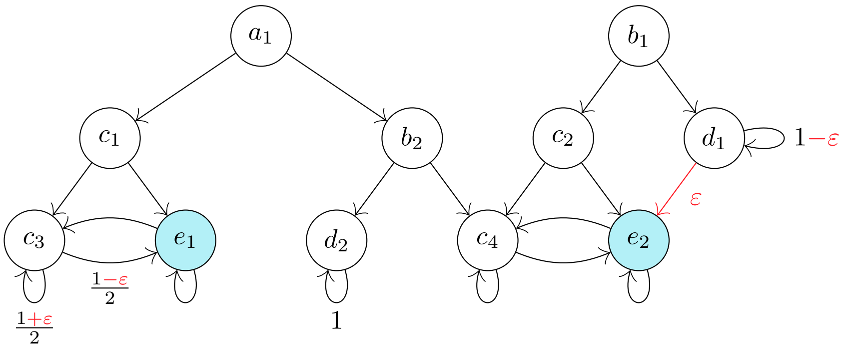
The bisimilarity distance function $\delta\_(s, t)$ is *lower semi-continuous* at τ if for any sequence $(\tau_n)_n$ converging to τ we have $\liminf_n \delta_{\tau_n}(s, t) \geq \delta_\tau(s, t)$ and *upper semi-continuous* at τ if we have $\limsup_n \delta_{\tau_n}(s, t) \leq \delta_\tau(s, t)$. Lastly, $\delta\_(s, t)$ is *continuous* at τ if it is both lower semi-continuous and upper semi-continuous at τ. The function $\delta\_(s, t)$ is lower semi-continuous (upper semi-continuous, continuous, resp.) if it is lower semi-continuous (upper semi-continuous, continuous, resp.) at τ for all transition probability functions τ.

▶ **Proposition 8** ([15, Proposition 1])**.** *For all* $s, t \in S$*, the function* $\delta\_(s, t) : (S \to \mathcal{D}(S)) \to [0, 1]$ *is lower semi-continuous.*

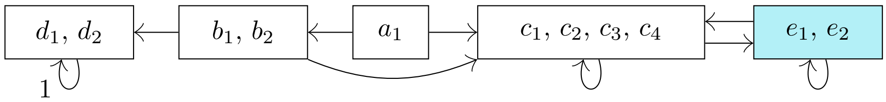
The following subsets of $S \times S$ play a key role in the subsequent discussion.

▶ **Definition 9.** *The sets* S^2_Δ*,* $S^2_{0,\tau}$*,* S^2_1*,* $S^2_{?,\tau}$*, and* $S^2_{0?}$ *are defined by*

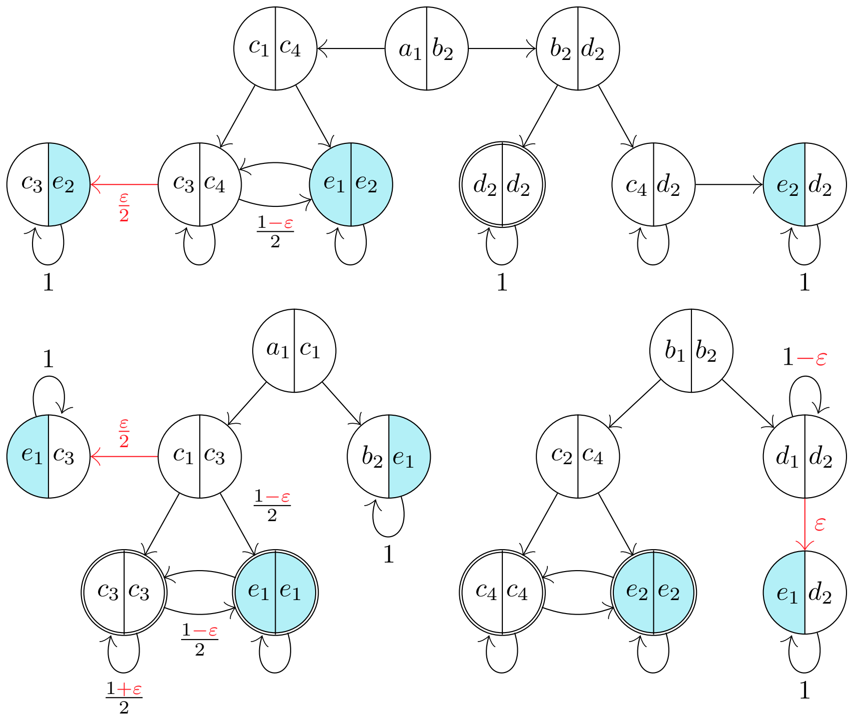
$$\begin{aligned} S^2_\Delta &= \{ (s, s) \mid s \in S \} \\ S^2_{0,\tau} &= \{ (s, t) \in S \times S \mid s \neq t \wedge s \sim t \} \\ S^2_1 &= \{ (s, t) \in S \times S \mid \ell(s) \neq \ell(t) \} \\ S^2_{?,\tau} &= (S \times S) \setminus (S^2_\Delta \cup S^2_{0,\tau} \cup S^2_1) \\ S^2_{0?} &= S^2_{0,\tau} \cup S^2_{?,\tau} \end{aligned}$$



(a) A family of labelled Markov chains for $\varepsilon \in [0,1]$. We denote the transition probability function for $\varepsilon = \frac{1}{n}$ by τ_n for each $n \in \mathbb{N}$ and for $\varepsilon = 0$ by τ. Note that $(\tau_n)_n$ converges to τ.



(b) The equivalence classes under traditional bisimilarity for τ. For each equivalence class, the outgoing transitions shown are those induced by the transitions of its constituent states.



(c) Part of the Markov chain $\langle S \times S, P_\varepsilon \rangle$ induced by a policy P_ε that is optimal for all $\varepsilon \in [0,1]$.

Figure 2 The running example. All omitted transition probabilities are $\frac{1}{2}$.

The first four sets form a partition of $S \times S$. Observe that the sets $S^2_{0,\tau}$ and $S^2_{?,\tau}$ depend on τ and may, therefore, change when we perturb τ, whereas the sets S^2_Δ and S^2_1 stay the same. Note that $S^2_{0?} = (S \times S) \setminus (S^2_\Delta \cup S^2_1)$. Hence, this set also stays the same if we perturb τ. Furthermore, note that $\sim\, = S^2_\Delta \cup S^2_{0,\tau}$ and for all $(s,t) \in S^2_1$, we have $\delta_\tau(s,t) = 1$.

▶ **Definition 10.** *Let $\tau : S \to \mathcal{D}(S)$. The set $\mathcal{P}_\tau$ of* policies *for τ is defined by*

$$\mathcal{P}_\tau = \left\{ P : (S \times S) \to \mathcal{D}(S \times S) \;\middle|\; \begin{array}{l} \forall (s,t) \in S^2_\Delta \cup S^2_{0?} : P(s,t) \in \Omega(\tau(s), \tau(t)) \\ \forall (s,t) \in S^2_1 : \text{support}(P(s,t)) = \{(s,t)\} \end{array} \right\}.$$

Note that a policy $P \in \mathcal{P}_\tau$ induces a Markov chain $\langle S \times S, P \rangle$. The subscript τ is omitted when clear from the context.

Intuitively, a policy chooses, for each pair of states $(s,t) \in S^2_\Delta \cup S^2_{0?}$, a joint transition distribution, which allows us to reason about reachability probabilities in the induced Markov chain on the product state space $S \times S$.

The following proposition characterizes δ_τ in terms of policies.

▶ **Proposition 11.** *For all s, $t \in S$, $\delta_\tau(s,t) = \min_{P \in \mathcal{P}} \rho^P_1(s,t)$, where $\rho^P_1(s,t)$ is the probability with which (s,t) reaches S^2_1 in $\langle S \times S, P \rangle$.*

The proof of Proposition 11 follows from [4, Theorem 8] and [2, Theorem 10.15]. This characterization reduces the distance to a reachability problem in a suitably constructed Markov chain.

We say that a policy $P \in \mathcal{P}$ is *optimal* if for all s, $t \in S$, we have $\rho^P_1(s,t) = \delta_\tau(s,t)$. For s, $t \in S$, we say that $\omega \in \Omega(\tau(s), \tau(t))$ is *optimal* if $\delta_\tau(s,t) = \omega \cdot \delta_\tau \stackrel{\text{def}}{=} \sum_{u,v \in S} \omega(u,v)\, \delta_\tau(u,v)$. For s, $t \in S$, we denote the set of optimal couplings of $\tau(s)$ and $\tau(t)$ by $\Omega_{\text{opt}}(\tau(s), \tau(t))$. Note that $\Omega_{\text{opt}}(\tau(s), \tau(t))$ is nonempty (see [16, Remark 42]).

4 Robust Probabilistic Bisimilarity

Robust bisimilarity [15] is an equivalence relation that is a subset of bisimilarity, as it also accounts for uncertainty in the transition probabilities. Two states that are robustly bisimilar remain almost behaviourally equivalent under small perturbations of the transition probabilities, that is, their bisimilarity distance remains small.

▶ **Definition 12** ([15, Definition 5])**.** Robust bisimilarity*, denoted $\simeq$, is defined for s, $t \in S$ as $s \simeq t$ if there exists a policy $P \in \mathcal{P}$ such that $\rho^P_\Delta(s,t) = 1$, where $\rho^P_\Delta(s,t)$ is the probability of reaching S^2_Δ from (s,t) in $\langle S \times S, P \rangle$.*

The following theorem states that robust bisimilarity is a sufficient condition for the continuity of the bisimilarity distance function for bisimilar pairs of states.

▶ **Theorem 13** ([15, Theorem 2])**.** *For all s, $t \in S$, if $s \simeq t$ then $s \sim t$, and $\delta\_(s,t) : (S \to \mathcal{D}(S)) \to [0,1]$ is continuous at τ.*

The following theorem provides a converse, closing a gap identified in [15].

▶ **Theorem 14.** *For all s, $t \in S$, if $s \sim t$ then $s \simeq t$ holds if and only if $\delta\_(s,t) : (S \to \mathcal{D}(S)) \to [0,1]$ is continuous at τ.*

We prove Theorem 14 in Section 5.

▶ **Example 15.** Consider the family of labelled Markov chains in Figure 2a. Let $\varepsilon = 0$. The equivalence classes under traditional bisimilarity for τ are displayed in Figure 2b. Among these, the only non-trivial robustly bisimilar sets of states are $\{c_1, c_3\}$ and $\{c_2, c_4\}$. Observe that each of these pairs reach S^2_Δ with probability 1 in $\langle S \times S, P_0 \rangle$, where the relevant part of $P_0 \in \mathcal{P}_\tau$ is illustrated in Figure 2c.

We analyse the continuity of the bisimilarity distance function at τ for three pairs of bisimilar states; see Figure 3 for an illustration. For the robustly bisimilar pair (c_1, c_3), we have $\delta_{\tau_n}(c_1, c_3) = \frac{\varepsilon}{2} = \frac{1}{2n}$ for all $n \in \mathbb{N}$. Note that if n is large then the distance is small and $\lim_n \delta_{\tau_n}(c_1, c_3) = 0 = \delta_\tau(c_1, c_3)$. Thus, $\delta\_(c_1, c_3) : (S \to \mathcal{D}(S)) \to [0, 1]$ is continuous at τ.

On the other hand, the pairs (b_1, b_2) and (c_1, c_4) are bisimilar but not robustly so. Indeed, the pair (b_1, b_2) reaches S^2_Δ with probability only $\frac{1}{2}$ in $\langle S \times S, P_0\rangle$, and no policy can improve this. Due to the ε transition from state d_1 to e_2, we have $\delta_{\tau_n}(d_1, d_2) = 1$ for all $n \in \mathbb{N}$. It follows that $\delta_{\tau_n}(b_1, b_2) = \frac{1}{2}$ for all $n \in \mathbb{N}$. Consequently, the sequence $\delta_{\tau_n}(b_1, b_2)$ does not converge to $\delta_\tau(b_1, b_2) = 0$.

Similarly, the pair (c_1, c_4) cannot reach S^2_Δ under any policy. Due to the perturbation of the outgoing transitions from c_3, we have $\delta_{\tau_n}(c_3, c_4) = 1$ and $\delta_{\tau_n}(e_1, e_2) = 1$ for all $n \in \mathbb{N}$. Hence, $\delta_{\tau_n}(c_1, c_4) = 1$ for all $n \in \mathbb{N}$. Therefore, $\delta\_(b_1, b_2) : (S \to \mathcal{D}(S)) \to [0, 1]$ and $\delta\_(c_1, c_4) : (S \to \mathcal{D}(S)) \to [0, 1]$ are discontinuous at τ. ⌟

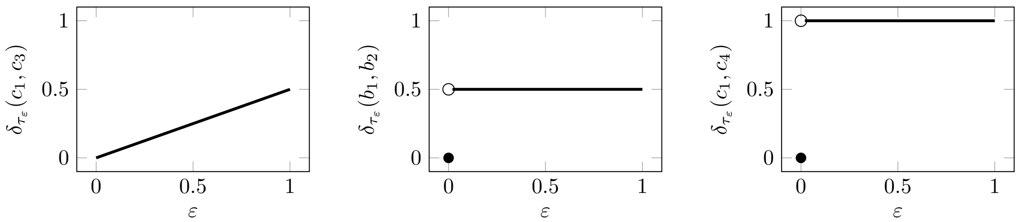


Figure 3 An illustration of the distances between the state pairs (c_1, c_3), (b_1, b_2), and (c_1, c_4) from the labelled Markov chain in Figure 2. All three state pairs are bisimilar, although only (c_1, c_3) is robustly bisimilar. It follows that the distance is continuous at τ for (c_1, c_3) only.

▶ **Theorem 16.** *Continuity for bisimilar state pairs can be decided in polynomial time.*

Proof. Robust bisimilarity can be computed in polynomial time [15]. Then the result follows from Theorem 14. ◀

5 Continuity

In this section, we extend the characterization of the continuity of the bisimilarity distance function to non-bisimilar state pairs.

▶ **Theorem 17.** *Let $s, t \in S$.*

a. *For all $P \in \mathcal{P}$, we have $\rho^P_\Delta(s, t) \le 1 - \delta_\tau(s, t)$.*

b. *The function $\delta\_(s, t) : (S \to \mathcal{D}(S)) \to [0, 1]$ is continuous at τ if and only if there exists a policy $P \in \mathcal{P}$ such that $\rho^P_\Delta(s, t) = 1 - \delta_\tau(s, t)$.*

Proof Sketch. Let $s, t \in S$.

a. Towards a contradiction, assume that there exists a policy $P \in \mathcal{P}$ such that $\rho^P_\Delta(s, t) > 1 - \delta_\tau(s, t)$. One could then modify such a policy so that pairs of states in S^2_Δ never leave S^2_Δ. This would produce a policy $P' \in \mathcal{P}$ where $\rho^{P'}_1(s, t) \le 1 - \rho^P_\Delta(s, t) < \delta_\tau(s, t)$, contradicting Proposition 11.

b. "$\Leftarrow$": The proof of this direction is similar to the one of Theorem 13. Let us give a sketch here.

Let $(\tau_n)_n$ be a sequence converging to τ. Assume that $P \in \mathcal{P}_\tau$ is a policy such that $\rho^P_\Delta(s, t) = 1 - \delta_\tau(s, t)$. Then P is optimal for (s, t), that is, $\rho^P_1(s, t) = \delta_\tau(s, t)$. We construct a graph consisting of the communication classes of $\langle S \times S, P\rangle$ that are reachable from (s,t). Let $P_n \in \mathcal{P}_{\tau_n}$. We then show that for all communication classes C reachable from (s,t) and for all pairs $(u,v) \in C$, it holds that $\lim_n \rho_1^{P_n}(u,v) = \rho_1^P(u,v)$, by induction on the length of a longest path from C in the above-mentioned graph.

By Proposition 11, $\limsup_n \delta_{\tau_n}(s,t) \leq \limsup_n \rho_1^{P_n}(s,t)$. Using the above results, we conclude that $\limsup_n \delta_{\tau_n}(s,t) \leq \rho_1^P(s,t) = \delta_\tau(s,t)$.

"$\Rightarrow$": Assume that for all policies $P \in \mathcal{P}$, we have $\rho_\Delta^P(s,t) < 1 - \delta_\tau(s,t)$.

We construct a sequence $(\tau_n)_n$ converging to τ as follows. Assign each state a unique positive integer *index*. Let $n \in \mathbb{N}$. For each state, move *index*$\cdot\varepsilon_n$ probability from an existing successor to a state with a different label, where $\varepsilon_n > 0$ is a small number. This forces each pair of bisimilar states to transition to S_1^2 with positive probability, thereby destroying $\sim$. See Figure 4a for an example of such a perturbation.

Since the state space is finite, there exists a subsequence $(P_{f(n)})_n$ of optimal policies with $P_{f(n)} \in \mathcal{P}_{\tau_{f(n)}}$ that all share the same support graph. Then the integer indices ensure that there exist a limit policy $P \in \mathcal{P}_\tau$ and a function $D : (S \times S) \to (S \times S) \to \mathbb{Z}$ such that $P_{f(n)} = P + \varepsilon_n D$. A visualisation of these policies appears in Figure 4b. Since $S \times S \to [0,1]$ is compact, the sequence $\left(\rho_1^{P_{f(n)}}\right)_n$ has a converging subsequence $\left(\rho_1^{P_{g(n)}}\right)_n$.

By Proposition 11, it suffices to consider the case where P is optimal. By the assumption above (next to "$\Rightarrow$"), we have $\rho_\Delta^P(s,t) < 1 - \delta_\tau(s,t)$, which then implies that (s,t) can reach a closed communication class C of $\langle S \times S, P\rangle$ with $C \subseteq S_{0,\tau}^2$. Due to the nature of the perturbation, for all $(u,v) \in C$ we have $\rho_1^P(u,v) = 0 < \liminf_n \rho_1^{P_{g(n)}}(u,v)$. Consider a path from (s,t) to C. For each (x,y) on this path, we prove that $\rho_1^P(x,y) < \lim_n \rho_1^{P_{g(n)}}(x,y)$, by induction on the length of the path from (x,y) to C.

Since every $P_{g(n)}$ is optimal, from the above we can conclude that $\delta_\tau(s,t) \leq \rho_1^P(s,t) < \lim_n \rho_1^{P_{g(n)}}(s,t) = \lim_n \delta_{\tau_{\varepsilon_{g(n)}}}(s,t)$. Hence, $\delta\_(s,t) : (S \to \mathcal{D}(S)) \to [0,1]$ is discontinuous at τ.

As an example, consider the bisimilar pair (u,v) in Figure 4b, which forms a closed communication class of $\langle S \times S, P\rangle$. We have $\rho_1^P(u,v) = 0$ and $\liminf_n \rho_1^{P_{g(n)}}(u,v) = \frac{1}{4}$. It follows that $\rho_1^P(s,t) = \frac{1}{4} < \lim_n \rho_1^{P_{g(n)}}(s,t) = \frac{3}{8}$. ◀

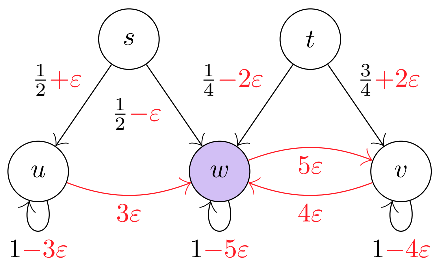
Proof of Theorem 14. By Theorem 13, it is sufficient to show that for all $s,\, t \in S$, if $s \sim t$ and $s \not\simeq t$ then $\delta\_(s,t) : (S \to \mathcal{D}(S)) \to [0,1]$ is discontinuous at τ.

Let $s,\, t \in S$. Assume that $s \sim t$ and $s \not\simeq t$. As $s \sim t$, by Theorem 7, we have $\delta_\tau(s,t) = 0$. Since $s \not\simeq t$, for all policies $P \in \mathcal{P}$, we have $\rho_\Delta^P(s,t) < 1 = 1 - \delta_\tau(s,t)$. Hence, by Theorem 17b, $\delta\_(s,t) : (S \to \mathcal{D}(S)) \to [0,1]$ is discontinuous at τ. ◀

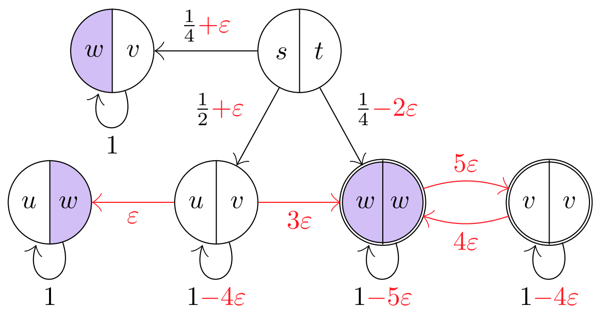
▶ **Example 18.** We revisit our running example, the family of labelled Markov chains in Figure 2a, to investigate how the bisimilarity distance changes when the transition function varies; see Figure 5 for an illustration.

Consider the pair of states (a_1, b_2). The policy $P_0 \in \mathcal{P}_\tau$, partially illustrated in Figure 2c, is optimal for (a_1, b_2). Hence, we have $\delta_\tau(a_1, b_2) = \frac{1}{4}$. Observe that $\rho_\Delta^{P_0}(a_1, b_2) = \frac{1}{4}$ and no policy can improve the probability of reaching S_Δ^2 from (a_1, b_2). Thus, $\delta\_(a_1, b_2) : (S \to \mathcal{D}(S)) \to [0,1]$ is discontinuous at τ. Indeed, we have $\delta_{\tau_n}(a_1, b_2) = \frac{3}{4}$ for all $n \in \mathbb{N}$, so small changes in the transition probabilities cause the bisimilarity distance to jump up by $\frac{1}{2}$.

In contrast, consider the pair of states (a_1, c_1). The policy $P_0 \in \mathcal{P}_\tau$ is also optimal for (a_1, c_1), hence we have $\delta_\tau(a_1, c_1) = \frac{1}{2}$. Since $\rho_\Delta^{P_0}(a_1, c_1) = \frac{1}{2} = 1 - \delta_\tau(a_1, c_1)$, we know that $\delta\_(a_1, c_1) : (S \to \mathcal{D}(S)) \to [0,1]$ is continuous at τ. Note that $\lim_n \delta_{\tau_n}(a_1, c_1) = \lim_n \left(\frac{1}{2} + \frac{1}{4n}\right) = \frac{1}{2}$. Moreover, for all sequences $(\sigma_n)_n$ converging to τ, we have that $\lim_n \delta_{\sigma_n}(a_1, c_1) = \frac{1}{2}$. ⌟



(a) A family of labelled Markov chains for $\varepsilon \in [0, \frac{1}{8}]$. We denote the transition probability function for $\varepsilon = 0$ by τ.



(b) The portion of the Markov chain $\langle S \times S, P_\varepsilon \rangle$ reachable from (s, t), induced by a policy P_ε that is optimal for all $\varepsilon \in [0, \frac{1}{8}]$. We denote the limit policy $P_0 \in \mathcal{P}_\tau$ by P.

Figure 4 An illustration for the proof of Theorem 17.

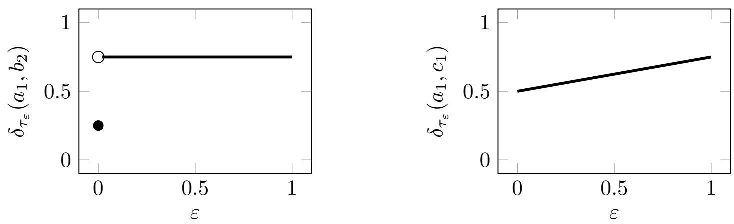


Figure 5 An illustration of the distances between the state pairs (a_1, b_2) and (a_1, c_1) from the labelled Markov chain in Figure 2. The pairs of states have distance $\frac{1}{4}$ and $\frac{1}{2}$, respectively, when $\varepsilon = 0$. For all policies $P \in \mathcal{P}_\tau$, $\rho^P_\Delta(a_1, b_2) \leq \frac{1}{4} < \frac{3}{4} = 1 - \delta_\tau(a_1, b_2)$. In contrast, there exists a policy $P \in \mathcal{P}_\tau$ such that $\rho^P_\Delta(a_1, c_1) = \frac{1}{2} = 1 - \delta_\tau(a_1, c_1)$. Hence, the distance is continuous at τ for (a_1, c_1) only.

Bisimilarity and robust bisimilarity coincide if and only if the bisimilarity distance is continuous for all pairs of states.

▶ **Corollary 19.** $\sim\, = \,\simeq \iff$ *for all* s, $t \in S$, $\delta\_(s, t) : (S \to \mathcal{D}(S)) \to [0, 1]$ *is continuous at* τ.

6 Algorithm for Deciding Continuity

Theorem 16 establishes polynomial time decidability of continuity for bisimilar pairs of states. In this section, we extend this result. We show that one can decide in polynomial time whether $\delta_\tau(s, t)$ is continuous regardless of whether s and t are bisimilar.

To this end, we first introduce two auxiliary functions, which allow us to express the algorithm as a fixed point computation.

Let $R \subseteq S \times S$. We use the notation $\overline{R} = (S \times S) \setminus R$. We define a function $A : 2^{S\times S} \to 2^{S\times S}$ that, intuitively, collects those pairs of states that are incompatible with the relation R. More precisely, for states s and t, the pair (s,t) is included in $A(R)$ either if it is already known to violate robustness, or if every optimal way of matching the transitions of s and t necessarily results in a matched pair that lies outside R. Formally,

$$A(R) = (\sim \setminus \simeq) \cup \{\, (s,t) \in S^2_{?,\tau} \mid \forall \omega \in \Omega_{\mathrm{opt}}(\tau(s), \tau(t)) : \mathrm{support}(\omega) \cap R \neq \varnothing \,\}.$$

Dually, we define a function $B : 2^{S\times S} \to 2^{S\times S}$ that, intuitively, identifies those pairs of states that are compatible with R. That is, (s,t) belongs to $B(R)$ either if it is already known to be robust, or if there exists an optimal way of matching the transitions of s and t such that all matched pairs are contained in R. Formally, $B(R) = \overline{A(\overline{R})}$ or, equivalently,

$$B(R) = \simeq \cup\, S^2_1 \cup \{\, (s,t) \in S^2_{?,\tau} \mid \exists \omega \in \Omega_{\mathrm{opt}}(\tau(s), \tau(t)) : \mathrm{support}(\omega) \subseteq R \,\}.$$

▶ **Proposition 20.** *A and B are monotone with respect to $\subseteq$.*

Therefore, A and B are monotone functions from the complete lattice $2^{S\times S}$ to itself. According to the Knaster-Tarski fixed point theorem, A has a least fixed point, which we denote by α and B has a greatest fixed point, which we denote by β.

The following two propositions state that β is the set of pairs of states for which the distance function is continuous at τ, while α is its complement.

▶ **Proposition 21.** *$\beta = \overline{\alpha}$.*

Proof. First we show that for all $R \subseteq S \times S$, R is a fixed point of A if and only if $\overline{R}$ is a fixed point of B. Let $R \subseteq S \times S$.

$$A(R) = R \iff \overline{A(R)} = \overline{R} \iff \overline{A(\overline{\overline{R}})} = \overline{R} \iff B(\overline{R}) = \overline{R}$$

We now prove the two inclusions. Since β is a fixed point of B, we can conclude that $\overline{\beta}$ is a fixed point of A. Thus, $\alpha \subseteq \overline{\beta}$, as α is the least fixed point of A. It follows that $\beta \subseteq \overline{\alpha}$.

Similarly, since α is a fixed point of A, we can conclude that $\overline{\alpha}$ is a fixed point of B. Thus, $\overline{\alpha} \subseteq \beta$, as β is the greatest fixed point of B. Hence, $\beta = \overline{\alpha}$. ◀

▶ **Proposition 22.** *$\beta = \{\, (s,t) \in S \times S \mid \delta\_(s,t)$ is continuous at $\tau \,\}$.*

Proof Sketch. Let $s, t \in S$. We prove the two implications.

Suppose that $(s,t) \in \beta$. Using the definition of B, we construct a policy P such that $\rho^P_\Delta(s,t) = 1 - \delta_\tau(s,t)$. Thus, by Theorem 17, $\delta\_(s,t)$ is continuous at τ.

Suppose that $\delta\_(s,t)$ is continuous at τ. Then, by Theorem 17, there exists a policy P such that $\rho^P_\Delta(s,t) = 1 - \delta_\tau(s,t)$. Towards a contradiction, assume that $(s,t) \notin \beta$. It follows from Proposition 21 that $(s,t) \in \alpha$. Then, using the definition of A, we show that (s,t) reaches $\sim \setminus \simeq$ in $\langle S \times S, P\rangle$. It follows that $\rho^P_\Delta(s,t) < 1 - \delta_\tau(s,t)$, a contradiction. ◀

We now introduce the algorithm to decide continuity in polynomial time. Let $T \subseteq S \times S$. We use the following notation below: $\mathrm{Pre}(T) = \{\, (s,t) \in S \times S \mid (\mathrm{support}(\tau(s)) \times \mathrm{support}(\tau(t))) \cap T \neq \varnothing \,\}$.

▶ **Proposition 23.** *Algorithm 1 computes the set $D = \alpha$.*

Algorithm 1 Continuity.

Input: A labelled Markov chain $\langle S, L, \tau, \ell \rangle$, the bisimilarity distance function $\delta_\tau : S \times S \to [0,1]$, the set of bisimilar states $\sim$, and the set of robustly bisimilar states $\simeq$

Output: $\{ (s,t) \in S \times S \mid \delta\_(s,t) \text{ is continuous at } \tau \}$

$D_{\text{old}} \leftarrow \varnothing$
$D \leftarrow \sim \setminus \simeq$
repeat
  $D_{\text{old}} \leftarrow D$
  foreach $(s,t) \in (S^2_{?,\tau} \cap \text{Pre}(D_{\text{old}})) \setminus D_{\text{old}}$ **do**
    if $\forall \omega \in \Omega_{\text{opt}}(\tau(s), \tau(t))$ *we have* $\text{support}(\omega) \cap D_{\text{old}} \neq \varnothing$ **then**
      $D \leftarrow D \cup \{(s,t)\}$
    end
  end
until $D = D_{\text{old}}$
return $(S \times S) \setminus D$

Proof Sketch. We show the following loop invariant of Algorithm 1: $D = A(D_{\text{old}}) \subseteq \alpha$.

The loop terminates when a fixed point is reached, therefore, by the loop invariant we know that $D = A(D) \subseteq \alpha$. Thus, D is a fixed point of A less than or equal to α. Since α is the least fixed point of A, we can conclude that Algorithm 1 computes α. ◀

▶ **Theorem 24.** *Algorithm 1 returns the set* $\{ (s,t) \in S \times S \mid \delta\_(s,t) \textit{ is continuous at } \tau \}$.

Proof. This is immediate from Propositions 21–23. ◀

In the remainder of this section, we show that Algorithm 1 runs in polynomial time. Note that $\sim$, $\simeq$, and δ_τ can be computed in polynomial time [4, 15].

▶ **Proposition 25.** *For all s, $t \in S$ and $X \subseteq S \times S$, we can determine in polynomial time whether there exists $\omega \in \Omega_{\text{opt}}(\tau(s), \tau(t))$ such that* $\text{support}(\omega) \subseteq X$.

Proof Sketch. We reduce the problem to a minimum cost maximum flow instance [17].

We construct the network, with source s and sink t, as follows. For each $x \in S$, we add two vertices $(x,0)$ and $(x,1)$, an edge from s to $(x,0)$ with capacity $\tau(s)(x)$, and an edge from $(x,1)$ to t with capacity $\tau(t)(x)$. These edges have cost 0. For each $(x,y) \in X$, we add an edge from $(x,0)$ to $(y,1)$ with capacity 1 and cost $\delta_\tau(x,y)$.

Let $f : E \to \mathbb{R}$ denote the flow on the edges. The value of the flow, $|f|$ is defined as the net flow entering the sink, or equivalently, the net flow exiting the source. The cost of the flow is defined as the sum of the cost per unit flow over all edges. We can use standard network flow algorithms to solve the problem. For example, Dinic's maximum flow algorithm maximizes the value of the flow in polynomial time [10] and the cycle-cancelling algorithm then minimizes the cost of the flow, without changing the value of the flow [25], in polynomial time [19]. Note that the problem can also be solved in polynomial time using linear programming.

We show that there exists $\omega \in \Omega_{\text{opt}}(\tau(s), \tau(t))$ such that $\text{support}(\omega) \subseteq X$ if an only if the resulting value of the flow is 1 and the cost of the flow is $\delta_\tau(s,t)$. Intuitively, a flow of value 1 corresponds exactly to a coupling $\omega \in \Omega(\tau(s), \tau(t))$, with $\omega(x,y) = f((x,0),(y,1))$, and the construction of the network ensures that $\text{support}(\omega) \subseteq X$. Conversely, any such coupling defines a feasible flow of value 1. It follows that the cost of the flow equals $\sum_{x \in S} \sum_{y \in S} \delta_\tau(x, y)\, \omega(x, y)$. Hence, the cost of the flow is $\delta_\tau(s, t)$ if and only if ω is an optimal coupling. ◀

▶ **Proposition 26.** *Algorithm 1 runs in polynomial time.*

Proof. Let $|S| = m$. Since at least one pair of states is added to D in each iteration, the algorithm requires at most m^2 iterations. Furthermore, during each iteration, at most m^2 pairs of states are checked in the inner loop.

The check on line 6 is equivalent to determining whether there exists $\omega \in \Omega_{\mathrm{opt}}(\tau(s), \tau(t))$ such that $\mathrm{support}(\omega) \subseteq \overline{D_{\mathrm{old}}}$. By Proposition 25, this can be done in polynomial time. ◀

7 Experiments

To evaluate the efficiency and practical usefulness of our algorithm to decide continuity, we implemented it in the widely used probabilistic model checker PRISM [27], an open-source tool for quantitative verification and analysis of probabilistic models, including labelled Markov chains.

7.1 Implementation

Our algorithm requires as input the set of bisimilar states $\sim$, the set of robustly bisimilar states $\simeq$, and the bisimilarity distance function $\delta_\tau : S \times S \to [0, 1]$. We first briefly mention how they are computed.

PRISM's implementation of the traditional (i.e., non-robust) bisimilarity algorithm is a standard partition-refinement approach which uses the signature-based method of Derisavi [7]. The initial partition is based on the labelling of the states. Let Π be the current partition and E_Π be the set of equivalence classes in Π. Then the new partition is computed as $\{ (s, t) \in \Pi \mid \forall B \in E_\Pi : \tau(s)(B) = \tau(t)(B) \}$. Each state and equivalence class (referred to as a block) is represented by an integer ID. The current partition of the state space is tracked by an array that is indexed by state IDs and contains the corresponding block IDs.

PRISM's implementation of the robust bisimilarity algorithm was described in [15]. Given a policy $P \in \mathcal{P}$, a set $R \subseteq S \times S$ *supports* a path $(u_1, v_1) \dots (u_n, v_n)$ in $\langle S \times S, P \rangle$ if for all $1 \le i \le n$ we have $(u_i, v_i) \in R$ and $\mathrm{support}(P(u_i, v_i)) \subseteq R$. Let Π be the current partition. Then the new partition is the largest bisimulation such that for every $(s, t) \in \Pi$ there exists a policy P such that Π supports a path from (s, t) to S^2_Δ in $\langle S \times S, P \rangle$.

An algorithm to compute the bisimilarity distance, based on Condon's simple policy iteration algorithm [5], was proposed in [1, 35, 37]. Distance zero (i.e. bisimilarity) and one are computed first. Then, starting from an arbitrary vertex policy, the algorithm repeatedly improves the policy with a locally optimal vertex coupling, until no further local improvement is possible. For the bisimilarity distance, Tang and van Breugel [36] showed that, in practice, policy iteration approaches are more efficient than the theoretically superior linear programming algorithm. Hence, we adapt an implementation of the aforementioned algorithm, provided in [34]. We rely on Google's OR-Tools [32] to compute optimal couplings. The OR-Tools suite is an open source software for combinatorial optimization and includes solvers for linear programming.

We implemented Algorithm 1 in Java as part of PRISM's explicit-state model checking engine.[1] The transition probabilities of the labelled Markov chain are stored as a two-dimensional array and the set D of discontinuous pairs is represented as a flat boolean array where a state pair (s, t) is indexed by $s \cdot |S| + t$. Throughout the algorithm, we exploit the symmetry of D by setting both (s, t) and (t, s) simultaneously. It is initialised to $\sim \setminus \simeq$, that is, those pairs of states that are bisimilar but not robustly so.

The main loop iterates over all pairs $(s, t) \notin D$ with non-zero distance and matching labels, restricting attention to those that are predecessors of D. For each such candidate pair, we attempt to find an optimal coupling of $\tau(s)$ and $\tau(t)$ whose support lies entirely outside the known discontinuous pairs D, as described in the proof of Proposition 25, using Google's OR-Tools. If no such coupling exists, the pair is added to D. This process continues until no new pairs are added to D, at which point a fixed point has been reached.

7.2 Experimental Setup

All experiments were run on a MacBook with an M1 chip and 16GB memory.

We evaluated our algorithm by applying it to all (discrete-time) labelled Markov chains from the Quantitative Verification Benchmark Set (QVBS) [20] for which the bisimilarity distance algorithm could complete within the available memory. QVBS is a comprehensive collection of probabilistic models which is designed as a benchmark suite for quantitative verification and analysis tools and is the foundation of the Quantitative Verification Competition (QComp), which compares the performance, versatility, and usability of such tools.

For an additional source of models, we also use jpf-probabilistic [13]. Java PathFinder (JPF) [38] is a popular model checker for Java code, and the JPF extension jpf-probabilistic provides Java implementations of sixty randomized algorithms [13]. As shown in [13], JPF, extended by jpf-probabilistic and jpf-label [12], can be used in tandem with PRISM to check properties of these algorithms and supplement JPF's qualitative results with quantitative information. A description of the subset of these algorithms utilized in our study is provided in [16, Appendix H].

7.3 Results

Table 1 reports the results for benchmarks where the minimized models obtained via traditional bisimilarity and robust bisimilarity differ. These are of particular interest because, according to Corollary 19, they are instances where our algorithm identifies pairs of states for which the distance is discontinuous. Note that the property used for each benchmark determines the labelling of the model. In the table, *Time* denotes the runtime (in seconds) and *%* denotes the percentage of state pairs for which the distance is continuous. The percentage of continuous pairs with distance less than one is given in parentheses.

In terms of the size of the state space, we are limited by what the bisimilarity distance algorithm can handle, constrained by 16GB of memory. This is similar to the model sizes considered in [36]. Nevertheless, the results are promising: the additional cost of deciding continuity is, on all but one benchmark, less than that for computing the bisimilarity distance and in some cases substantially so, demonstrating that our approach is feasible in practice. Moreover, our results suggest that when the bisimilarity distance is continuous for most state pairs, the continuity check is particularly efficient.

[1] `https://github.com/zainabfatmi/prism/blob/concur/prism/src/explicit/bisim/distances/Continuity.java`

Table 1 The results of the benchmarks for which continuity is non-trivial.

Benchmark			Distance	Continuity	
Name (Property)	Parameters	States	Time (s)	Time (s)	%
brp (p1)	N=2, M=2	89	17.276	16.164	34.6 (6.5)
	N=2, M=3	116	58.826	78.724	27.5 (5.0)
	N=2, M=4	143	156.344	252.931	22.9 (4.2)
brp (p4)	N=2, M=2	89	2.396	2.042	15.8 (3.3)
	N=2, M=3	116	7.987	7.834	12.2 (2.4)
	N=2, M=4	143	20.332	19.645	9.9 (1.9)
	N=2, M=5	170	40.691	38.109	8.4 (1.6)
	N=4, M=2	173	21.533	18.225	8.3 (1.7)
crowds (positive)	R=2, S=2	77	126.911	8.450	16.7 (4.9)
oscillators (power)	T=6, N=3	57	5.059	0.002	94.6 (82.2)
oscillators (time)	T=6, N=3	57	4.687	0.002	94.6 (82.2)

Table 2 The results of some benchmarks for which continuity is trivial.

Benchmark			Distance		Continuity
Name (Property)	Parameters	States	Time (s)	Trivial	Time (s)
fair biased coin (heads)	p=0.9	13	0.161	no	0.000
herman (steps)	N=5	32	7.015	no	0.001
leader-sync (elected)	N=3, K=4	147	0.988	yes	0.004
majority element (incorrect)	s=50, t=5	101	26.608	no	0.011
haddad-monmege (target)	N=20, p=0.7	41	312.192	no	0.001
pollards factorization (input)	i=4004	5	0.000	yes	0.000
queens (success)	n=4	16	0.001	yes	0.000

Table 2 presents the largest feasible model per benchmark where the minimized models obtained by traditional bisimilarity and robust bisimilarity coincide. In these cases, by Corollary 19, the distance is continuous for all pairs of states and our algorithm terminates almost immediately. The column *Trivial* denotes whether the distance is an integer, that is 0 or 1, for all state pairs. Evidently, this does not affect the time to decide continuity.

Across all benchmarks, the experiments demonstrate that the additional computation for deciding continuity is comparable to or faster than calculating the bisimilarity distance. Hence the extra step of deciding continuity is reasonable in practice, while providing additional robustness guarantees. As discussed in the introduction, in applications where transition probabilities are subject to approximation or uncertainty, continuity guarantees that sufficiently small errors in these values lead only to small changes in the bisimilarity distance. Conversely, discontinuity identifies situations in which arbitrarily small perturbations may cause a non-negligible change in the distance, highlighting that conclusions drawn from the computed distance should be interpreted with caution.

8 Conclusions and Future Work

The concept of robust bisimilarity was proposed in [15] to address the lack of robustness of bisimilarity, where it was shown to be a sufficient condition for continuity. In this paper, we have strengthened that result by proving that robust bisimilarity is also a necessary condition for continuity. We have completed the theoretical characterization of robustness in Theorem 17. Moreover, we have shown that continuity is decidable in polynomial time for all pairs of states and we have extended the experimental study done in [15], with particular emphasis on the quantitative setting.

As future work, we aim to provide a bound on how much the distance can increase to, under small perturbations of the transition probabilities. Additionally, we plan to investigate the continuity of the distance function when we restrict to perturbations that preserve the underlying graph structure, as this is often known, even when the transition probabilities are subject to approximation.

References

1. Giorgio Bacci, Giovanni Bacci, Kim G. Larsen, and Radu Mardare. On-the-fly exact computation of bisimilarity distances. In Nir Piterman and Scott A. Smolka, editors, *Proceedings of the 19th International Conference on Tools and Algorithms for the Construction and Analysis of Systems*, volume 7795 of *Lecture Notes in Computer Science*, pages 1–15, Rome, Italy, March 2013. Springer. doi:10.1007/978-3-642-36742-7_1.
2. Christel Baier and Joost-Pieter Katoen. *Principles of model checking*. The MIT Press, Cambridge, MA, USA, 2008.
3. Krishnendu Chatterjee, Luca de Alfaro, Rupak Majumdar, and Vishwanath Raman. Algorithms for game metrics (full version). *Logical Methods in Computer Science*, 6(3):1–27, 2010. doi:10.48550/arXiv.0809.4326.
4. Di Chen, Franck van Breugel, and James Worrell. On the complexity of computing probabilistic bisimilarity. In Lars Birkedal, editor, *Proceedings of the 15th International Conference on Foundations of Software Science and Computational Structures*, volume 7213 of *Lecture Notes in Computer Science*, pages 437–451, Tallinn, Estonia, March/April 2012. Springer-Verlag. doi:10.1007/978-3-642-28729-9_29.
5. Anne Condon. On algorithms for simple stochastic games. In Jin-Yi Cai, editor, *Advances In Computational Complexity Theory*, volume 13 of *Series in Discrete Mathematics and Theoretical Computer Science*, pages 51–71, New Jersey, USA, December 1990. DIMACS/AMS. doi:10.1090/DIMACS/013/04.
6. Luca de Alfaro and Rupak Majumdar. Quantitative solution of omega-regular games. In Jeffrey Scott Vitter, Paul G. Spirakis, and Mihalis Yannakakis, editors, *Proceedings of the 33rd Annual Symposium on Theory of Computing*, pages 675–683, Heraklion, Crete, Greece, July 2001. ACM. doi:10.1145/380752.380871.
7. Salem Derisavi. Signature-based symbolic algorithm for optimal Markov chain lumping. In *Proceedings of the 4th International Conference on the Quantitative Evaluation of Systems*, pages 141–150, Edinburgh, Scotland, UK, September 2007. IEEE Computer Society. doi:10.1109/QEST.2007.27.
8. Josée Desharnais, Vineet Gupta, Radha Jagadeesan, and Prakash Panangaden. Metrics for labeled Markov systems. In Jos C. M. Baeten and Sjouke Mauw, editors, *Proceedings of the 10th International Conference on Concurrency Theory*, volume 1664 of *Lecture Notes in Computer Science*, pages 258–273, Eindhoven, The Netherlands, August 1999. Springer-Verlag. doi:10.1007/3-540-48320-9_19.
9. Josée Desharnais, Vineet Gupta, Radha Jagadeesan, and Prakash Panangaden. Metrics for labelled Markov processes. *Theoretical Computer Science*, 318(3):323–354, June 2004. doi:10.1016/J.TCS.2003.09.013.

10 Yefim Dinitz. Dinitz' algorithm: the original version and Even's version. In Arnold L. Rosenberg Oded Goldreich and Alan L. Selman, editors, *Theoretical Computer Science: Essays in Memory of Shimon Even*, pages 218–240. Springer-Verlag, Berlin, Germany, 2006. doi:10.1007/11685654_10.

11 J. Ronald Eastman and Jiena He. A regression-based procedure for Markov transition probability estimation in land change modeling. *Land*, 9(11), October 2020. doi:10.3390/land9110407.

12 Syyeda Zainab Fatmi. Probabilistic model checking of randomized Java code. Master's thesis, York University, Toronto, Canada, September 2020. URL: http://hdl.handle.net/10315/38445.

13 Syyeda Zainab Fatmi, Xiang Chen, Yash Dhamija, Maeve Wildes, Qiyi Tang, and Franck van Breugel. Probabilistic model checking of randomized Java code. In Alfons Laarman and Ana Sokolova, editors, *Proceedings of the 27th International Symposium on Model Checking Software, SPIN*, volume 12864 of *Lecture Notes in Computer Science*, pages 157–174. Springer, July 2021. doi:10.1007/978-3-030-84629-9_9.

14 Syyeda Zainab Fatmi, Stefan Kiefer, David Parker, and Franck van Breugel. PRISM extension for deciding continuity of the probabilistic bisimilarity distance. Software, swhId: swh:1:dir:4efb0e4307476e883191835aed030dce0af9782c (visited on 2026-08-12). URL: https://github.com/zainabfatmi/prism/tree/concur, doi:10.4230/artifacts.27642.

15 Syyeda Zainab Fatmi, Stefan Kiefer, David Parker, and Franck van Breugel. Robust probabilistic bisimilarity for labelled Markov chains. In Ruzica Piskac and Zvonimir Rakamaric, editors, *Proceedings of the 37th International Conference on Computer Aided Verification*, volume 15932 of *Lecture Notes in Computer Science*, pages 254–275, Zagreb, Croatia, July 2025. Springer-Verlag. doi:10.1007/978-3-031-98679-6_12.

16 Syyeda Zainab Fatmi, Stefan Kiefer, David Parker, and Franck van Breugel. On the continuity of the probabilistic bisimilarity distance (full version). *CoRR*, June 2026. arXiv:2606.27209.

17 Lester Randolph Ford and Delbert Ray Fulkerson. *Flows in networks*. Princeton University Press, Princeton, NJ, USA, 1962.

18 Alessandro Giacalone, Chi-Chang Jou, and Scott A. Smolka. Algebraic reasoning for probabilistic concurrent systems. In Manfred Broy and Cliff B. Jones, editors, *Proceedings of the Working Conference on Programming Concepts and Methods*, pages 443–458, Sea of Galilee, Israel, April 1990. North-Holland.

19 Andrew V. Goldberg and Robert Endre Tarjan. Finding minimum-cost circulations by canceling negative cycles. *Journal of the ACM*, 36(4):873–886, October 1989. doi:10.1145/76359.76368.

20 Arnd Hartmanns, Michaela Klauck, David Parker, Tim Quatmann, and Enno Ruijters. The quantitative verification benchmark set. In Tomás Vojnar and Lijun Zhang, editors, *Proceedings of the 25th International Conference on Tools and Algorithms for the Construction and Analysis of Systems*, volume 11427 of *Lecture Notes in Computer Science*, pages 344–350, Prague, Czech Republic, April 2019. Springer. doi:10.1007/978-3-030-17462-0_20.

21 Manfred Jaeger, Hua Mao, Kim Guldstrand Larsen, and Radu Mardare. Continuity properties of distances for Markov processes. In Gethin Norman and William H. Sanders, editors, *Proceedings of the 11th International Conference on Quantitative Evaluation of Systems*, volume 8657 of *Lecture Notes in Computer Science*, pages 297–312, Florence, Italy, September 2014. Springer-Verlag. doi:10.1007/978-3-319-10696-0_24.

22 Bengt Jonsson and Kim Larsen. Specification and refinement of probabilistic processes. In *Proceedings of the 6th Annual Symposium on Logic in Computer Science*, pages 266–277, Amsterdam, The Netherlands, July 1991. IEEE. doi:10.1109/LICS.1991.151651.

23 Joost-Pieter Katoen, Tim Kemna, Ivan S. Zapreev, and David N. Jansen. Bisimulation minimisation mostly speeds up probabilistic model checking. In *Proceedings of the 13th International Conference on Tools and Algorithms for the Construction and Analysis of Systems*, volume 4424 of *Lecture Notes in Computer Science*, pages 87–101, Braga, Portugal, March 2007. Springer. doi:10.1007/978-3-540-71209-1_9.

24 John G. Kemeny and J. Laurie Snell. *Finite Markov chains*. Springer-Verlag, Heidelberg, Germany, 1960.

25 Morton Klein. A primal method for minimal cost flows with applications to the assignment and transportation problems. *Management Science*, 14(3):205–220, November 1967. doi:10.1287/mnsc.14.3.205.

26 Dexter Kozen. A probabilistic PDL. In David S. Johnson, Ronald Fagin, Michael L. Fredman, David Harel, Richard M. Karp, Nancy A. Lynch, Christos H. Papadimitriou, Ronald L. Rivest, Walter L. Ruzzo, and Joel I. Seiferas, editors, *Proceedings of the 15th Annual Symposium on Theory of Computing*, pages 291–297, Boston, Massachusetts, USA, April 1983. ACM. doi:10.1145/800061.808758.

27 Marta Kwiatkowska, Gethin Norman, and David Parker. PRISM 4.0: Verification of probabilistic real-time systems. In Ganesh Gopalakrishnan and Shaz Qadeer, editors, *Proceedings of the 23rd International Conference on Computer Aided Verification*, volume 6806 of *Lecture Notes in Computer Science*, pages 585–591, Snowbird, Utah, USA, July 2011. Springer-Verlag. doi:10.1007/978-3-642-22110-1_47.

28 Kim Larsen and Arne Skou. Bisimulation through probabilistic testing. In *Proceedings of the 16th Annual ACM Symposium on Principles of Programming Languages*, pages 344–352, Austin, TX, USA, January 1989. ACM. doi:10.1145/75277.75307.

29 Annabelle McIver and Carroll Morgan. *Abstraction, Refinement and Proof for Probabilistic Systems*. Monographs in Computer Science. Springer, New York, NY, USA, 2004. doi:10.1007/B138392.

30 Daijiro Mizutani, Nam Lethanh, Bryan T. Adey, and Kiyoyuki Kaito. Improving the estimation of Markov transition probabilities using mechanistic-empirical models. *Frontiers in Built Environment*, 3:58, October 2017. doi:10.3389/fbuil.2017.00058.

31 Elena Olariu, Kevin K. Cadwell, Elizabeth Hancock, David Trueman, and Helene Chevrou-Severac. Current recommendations on the estimation of transition probabilities in Markov cohort models for use in health care decision-making: a targeted literature review. *ClinicoEconomics and Outcomes Research*, 9:537–546, September 2017. doi:10.2147/CEOR.S135445.

32 Laurent Perron and Vincent Furnon. OR-Tools. URL: https://developers.google.com/optimization/.

33 Tushar Srivastava, Nicholas R. Latimer, and Paul Tappenden. Estimation of transition probabilities for state-transition models: a review of NICE appraisals. *PharmacoEconomics*, 39(8):869–878, August 2021. doi:10.1007/s40273-021-01034-5.

34 Qiyi Tang. *Computing probabilistic bisimilarity distances*. PhD thesis, York University, Toronto, Canada, August 2018. URL: http://hdl.handle.net/10315/35586.

35 Qiyi Tang and Franck van Breugel. Computing probabilistic bisimilarity distances via policy iteration. In Josée Desharnais and Radha Jagadeesan, editors, *Proceedings of the 27th International Conference on Concurrency Theory*, volume 59 of *LIPIcs*, pages 22:1–22:15, Québec City, Canada, August 2016. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.CONCUR.2016.22.

36 Qiyi Tang and Franck van Breugel. Algorithms to compute probabilistic bisimilarity distances for labelled Markov chains. In Roland Meyer and Uwe Nestmann, editors, *Proceedings of the 28th International Conference on Concurrency Theory*, volume 85 of *LIPIcs*, pages 27:1–27:16, Berlin, Germany, September 2017. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.CONCUR.2017.27.

37 Qiyi Tang and Franck van Breugel. Deciding probabilistic bisimilarity distance one for labelled Markov chains. In Hana Chockler and Georg Weissenbacher, editors, *Proceedings of the 30th International Conference on Computer Aided Verification*, Lecture Notes in Computer Science, pages 681–699, Oxford, UK, July 2018. Springer. doi:10.1007/978-3-319-96145-3_39.

38 Willem Visser, Klaus Havelund, Guillaume Brat, Seungjoon Park, and Flavio Lerda. Model checking programs. *Automated Software Engineering*, 10(2):203–232, April 2003. doi:10.1023/A:1022920129859.