Under consideration for publication in Math. Struct. in Comp. Science

Converging from Branching to Linear Metrics on Markov Chains†

GIORGIO BACCI, GIOVANNI BACCI, KIM G. LARSEN,

and RADU MARDARE

Department of Computer Science, Aalborg University, DK.
Email: {grbacci,giovbacci,kgl,mardare}@cs.aau.dk

Received 17 April 2017

We study two well known linear-time metrics on Markov chains (MCs), namely, the strong and strutter trace distances. Our interest in these metrics is motivated by their relation to the probabilistic LTL-model checking problem: we prove that they correspond to the maximal differences in the probability of satisfying the same LTL and $\text{LTL}^{-\text{x}}$ (LTL without next operator) formulas, respectively.
The threshold problem for these distances (whether their value exceeds a given threshold) is NP-hard and not known to be decidable. Nevertheless, we provide an approximation schema where each lower and upper-approximant is computable in polynomial time in the size of the MC.
The upper-approximants are bisimilarity-like pseudometrics (hence, branching-time distances) that converge point-wise to the linear-time metrics. This convergence is interesting in itself, because it reveals a nontrivial relation between branching and linear-time metric-based semantics that does not hold in equivalence-based semantics.

1. Introduction

The growing interest in quantitative systems, e.g. probabilistic and real-time systems, motivated the introduction of new techniques for studying their operational semantics. For the comparison of the behavior of quantitative systems, metrics are preferred to equivalences since the latter are not robust with respect to small variations of the numerical values. Such metrics are said *behavioral* if they generalize the concept of behavioral equivalence (e.g., bisimilarity or trace equivalence) by measuring the dissimilarities of two states in terms of their operational behavior.

Several of the behavioral distances that have been proposed in the literature, e.g.

† Work supported by the European Union 7th Framework Programme (FP7/2007-2013) under Grants Agreement nr. 318490 (SENSATION), nr. 601148 (CASSTING) and by the Sino-Danish Basic Research Center IDEA4CPS funded by the Danish National Research Foundation and the National Science Foundation China.

(DGJP99; DJGP02; dAFS04; Mio14; FPK14), come with logical characterizations according to the general schema $d(u,v) = \sup_{\varphi \in \Phi} |\langle\varphi\rangle(u) - \langle\varphi\rangle(v)|$, where Φ is a suitable set of logical properties of interest over which the states u, v are compared, and $\langle\varphi\rangle(u) \in \mathbb{R}$ denotes the value of the formula φ at state u. The interest in such logical descriptions is two fold. The first reason is purely theoretical, as it accounts for the expressiveness of the metric in terms of a set Φ of logical properties. The second one is that such characterization relates the metric to the verification problem over logical specifications.

In this paper we focus our attention on probabilistic systems, namely, Markov chains (MCs), and we consider two behavioral metrics on them: the *strong* and *stutter trace metrics*, respectively lifting the strong and stutter probabilistic trace equivalences to distances. As a first main result we show, via a logical characterization, how these metrics are related to the problem of model checking LTL formulas against MCs, a.k.a. the *probabilistic* LTL*-model checking problem* (Var85; CY90; Var99).

The model checking problem against non-probabilistic systems checks whether a logical formula φ is satisfied by all the execution runs from a certain state u. For probabilistic systems the same problem amounts to asking what is the *probability* that an execution run from u satisfies φ. By denoting this probability by $\mathbb{P}(u)(\llbracket\varphi\rrbracket) \in [0,1]$, we show that the strong trace distance, denoted by δ_t, is related to the LTL-probabilistic model checking problem as follows:

$$\delta_t(u,v) = \sup_{\varphi \in \mathrm{LTL}} |\mathbb{P}(u)(\llbracket\varphi\rrbracket) - \mathbb{P}(v)(\llbracket\varphi\rrbracket)| \,.$$

An immediate application of this result is that one may turn any probabilistic verification problem at u into one at a state v, ensuring that (i) the difference in the results obtained by doing probabilistic verification on v rather than u is always bounded by $\delta_t(u,v)$ for any LTL-formula; and (ii) that $\delta_t(u,v)$ is the least upper-bound for this difference.

A similar logical characterization is shown to hold for the stutter trace distance too with the key difference that now logical formulas are from sub-fragment of LTL without next operator. As observed by Lamport (Lam83), program specifications using the next operator are too specific to the actual implementation of programs, rendering them unusable for any practical verification purpose. This is particularly evident for systems that cannot be internally accessed, as it is typical e.g. in systems biology and machine learning. This motivates our interest in a stutter-variant for the trace distance.

The above logical characterizations encourage for the study of efficient methods for computing trace distances. Unfortunately, in (LP02; CK14) the threshold problem for the trace distance is proven to be NP-hard and, to the best of our knowledge, its decidability is still an open problem. Nevertheless, in (CK14) it is shown that the problem of approximating this distance with arbitrary precision is decidable. This is done by providing two effective methods to converge from below and above to the trace distance. In this paper we provide an alternative approximation schema that, differently from (CK14), is formed by sequences of lower and upper-approximants that are shown to be computable in *polynomial time* in the size of the MC. With respect to (CK14), our approach is more general with the nice consequence that one obtains a similar approximating schema also for the stutter trace distance.

Notably, in our construction the upper-approximants are bisimilarity-like pseudometrics, i.e., branching-time distances. We show that these metrics form a net —the topological generalization of infinite sequences that use directed poset as set of indices instead of the usual increasing chain of natural numbers— that converges point-wise to the linear-time metrics. The result is interesting in itself, since it relates branching- and linear-time metric-based semantics by means of a topological limit argument. It is worth noticing that such a limit argument does occur between equivalence-based semantics (see Remark 5). This opens new perspectives in the study of quantitative systems, and suggests that relating linear- and branching-time distances by means of converging nets may lead to new ways to cope with the decision problem of computing linear-time metrics.

The technical contributions of the paper can be summarized as follows.

(i) We provide a logical characterization of the trace distance in terms of LTL. This result, differently from previous proposals (e.g. (dAFS04; DGJP99)), explicitly relates the trace distance to the probabilistic model checking problem of LTL formulas. We show that a similar characterization holds also for the stutter trace distance on the fragment of LTL without next operator.

(ii) We construct two nets of bisimilarity-like distances that converge to the strong and stutter trace distance. This construction leverages a classical duality result that characterizes the total variation distance between two measures as the minimal discrepancy associated with their couplings. To do so we generalize and improve the applicability of two results in (CvBW12), namely Theorem 8 and Corollary 11.

(iii) We demonstrate that each element of the proposed converging nets is computable in polynomial time in the size of the MC. Moreover, we provide an other pair of sequences of lower approximant pseudometrics that converge from below to the strong and stutter trace distances, respectively. Also these lower approximants are proven to be polynomially computable. The pairs of converging sequences of upper and lower approximants form the approximation schemata for the problem of computing the strong and stutter trace distances. This approximation technique for the trace distance improves the one proposed in (CK14).

2. Preliminaries and Notation

The set of functions from X to Y is denoted by Y^X. Any preorder $\sqsubseteq$ on Y is extended to Y^X as $f \sqsubseteq g$ iff $f(x) \sqsubseteq g(x)$, for all $x \in X$. For $f \in Y^X$, let $\equiv_f = \{(x, x') \mid f(x) = f(x')\}$. For an equivalence relation $R \subseteq X \times X$, let $X/_R$ denote the quotient set, $[x]_R$ denote the R-equivalence class of x, and for $A \subseteq X$, let $[A]_R = \bigcup_{x \in A} [x]_R$.

Measure theory. A *field* over a set X is a nonempty family $\Sigma \subseteq 2^X$ closed under complement and finite union. Σ is a σ-algebra if, in addition, it is closed under countable union; in this case (X, Σ) is called a *measurable space* and the elements of Σ *measurable sets*. For $\mathcal{F} \subseteq 2^X$, $\sigma(\mathcal{F})$ denotes the smallest σ-algebra containing $\mathcal{F}$. For (X, Σ),(Y, Θ) measurable spaces, $f \colon X \to Y$ is *measurable* if for all $E \in \Theta$, $f^{-1}(E) = \{x \mid f(x) \in E\} \in \Sigma$. The *product space*, $(X, \Sigma) \otimes (Y, \Theta)$, is the measurable space $(X \times Y, \Sigma \otimes \Theta)$, where $\Sigma \otimes \Theta$ is the σ-algebra generated by the *rectangles* $E \times F$, for $E \in \Sigma$ and $F \in \Theta$. A *measure* on (X, Σ) is a σ-additive function $\mu \colon \Sigma \to \mathbb{R}_+$, i.e., such that $\mu(\bigcup_{i \in \mathbb{N}} E_i) = \sum_{i \in \mathbb{N}} \mu(E_i)$ for all of pairwise disjoint $E_i \in \Sigma$; it is a *probability measure* if, in addition, $\mu(X) = 1$. Hereafter $\Delta(X, \Sigma)$ denotes the set of probability measures on (X, Σ); and $\Delta(X)$ the set of (discrete) probability measures on $(X, 2^X)$. Given a measurable function $f: (X, \Sigma) \to (Y, \Theta)$ and measure $\mu \in \Delta(X, \Sigma)$, define *push forward of* μ *under* f as the measure $\mu[f] \in \Delta(Y, \Theta)$ given by $\mu[f](E) = \mu(f^{-1}(E))$, for all $E \in \Theta$.

Given μ and ν measures on (X, Σ) and (Y, Θ), respectively, the *product measure* $\mu \times \nu$ on $(X, \Sigma) \otimes (Y, \Theta)$ is *uniquely* defined by $(\mu \times \nu)(E \times F) = \mu(E) \cdot \nu(E)$, for all $(E, F) \in \Sigma \times \Theta$.

A measure ω on $(X, \Sigma) \otimes (Y, \Theta)$ is a *coupling* for the pair (μ, ν) if for all $E \in \Sigma$ and $F \in \Theta$, $\omega(E \times Y) = \mu(E)$ and $\omega(X \times F) = \nu(F)$ (i.e., when μ and ν are, respectively, is the *left* and *right marginal* of the measure ω). The set of couplings for a pair (μ, ν) is denoted by $\Omega(\mu, \nu)$. Note that the product measure $\mu \times \nu$ is a coupling for (μ, ν).

Metric spaces. For a set X, $d: X \times X \to \mathbb{R}_+$ is a *pseudometric* on X if for any $x, y, z \in X$, $d(x, x) = 0$, $d(x, y) = d(y, x)$ and $d(x, y) + d(y, z) \geq d(x, z)$; d is a *metric* if, in addition, $d(x, y) = 0$ implies $x = y$. If d is a (pseudo)metric on X, (X, d) is called a *(pseudo)metric space.* A (pseudo)metric space is *complete* if all Cauchy sequences converge; and is *separable* if it has countable dense subset. Define $\ker(d) = \{(u, v) \mid d(u, v) = 0\}$; this set will be called the *kernel of* d.

For (X, Σ) a measurable space, the set of probability measures $\Delta(X, \Sigma)$ can be equipped with the *total variation distance* $\|\mu - \nu\| = \sup_{E \in \Sigma} |\mu(E) - \nu(E)|$. If (X, d) is a separable and complete (pseudo)metric space one can turn $\Delta(X, \Sigma_d)$ into a (pseudo)metric space by using the *Kantorovich (pseudo-)metric* $\mathcal{K}(d)(\mu, \nu) = \min \left\{ \int d \, \mathrm{d}\omega \mid \omega \in \Omega(\mu, \nu) \right\}$, where Σ_d is the Borel σ-algebra induced by the (pseudo)metric d.

The space of words. Let X^n be the set of words on X of length $n \in \mathbb{N}$, $X^* = \bigcup_{n \in \mathbb{N}} X^n$, $AB = \{ab \in X^* \mid a \in A, b \in B\}$ $(A, B \subseteq X^*)$ and $X^+ = XX^*$.

An infinite word $\pi = x_0 x_1 \dots$ over X is an element in X^ω. For $i \in \mathbb{N}$, define $\pi[i] = x_i$, $\pi|^i = x_0 \dots x_{i-1} \in X^i$, and $\pi|_i = x_i x_{i+1} \dots \in X^\omega$. For $A \subseteq X^n$, the *cylinder set for* A (of rank n) is defined as $\mathfrak{C}(A) = \{\pi \in X^\omega \mid \pi|^n \in A\} \subseteq X^\omega$. For an arbitrary family $\mathcal{F} \subseteq 2^X$, let $\mathfrak{C}^n(\mathcal{F}) = \{\mathfrak{C}(X_1 \cdots X_n) \mid X_i \in \mathcal{F}\}$, for $n \geq 1$, and $\mathfrak{C}(\mathcal{F}) = \bigcup_{n \geq 1} \mathfrak{C}^n(\mathcal{F})$.

If (X, Σ) is a measurable space, $(X, \Sigma)^n$ denotes the product space over X^n, and $(X, \Sigma)^\omega$ the measurable space over X^ω with σ-algebra generated by $\mathfrak{C}(\Sigma)$ (i.e., the smallest such that, for all $n \in \mathbb{N}$, the prefix $(\cdot)|^n$ and tail $(\cdot)|_n$ functions are measurable). Note that, the stepwise extension $f^\omega: X^\omega \to Y^\omega$ of the function $f: X \to Y$ is measurable if f is so. Often, X^n and X^ω will also denote $(X, 2^X)^n$ and $(X, 2^X)^\omega$, respectively.

3. Markov Chains and Linear-time Equivalences

In this section we recall the definition discrete-time Markov chains and the notions of strong and stutter probabilistic trace equivalences on them.

In the rest of the paper we fix a finite set $\mathbb{A}$ of atomic propositions.

Definition. A *Markov chain* is a tuple $\mathcal{M} = (S, \tau, \ell)$ consisting of a countable set S of *states*, a *transition probability function* $\tau: S \to \Delta(S)$ and a *labeling function* $\ell: S \to 2^{\mathbb{A}}$.

Intuitively, if $\mathcal{M}$ is in the state u, it moves to a state $v \in S$ with probability $\tau(u)(v)$. We say that $p \in \mathbb{A}$ holds in u if $p \in \ell(u)$. We will use $\mathcal{M} = (S, \tau, \ell)$ to range over the class of MCs and we will often refer to it and its constituents implicitly.

An MC can be thought of as a stochastic process that, from an initial state u, emits execution runs distributed according to the probability $\mathbb{P}(u)$ given below.

Definition 1. Let $\mathbb{P} \colon S \to \Delta(S^\omega)$ be such that, for all $u \in S$, $\mathbb{P}(u)$ is the unique probability measure† on S^ω such that, for all $n \geq 1$ and $U_i \subseteq S$ $(i = 0..n)$

$$\mathbb{P}(u)(\mathfrak{C}(U_0 \cdots U_n)) = \mathbb{1}_{U_0}(u) \cdot \textstyle\int \mathbb{P}(\cdot)(\mathfrak{C}(U_1 \cdots U_n)) \,\mathrm{d}\tau(u)\,,$$

where $\mathbb{1}_A$ denotes the indicator function for a set A.

Intuitively, $\mathbb{P}(u)(E)$ is the probability that, starting from u, the MC executes a run in $E \subseteq S^\omega$. Given $u_i \in S$ $(i \in \{0, \dots n\})$, hereafter the cylinder set $\mathfrak{C}(\{u_0\} \cdots \{u_n\})$ will be simply denoted by $\mathfrak{C}(u_0..u_n)$. For example, $\mathbb{P}(u)(\mathfrak{C}(u_0..u_n)) = \mathbb{1}_{\{u_0\}}(u) \cdot \prod_{i=0}^{n-1} \tau(u_i)(u_{i+1})$.

Remark 1. In Definition 1, since $\mathfrak{C}(U_0) = \mathfrak{C}(U_0 S)$, the case $\mathbb{P}(u)(\mathfrak{C}(U_0))$ is covered implicitly. Indeed,

$$\mathbb{P}(u)(\mathfrak{C}(U_0 S)) = \mathbb{1}_{U_0}(u) \cdot \int \mathbb{P}(\cdot)(\mathfrak{C}(S)) \,\mathrm{d}\tau(u) = \mathbb{1}_{U_0}(u) \cdot \int 1 \,\mathrm{d}\tau(u) = \mathbb{1}_{U_0}(u)$$

since $\mathfrak{C}(S) = S^\omega$ and, for all $v \in S$, $\mathbb{P}(v)$ is a probability measure.

Two states of an MC are considered equivalent if they exhibit the same observable behavior. Since in this work we are particularly interested to linear-time properties, we recall the most used linear-time equivalences on MCs: *strong* and *stutter probabilistic trace equivalences*.

Definition 2. Two states $u, v \in S$ are *probabilistic trace equivalent*, written $u \sim_t v$, if for all $T \in \mathfrak{C}(S/_{\equiv_\ell})$, $\mathbb{P}(u)(T) = \mathbb{P}(v)(T)$.

Probabilistic trace equivalence tests two states w.r.t. all linear-time events, considered up to label equivalence. This is in accordance to the fact that the only things that we observe in a state are the atomic properties (labels). Hereafter, $\mathcal{T}$ denotes $\mathfrak{C}(S/_{\equiv_\ell})$ and its elements are called *trace cylinders*.

The *stutter* (or *weak*) variant of the probabilistic trace equivalence considers a transition step as "visible" only when a change of the current behavior occurs. The guiding idea to define stutter events is to replace the notion of "step" with that of "stutter step". Formally, this corresponds to change the definitions of the tail (i.e., the "next step") and prefix functions over infinite words. Let X be a set and $R \subseteq X \times X$ equivalence. For $n \geq 1$, define the n-th R*-stutter tail* function $\mathsf{tl}^n_R \colon X^\omega \to X^\omega$, by induction on n, as

† Existence follows by the Hahn-Kolmogorov extension theorem; uniqueness by $\tau(u)$ being σ-finite.

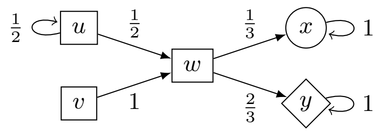

Fig. 1. The states u and v are stutter trace equivalent but neither bisimilar nor trace equivalent. States are labeled by different node shapes.

follows

$$\mathsf{tl}^1_R(\pi) = \begin{cases} \pi|_j & \text{if } \exists j \text{ s.t. } (\pi[0], \pi[j]) \notin R \text{ and } \forall i < j,\ (\pi[0], \pi[i]) \in R \\ \pi & \text{otherwise (i.e., } \pi \text{ is } R\text{-constant)}, \end{cases}$$
$$\mathsf{tl}^{n+1}_R(\pi) = \mathsf{tl}^1_R(\mathsf{tl}^n_R(\pi)) .$$

Intuitively, applying tl^1_R to a sequence π it leaves the sequence the same if all its elements are R-equivalent to its head $\pi[0]$, otherwise it removes the longest prefix of elements that are R-equivalent to its head; $\mathsf{tl}^n_R(\pi)$ is the n-th composition of it. For example, let $\pi = aaabbbc^\omega$, then $\mathsf{tl}^1_=(\pi) = bbbc^\omega$ and, for all $n > 1$, $\mathsf{tl}^n_=(\pi) = c^\omega$. The n-th *R-stutter prefix* function $\mathsf{pf}^n_R \colon X^\omega \to X^n$ is defined, by induction on $n \geq 1$, as $\mathsf{pf}^1_R(\pi) = \pi[0]$ and $\mathsf{pf}^{n+1}_R(\pi) = \pi[0]\mathsf{pf}^n_R(\mathsf{tl}^1_R(\pi))$.

Now, the standard definition of cylinder set for $A \subseteq X^n$ can be turned to that of *R-stutter cylinder set for A* (of rank n) as $\mathfrak{C}_R(A) = \{\pi \in X^\omega \mid \mathsf{pf}^n_R(\pi) \in A\}$. For a family $\mathcal{F} \subseteq 2^X$, denote by $\mathfrak{C}^n_R(\mathcal{F}) = \{\mathfrak{C}_R(E_1 \cdots E_n) \mid E_i \in \mathcal{F}\}$ the set of all R-stutter cylinders of rank n over $\mathcal{F}$ and $\mathfrak{C}_R(\mathcal{F}) = \bigcup_{n \geq 1} \mathfrak{C}^n_R(\mathcal{F})$. If (X, Σ) a measurable space, we denote by $(X, \Sigma)^\omega_R$ the measurable space of infinite words over X with σ-algebra generated by $\sigma(\mathfrak{C}_R(\Sigma))$ (i.e., the smallest σ-algebra such that, for all $n \geq 1$, the n-th R-stutter prefix and tail functions are measurable).

Definition 3. Two states $u, v \in S$ are *probabilistic stutter trace equivalent*, written $u \sim_{st} v$, if for all $T \in \mathfrak{C}_{\equiv_\ell}(S/_{\equiv_\ell})$, $\mathbb{P}(u)(T) = \mathbb{P}(v)(T)$.

Intuitively, $\sim_{st}$ equates the states that have the same probability on all the $\equiv_\ell$-stutter linear-time events, considered up to label equivalence. Hereafter, $\mathcal{ST}$ denotes $\mathfrak{C}_{\equiv_\ell}(S/_{\equiv_\ell})$ and its elements will be called *stutter trace cylinders*.

By σ-additivity of the measures $\mathbb{P}(u)$, for all $u \in S$, it is easy to show that $\sim_t \subseteq \sim_{st}$. Note that, $\sim_{st} \not\subseteq \sim_t$ (see Fig. 1 for a counterexample).

4. Trace Distances and Probabilistic Model Checking

We give the definitions of *strong* and *stutter trace distances* and provide logical characterizations to both of them in terms of suitable fragments of LTL, relating the two behavioral distances to the probabilistic LTL-model checking problem (Var85).

Linear Distances. The strong and stutter probabilistic trace equivalences on MCs are naturally lifted to pseudometrics $\delta_t, \delta_{st} \colon S \times S \to [0,1]$ as follows

$$\delta_t(u,v) = \sup_{E \in \sigma(\mathcal{T})} |\mathbb{P}(u)(E) - \mathbb{P}(v)(E)| \,, \qquad \text{(STRONG TRACE DISTANCE)}$$

$$\delta_{st}(u,v) = \sup_{E \in \sigma(\mathcal{ST})} |\mathbb{P}(u)(E) - \mathbb{P}(v)(E)| \,. \qquad \text{(STUTTER TRACE DISTANCE)}$$

Observe that two states $u, v \in S$ are strong (resp. stutter) trace equivalent if and only if $\delta_t(u,v) = 0$ (resp. $\delta_{st}(u,v) = 0$). Moreover, by $\sigma(\mathcal{ST}) \subseteq \sigma(\mathcal{T})$, it holds $\delta_{st} \leq \delta_t$.

Note that, the above distances are total variation distances between two measures, namely the restriction of $\mathbb{P}(u)$ and $\mathbb{P}(v)$, on $\sigma(\mathcal{T})$ and $\sigma(\mathcal{ST})$, respectively.

Linear Temporal Logic (LTL) is a formalism for reasoning about sequences of events (Pnu77). The LTL formulas are generated by the following grammar

$$\varphi ::= p \mid \bot \mid \varphi \to \varphi \mid \mathsf{X}\varphi \mid \varphi \,\mathsf{U}\, \varphi \,, \qquad \text{where } p \in \mathbb{A} \,.$$

Where $\bot$ is constant false formula, $\phi \to \psi$ is implication, $\mathsf{X}\varphi$ is the next modal operator, and $\varphi \,\mathsf{U}\, \psi$ is the until temporal modality. Let LTL$^{\text{-u}}$ and LTL$^{\text{-x}}$ be the fragments of LTL without until (U) and next (X) operators, respectively.

As usual, negation, disjunction, conjunction, and double implication are derived as: $\neg\varphi = \varphi \to \bot$; $\varphi \vee \psi = \neg\varphi \to \psi$; $\varphi \wedge \psi = \neg(\neg\varphi \vee \neg\psi)$; and $\varphi \leftrightarrow \psi = (\varphi \to \psi) \wedge (\psi \to \varphi)$.

The semantics of the formulas is given by means of a satisfiability relation defined, for an MC $\mathcal{M}$ and $\pi \in S^\omega$, as follows

$$\begin{array}{ll}
\mathcal{M}, \pi \models p & \text{if } p \in \ell(\pi[0]) \,, \\
\mathcal{M}, \pi \models \bot & \text{never} \,, \\
\mathcal{M}, \pi \models \varphi \to \psi & \text{if } \mathcal{M}, \pi \models \psi \text{ whenever } \mathcal{M}, \pi \models \varphi \,, \\
\mathcal{M}, \pi \models \mathsf{X}\varphi & \text{if } \mathcal{M}, \pi|_1 \models \varphi \,, \\
\mathcal{M}, \pi \models \varphi \,\mathsf{U}\, \psi & \text{if } \exists i \geq 0 \text{ s.t. } \mathcal{M}, \pi|_i \models \psi, \text{ and } \forall\, 0 \leq j < i,\, \mathcal{M}, \pi|_j \models \varphi \,.
\end{array}$$

Define $[\![\varphi]\!] = \{\pi \mid \mathcal{M}, \pi \models \varphi\}$ and $[\![\mathcal{L}]\!] = \{[\![\varphi]\!] \mid \varphi \in \mathcal{L}\}$, for any $\mathcal{L} \subseteq$ LTL. The probabilistic model checking problem against MCs over LTL formulas consists in determining the probability $\mathbb{P}(u)([\![\varphi]\!])$ for an initial state u and $\varphi \in$ LTL. For any $\mathcal{L} \subseteq$ LTL, the pseudometric

$$\delta_{\mathcal{L}}(u,v) = \sup_{\varphi \in \mathcal{L}} |\mathbb{P}(u)([\![\varphi]\!]) - \mathbb{P}(v)([\![\varphi]\!])|$$

measures the maximal difference that can be observed between the states u and v by model checking them over a subset $\mathcal{L}$ of linear temporal logic formulas of interest.

In the rest of the section we show that δ_t and δ_{st} can be logically characterized as δ_{LTL} (or $\delta_{\text{LTL}^{\text{-u}}}$) and $\delta_{\text{LTL}^{\text{-x}}}$, respectively. For the proof of this result we will exploiting the following technical result from (BBLM15).

Lemma 1. Let μ and ν be two finite measures on a measurable space (X, Σ). If Σ is generated by a field $\mathcal{F}$, then $\|\mu - \nu\| = \sup_{E \in \mathcal{F}} |\mu(E) - \nu(E)|$.

Indeed, as we already noted after its definition, the distance $\delta_t(u,v)$ can be seen as a total variation distance between measures $\mathbb{P}(u)$ and $\mathbb{P}(v)$ suitably restricted to the

σ-algebra $\sigma(\mathcal{T})$ generated from the trace cylinders. So that, by Lemma 1, to provide a logical characterization for δ_t in terms of LTL (or LTL$^{\text{-u}}$) formulas it suffices to show that the σ-algebra $\sigma(\mathcal{T})$ is generated by $[\![\text{LTL}]\!]$ (or $[\![\text{LTL}^{\text{-u}}]\!]$).

Theorem 1. (i) $\sigma(\mathcal{T}) = \sigma([\![\text{LTL}]\!]) = \sigma([\![\text{LTL}^{\text{-u}}]\!])$, (ii) $\delta_t = \delta_{\text{LTL}} = \delta_{\text{LTL}^{\text{-u}}}$.

Proof. (ii) is a direct consequence of (i) and Lemma 1, since both $[\![\text{LTL}]\!]$ and $[\![\text{LTL}^{\text{-u}}]\!]$ are fields. To prove (i) it suffices to show (a) $[\![\text{LTL}]\!] \subseteq \sigma(\mathcal{T})$ and (b) $\mathcal{T} \subseteq \sigma([\![\text{LTL}^{\text{-u}}]\!])$.

(a) By structural induction on the formula $\varphi \in \text{LTL}$ we show that $[\![\varphi]\!] \in \sigma(\mathcal{T})$.

(Case $\varphi = p \in \mathbb{A}$). $[\![p]\!] = \bigcup \{ \mathfrak{C}([u]_{\equiv_\ell}) \mid u \in S, p \in \ell(u) \}$. Since S is assumed to be countable and, for all $u \in S$, $\mathfrak{C}([u]_{\equiv_\ell}) \in \mathcal{T}$, then we have that $[\![p]\!] \in \sigma(\mathcal{T})$.

(Case $\varphi = \bot$). $[\![\bot]\!] = \emptyset \in \sigma(\mathcal{T})$.

(Case $\varphi = \phi \to \psi$). $[\![\phi \to \psi]\!] = [\![\neg\phi \vee \psi]\!] = [\![\phi]\!]^c \cup [\![\psi]\!]$. By inductive hypothesis we know that $[\![\phi]\!], [\![\psi]\!] \in \sigma(\mathcal{T})$, therefore $[\![\phi \to \psi]\!] \in \sigma(\mathcal{T})$.

(Case $\varphi = \mathsf{X}\phi$). The following holds:

$$
\begin{aligned}
[\![\mathsf{X}\phi]\!] &= \{\pi \mid \mathcal{M}, \pi|_1 \models \phi\} && \text{(by def. of } \mathsf{X}) \\
&= \{\pi \mid \pi|_1 \in [\![\phi]\!]\} && \text{(by def. of } [\![\cdot]\!]) \\
&= (\cdot)|_1^{-1}([\![\phi]\!]) && \text{(by def. of } (\cdot)|_1)
\end{aligned}
$$

By inductive hypothesis and $(\cdot)|_1$ being measurable, it follows that $[\![\mathsf{X}\phi]\!] \in \sigma(\mathcal{T})$.

(Case $\varphi = \phi \,\mathsf{U}\, \psi$). The following holds:

$$
\begin{aligned}
[\![\phi \,\mathsf{U}\, \psi]\!] &= \{\pi \mid \exists i \geq 0.\, \mathcal{M}, \pi|_i \models \psi \text{ and } \forall 0 \leq j < i.\, \mathcal{M}, \pi|_j \models \phi\} && \text{(by def. } \mathsf{U}) \\
&= \{\pi \mid \exists i \geq 0.\, \pi|_i \in [\![\psi]\!] \text{ and } \forall 0 \leq j < i.\, \pi|_j \in [\![\phi]\!]\} && \text{(by def. } [\![\cdot]\!]) \\
&= \bigcup_{i \geq 0} \bigcap_{0 \leq j < i} ((\cdot)|_i^{-1}([\![\psi]\!]) \cap (\cdot)|_j^{-1}([\![\phi]\!]))\,. && \text{(by def. } (\cdot)|_k)
\end{aligned}
$$

By inductive hypothesis on ϕ, ψ and measurability of $(\cdot)|_k$ for arbitrary $k \in \mathbb{N}$, it follows $[\![\phi \,\mathsf{U}\, \psi]\!] \in \sigma(\mathcal{T})$.

(b) To prove $\sigma(\mathcal{T}) \subseteq \sigma([\![\text{LTL}^{\text{-u}}]\!])$ it suffices to show $\mathcal{T} \subseteq \sigma([\![\text{LTL}^{\text{-u}}]\!])$. Define the functions $A \colon \mathbb{A} \times 2^S \to \text{LTL}^{\text{-u}}$ and $B \colon \mathcal{T} \to \text{LTL}^{\text{-u}}$ as follows

$$
A(p, C) = \begin{cases} p & \text{if } \exists u \in C \text{ s.t. } p \in \ell(u) \\ \neg p & \text{otherwise} \end{cases}
$$

$$
\begin{aligned}
B(\mathfrak{C}(C_0)) &= \textstyle\bigwedge_{p \in \mathbb{A}} A(p, C_0) \\
B(\mathfrak{C}(C_0 \cdots C_{n+1})) &= B(\mathfrak{C}(C_0)) \wedge \mathsf{X} B(\mathfrak{C}(C_1 \cdots C_{n+1}))\,.
\end{aligned}
$$

For $T = \mathfrak{C}(C_0 \cdots C_n) \in \mathcal{T}$, by induction on $n \in \mathbb{N}$, it is easy to prove that $[\![B(T)]\!] = T$. This implies that $\mathcal{T} \subseteq [\![\text{LTL}^{\text{-u}}]\!] \subseteq \sigma([\![\text{LTL}^{\text{-u}}]\!])$. □

Remark 2. We would like to remark that the equality $\delta_t = \delta_{\text{LTL}}$ is not trivial. In (CK14) it is proven that for the MC in Fig. 2, the value $\delta_t(u, x)$ is obtained as a supremum over $\sigma(\mathcal{T})$ only on a maximizing measurable set that *is not* ω-regular. Since the properties

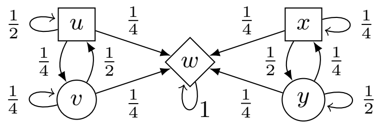

Fig. 2. The trace distance between u and x is irrational: $\delta_t(u,x) = \sqrt{2}/4$ (cf. (CK14)).

definable by LTL formulas are restricted to the star-free regular languages and the star-free languages are a strict subset of the ω-regular languages, this means that such a supremum cannot be achieved by a single LTL formula or trivial unions of families of LTL formulas.

In the reminder of the section we provide a logical characterization for the stutter trace distance δ_{st} similar to the one proven above, in terms of LTL formulas without next operator. The key for proving this result will still be Lemma 1. However, for the proof of $\sigma(\mathcal{T}) \subseteq \sigma([\![\mathrm{LTL}]\!])$ in Theorem 1, we needed to use the measurability of the n-th tail function $(\cdot)|_n$ w.r.t. $\sigma(\mathcal{T})$. Unfortunately, $(\cdot)|_n$ is not measurable w.r.t. $\sigma(\mathcal{ST})$, so the logical characterization does not carry over easily to the stutter case.

We solve this problem by giving a coinductive characterization to Lamport's *stutter equivalence* (Lam83) (for a standard definition see e.g. (BK08, §7.7.1)).

Definition 4. For a relation $R \subseteq S^\omega \times S^\omega$, $\pi \in S^\omega$ is said *R-constant* if, for all $i \in \mathbb{N}$, $\pi \; R \; \pi|_i$. A relation $R \subseteq S^\omega \times S^\omega$ is a *stutter relation* if whenever $\pi \; R \; \rho$

(i) $\pi[0] \equiv_\ell \rho[0]$;
(ii) π is R-constant iff ρ is R-constant;
(iii) $\pi|_1 \; R \; \rho$ or $\pi \; R \; \rho|_1$ or $\pi|_1 \; R \; \rho|_1$.

Two traces $\pi, \rho \in S^\omega$ are *stutter equivalent*, written $\pi \simeq \rho$, if they are related by some stutter relation.

It is immediate to see that stutter relations are closed under union and reflexive/symmetric/transitive closure, therefore $\simeq$ is an equivalence and a stutter relation.

Proposition 1. $\pi \simeq \rho$ iff $\forall \varphi \in \mathrm{LTL}^{\text{-X}}. \, (\mathcal{M}, \pi \models \varphi \Leftrightarrow \mathcal{M}, \rho \models \varphi)$.

Proof. ($\Rightarrow$) Let R be a stutter relation such that $\pi \; R \; \rho$. Without loss of generality assume R to be an equivalence relation (indeed if R is a stutter relation so is the smallest equivalence containing it). It suffices to prove that, for any $\varphi \in \mathrm{LTL}^{\text{-X}}$, $\mathcal{M}, \pi \models \varphi$ implies $\mathcal{M}, \rho \models \varphi$. We proceed by induction on the structure of the formula φ.

(Case $\varphi = p \in \mathbb{A}$). Assume $\mathcal{M}, \pi \models p$. By $\pi[0] \equiv_\ell \rho[0]$, then $\mathcal{M}, \rho \models p$.
(Case $\varphi = \bot$). Immediate, by the semantics of $\bot$.
(Case $\varphi = \phi \to \psi$). Assume $\mathcal{M}, \rho \models \phi$. By symmetry of R, $\rho \; R \; \pi$. By inductive hypothesis, $\mathcal{M}, \pi \models \phi$. By hypothesis $\mathcal{M}, \pi \models \phi \to \psi$, hence $\mathcal{M}, \pi \models \psi$. By inductive hypothesis, $\mathcal{M}, \rho \models \psi$, hence $\mathcal{M}, \rho \models \phi \to \psi$.
(Case $\varphi = \phi \, \mathsf{U} \, \psi$). For $\pi' \in [\![\phi \, \mathsf{U} \, \psi]\!]$, define $i^*(\pi') = \min\{i \mid \mathcal{M}, \pi'|_i \models \psi\}$. We proceed by induction on $i^*(\pi)$. [Base case: $i^*(\pi) = 0$] By definition of i^*, $\mathcal{M}, \pi \models \psi$. By inductive hypothesis on the formula, $\mathcal{M}, \rho \models \psi$. Therefore, $\mathcal{M}, \rho \models \phi \mathsf{U} \psi$. [Inductive step: $i^*(\pi) > 0$]

By definition of i^*, $\mathcal{M}, \pi|_1 \models \phi \mathsf{U} \psi$. Since $i^*(\pi|_1) = i^*(\pi) - 1$, by inductive hypothesis on $\pi|_1$, for every ρ' such that $\pi|_1 \; R \; \rho'$, it holds $\mathcal{M}, \rho' \models \phi \mathsf{U} \psi$. By $\pi \; R \; \rho$, one of the following cases holds:

(i) Case $\pi|_1 \; R \; \rho$. Then, by the above, $\mathcal{M}, \rho \models \phi \mathsf{U} \psi$.

(ii) Case $\pi|_1 \; R \; \rho|_1$. Then $\mathcal{M}, \rho|_1 \models \phi \mathsf{U} \psi$. By $i^*(\pi) > 0$ and $\mathcal{M}, \pi \models \phi \mathsf{U} \psi$ we have $\mathcal{M}, \pi \models \phi$. By inductive hypothesis on the formula, $\mathcal{M}, \rho \models \phi$. From this and $\mathcal{M}, \rho|_1 \models \phi \mathsf{U} \psi$ we conclude that $\mathcal{M}, \rho \models \phi \mathsf{U} \psi$.

(iii) Case $\pi \; R \; \rho|_1$. First we prove that π is not R-constant. By contradiction assume π is R-constant. Then, $\pi[0] \equiv_\ell \pi[i]$, for all $i \geq 0$, and, in particular $\pi|_0 \equiv_\ell^\omega \pi|_{i^*(\pi)}$. By $\mathcal{M}, \pi|_{i^*(\pi)} \models \psi$ and $\pi|_0 \equiv_\ell^\omega \pi|_{i^*(\pi)}$, we have $\mathcal{M}, \pi|_0 \models \psi$. But this contradicts the assumption that $i^*(\pi) > 0$, hence π is not R-constant. By $\pi \; R \; \rho$ and the fact that π is not R-constant, then also ρ is not R-constant. This means that there exists $j \geq 0$ such that $\pi \not R \; \rho|_j$; and let j^* be the minimal one. Clearly $j^* > 1$, because we assumed $\pi \; R \; \rho|_1$. Thus we have that for all $j < j^*$ $\pi \; R \; \rho|_j$, and by inductive hypothesis on the formula $\mathcal{M}, \rho|_j \models \phi$. In particular from $\pi \; R \; \rho|_{j^*-1}$ we have have two possible cases: either $\pi|_1 \; R \; \rho|_{j-1}$ or $\pi|_1 \; R \; \rho|_j$ (note that by definition of j^* the case $\pi \; R \; \rho|_j$ does not hold). Similarly to (i) and (ii) above, in both cases we can prove $\mathcal{M}, \rho \models \phi \mathsf{U} \psi$.

($\Leftarrow$) We show that $\equiv_{\text{LTL}^{\text{-x}}} = \{(\pi, \rho) \mid \forall \varphi \in \text{LTL}^{\text{-x}}, \mathcal{M}, \pi \models \varphi \text{ iff } \mathcal{M}, \rho \models \varphi\}$ is a stutter relation. Assume $\pi \equiv_{\text{LTL}^{\text{-x}}} \rho$. We check that the three conditions of Definition 4 hold.

(i) $p \in \ell(\pi[0])$ iff $\mathcal{M}, \pi \models p$ iff $\mathcal{M}, \rho \models p$ iff $p \in \ell(\rho[0])$. Hence $\pi[0] \equiv_\ell \rho[0]$.

(ii) Assume π is not $\equiv_{\text{LTL}^{\text{-x}}}$-constant. Then, there exists $i > 0$ such that $\pi \not\equiv_{\text{LTL}^{\text{-x}}} \pi|_i$. Let k be the least of such indices. Then, there exists φ such that $\mathcal{M}, \pi \models \varphi$, $\mathcal{M}, \pi|_k \models \neg\varphi$ and, for all $j < k$, $\mathcal{M}, \pi|_j \models \varphi$. From this we get that $\mathcal{M}, \pi \models \varphi \mathsf{U} \neg\varphi$. By $\pi \equiv_{\text{LTL}^{\text{-x}}} \rho$, we have $\mathcal{M}, \rho \models \phi \mathsf{U} \neg\varphi$ and $\mathcal{M}, \rho \models \varphi$. This implies that there exists $j > 0$ such that $\rho|_j \models \neg\varphi$. Therefore ρ is not $\equiv_{\text{LTL}^{\text{-x}}}$-constant.

(iii) Let $\pi \not\equiv_{\text{LTL}^{\text{-x}}} \rho|_1$ and $\pi|_1 \not\equiv_{\text{LTL}^{\text{-x}}} \rho$. We prove $\pi|_1 \equiv_{\text{LTL}^{\text{-x}}} \rho|_1$. By $\pi \equiv_{\text{LTL}^{\text{-x}}} \rho$, we have $\pi \not\equiv_{\text{LTL}^{\text{-x}}} \pi|_1$ and $\rho \not\equiv_{\text{LTL}^{\text{-x}}} \rho|_1$. Hence there exist $\alpha, \beta \in \text{LTL}^{\text{-x}}$ such that

$$\mathcal{M}, \pi \models \alpha \quad \text{and} \quad \mathcal{M}, \pi|_1 \models \neg\alpha\,, \tag{1}$$

$$\mathcal{M}, \rho \models \beta \quad \text{and} \quad \mathcal{M}, \rho|_1 \models \neg\beta\,. \tag{2}$$

Let $\varphi \in \text{LTL}^{\text{-x}}$ be such that $\mathcal{M}, \pi|_1 \models \varphi$. We show $\mathcal{M}, \rho|_1 \models \varphi$.

(Case $\mathcal{M}, \pi \models \varphi$) Assume by contradiction that $\mathcal{M}, \rho|_1 \models \neg\varphi$. By $\pi \equiv_{\text{LTL}^{\text{-x}}} \rho$ and (1), $\mathcal{M}, \rho \models \alpha$. Similarly, $\mathcal{M}, \rho \models \varphi$. By $\mathcal{M}, \rho \models \alpha$ and $\mathcal{M}, \rho|_1 \models \neg\varphi$ it holds $\mathcal{M}, \rho \models \alpha \mathsf{U} \neg\varphi$, hence, by $\pi \equiv_{\text{LTL}^{\text{-x}}} \rho$, we have $\mathcal{M}, \pi \models \alpha \mathsf{U} \neg\varphi$. By (1) $\mathcal{M}, \pi|_1 \models \neg\alpha$, hence the only possibility is that $\mathcal{M}, \pi|_1 \models \neg\varphi$, so that we get a contradiction.

(Case $\mathcal{M}, \pi \models \neg\varphi$) By (2) and $\pi \equiv_{\text{LTL}^{\text{-x}}} \rho$, $\mathcal{M}, \pi \models \beta$. By this and hypothesis on $\pi|_1$, we have $\mathcal{M}, \pi \models \beta \mathsf{U} \varphi$. Then, by $\pi \equiv_{\text{LTL}^{\text{-x}}} \rho$, we have $\mathcal{M}, \rho \models \beta \mathsf{U} \varphi$ and similarly, by the hypothesis made on π, we also have $\mathcal{M}, \rho \models \neg\varphi$. This means that $\mathcal{M}, \rho|_1 \models \varphi$. □

The above states that $\simeq$ characterizes the logical equivalence w.r.t. $\text{LTL}^{\text{-x}}$.

Proposition 2. Define $q\colon S^\omega \to S^\omega$ as follows, for $\pi \in S^\omega$,

$$q(\pi) = \begin{cases} \pi[0]q(\pi|_k) & \text{if } \exists k \text{ s.t. } \pi[0] \not\equiv_\ell \pi[k] \text{ and } \forall j < k,\, \pi[0] \equiv_\ell \pi[j] \\ \pi & \text{otherwise (i.e., } \pi \text{ is } \equiv_{\ell^\omega}\text{-constant)} \end{cases}$$

Then q is $\sigma(\mathcal{ST})$-$\sigma(\mathcal{T})$ measurable and $R = \{(\pi, \rho) \mid q(\pi) \equiv_{\ell^\omega} q(\rho)\}$ is a stutter relation.

Now we are ready to prove the logical characterization of the stutter trace distance. Note that Definition 4 and Propositions 1, 2 are essential in the proof of this result.

Theorem 2. (i) $\sigma(\mathcal{ST}) = \sigma(\llbracket \text{LTL}^{\text{-x}} \rrbracket)$, (ii) $\delta_{st} = \delta_{\text{LTL}^{\text{-x}}}$.

Proof. (ii) is a direct consequence of (i) and Lemma 1, since $\llbracket \text{LTL}^{\text{-x}} \rrbracket$ is a field. To prove (i) it suffices to show (a) $\llbracket \text{LTL}^{\text{-x}} \rrbracket \subseteq \sigma(\mathcal{ST})$ and (b) $\mathcal{ST} \subseteq \sigma(\llbracket \text{LTL}^{\text{-x}} \rrbracket)$.

(a) By induction on the structure of $\varphi \in \text{LTL}^{\text{-x}}$ we prove $\llbracket \varphi \rrbracket \in \sigma(\mathcal{ST})$. We show only the case $\varphi = \phi \mathsf{U} \psi$. Consider the function $q\colon S^\omega \to S^\omega$ defined in Proposition 2. By Proposition 2, the relation $R = \{(\pi, \rho) \mid q(\pi) \equiv_{\ell^\omega} q(\rho)\}$ is a stutter relation. One can easily prove that $q(\pi) \equiv_{\ell^\omega} q(q(\pi))$ for all $\pi \in S^\omega$, hence $\pi \simeq q(\pi)$. Then,

$$\begin{aligned} \llbracket \phi \mathsf{U} \psi \rrbracket &= \{\pi \mid \exists i \geq 0.\, \mathcal{M}, \pi|_i \models \psi \text{ and } \forall 0 \leq j < i.\, \mathcal{M}, \pi|_j \models \phi\} && \text{(by def. } \mathsf{U}) \\ &= \{\pi \mid \exists i \geq 0.\, \mathcal{M}, q(\pi)|_i \models \psi \text{ and } \forall 0 \leq j < i.\, \mathcal{M}, q(\pi)|_j \models \phi\} && \text{(Prop.1)} \\ &= \{\pi \mid \exists i \geq 0.\, q(\pi)|_i \in \llbracket \psi \rrbracket, \forall 0 \leq j < i.\, q(\pi)|_j \in \llbracket \phi \rrbracket\} && \text{(by def. } \llbracket\cdot\rrbracket) \\ &= \bigcup_{i \geq 0} \bigcap_{0 \leq j < i} (((\cdot)|_i \circ q)^{-1}(\llbracket \psi \rrbracket) \cap ((\cdot)|_j \circ q)^{-1}(\llbracket \phi \rrbracket))\,. && \text{(preimage)} \end{aligned}$$

By Proposition 2 the function q is $\sigma(\mathcal{ST})$-$\sigma(\mathcal{T})$ measurable, hence, for any $k \in \mathbb{N}$, the composite $(\cdot)|_k \circ q$ is $\sigma(\mathcal{ST})$-measurable. By inductive hypothesis on ϕ, ψ and $\sigma(\mathcal{ST})$-measurability of $(\cdot)|_k \circ q$, it follows that $\llbracket \phi \mathsf{U} \psi \rrbracket \in \sigma(\mathcal{ST})$.

(b) To prove $\sigma(\mathcal{ST}) \subseteq \sigma(\llbracket \text{LTL}^{\text{-x}} \rrbracket)$ it suffices to show $\mathcal{ST} \subseteq \sigma(\llbracket \text{LTL}^{\text{-x}} \rrbracket)$. To this end, define $A\colon \mathbb{A} \times 2^S \to \text{LTL}^{\text{-x}}$ and $B\colon \mathcal{ST} \to \text{LTL}^{\text{-x}}$ as follows, for $i = 1..n$ and $C_i \in S/_{\equiv_\ell}$ s.t. $C_i \neq C_{i+1}$,

$$A(p, C) = \begin{cases} p & \text{if } \exists u \in C \text{ s.t. } p \in \ell(u) \\ \neg p & \text{otherwise} \end{cases}$$

$$\begin{aligned} B(\mathfrak{C}_{\equiv_\ell}(C_1)) &= \textstyle\bigwedge_{p \in \mathbb{A}} A(p, C_0) \\ B(\mathfrak{C}_{\equiv_\ell}(C_1 \cdots C_{n+1})) &= \big(B(\mathfrak{C}_{\equiv_\ell}(C_1)) \wedge \neg B(\mathfrak{C}_{\equiv_\ell}(C_2)\big) \mathsf{U}\, B(\mathfrak{C}_{\equiv_\ell}(C_2 \cdots C_{n+1}))\,, \end{aligned}$$

Note that, by definition of $\mathsf{pf}^n_{\equiv_\ell}$ and $\mathsf{tl}^1_{\equiv_\ell}(\pi)$, every stutter cylinder in $\mathcal{ST}$ can always be represented as a set of the form $\mathfrak{C}_{\equiv_\ell}(C_1 \cdots C_n)$, where, for all $i = 1..n$, $C_i \in S/_{\equiv_\ell}$ and $C_i \neq C_{i+1}$. Now, for a stutter cylinder $T \in \mathcal{ST}$ of this form, it is easy to prove that $\llbracket B(T) \rrbracket = T$. This implies that $\mathcal{ST} \subseteq \llbracket \text{LTL}^{\text{-x}} \rrbracket \subseteq \sigma(\llbracket \text{LTL}^{\text{-x}} \rrbracket)$. □

5. Convergence from Branching to Linear Distances

We provide two nets of pseudometrics that converge, respectively, to the strong and stutter trace distances. The pseudometrics are shown to be liftings of multi-step extensions of probabilistic bisimilarity and a suitable stutter variant of it.

Our construction is inspired by (CvBW12, Cor. 11), where the bisimilarity pseudometric δ_b of Desharnais et al. (DGJP04) is shown to be an upper bound for the trace distance δ_t. Their result is based on an alternative characterization of δ_b by means of the notion of "coupling structure" (CvBW12, Th. 8). The proof of $\delta_t \leq \delta_b$ uses a classic duality result (Lemma 2 below) asserting that *the total variation of two measures coincides to the minimal discrepancy measured among all their couplings*. Formally, given $\mu, \nu \in \Delta(X, \Sigma)$, the *discrepancy* of $\omega \in \Omega(\mu, \nu)$ is the value $\omega(\ncong_\Sigma)$, where $\ncong_\Sigma = \bigcup \{E \times E^c \mid E \in \Sigma\}$ is the *separability relation* w.r.t. Σ. Note that $x \ncong_\Sigma y$ if and only if, there exists $E \in \Sigma$ such that $x \in E$ and $y \notin E$. Conversely, the *inseparability relation* w.r.t. Σ, denoted $\cong_\Sigma$, has the property that $x \cong_\Sigma y$ if and only if, for all $E \in \Sigma$, $x \in E$ iff $y \in E$.

Lemma 2 ((Lin92, Th.5.2)). Let μ, ν be probability measures on (X, Σ). Then, provided that $\ncong_\Sigma$ is measurable in $\Sigma \otimes \Sigma$, $\|\mu - \nu\| = \min \{\omega(\ncong_\Sigma) \mid \omega \in \Omega(\mu, \nu)\}$.

For a proof of the above result see the Appendix.

Along the way to obtain our construction, we nontrivially extend (and improve the proofs of) both Corollary 11 and Theorem 8 in (CvBW12). Moreover, this construction reveals a nontrivial relation between branching and linear-time metric-based semantics (by means of a convergence of the observable behaviors) that does not hold by using the standard equivalence-based semantics.

5.1. *The Strong Case*

We start by introducing a multi-step variant of probabilistic bisimulation.

Definition 5. Let $k \geq 1$. An equivalence relation $R \subseteq S \times S$ is a *k-probabilistic bisimulation* on $\mathcal{M}$ if whenever $u \, R \, v$, then, for all $E_i \in S/_{\equiv_\ell}$ and $C \in S/_R$,

$$\mathbb{P}(u)(\mathfrak{C}(E_0 \cdots E_{k-1} C)) = \mathbb{P}(v)(\mathfrak{C}(E_0 \cdots E_{k-1} C)) \, .$$

Two states $u, v \in S$ are *k-probabilistic bisimilar*, written $u \sim_b^k v$, if they are related by some k-probabilistic bisimulation.

The notion of k-bisimulation weakens that of probabilistic bisimulation of Larsen and Skou (LS91) by equating states that have the same probability to move to the same k-bisimilarity class after having observed the same labels within the first k-steps. Note that $\sim_b^1$ coincides with Larsen and Skou probabilistic bisimilarity. Moreover, for all $k \geq 1$, $\sim_b^k$ is a k-bisimulation and, by σ-additivity of measures, $\sim_b^1 \subseteq \sim_b^k \subseteq \sim_t$.

Remark 3. Clearly, $\bigcup_{k \geq 1} \sim_b^k \subseteq \sim_t$. However, the converse inclusion does not hold. A counterexample is shown in Fig. 3(left), where the states u and v are probabilistic trace equivalent, but they are not probabilistic k-bisimilar for any $k \geq 1$.

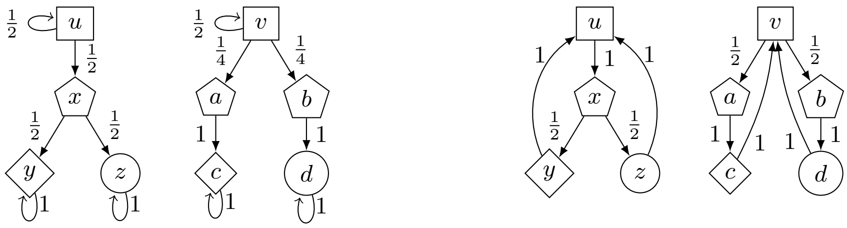

Fig. 3. (Left) the states u and v are trace equivalent but they are not k-bisimilar, for any $k \geq 1$; (Right) $u \sim_b^3 v$ but $u \not\sim_b^4 v$. States are labeled by different node shapes.

Remark 4. Differently from what one may expect, the k-bisimilarities do not necessarily get weaker by increasing k, i.e., for an arbitrary $k \geq 1$, it does not hold $\sim_b^k \subseteq \sim_b^{k+1}$. An example is shown in Fig. 3(right), where $u \sim_b^3 v$ but $u \not\sim_b^4 v$, hence $\sim_b^3 \not\subseteq \sim_b^4$.

Next we show how to "lift" the above equivalences to behavioral pseudometrics. A pseudometric that lifts bisimilarity is δ_b (DGJP04), defined as the least fixed point of the following operator on 1-bounded pseudometrics $d \colon S \times S \to [0,1]$

$$\Theta(d)(u,v) = \begin{cases} 1 & \text{if } u \not\equiv_\ell v \\ \mathcal{K}(d)(\tau(u), \tau(v)) & \text{otherwise}. \end{cases} \qquad \text{(KANTOROVICH OPERATOR)}$$

Intuitively, two states at maximal distance if they have different labels, otherwise the difference is given by Kantorovich distance of their transition probabilities.

Analogously, for $k \geq 1$, define the *k-steps transition probability function* $\tau^k \colon S \to \Delta(S^k)$ as the function such that $\tau^k(u)$ is the unique probability measure on S^k that, for all $U_i \subseteq S$ $(i = 1..k)$, $\tau^k(u)(U_1 \cdots U_k) = \mathbb{P}(u)(\mathfrak{C}(uU_1 \cdots U_k))$ (i.e., $\tau^k(u) = \mathbb{P}(u)[(\cdot)|^k \circ (\cdot)|_1]$). Note that, $\tau = \tau^1$. Then Θ is generalized by

$$\Theta^k(d)(u,v) \begin{cases} 1 & \text{if } u \not\equiv_\ell v \\ \mathcal{K}(\Lambda^k(d))(\tau^k(u), \tau^k(v)) & \text{otherwise}. \end{cases}$$

where $\Lambda^k(d)(u_1..u_k, v_1..v_k) = 1$ if $u_i \not\equiv_\ell v_i$ for some $i = 1..k$, otherwise $d(u_k, v_k)$. We call the above *k-Kantorovich operator*. It is easy to see that Θ^k is monotonic, so that, by Tarski's fixed point theorem, it has least fixed point, hereafter denoted by δ_b^k. Note that $\delta_b^1 = \delta_b$, moreover due to the following result, we call δ_b^k the *k-bisimilarity pseudometric*.

Lemma 3 (k-Bisimilarity Distance). $u \sim_b^k v$ iff $\delta_b^k(u,v) = 0$.

Proof. ($\Rightarrow$) Define the pseudometric d as $d(u,v) = 0$ if $u \sim_b^k v$, $d(u,v) = 1$, otherwise. By Tarski's fixed point theorem, it suffices to show that $\Theta^k(d) \leq d$. The case when $u \not\sim_b^k v$ holds trivially. Let $u \sim_b^k v$, we prove that $\Theta^k(d)(u,v) = 0$. By definition of k-probabilistic bisimulation we have that $u \equiv_\ell v$, therefore $\Theta^k(d)(u,v) = \mathcal{K}(\Lambda^k(d))(\tau^k(u), \tau^k(v))$. By (FPP04, Lemma 3.1) we have that if for all $E \in S^k/_{\ker(\Lambda^k(d))}$, $\tau^k(u)(E) = \tau^k(v)(E)$ then $\mathcal{K}(\Lambda^k(\delta_b^k))(\tau^k(u), \tau^k(v)) = 0$. It remains to show that $\tau^k(u)(E) = \tau^k(v)(E)$ for all

$E \in S^k/_{\ker(\Lambda^k(d))}$. By definition of Λ^k,

$$S^k/_{\ker(\Lambda^k(d))} = \left\{ E_1 \cdots E_{k-1} C \mid E_i \in S/_{\equiv_\ell},\, C \in S/_{\sim_b^k} \right\} .$$

Since $u \sim_b^k v$, $\mathbb{P}(u)(\mathfrak{C}(E_0 \cdots E_{k-1}C)) = \mathbb{P}(v)(\mathfrak{C}(E_0 \cdots E_{k-1}C))$, for all $E_i \in S/_{\equiv_\ell}$ $(i < k)$ and $C \in S/_{\sim_b^k}$. Then the following hold:

$$\begin{aligned}
\tau^k(u)(E_1 \cdots E_{k-1}C) &= \mathbb{P}(u)(\mathfrak{C}(uE_1 \cdots E_{k-1}C)) && (\text{def. } \tau^k)\\
&= \mathbb{P}(u)(\mathfrak{C}([u]_{\equiv_\ell}, E_1 \cdots E_{k-1}C)) && (\text{def. } \mathbb{P})\\
&= \mathbb{P}(v)(\mathfrak{C}([v]_{\equiv_\ell}, E_1 \cdots E_{k-1}C)) && (u \sim_b^k v)\\
&= \tau^k(v)(E_1 \cdots E_{k-1}C)\,. && (\text{def. } \tau^k \text{ and } \mathbb{P})
\end{aligned}$$

($\Leftarrow$) Let $R = \ker(\Lambda^k(\delta_b^k))$ and $R' = \ker(\delta_b^k)$. We show that R' is a k-probabilistic bisimulation. Clearly, R and R' are equivalence relations, since δ_b^k is a pseudometric and Λ^k preserves pseudometrics (i.e., $\Lambda^k(\delta_b^k)$ is a pseudometric). By definition of Λ^k,

$$S^k/_R = \left\{ E_1 \cdots E_{k-1} C \mid E_i \in S/_{\equiv_\ell},\, C \in S/_{R'} \right\} .$$

Let $u, v \in S$ be such that $u \; R' \; v$. By definition of Θ^k, we have that $u \equiv_\ell v$ and $\mathcal{K}(\Lambda^k(\delta_b^k))(\tau^k(u), \tau^k(v)) = 0$. By (FPP04, Lemma 3.1), $\mathcal{K}(\Lambda^k(\delta_b^k))(\tau^k(u), \tau^k(v)) = 0$ implies that $\tau^k(u)(E) = \tau^k(v)(E)$, for all $E \in S^k/_R$. Consider $E_i \in S/_{\equiv_\ell}$ $(0 \leq i < k)$ and $C \in S/_{R'}$. We consider two cases. If $u \notin E_0$ we have

$$\begin{aligned}
\mathbb{P}(u)(\mathfrak{C}(E_0 \cdots E_{k-1}C)) &= 0 && (\text{def. } \mathbb{P})\\
&= \mathbb{P}(v)(\mathfrak{C}(E_0 \cdots E_{k-1}C))\,. && (u \equiv_\ell v \text{ and def. } \mathbb{P})
\end{aligned}$$

If $u \in E_0$, the following hold:

$$\begin{aligned}
\mathbb{P}(u)(\mathfrak{C}(E_0 \cdots E_{k-1}C)) &= \mathbb{P}(u)(\mathfrak{C}(uE_1 \cdots E_{k-1}C)) && (\text{def. } \mathbb{P})\\
&= \tau^k(u)(E_1 \cdots E_{k-1}C) && (\text{def. } \tau^k)\\
&= \tau^k(v)(E_1 \cdots E_{k-1}C) && (\text{as shown before})\\
&= \mathbb{P}(v)(\mathfrak{C}(E_0 \cdots E_{k-1}C))\,. && (u \equiv_\ell v, \text{ def. } \tau^k \text{ and } \mathbb{P})
\end{aligned}$$

This proves that R' is a k-probabilistic bisimulation. □

In the reminder of the section we show that the k-bisimilarity pseudometrics δ_b^k form a net that converges point-wise to the trace distance δ_t. Such a convergence will be obtained by means of two key results, namely, Theorem 3 and Lemma 4.

Recall that a poset is *directed* if all its finite subsets have an upper bound. A *net* over a topological space X is a function from a directed poset $(D, \preceq)$ to X. We denote a net as $(x_i)_{i \in D}$, meaning that $i \in D$ is mapped to x_i. A net $(x_i)_{i \in D}$ over X converges to $x \in X$, written $(x_i)_{i \in D} \to x$, if for every open subset $A \subseteq X$ such that $x \in A$, there exits $h \in D$ such that, for all $j \succeq h$, $x_j \in A$.

Theorem 3. Let (X, Σ) be a measurable space s.t. $\not\cong_\Sigma \in \Sigma \otimes \Sigma$, μ, ν be probability measures on it, $(D, \preceq)$ be a directed poset and $\Omega \colon D \to 2^{\Omega(\mu,\nu)}$ be a monotone map such that $\bigcup_{i \in D} \Omega(i)$ is dense in $\Omega(\mu, \nu)$ w.r.t. the total variation distance. Then, the net $(u_i)_{i \in D}$ over $\mathbb{R}_+$ defined by $u_i = \inf \{\omega(\not\cong_\Sigma) \mid \omega \in \Omega(i)\}$, converges to $\|\mu - \nu\|$.

Proof. By Lemma 2, for all $i \in D$, $u_i \geq \|\mu - \nu\|$. Moreover, by monotonicity of Ω, $i \preceq j$ implies $u_i \geq u_j$. Hence, to prove $(u_i)_{i \in D} \to \|\mu - \nu\|$, it suffices to show $\inf_{i \in D} u_i = \|\mu - \nu\|$. Recall that for $Y \neq \emptyset$ and $f \colon Y \to \mathbb{R}$ bounded and continuous, if $D \subseteq Y$ is dense then $\inf f(D) = \inf f(Y)$. By hypothesis $\bigcup_{i \in D} \Omega(i) \subseteq \Omega(\mu, \nu)$ is dense; moreover, $\mu \times \nu \in \Omega(\mu, \nu) \neq \emptyset$. We show that $ev_{\not\cong} \colon \Omega(\mu, \nu) \to \mathbb{R}$, defined by $ev_{\not\cong}(\omega) = \omega(\not\cong)$ is bounded and continuous. It is bounded since all $\omega \in \Omega(\mu, \nu)$ are probability measures. It is continuous because $\|\omega - \omega'\| \geq |\omega(\not\cong) - \omega'(\not\cong)| = |ev_{\not\cong}(\omega) - ev_{\not\cong}(\omega')|$ (1-Lipschitz continuity). Now, applying Lemma 2, we derive our result. □

Recall that, $\delta_t(u, v)$ is the total variation distance between $\mathbb{P}(u)$ and $\mathbb{P}(v)$ restricted to $\sigma(\mathcal{T})$. Observe that the separability relation w.r.t. $\sigma(\mathcal{T})$ is $\not\equiv_{\ell^\omega}$, moreover it is measurable in $\sigma(\mathcal{T}) \otimes \sigma(\mathcal{T})$ (see Proposition 5 in Appendix). Therefore, by Lemma 2,

$$\delta_t(u, v) = \min \{\omega(\not\equiv_{\ell^\omega}) \mid \omega \in \Omega(\mathbb{P}(u), \mathbb{P}(v))\} . \tag{3}$$

To show that the k-bisimilarity distances converge to δ_t, it suffices to provide a net of the form $(\delta_b^k)_{k \in \mathbb{K}}$ complying with the requirements of Theorem 3. To this end we will characterize δ_b^k by means of the notion of *coupling structure* of rank k.

A coupling structure may be thought of as a stochastic process generating infinite traces of pairs of states starting from a distinguished initial pair (u, v) and distributed according to a coupling in $\Omega(\mathbb{P}(u), \mathbb{P}(v))$. The traces of pairs of states are generated by multi-steps of length k.

Definition 6 (Coupling Structure). A function $\mathcal{C} \colon S \times S \to \Delta(S^k \otimes S^k)$ is a *coupling structure of rank* $k \geq 1$ if for all $u, v \in S$, $\mathcal{C}(u, v) \in \Omega(\tau^k(u), \tau^k(v))$.

The set of coupling structures of rank k is denoted by $\mathbb{C}_k$.

Definition 7. For $k \geq 1$ and $\mathcal{C} \in \mathbb{C}_k$, let $\mathbb{P}_{\mathcal{C}} \colon S \times S \to \Delta(S^\omega \otimes S^\omega)$ be such that, for all $u, v \in S$, $\mathbb{P}_{\mathcal{C}}(u, v)$ is the unique probability measure on $S^\omega \otimes S^\omega$ such that, for all, $n \geq 1$ and $U_i, V_i \subseteq S$ $(i = 0..nk)$

$$\mathbb{P}_{\mathcal{C}}(u, v)(\mathfrak{C}(U_{0,nk}) \times \mathfrak{C}(V_{0,nk})) = \mathbb{1}_{U_0 \times V_0}(u, v) \cdot \int \mathbb{P}_{\mathcal{C}}(\cdot)(\mathfrak{C}(U_{k,nk}) \times \mathfrak{C}(V_{k,nk})) \, \mathrm{d}\omega \,,$$

where, $U_{i,j} = U_i \cdots U_j$ (similarly for V)[‡] and ω is the unique (subprobability) measure on $S \otimes S$ s.t., for all $A, B \subseteq S$, $\omega(A \times B) = \mathcal{C}(u, v)(U_{1,k-1}A \times V_{1,k-1}B)$.

The following lemma extends (CvBW12, Th. 8) to k-bisimilarity pseudometrics and provides the alternative characterization of δ_b^k in terms of coupling structures.

Lemma 4 (Coupling Lemma). $\delta_b^k(u, v) = \inf \{\mathbb{P}_{\mathcal{C}}(u, v)(\not\equiv_{\ell^\omega}) \mid \mathcal{C} \in \mathbb{C}_k\}$.

Proof. For $\mathcal{C} \in \mathbb{C}_k$, define $\Gamma_{\mathcal{C}} \colon [0, 1]^{S \times S} \to [0, 1]^{S \times S}$ as

$$\Gamma_{\mathcal{C}}(d)(u, v) = \begin{cases} 1 & \text{if } u \not\equiv_\ell v \\ \int \Lambda^k(d) \, \mathrm{d}\mathcal{C}(u, v) & \text{otherwise} \end{cases}$$

[‡] We assume that $U_{i,j} = \{\epsilon\}$ whenever $i > j$.

$\Gamma_{\mathcal{C}}$ is readily seen to be monotonic, thus, by Tarski's fixed point theorem, it has least fixed point, denoted by $d_{\mathcal{C}}$.

We split the proof in two parts. First we show that the following equality holds

$$\delta_b^k = \inf \{ d_{\mathcal{C}} \mid \mathcal{C} \in \mathbb{C}_k \} \; . \tag{4}$$

Then we prove that for any $\mathcal{C} \in \mathbb{C}_k$ and $u, v \in S$

$$d_{\mathcal{C}}(u,v) = \mathbb{P}_{\mathcal{C}}(u,v)(\not\equiv_{\ell^\omega}) \tag{5}$$

(Part 1) For an arbitrary $\mathcal{C} \in \mathbb{C}_k$, and $u \equiv_\ell v$ the following hold

$$\begin{aligned}
d_{\mathcal{C}}(u,v) &= \textstyle\int \Lambda^k(d_{\mathcal{C}}) \, \mathrm{d}\mathcal{C}(u,v) && \text{(def. } d_{\mathcal{C}} \text{ and } \Gamma_{\mathcal{C}}) \\
&\geq \mathcal{K}(\Lambda^k(d_{\mathcal{C}}))(\tau^k(u), \tau^k(v)) && (\mathcal{C}(u,v) \in \Omega(\tau^k(u), \tau^k(v))) \\
&= \Theta^k(d_{\mathcal{C}})(u,v) && \text{(def. } \Theta^k) \\
&\geq \delta_b^k(u,v) && \text{(Tarski's fixed point theorem)}
\end{aligned}$$

Therefore $\delta_b^k \leq \inf \{ d_{\mathcal{C}} \mid \mathcal{C} \in \mathbb{C}_k \}$. For the other inequality, note that for any $d \in [0,1]^{S \times S}$ there exists $\mathcal{C} \in \mathbb{C}_k$ such that $\Gamma_{\mathcal{C}}(d) = \Theta^k(d)$. Indeed one can take $\mathcal{C}$ such that, for all $u, v \in S$, $\mathcal{C}(u,v) = \operatorname{argmin} \left\{ \int \Lambda^k(d) \, \mathrm{d}\omega \mid \omega \in \Omega(\tau^k(u), \tau^k(v)) \right\}$. Let $\mathcal{D}$ be a coupling structure such that $\Gamma_{\mathcal{D}}(\delta_b^k) = \Theta^k(\delta_b^k)$. By definition of δ_b^k, $\Theta^k(\delta_b^k) = \delta_b^k$, thus also have that $\Gamma_{\mathcal{D}}(\delta_b^k) = \delta_b^k$ (i.e., δ_b^k is a fixed point of $\Gamma_{\mathcal{D}}$). Since $d_{\mathcal{D}}$ is the least fixed point of $\Gamma_{\mathcal{D}}$ we obtain that $d_{\mathcal{D}} \leq \delta_b^k$, from which we conclude $\delta_b^k \geq \inf \{ d_{\mathcal{C}} \mid \mathcal{C} \in \mathbb{C}_k \}$. This proves (4).
(Part 2) For the sake of readability, we will write d to refer to $\mathbb{P}_{\mathcal{C}}(.)(\not\equiv_{\ell^\omega})$. Let us define the following functions

$$\begin{aligned}
&tl \colon S^\omega \times S^\omega \to S^\omega \times S^\omega && tl(\pi, \rho) = (\pi|_k, \rho|_k) \\
&lst \colon S^k \times S^k \to S \times S && lst(u_1..u_k, v_1..v_k) = (u_k, v_k)
\end{aligned}$$

and let us define the following relations

$$\begin{aligned}
A &= \left\{ (u_1..u_k, v_1..v_k) \in S^k \times S^k \mid \exists 1 \leq i \leq k \,.\, u_i \not\equiv_\ell v_i \right\} \\
\equiv_\ell^k &= \{ (\pi, \rho) \in S^\omega \times S^\omega \mid \forall 0 \leq i \leq k \,.\, \pi[i] \equiv_\ell \rho[i] \}
\end{aligned}$$

Let $u, v \in S$, we consider two possible cases. If $u \not\equiv_\ell v$, then the following holds

$$\begin{aligned}
d(u,v) &= 1 - \mathbb{P}_{\mathcal{C}}(u,v)(\equiv_{\ell^\omega}) && \text{(additivity)} \\
&= 1 - 0 && (u \not\equiv_\ell v \text{ and def. } \mathbb{P}_{\mathcal{C}}) \\
&= \Gamma_{\mathcal{C}}(d)(u,v) \, . && \text{(def. } \Gamma_{\mathcal{C}})
\end{aligned}$$

If $u \equiv_\ell v$, then the following hold

$$\begin{aligned}
\Gamma_{\mathcal{C}}(d)(u,v) &= \textstyle\int \Lambda^k(d) \, \mathrm{d}\mathcal{C}(u,v) && \text{(def. } \Gamma_{\mathcal{C}}) \\
&= \mathcal{C}(u,v)(A) + \textstyle\int_{A^c} (d \circ lst) \, \mathrm{d}\mathcal{C}(u,v) && \text{(def. } \Lambda^k) \\
&= \mathbb{P}_{\mathcal{C}}(u,v)(\not\equiv_\ell^k) + \textstyle\int_{A^c} (d \circ lst) \, \mathrm{d}\mathcal{C}(u,v) && \text{(def. } \mathcal{C}(u,v)) \\
&= \mathbb{P}_{\mathcal{C}}(u,v)(\not\equiv_\ell^k) + \mathbb{P}_{\mathcal{C}}(u,v)(\equiv_\ell^k \cap tl_k^{-1}(\not\equiv_{\ell^\omega})) && \text{(def. } \mathbb{P}_{\mathcal{C}}) \\
&= \mathbb{P}_{\mathcal{C}}(u,v)(\not\equiv_{\ell^\omega}) = d(u,v) && (\not\equiv_{\ell^\omega} = \not\equiv_\ell^k \uplus (\equiv_\ell^k \cap tl_k^{-1}(\not\equiv_{\ell^\omega})))
\end{aligned}$$

This proves that d is a fixed point of $\Gamma_{\mathcal{C}}$. To prove that d is actually the least one, we proceed by contradiction. Assume that d' is the least fixed point of $\Gamma_{\mathcal{C}}$ and $d' < d$. Let define the relation $M \subseteq S \times S$ as

$$u \; M \; v \quad \text{iff} \quad d(u,v) - d'(u,v) = \|d - d'\|_\infty \,,$$

where we recall that $\|d-d'\|_\infty = \max_{u,v \in S} |d'(u,v) - d(u,v)|$. Take $u \; M \; v$, then we have

$$\begin{aligned}
\|d - d'\|_\infty &= d(u,v) - d'(u,v) \\
&= \textstyle\int \Lambda^k(d) \, \mathrm{d}\mathcal{C}(u,v) - \int \Lambda^k(d') \, \mathrm{d}\mathcal{C}(u,v) && (\text{def. } \Gamma_{\mathcal{C}}) \\
&= \textstyle\int (\Lambda^k(d) - \Lambda^k(d')) \, \mathrm{d}\mathcal{C}(u,v) && (\text{linearity}) \\
&= \textstyle\int_{A^c} (d - d') \circ lst \, \mathrm{d}\mathcal{C}(u,v) && (\text{def. } \Lambda^k)
\end{aligned}$$

define the sub-probability ω over $S \times S$ as $\omega(X \times Y) = \mathcal{C}(u,v)(A^c \cap lst^{-1}(X \times Y))$, then

$$\begin{aligned}
&= \textstyle\int_{S \times S} (d - d') \, \mathrm{d}\omega && (\text{def. } \omega) \\
&\leq \|d - d'\|_\infty \,.
\end{aligned}$$

Since $\|d - d'\|_\infty > 0$, the inequality $\|d - d'\|_\infty = \int_{S \times S} (d - d') \, \mathrm{d}\omega \leq \|d - d'\|_\infty$ proved above implies that the support of ω has to be included in M (i.e., $\omega(u', v') > 0$ implies $u' \; M \; v'$). Thus, whenever $u \; M \; v$ the following holds

$$\mathcal{C}(u,v)(A) = 0 \qquad \text{and,} \qquad \mathcal{C}(u,v)(A^c \cap lst^{-1}(M)) = 1 \,. \tag{6}$$

Equation (6) shall be read as follows. Given that the coupling structure $\mathcal{C}$ in a pair in M, it has probability 0 of preforming a multi-step of length k in A. Moreover, after having preformed a multi-step it moves with probability 1 to another pair in M. Thus, from (6) and def. of $\mathbb{P}_{\mathcal{C}}$ one can prove that for all $(u,v) \in M$, $\mathbb{P}_{\mathcal{C}}(u,v)(\not\equiv_{\ell^\omega}) = 0$. But this implies that $\|d - d'\|_\infty = 0$, contradicting the assumption that $d' < d$. □

The next lemma shows that (i) any coupling structure $\mathcal{C}$ induces a probability measure $\mathbb{P}_{\mathcal{C}}(u,v)$ which is a proper coupling for the pair $(\mathbb{P}(u), \mathbb{P}(v))$; (ii) the set of couplings constructed via the coupling structures grows by multiples of the rank k; and (iii) their union is dense in $\Omega(\mathbb{P}(u), \mathbb{P}(v))$.

Lemma 5. Let $u, v \in S$ be a pair of states of an MC $\mathcal{M}$. Then,

(i) for $k \geq 1$ and $\mathcal{C} \in \mathbb{C}_k$, $\mathbb{P}_{\mathcal{C}}(u,v) \in \Omega(\mathbb{P}(u), \mathbb{P}(v))$;
(ii) for $k, h \geq 1$, $\{\mathbb{P}_{\mathcal{C}}(u,v) \mid \mathcal{C} \in \mathbb{C}_k\} \subseteq \{\mathbb{P}_{\mathcal{C}}(u,v) \mid \mathcal{C} \in \mathbb{C}_{hk}\}$;
(iii) $\bigcup_{k \geq 1} \{\mathbb{P}_{\mathcal{C}}(u,v) \mid \mathcal{C} \in \mathbb{C}_k\}$ is dense in $\Omega(\mathbb{P}(u), \mathbb{P}(v))$ w.r.t. the total variation.

Proof. (i) Let $k \geq 1$ and $\mathcal{C} \in \mathbb{C}_k$. For proving $\mathbb{P}_{\mathcal{C}}(u,v) \in \Omega(\mathbb{P}(u), \mathbb{P}(v))$, it suffices to show the following equalities, for arbitrary $n \geq 1$ and $E \subseteq S^{nk+1}$:

$$\begin{aligned}
\mathbb{P}_{\mathcal{C}}(u,v)(\mathfrak{C}(E) \times S^\omega) &= \mathbb{P}(u)(\mathfrak{C}(E)) \,, && (\text{left marginal}) \\
\mathbb{P}_{\mathcal{C}}(u,v)(S^\omega \times \mathfrak{C}(E)) &= \mathbb{P}(v)(\mathfrak{C}(E)) \,. && (\text{right marginal})
\end{aligned}$$

The above follow immediately by definition of $\mathbb{P}_{\mathcal{C}}$ and definitional conditions of coupling structures, via a routine induction on $n \geq 1$.

(ii) Let $k, h \geq 1$ and $\mathcal{C} \in \mathbb{C}_k$. Define $\mathcal{D}(u,v)$ as the unique measure on $S^{hk} \otimes S^{hk}$ such that, for all $E, F \subseteq S^{hk}$,

$$\mathcal{D}(u,v)(E \times F) = \mathbb{P}_{\mathcal{C}}(u,v)(\mathfrak{C}(uE) \times \mathfrak{C}(vF)),$$

i.e., $\mathcal{D}(u,v) = \mathbb{P}_{\mathcal{C}}(u,v)[(\cdot)|^{hk} \circ (\cdot)|_1 \times (\cdot)|^{hk} \circ (\cdot)|_1]$. Then, $\mathcal{D} \in \mathbb{C}_{hk}$. Indeed, for any $E \subseteq S^{hk}$

$$\begin{aligned}
\mathcal{D}(u,v)(E \times S^{hk}) &= \mathbb{P}_{\mathcal{C}}(u,v)[(\cdot)|^{hk} \circ (\cdot)|_1 \times (\cdot)|^{hk} \circ (\cdot)|_1](E \times S^{hk}) && \text{(def. } \mathcal{D}(u,v)) \\
&= \mathbb{P}_{\mathcal{C}}(u,v)(((\cdot)|^{hk} \circ (\cdot)|_1)^{-1}(E) \times S^{\omega}) && \text{(push forward)} \\
&= \mathbb{P}(u)[(\cdot)|^{hk} \circ (\cdot)|_1](E) && \text{(by (i))} \\
&= \tau^{hk}(u)(E). && \text{(def. } \tau^k(u))
\end{aligned}$$

Similarly, $\mathcal{D}(u,v)(S^{hk} \times E) = \tau^{hk}(v)(E)$. Next we prove that $\mathbb{P}_{\mathcal{C}}(u,v) = \mathbb{P}_{\mathcal{D}}(u,v)$. Since any cylinder of rank n can be equivalently expressed as a cylinder of rank $m \geq n$, the equality between the two measures follows by the fact that, for any $U, V \subseteq S^{hk+1}$

$$\mathbb{P}_{\mathcal{C}}(u,v)(\mathfrak{C}(U) \times \mathfrak{C}(V)) = \mathbb{P}_{\mathcal{D}}(u,v)(\mathfrak{C}(U) \times \mathfrak{C}(V)),$$

which follows by definition of $\mathcal{D}(u,v)$.

(iii) We prove the following preliminary result from which we will obtain (iii).

Let (X, Σ) be a measurable space such that $\mathcal{F}$ is a field generating Σ and let $D \subseteq \Delta(X)$ be such that, for all $\mu \in \Delta(X)$ and $F \in \mathcal{F}$, there exists $\nu \in D$ such that $\nu(F) = \mu(F)$. Then D is dense in $\Delta(X)$ w.r.t. the total variation distance.

Let $E \in \Sigma$ be an arbitrary measurable set and $d_E \colon \Delta(X) \times \Delta(X) \to \mathbb{R}_+$ be the pseudometric defined as $d_E(\mu, \nu) = |\mu(E) - \nu(E)|$, for $\mu, \nu \in \Delta(X)$. Since $\|\mu - \nu\| = \sup_{E \in \Sigma} d_E(\mu, \nu)$, to prove that D is dense w.r.t. the total variation distance it suffices to show that D is dense w.r.t. d_E, for any $E \in \Sigma$ (see Proposition 9). Let $E \in \Sigma$ and $\varepsilon > 0$. For any $\mu \in \Delta(X)$ we have to provide $\nu \in D$ such that $d_E(\mu, \nu) < \varepsilon$. Define the measure $\tilde{\mu}$ as the least upper bound of $D \cup \{\mu\}$ w.r.t. the point-wise partial order between measures ($\nu \sqsubseteq \nu'$ iff $\nu(A) \leq \nu'(A)$, for all $A \in \Sigma$). The existence of $\tilde{\mu}$ is guaranteed by (DS88, Corr.6 pp.163) (note that $\tilde{\mu}$ is not necessarily finite). By (BBLM15, Lemma 5), $\mathcal{F} \subseteq \Sigma$ is dense in $(\Sigma, d_{\tilde{\mu}})$, where for $E, F \in \Sigma$, $d_{\tilde{\mu}}(E, F) = \mu(E \triangle F)$ is the Fréchet-Nikodym pseudometric[§], where $E \triangle F$ denotes the symmetric difference of sets. So there exists $F \in \mathcal{F}$ such that $d_{\tilde{\mu}}(E, F) < \frac{\varepsilon}{2}$. By hypothesis, there exists $\nu \in D$, such that $\nu(F) = \mu(F)$. Let $\omega \in \{\mu, \nu\}$ then

$$\begin{aligned}
\omega(E) &= \omega(E \setminus F) + \omega(E \cap F) && (\omega \text{ additive}) \\
&\leq \omega((E \setminus F) \cup (F \setminus E)) + \omega(F) && (\omega \text{ monotone}) \\
&= \omega(E \triangle F) + \omega(E) && \text{(by def)} \\
&\leq \tilde{\mu}(E \triangle F) + \omega(F) && (\omega \sqsubseteq \tilde{\mu}) \\
&= d_{\tilde{\mu}}(E, F) + \omega(F). && \text{(by def)}
\end{aligned}$$

This implies $|\omega(E) - \omega(F)| \leq d_{\tilde{\mu}}(E, F)$, and in particular that $|\mu(E) - \mu(F)| < \frac{\varepsilon}{2}$ and

[§] Notice that (BBLM15, Lemma 5) does not assume the measure to be finite, hence it can be safely applied to $\tilde{\mu}$.

$|\nu(E) - \nu(F)| < \frac{\varepsilon}{2}$. Then, the density of D follows by

$$\begin{aligned}
d_E(\mu,\nu) &= |\mu(E) - \nu(E)| && \text{(def. } d_E\text{)}\\
&\le |\mu(E) - \mu(F)| + |\mu(F) - \nu(E)| && \text{(triangular ineq.)}\\
&= |\mu(E) - \mu(F)| + |\nu(F) - \nu(E)| && (\nu(F) = \mu(F))\\
&< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon\,.
\end{aligned}$$

This concludes the proof of the preliminary result.

Let $u, v \in S$, $\Omega = \bigcup_{k\ge 1} \{\mathbb{P}_{\mathcal{C}}(u,v) \mid \mathcal{C} \in \mathbb{C}_k\}$. Given the result above, to prove (iii) it is sufficient to provide a field $\mathcal{F}$ generating the σ-algebra of $S^\omega \otimes S^\omega$ and show that, for every $\mu \in \Omega(\mathbb{P}(s), \mathbb{P}(s'))$ and $F \in \mathcal{F}$, there exists $\omega \in \Omega$ such that $\omega(F) = \mu(F)$.

Define $\mathcal{F} = \bigcup_{k\ge1} \mathcal{F}_k$, where $\mathcal{F}_k$ denotes the collection of all *finite* union of measurable sets of the form $\mathfrak{C}(E) \times \mathfrak{C}(F)$, for some $E, F \subseteq S^k$. It holds that $\mathcal{F}_k \subseteq \mathcal{F}_{k+1}$ and $\mathcal{F}_k$ is a field, for all $k \ge 1$. Therefore $\mathcal{F}$ is a field that generates the σ-algebra of $S^\omega \otimes S^\omega$.

Let $\mu \in \Omega(\mathbb{P}(u), \mathbb{P}(v))$, $k \ge 1$ and $\mathcal{D} \in \mathbb{C}_k$. We define the measure $\omega_k = \mathbb{P}_{\mathcal{C}_k}(u, v)$, where $\mathcal{C}_k \colon S \times S \to \Delta(S^k \otimes S^k)$ is defined by

$$\mathcal{C}_k(u',v') = \begin{cases} \mu[(\cdot)|^k \circ (\cdot)|_1 \times (\cdot)|^k \circ (\cdot)|_1] & \text{if } (u,v) = (u',v') \\ \mathcal{D}(u',v') & \text{otherwise} \end{cases}$$

Note that, since $\mathbb{C}_k$ is nonempty, $\mathcal{C}_k$ is well defined. We show $\mathcal{C}_k \in \mathbb{C}_k$. To this end, since $\mathcal{D} \in \mathbb{C}_k$, we just need to check that $\mu[(\cdot)|^k \circ (\cdot)|_1 \times (\cdot)|^k \circ (\cdot)|_1](E \times S^k) = \tau^k(u)(E)$ and $\mu[(\cdot)|^k \circ (\cdot)|_1 \times (\cdot)|^k \circ (\cdot)|_1](S^k \times E) = \tau^k(v)(E)$, for arbitrary $E \subseteq S^k$ (we check one equality, the other is analogous):

$$\begin{aligned}
\mu[(\cdot)|^k \circ (\cdot)|_1 \times (\cdot)|^k \circ (\cdot)|_1](E \times S^k) &= \mu(\mathfrak{C}(SE) \times S^\omega) && \text{(preimage)}\\
&= \mathbb{P}(u)(\mathfrak{C}(SE)) && (\mu \in \Omega(\mathbb{P}(u), \mathbb{P}(v)))\\
&= \tau^k(u)(E)\,. && \text{(def. } \tau^k\text{)}
\end{aligned}$$

Next we prove that for all $A \in \mathcal{F}_k$, $\omega_k(A) = \mu(A)$. Note that since $\mathcal{F}_k \subseteq \mathcal{F}_{k+1}$, this suffices to show that $\omega_k(B) = \mu(B)$ holds for all $B \in \mathcal{F}_j$ such that $j \le k$. Let $A = \bigcup_{i=1}^n \mathfrak{C}(E_i) \times \mathfrak{C}(F_i) \in \mathcal{F}_k$, for some $n \in \mathbb{N}$ and $E_i, F_i \subseteq S^k$ $(i = 1..n)$. Without loss of generality we can assume that the $\mathfrak{C}(E_i) \times \mathfrak{C}(F_i)$'s forming A are pairwise disjoint (indeed, $\mathcal{F}_k$ is a field, hence we can simply replace any two "overlapping" sets by taking the intersection and their symmetric difference).

$$\begin{aligned}
\omega_k(A) &= \mathbb{P}_{\mathcal{C}_k}(u,v)(A) && \text{(def. } \omega_k\text{)}\\
&= \textstyle\sum_{i=1}^n \mathbb{P}_{\mathcal{C}_k}(u,v)(\mathfrak{C}(E_i) \times \mathfrak{C}(F_i)) && \text{(additivity)}\\
&= \textstyle\sum_{i=1}^n \mathcal{C}_k(u,v)(E_i \times F_i) && \text{(def. } \mathbb{P}_{\mathcal{C}_k}\text{)}\\
&= \textstyle\sum_{i=1}^n \mu(\mathfrak{C}(E_i) \times \mathfrak{C}(F_i)) && \text{(def. } \mathcal{C}_k\text{)}\\
&= \mu(A)\,. && \text{(additivity)}
\end{aligned}$$

To conclude the proof, observe that, given $\mu \in \Omega(\mathbb{P}(u), \mathbb{P}(v))$ and $F \in \mathcal{F}$, there exists $k \ge 1$ such that $F \in \mathcal{F}_k$, and for ω_k given as above, $\omega_k(F) = \mu(F)$ and $\omega_k \in \Omega$. □

Note that Equation (3) and Lemmas 4 and 5(i) imply that, for all $k \ge 1$, $\delta_b^k \ge \delta_t$. This generalizes (CvBW12, Cor. 11) to arbitrary k-bisimilarity distances. Moreover, multi-steps bisimilarity distances are ordered by divisibility. Indeed, by Lemma 5(ii), it holds that, for all $k, h \geq 1$, $\delta_b^k \geq \delta_b^{hk}$ (hence, by Lemma 3, $\sim_b^{hk} \subseteq \sim_b^k$).

Denote by $\mathbb{K}$ the directed poset of positive integers ordered by divisibility, that is $\mathbb{K} = (\mathbb{N} \setminus \{0\}, \preceq)$ where $n \preceq m$ iff there exists $k \in \mathbb{N}$ such that $m = nk$.

Then, Theorem 3, Lemmas 4 and 5 suffice to prove the following net-convergence.

Theorem 4 (Convergence). The net $(\delta_b^k)_{k \in \mathbb{K}}$ converges point-wise to δ_t.

We already noticed in Remark 3 that there are examples of pairs of states that are trace equivalent but not k-bisimilar, for any $k \geq 1$. The above convergence theorem implies that in these cases, even though the k-bisimilar distance between the are non-zero, for all $k \geq 1$ (Lemma 3), they have to converge towards 0 for k tending to ∞.

Example 1. Consider the Markov chains in Fig. 3 (left). One can easily show that, for arbitrary $k \geq 1$, $\delta_t(x, a) = \delta_b^k(x, a) = \frac{1}{2}$ and, analogously, $\delta_t(x, b) = \delta_b^k(x, b) = \frac{1}{2}$. Now, it is not hard to see that $\delta_b^k(u, v)$ satisfies the following equalities

$$\delta_b^k(u,v) = \frac{1}{2^k}\delta_b^k(u,v) + \frac{1}{2^{k+1}}\delta_b^k(x,a) + \frac{1}{2^{k+1}}\delta_b^k(x,b) = \frac{1}{2^k}\delta_b^k(u,v) + \frac{1}{2^{k+1}} \,.$$

From the above we obtain that $\delta_b^k(u, v) = \frac{1}{2^{k+1}-2}$, thus the net $(\delta_b^k(u, v))_{k \in \mathbb{K}}$ converges to 0, in accordance with the fact that $\delta_t(u, v) = 0$.

Remark 5 (Equivalence vs Metric-based semantics). Although $\bigcup_{k \geq 1} \sim_b^k \neq \sim_t$ (see Remark 3), by Theorem 4, we have that $\inf_{k \geq 1} \delta_b^k = \delta_t$. Note that this is not in contradiction with Lemma 3. Actually it shows how much an equivalence and a metric-based semantics may differ. The explanation is topological, and it is due to the fact that equivalences (interpreted as functions) differ from 1-bounded pseudometrics by mapping pairs of states to the two-point space $\{0, 1\}$ (with the discrete topology) which is *disconnected*, whereas $[0, 1]$ is *connected*.

5.2. *The Stutter Case*

We show how the construction that led to Theorem 4 can be easily adapted to obtain a net that converges to the strutter trace distance δ_{st}. This proves that the method is general enough to accommodate nontrivial convergence results.

Definition 8. Let $k \geq 1$. An equivalence relation $R \subseteq S \times S$ is a $\equiv_\ell$*-stutter k-probabilistic bisimulation* on $\mathcal{M}$ if whenever $u \; R \; v$, then, for all $E_i \in S/_{\equiv_\ell}$ and $C \in S/_R$,

$$\mathbb{P}(u)(\mathfrak{C}_{\equiv_\ell}(E_0 \cdots E_{k-1} C)) = \mathbb{P}(v)(\mathfrak{C}_{\equiv_\ell}(E_0 \cdots E_{k-1} C)) \,.$$

Two states $u, v \in S$ are $\equiv_\ell$*-stutter k-probabilistic bisimilar*, written $u \sim_{sb}^k v$, if they are related by some $\equiv_\ell$-stutter k-probabilistic bisimulation.

The above definition weakens that of k-probabilistic bisimulation by testing the equality between the probabilities on u and v only on $\equiv_\ell$-stutter cylinders.

It is easy to show that, for all $k \geq 1$, $\sim_b^k \subseteq \sim_{sb}^k$. Note that, $\sim_{sb}^k \not\subseteq \sim_b^k$ (in Fig. 1,

$u \sim_{sb}^1 v$ but $u \not\sim_b^1 v$). In analogy with the strong case, for all $k \geq 1$, $\sim_{sb}^k$ is a $\equiv_\ell$-stutter k-bisimulation, $\sim_{sb}^1 \subseteq \sim_{sb}^k \subseteq \sim_{st}$.

Now we lift these equivalences to pseudometrics by means of a Kantorivich-like operator. For $k \geq 1$, define the *$\equiv_\ell$-stuttered k-steps transition probability function* $\tau_s^k \colon S \to \Delta(S^k)$ as the function s.t., $\tau_s^k(u)$ is the unique probability measure on S^k that, for all $U_i \subseteq S$, $\tau_s^k(u)(U_1 \cdots U_k) = \mathbb{P}(u)(\mathfrak{C}_{\equiv_\ell}(uU_1 \cdots U_k))$ (i.e., $\tau_s^k(u) = \mathbb{P}(u)[\mathsf{pf}^k_{\equiv_\ell} \circ \mathsf{tl}^1_{\equiv_\ell}]$). Define, for $d \colon S \times S \to [0,1]$ pseudometric,

$$\Psi^k(d)(u,v) = \begin{cases} 1 & \text{if } u \not\equiv_\ell v \\ \mathcal{K}(\Lambda^k(d))(\tau_s^k(u), \tau_s^k(v)) & \text{otherwise}. \end{cases}$$

The above extends to the stutter case the k-Kantorovich operator. Clearly, Ψ^k is monotonic, so that, by Tarski fixed point theorem, it has a least fixed point, denoted by δ_{sb}^k.

Due to the following result we call δ_{sb}^k the *$\equiv_\ell$-stutter k-bisimilarity distance.*

Lemma 6 (Stutter k-Bisimilarity Distance). *$u \sim_{sb}^k v$ iff $\delta_{sb}^k(u,v) = 0$.*

Proof. Similar to Lemma 3. □

Next we provide a characterization of δ_{sb}^k by means of the notion of coupling structure, now modified to accommodate the notion of $\equiv_\ell$-stutter step.

Definition 9. A function $\mathcal{C} \colon S \times S \to \Delta(S^k \otimes S^k)$ is a *stutter coupling structure of rank* $k \geq 1$ if, for all $u, v \in S$, $\mathcal{C}(u,v) \in \Omega(\tau_s^k(u), \tau_s^k(v))$.

Hereafter, $\mathbb{C}_k^s$ denotes the set of stutter coupling structures of rank k.

Denote by $st(S^\omega)$ the measurable space over S^ω with σ-algebra $\sigma(\mathfrak{C}_{\equiv_\ell}(2^S))$. The stutter coupling structures are used to define measures in the product space $st(S^\omega) \otimes st(S^\omega)$.

Definition 10. For $k \geq 1$ and $\mathcal{C} \in \mathbb{C}_k^s$, let $\mathbb{P}_\mathcal{C} \colon S \times S \to \Delta(st(S^\omega) \otimes st(S^\omega))$ be such that, for all $u, v \in S$, $\mathbb{P}_\mathcal{C}(u,v)$ is the unique probability measure on $st(S^\omega) \otimes st(S^\omega)$ such that, for all, $n \geq 1$ and $U_i, V_i \subseteq S$ $(i = 0..nk)$

$$\mathbb{P}_\mathcal{C}(u,v)(\mathfrak{C}_{\equiv_\ell}(U_{0,nk}) \times \mathfrak{C}_{\equiv_\ell}(V_{0,nk})) = \mathbb{1}_{U_0 \times V_0}(u,v) \cdot \int \mathbb{P}_\mathcal{C}(\cdot)(\mathfrak{C}_{\equiv_\ell}(U_{k,nk}) \times \mathfrak{C}_{\equiv_\ell}(V_{k,nk}))\, \mathrm{d}\omega\,,$$

where, $U_{i,j} = U_i \cdots U_j$ (similarly for V) and ω is the unique (subprobability) measure on $S \otimes S$ s.t., for all $A, B \subseteq S$, $\omega(A \times B) = \mathcal{C}(u,v)(U_{1,k-1}A \times V_{1,k-1}B)$.

The following gives a characterization of the k-stutter bisimilarity pseudometric δ_{sb}^k in terms of stutter coupling structures. Note that, by Proposition 1, $\simeq$ is the inseparability relation w.r.t. $\sigma(\mathcal{ST})$ and, since LTL$^{\text{-X}}$ is countable, it holds $\simeq \in \sigma(\mathcal{ST}) \otimes \sigma(\mathcal{ST})$.

Lemma 7 (Coupling Lemma). *$\delta_{sb}^k(u,v) = \inf \{\mathbb{P}_\mathcal{C}(u,v)(\not\simeq) \mid \mathcal{C} \in \mathbb{C}_k^s\}$.*

Proof. Similar to Lemma 4. □

According to Theorem 3 what follows suffices to prove the convergence.

Lemma 8. Let $u, v \in S$ be a pair of states of an MC $\mathcal{M}$. Then,

(i) for $k \geq 1$ and $\mathcal{C} \in \mathbb{C}_k^s$, $\mathbb{P}_\mathcal{C}(u,v) \in \Omega(\tilde{\mathbb{P}}(u), \tilde{\mathbb{P}}(v))$;

(ii) for $k, h \geq 1$, $\{\mathbb{P}_{\mathcal{C}}(u,v) \mid \mathcal{C} \in \mathbb{C}^s_k\} \subseteq \{\mathbb{P}_{\mathcal{C}}(u,v) \mid \mathcal{C} \in \mathbb{C}^s_{hk}\}$;
(iii) $\bigcup_{k\geq 1} \{\mathbb{P}_{\mathcal{C}}(u,v) \mid \mathcal{C} \in \mathbb{C}^s_k\}$ is dense in $\Omega(\tilde{\mathbb{P}}(u), \tilde{\mathbb{P}}(v))$ w.r.t. the total variation,

where $\tilde{\mathbb{P}}(u)$ is the restriction of $\mathbb{P}(u)$ on the sub-σ-algebra $\sigma(\mathfrak{C}_{\equiv_\ell}(2^S))$.

Proof. Similar to Lemma 5 □

The next result is a direct consequence of Theorem 3, Lemmas 7, and 8.

Theorem 5 (Convergence). The net $(\delta^k_{sb})_{k\in\mathbb{K}}$ converges point-wise to δ_{st}.

6. Approximation Schema for the Linear Distances

In this section we provide each of the two trace distances (strong and stutter) with an approximation schema, that is, a pair of sequences of pseudometrics that converge from below and above to them. We show that each lower- and upper- approximant is computable in polynomial time in the size of the MC.

In the following, we assume that $\mathcal{M}$ has a finite set of states and a rational transition probability function, that is, $\tau(u)(v) \in \mathbb{Q} \cap [0,1]$ for all $u, v \in S$. The size of $\mathcal{M}$ is determined by the sum of the size of the binary representation of its components. Under this restrictions the pseudometrics proposed in this section have finite domain and image in $\mathbb{Q}$. They are computable if they can be computed on all their domain.

6.1. *The Strong Case*

6.1.1. *Lower-Approximants.* The sequence of lower-approximants will be defined by restricting the set of measurable sets over which δ_t evaluates the differences in the probabilities. Formally, for $k \geq 1$, let $\mathcal{E}_k$ be the set of all finite unions of cylinders in $\mathfrak{C}^k(S/_{\equiv_\ell})$. We define the pseudometrics $l^k : S \times S \to [0,1]$ as follows

$$l^k(u,v) = \max_{E\in\mathcal{E}_k} |\mathbb{P}(u)(E) - \mathbb{P}(v)(E)|$$

The following lemma states that the sequence $(l^k)_{k\geq 1}$ is increasing and that it converges point-wise to the trace distance δ_t.

Lemma 9. For all $k \geq 1$, $l^k \leq l^{k+1}$ and $\delta_t = \sup_{k\geq 1} l^k$.

Proof. $l^k \leq l^{k+1}$ follows by $\mathcal{E}_k \subseteq \mathcal{E}_{k+1}$. The equality $\delta_t = \sup_{k\geq 1} l^k$ is a consequence of (BBLM15, Theorem 6) and the fact that $\bigcup_{k\geq 1} \mathcal{E}_k$ is a field generating $\sigma(\mathcal{T})$. □

By looking at its definition, it is not clear whether l^k can be computed in polynomial time in the size of $\mathcal{M}$. Indeed, the maximum ranges over a set whose cardinality may be exponential in $|S^k|$ in the worst case. The following characterization shows that to compute l^k we do not need to evaluate the probabilities on all the elements of $\mathcal{E}_k$ but only on the cylinders of rank k up to label equivalence, namely, $\mathfrak{C}^k(S/_{\equiv_\ell})$.

Proposition 3. $l^k(u,v) = \frac{1}{2} \sum_{C\in\mathfrak{C}^k(S/_{\equiv_\ell})} |\mathbb{P}(u)(C) - \mathbb{P}(v)(C))|$.

Proof. Note that $\mathcal{E}_k$ is finite and closed under complement. Let $\mathcal{F}$ be the family of cylinders $C \in \mathfrak{C}^k(S/_{\equiv_\ell})$ s.t. $\mathbb{P}(u)(C) \geq \mathbb{P}(v)(C)$. By Hahn's decomposition theorem, for $F = \bigcup \mathcal{F}$ we have $\mathbb{P}(u)(F) - \mathbb{P}(v)(F) = \max_{E \in \mathcal{E}_k} |\mathbb{P}(u)(E) - \mathbb{P}(v)(E)|$. Then

$$\begin{aligned}
2 \cdot l^k(u,v) &= \mathbb{P}(u)(F) - \mathbb{P}(v)(F) \\
&= 2 \cdot \textstyle\sum_{C \in \mathcal{F}} \mathbb{P}(u)(C) - \mathbb{P}(v)(C) && (\sigma\text{-additive}) \\
&= \textstyle\sum_{C \in \mathcal{F}} (\mathbb{P}(u)(C) - \mathbb{P}(v)(C)) + (\mathbb{P}(v)(C^c) - \mathbb{P}(u)(C^c)) && (\text{compl.}) \\
&= \textstyle\sum_{C \in \mathfrak{C}^k(S/_{\equiv_\ell})} |\mathbb{P}(u)(C) - \mathbb{P}(v)(C))| \,, && (\mathcal{F} \cup \mathcal{F}^c = \mathfrak{C}^k(S/_{\equiv_\ell}))
\end{aligned}$$

where the second equality holds since $\mathbb{P}(s)(C^c) = 1 - \mathbb{P}(s)(C)$ for all $s \in S$ and $C \in \mathfrak{C}^k(S/_{\equiv_\ell})$. □

The cylinders in $\mathfrak{C}^k(S/_{\equiv_\ell})$ are all those of the form $\mathfrak{C}(U_0..U_k)$, for some $U_i \in S/_{\equiv_\ell}$ $(i = 0..k)$, therefore the number such cylinders is bounded by $|S|^{k+1}$. Hence, if we show that given $s \in S$, we can to compute $\mathbb{P}(s)(\mathfrak{C}(U_0..U_k))$ in polynomial time, Proposition 3 tells us that also $l^k(u,v)$ can be computed in polynomial time.

To compute $\mathbb{P}(s)(\mathfrak{C}(U_0..U_k))$ we employ a variant of the forward algorithm, by defining the function $f : S \times \{0, \dots, k\} \to [0,1]$ as

$$f(u,i) = \mathbb{P}(s)(\mathfrak{C}(U_0..U_i) \cap (\cdot[i])^{-1}(u))$$

that is, the probability of having emitted a trace in $\mathfrak{C}(U_0..U_i)$ such that the i-th state is u by starting from the state s. The function f can be calculated using the following dynamic programming recurrence:

$$f(u,i) = \begin{cases} \mathbb{P}(s)(\mathfrak{C}(U_0) \cap \mathfrak{C}(u)) & \text{if } i = 0 \\ \mathbb{1}_{U_i}(u) \cdot \sum_{v \in S} f(v, i-1) \cdot \tau(v)(u) & \text{otherwise} \end{cases}$$

Finally, we can compute $\mathbb{P}(s)(\mathfrak{C}(U_0..U_k))$ as $\sum_{u \in S} f(u,k)$.

Theorem 6. l^k can be computed in polynomial time in the size of $\mathcal{M}$.

6.1.2. *Upper-Approximants.* The decreasing sequence $(u^k)_{k \geq 1}$ of upper-approximants converging to δ_t simply derives from the net of k-bisimilarity pseudometrics presented in Section 5. and is defined by $u^k = \delta_b^{2^{k-1}}$ (actually, any infinite subsequence of $(\delta^k)_{k \in \mathbb{K}}$ is fine). The actual contribution of this section is to show that, for all $k \geq 1$, the k-bisimilarity distance δ_b^k can be characterized as the optimal solution of a linear program that can be constructed and solved in polynomial time in the size of the MC.

Our linear program characterization leverages on a *dual* linear program characterization of the Kantorovich distance. For X finite, $d : X \times X \to [0,1]$ a pseudometric and $\mu, \nu \in \Delta(X)$, the value of $\mathcal{K}(d)(\mu,\nu)$ coincides with the optimal value of the following

$$
\begin{array}{ll}
\operatorname*{argmax}_{d,\alpha} \sum_{u,v \in S} d_{u,v} & \\
\quad d_{u,v} = 0 & \forall u,v \in S.\, u \sim_b^k v \\
\quad d_{u,v} = 1 & \forall u,v \in S.\, u \not\equiv_\ell v \\
\quad d_{u,v} = \sum_{x \in S^k} \big(\tau^k(u)(x) - \tau^k(v)(x)\big)\alpha_x^{u,v} & \forall u,v \in S.\, u \equiv_\ell v \text{ and } u \not\sim_b^k v \\
\quad \alpha_x^{u,v} - \alpha_y^{u,v} \leq d_{x_k,y_k} & \forall u,v \in S\, \forall x,y \in S^k.\, \forall i.\, x_i \equiv_\ell y_i \\
\quad \alpha_x^{u,v} - \alpha_y^{u,v} \leq 1 & \forall u,v \in S\, \forall x,y \in S^k.\, \exists i.\, x_i \not\equiv_\ell y_i
\end{array}
$$

Fig. 4. Linear program characterization of the k-bisimilarity distance δ_b^k.

linear programs.

PRIMAL	DUAL
$\min_\omega \sum_{x,y \in X} d(x,y) \cdot \omega_{x,y}$ $\sum_y \omega_{x,y} = \mu(x) \quad \forall x \in X$ $\sum_x \omega_{x,y} = \nu(y) \quad \forall y \in X$ $\omega_{x,y} \geq 0 \quad \forall x,y \in X$	$\max_\alpha \sum_{x \in X} (\mu(x) - \nu(x)) \cdot \alpha_x$ $\alpha_x - \alpha_y \leq d(x,y) \quad \forall x,y \in X$

Consider the linear program in Figure 4, hereafter denoted by D. Note that for an optimal solution of D the value of the unknown $d \in \mathbb{R}^{S \times S}$ is maximized at each component. Therefore, for an optimal solution of D it holds that, if $u \equiv_\ell v$ and $u \not\sim_b^k v$, the maximal value of $d_{u,v}$ is achieved at $\mathcal{K}(\Lambda^k(d))(\tau^k(u), \tau^k(v))$. Otherwise, $d_{u,v} = 1$ when $u \not\equiv_\ell v$, and $d_{u,v} = 0$ when $u \sim_b^k v$. Thus, any optimal solution of D induces a fixed point for Θ^k whose kernel coincides with $\sim_b^k$. In fact, an optimal solution of D characterizes the greatest fixed point of the operator $\Upsilon^k \colon [0,1]^{S \times S} \to [0,1]^{S \times S}$ defined as

$$
\Upsilon^k(d)(u,v) = \begin{cases} 0 & \text{if } u \sim_b^k v \\ \Theta^k(d)(u,v) & \text{otherwise}\,. \end{cases}
$$

Lemma 10. Υ^k has a unique fixed point that coincides with δ_b^k.

Proof. We fist prove that Υ^k has a unique fixed point. Assume that d and d' are two fixed points for Υ^k such that $d' > d$. Consider the relation $R \subseteq S \times S$ defined as

$$
u \mathrel{R} v \quad \text{iff} \quad d'(u,v) - d(u,v) = \|d - d'\|_\infty\,.
$$

Note that $R \cap \sim_b^k = \emptyset$, otherwise we would have the following contradiction

$$
\begin{aligned}
\|d - d'\|_\infty &= d'(u,v) - d(u,v) && \text{(def. } R\text{)} \\
&= \Upsilon^k(d')(u,v) - \Upsilon^k(d)(u,v) && \text{(by } d' = \Upsilon^k(d') \text{ and } d = \Upsilon^k(d)\text{)} \\
&= 0 - 0 && \text{(def. } \Upsilon^k\text{)} \\
&< \|d - d'\|_\infty\,. && (d' > d)
\end{aligned}
$$

With a similar argument one can also show that $R \cap \not\equiv_\ell = \emptyset$.

Consider $u, v \in S$ such that $u \; R \; v$, then the following inequality hold

$$\begin{aligned}
&\|d - d'\|_\infty \\
&= d'(u,v) - d(u,v) && \text{(def. } R\text{)} \\
&= \Upsilon^k(d')(u,v) - \Upsilon^k(d)(u,v) && (d' = \Upsilon^k(d') \text{ and } d = \Upsilon^k(d)) \\
&= \Theta^k(d')(u,v) - \Theta^k(d)(u,v) && \text{(def. } \Upsilon^k \text{ and } R \cap \sim_b^k = \emptyset) \\
&= \mathcal{K}(\Lambda^k(d'))(\tau^k(u), \tau^k(v)) - \mathcal{K}(\Lambda^k(d))(\tau^k(u), \tau^k(v)) && \text{(def. } \Theta^k \text{ and } R \cap \not\equiv_\ell = \emptyset)
\end{aligned}$$

Let $\omega \in \Omega(\tau^k(u), \tau^k(v))$ be s.t. $\mathcal{K}(\Lambda^k(d))(\tau^k(u), \tau^k(v)) = \int \Lambda^k(d) \, \mathrm{d}\omega$, thus we have

$$\begin{aligned}
&= \mathcal{K}(\Lambda^k(d'))(\tau^k(u), \tau^k(v)) - \int \Lambda^k(d) \, \mathrm{d}\omega \\
&\leq \int \Lambda^k(d') \, \mathrm{d}\omega - \int \Lambda^k(d) \, \mathrm{d}\omega && \text{(def. } \mathcal{K}\text{)} \\
&\leq \int (\Lambda^k(d') - \Lambda^k(d)) \, \mathrm{d}\omega && \text{(linearity)}
\end{aligned}$$

let $E = \big\{(u_1..u_k, v_1..v_k) \in S^k \times S^k \mid \forall i \leq n \,.\, u_i \equiv_\ell v_i\big\}$, then

$$\begin{aligned}
&= \textstyle\int_E (\Lambda^k(d') - \Lambda^k(d)) \, \mathrm{d}\omega && \text{(def. } \Lambda^k\text{)} \\
&\leq \textstyle\int_E \|d - d'\|_\infty \, \mathrm{d}\omega && \text{(def. } \Lambda^k\text{)} \\
&\leq \|d - d'\|_\infty
\end{aligned}$$

Summarizing, we have that

$$\|d - d'\|_\infty \leq \textstyle\int_E \|d - d'\|_\infty \, \mathrm{d}\omega \leq \|d - d'\|_\infty$$

Since $\|d - d'\|_\infty > 0$, the above inequality implies that

$$\mathit{support}(\omega) \subseteq E \qquad \text{and} \qquad \{(u_k, v_k) \mid (u_1..u_k, v_1..v_k) \in E\} \subseteq R \,. \tag{7}$$

Let $R^* \subseteq S \times S$ be the smallest equivalence relation s.t. $R \subseteq R^*$. By the fact $\omega \in \Omega(\tau^k(u), \tau^k(v))$ and that $\mathit{support}(\omega) \subseteq E$ we have that, for all $E_i \in S/_{\equiv_\ell}$ $(i = 0 \dots k-1)$ and $C \in S/_{R^*}$ the following holds

$$\mathbb{P}(u)(\mathfrak{C}(E_0 \cdots E_{k-1} C)) = \mathbb{P}(v)(\mathfrak{C}(E_0 \cdots E_{k-1} C)) \,. \tag{8}$$

Before proving Equation (8), note that $R \subseteq \equiv_\ell$ (otherwise $\|d - d'\|_\infty > 0$), thus we have that either $\{u, v\} \in E_0$ or $\{u, v\} \notin E_0$. On the one hand, if $\{u, v\} \notin E_0$, by definition of $\mathbb{P}$, we have $\mathbb{P}(u)(\mathfrak{C}(E_0 \cdots E_{k-1} C)) = \mathbb{P}(v)(\mathfrak{C}(E_0 \cdots E_{k-1} C)) = 0$.

On the other hand, if $\{u, v\} \in E_0$ we have

$$\begin{aligned}
&\mathbb{P}(u)(\mathfrak{C}(E_0 \cdots E_{k-1} C)) && \text{(def. } \mathbb{P}\text{)} \\
&= \tau^k(u)(E_1..E_{k-1} C) && \text{(def. } \tau^k\text{)} \\
&= \omega(E_1..E_{k-1} C, S^k) && \text{(left marginal)} \\
&= \omega(E_1..E_{k-1} C, E_1..E_{k-1} C) && \text{(by (7))} \\
&= \omega(S^k, E_1..E_{k-1} C) && \text{(by (7))} \\
&= \tau^k(v)(E_1..E_{k-1} C) && \text{(right marginal)} \\
&= \mathbb{P}(v)(\mathfrak{C}(E_0 \cdots E_{k-1} C)) && \text{(def. } \mathbb{P} \text{ and } \tau^k\text{)}
\end{aligned}$$

Therefore R^* is a k-probabilistic bisimulation. This means that $u \sim_b^k v$, which is in contradiction with the fact that $R \cap \sim_b^k = \emptyset$.

It remains to prove that δ_k is a fixed point for Υ^k, namely $\Upsilon^k(\delta_k)(u,v) = \delta_k(u,v)$ for all $u, v \in S$. Take $u, v \in S$. On the one hand, if $u \sim_b^k v$ the following equalities hold

$$\begin{aligned} \Upsilon^k(\delta_k)(u,v) &= 0 && \text{(def. } \Upsilon^k\text{)} \\ &= \delta_k(u,v)\,. && \text{(Lemma 3)} \end{aligned}$$

On the other hand, if $u \not\sim_b^k v$ the following equalities hold

$$\begin{aligned} \Upsilon^k(\delta_k)(u,v) &= \Theta^k(\delta_k)(u,v) && \text{(def. } \Upsilon^k\text{)} \\ &= \delta_k(u,v)\,. && \text{(def. } \delta_k\text{)} \end{aligned}$$

This completes the proof. □

Lemma 10 implies that for any optimal solution of D, $d_{u,v} = \delta_b^k(u,v)$, for all $u, v \in S$.

Note that D has a number of constraints bounded by $O(|S|^2 + |S|^{2k+2})$ and a number of unknowns bounded by $O(|S|^2 + |S|^{k+2})$. Moreover, the following lemma ensures that the linear program D can be constructed in polynomial time, provided that k is a constant.

Lemma 11. $\sim_b^k$ can be computed in polynomial time in the size of $\mathcal{M}$.

Proof. Let $\chi \colon \{0,1\}^{S\times S} \to \{0,1\}^{S\times S}$ be defined as

$$\chi(d)(u,v) = \begin{cases} 1 & \text{if } \Theta^k(d)(u,v) > 0 \\ 0 & \text{otherwise} \end{cases}$$

The domain $\{0,1\}^{S\times S}$ endowed with the usual point-wise preorder is a finite lattice where any (strictly) increasing chain has at most $|S|^2$ elements. The operator χ is monotonic since Θ^k is so, thus it has least fixed point, say d.

We show that $\ker(d) = \sim_b^k$. By Lemma 3 it suffices to prove $\ker(d) = \ker(\delta_b^k)$.

($\subseteq$) By def. of χ, $\Theta^k(d) \leq d$ thus, by Knaster-Tarski's fixed point theorem, $\delta_b^k \leq d$, hence $\ker(d) \subseteq \ker(\delta_b^k)$.

($\supseteq$) Let $d' \colon S \times S \to \{0,1\}$ be defined as $d'(u,v) = 0$ if $\delta_b^k(u,v) = 0$, and 1 otherwise. From the definition of χ we easily obtain $\chi(d') \leq d'$, which implies $d \leq d'$. Therefore $\ker(\delta_b^k) = \ker(d') \subseteq \ker(d)$.

Having established that $\ker(d) = \sim_b^k$, we notice that by Kleene fixed-point theorem d can be computed by iterating the application of χ at most $|S|^2$ times starting from the bottom element, $\bot(u,v) = 0$, for all $u, v \in S$, that is to say $d = \chi^{|S|^2}(\bot)$. We show that each application of χ can be performed in polynomial time in the size of $\mathcal{M}$. Given $d' \colon S \times S \to [0,1]$ and $u, v \in S$, $\Theta^k(d')(u,v)$ is computed in constant time if $u \not\equiv_\ell v$, whereas, for $u \equiv_\ell v$ it coincides with the optimal value of the following linear program:

$$\begin{aligned} \mathcal{K}(\Lambda^k(d'))(\tau^k(u), \tau^k(v)) = \max_{\alpha} & \textstyle\sum_{x \in S^k} (\tau^k(u)(x) - \tau^k(v)(x)) \cdot \alpha_x && (9) \\ & \alpha_x - \alpha_y \leq \Lambda^k(d')(x,y) \qquad \forall x, y \in S^k\,. \end{aligned}$$

Note that the above linear program has $|S|^k$ unknowns and $|S|^{2k}$ constraints. Since k is constant, (9) can be computed in polynomial time in the size of $\mathcal{M}$ using e.g., the ellipsoid method. Finally, $\chi(d')$ can be computed within the time it takes to compute $\Theta^k(d)(u,v)$ for each $u,v \in S$, and checking whether its value is greater than zero. □

Remark 6. Providing an efficient algorithm for computing for computing $\sim_b^k$ is out of the scope of the present paper. The interested reader may consider to look at more efficient techniques such as partition refinement (BEM00; PT87).

The following result states that the k-bisimilarity distance can be computed in polynomial time, provided that k is a constant.

Theorem 7. δ_b^k can be computed in polynomial time in the size of $\mathcal{M}$.

Proof. By Lemma 11, D can be constructed in polynomial time. The number of constraints and unknowns in D is bounded by a polynomial in the size of $\mathcal{M}$. Hence, the linear program D can be solved in polynomial time using the ellipsoid method. □

6.2. *The Stutter Case*

As one may expect, the sequences $(l_{st}^k)_{k\geq 1}$ and $(u_{st}^k)_{k\geq 1}$ of lower- and upper-approximants for the stutter trace distance δ_{sb} can be defined similarly to those we have shown in the previous section for the strong case. Specifically, for $k \geq 1$

$$l_{st}^k(u,v) = \max_{E\in\mathcal{S}_k} |\mathbb{P}(u)(E) - \mathbb{P}(v)(E)| \qquad \text{and} \qquad u_{st}^k(u,v) = \delta_{st}^{2^{k-1}}\,,$$

where $\mathcal{S}_k$ is the set of all finite unions of stutter trace cylinders in $\mathfrak{C}_{\equiv_\ell}^k(S/_{\equiv_\ell})$.

Convergence and (anti)monotonicity of the sequences follow exactly as before. However, what is not immediate is the proof that, for all $k \geq 1$, l_{st}^k and u_{st}^k can actually be computed in polynomial time. The first difficulty arises, when for computing l_{st}^k, we try to apply the characterization provided by Lemma 9:

$$l^k(u,v) = \tfrac{1}{2} \textstyle\sum_{C\in\mathfrak{C}_{\equiv_\ell}^k(S)} |\mathbb{P}(u)(C) - \mathbb{P}(v)(C))|\,.$$

The thin cylinders in $\mathfrak{C}_{\equiv_\ell}^k(S)$ are of the form $\mathfrak{C}(w)$, for some $w \in A_1^* \cdots A_k^*$ and $A_i \in S/_{\equiv_\ell}$ ($i = 1..k$), hence $\mathfrak{C}_{\equiv_\ell}^k(S)$ is not finite (the word w can be arbitrarily long). Similarly, as for computing u_{st}^k, if we tried to apply directly the LP characterization in Figure 4 we would have an infinite number of constraints.

To cope with this problem, we propose a reduction from the stutter to the strong case. Formally, we show that, for $k \geq 1$, the problem of computing $\mathbb{P}(u)(\mathfrak{C}_{\equiv_\ell}(u_1..u_k))$ and the k-stutter bisimilarity distance δ_{sb}^k for an MC $\mathcal{M}$ can be reduced to computing $\mathbb{P}(u)(\mathfrak{C}(u_1..u_k))$ and δ_b^k for an MC $\mathcal{N}$ derived from $\mathcal{M}$.

The following lemma states that $\mathcal{N}$ is obtained by replacing the probability transition function τ in $\mathcal{M}$ with the (1-)stutter probability transition function τ_s^1.

Lemma 12. Let $\mathcal{M} = (S,\tau,\ell)$ and $\mathcal{N} = (S,\tau_s^1,\ell)$. Then, for all $k \geq 1$,

(i) $U_i \subseteq S$, $\mathbb{P}_{\mathcal{M}}(u)(\mathfrak{C}_{\equiv_\ell}(uU_1\cdots U_k)) = \mathbb{P}_{\mathcal{N}}(u)(\mathfrak{C}(uU_1\cdots U_k))$;

(ii) $\Psi_{\mathcal{M}}^k = \Theta_{\mathcal{N}}^k$.

Proof. We prove the two point separately.

(i) Let $U_i \subseteq S$ $(1 \leq i \leq k)$. We proceed by induction on k.
Base Case ($k = 1$):

$$\begin{aligned}
&\mathbb{P}_{\mathcal{M}}(u)(\mathfrak{C}_{\equiv_\ell}(uU_1)) = \tau_s^1(u)(U_1) && \text{(def. } \tau_s^1\text{)}\\
&= \textstyle\int \mathbb{1}_{U_1} \, \mathrm{d}\tau_s^1(u) \\
&= \textstyle\int \mathbb{P}_{\mathcal{N}}(\cdot)(\mathfrak{C}(U_1)) \, \mathrm{d}\tau_s^1(u) && \text{(by def. } \mathbb{P}(),\ \mathbb{P}_{\mathcal{N}}(\cdot)(\mathfrak{C}(U_1)) = \mathbb{1}_{U_1}\text{)}\\
&= \mathbb{P}_{\mathcal{N}}(u)(\mathfrak{C}(uU_1))\,. && \text{(def. } \mathbb{P}\text{)}
\end{aligned}$$

Inductive Step:

$$\begin{aligned}
&\mathbb{P}_{\mathcal{N}}(u)(\mathfrak{C}(uU_1 \cdots U_k)) \\
&= \textstyle\int \mathbb{P}_{\mathcal{N}}(\cdot)(\mathfrak{C}(U_1 \cdots U_k)) \, \mathrm{d}\tau_s^1(u) && \text{(def. } \mathbb{P}\text{)}\\
&= \textstyle\int \mathbb{P}_{\mathcal{M}}(\cdot)(\mathfrak{C}_{\equiv_\ell}(U_1 \cdots U_k)) \, \mathrm{d}\tau_s^1(u) && \text{(ind. hp.)}\\
&= \mathbb{P}_{\mathcal{M}}(u)(\mathfrak{C}_{\equiv_\ell}(uU_1 \cdots U_k)) && \text{(def. } \tau_s^k \text{ and } \mathbb{P}\text{)}
\end{aligned}$$

(ii) Given $k \geq 1$, by (i) we have that $(\tau(u)_s^1)^k = \tau(u)_s^k$, for all $u \in S$. Therefore, the equality $\Psi_{\mathcal{M}}^k = \Theta_{\mathcal{N}}^k$ follows by definition of Ψ^k and Θ^k. □

Remark 7. Lemma 12 may be used to provide an alternative proof for the convergence of the stutter k-bisimilarity distances to the stutter trace distance (Theorem 5) by reducing it to the convergence obtained in the strong case (Theorem 4). In this way Section 5.2 may be simplified significantly, avoiding the need of proving some preliminary lemmas that lead us to the proof of Theorem 5. However, we present the results for the stutter case as in Section 5.2 to demonstrate the generality of the proof technique that lead to Theorem 4 in a seemingly more complex case.

Next we show that $\mathcal{N}$ can be constructed in polynomial time and its size is polynomial in the size of $\mathcal{M}$. Consider the problem of computing $\tau_s^1(u)(v)$.

We consider two possible cases:

Case $u \not\equiv_\ell v$**.** By definition $\tau_s^1(u)(v) = \mathbb{P}_{\mathcal{M}}(u)(\mathfrak{C}([u]_{\equiv_\ell}^+ v))$. This is the probability of reaching the state v starting from u visiting only states in $[u]_{\equiv_\ell}$ prior to reaching v. Using LTL-like notations, this can be written as $\mathbb{P}_{\mathcal{M}}(u)([u]_{\equiv_\ell} \mathsf{U} \{v\})$. This is a well studied probabilistic model checking problem that can be solved in polynomial time in the size of $\mathcal{M}$ as the solution of a linear system of equations (see e.g. (BK08, §10.1.1 p.762)).

Case $u \equiv_\ell v$**.** By definition $\tau_s^1(u)(v) = \mathbb{P}_{\mathcal{M}}(u)(uv[v]_{\equiv_\ell}^\omega)$. This corresponds to the probability of making a transition from u to v and, from v, generating an infinite run that never escapes from the $\equiv_\ell$-equivalence class of v, i.e., $\tau(u)(v) \cdot \mathbb{P}(v)([v]_{\equiv_\ell}^\omega)$. The probability $\mathbb{P}_{\mathcal{M}}(v)([v]_{\equiv_\ell}^\omega)$ can be conveniently computed as $1 - \sum_{x \not\equiv_\ell u} \tau_s^1(v)(x)$, reusing the probabilities computed in the previous case.

Therefore $\mathcal{N}$ can be constructed in polynomial time in the size of $\mathcal{M}$.

Lemma 13. $\mathcal{N} = (S, \tau_s^1, \ell)$ has size polynomial in the size of $\mathcal{M}$.

Proof. It suffices to show that τ_s^1 is rational of size polynomial in the size of $\mathcal{M}$. Let $u, v \in S$. If $u \not\equiv_\ell v$ then $\tau_s^1(u)(v) = \mathbb{P}_{\mathcal{M}}(u)([u]_{\equiv_\ell} \mathsf{U} \{v\})$. Its value is the solution of a system of linear equations where the coefficients are some transition probabilities taken from $\mathcal{M}$ (or a sum of them). Therefore, $\tau_s^1(u)(v)$ is an intersection of hyperplanes given by some equalities with rational coefficients whose size is bounded in the size of $\mathcal{M}$. Thus, we conclude that $\tau_s^1(u)(v)$ is rational of size polynomial in size of $\mathcal{M}$. The case $u \not\equiv_\ell v$ follows by the previous one, since $\tau_s^1(u)(v) = \tau(u)(v) \cdot (1 - \sum_{x \not\equiv_\ell u} \tau_s^1(v)(x))$. □

By Lemmas 12 and 13, and Theorems 6 and 7, the following holds.

Theorem 8. l_{sb}^k and δ_{sb}^k can be computed in polynomial time in the size of $\mathcal{M}$.

Remark 8. Theorems 7 and 8 do not contradict the fact that the problem of approximating the trace distances up to a given precision $\epsilon > 0$ is NP-hard (CMR07). Indeed, this requires one to compute the lower and upper approximants l_*^k and δ_*^k ($* \in \{b, sb\}$), for increasing values of k, until $\delta_*^k - l_*^k < \epsilon$. Note that the time-complexity of this procedure increases exponentially in the value of k.

7. Conclusions and Future Work

In this paper we provide the strong and stutter trace distances with a logical characterization in terms of LTL and LTL$^{-\mathsf{X}}$ formulas, respectively. These characterizations, differently from other proposals, relate these behavioral distances to the probabilistic model checking problem over MCs.

Then, we proposed a family of behavioral equivalences, namely probabilistic k-bisimilarities, that weaken probabilistic bisimilarity of Larsen and Skou on MCs. These equivalences are in turn generalized to pseudometrics by means of a fixed point definition that uses a generalized Kantorovich operator. These pseudometrics are shown to form a net that converges point-wise to the trace distance. Remarkably, to prove this convergence we extended and improved two important results in (CvBW12), namely, Theorem 8 and Corollary 11. The proposed construction is shown to be general enough to accommodate a second nontrivial convergence result between a net of suitable stutter variants of k-bisimilarities pseudometrics and the stutter trace distance. These convergences are interesting, because they reveal a nontrivial relation between branching and linear-time metric-based semantics that in Remark 3 is shown not to hold when the standard equivalence-based semantics on MCs are used instead.

The above distances are then used to provide the strong and stutter trace distances with an approximation schema, that is, two sequences of pseudometrics that converge from above and below to the two respective linear distances. Each of these lower and under-approximants are shown to be computable in polynomial time in the size of the MC. Notably, for this proof the under-approximants of the trace distance (i.e., the k-bisimilarity pseudometrics) are given a characterization in terms of optimal solutions of a linear program that have size polynomial in the MC. The one we propose generalizes and improves the linear program characterization presented in (CvBW12, Eq. 8) for the (undiscounted) bisimilarity pseudometric of Desharnais et al. that, in contrast, has a number of constraints exponential in the size of the MC. Moreover, our approximation schema improves that in (CK14), both for the generality of its applicability and in terms of computational complexity.

Some natural questions now are: (i) to see if the on-the-fly algorithm for the computation of bisimilarity distance in (BBLM13) can be used to compute the k-bisimilarity distances and their stutter variants; (ii) whether this approximation technique carries over to models with non-determinism, such as MDPs (where a result by Fu (Fu12) gives new insight on how to obtain minimal information in case the distance is not a bisimilarity metric, and where the PSPACE-complexity result is sharpened to NP $\cap$ coNP); (iii) whether a similar construction can be applied to stochastic models with continuous time, such as CTMCs.

References

Giorgio Bacci, Giovanni Bacci, Kim G. Larsen, and Radu Mardare. On-the-Fly Exact Computation of Bisimilarity Distances. In *TACAS*, volume 7795 of *LNCS*, pages 1–15, 2013.

Giorgio Bacci, Giovanni Bacci, Kim G. Larsen, and Radu Mardare. On the Total Variation Distance of Semi-Markov Chains. In *FoSSaCS*, volume 9034 of *LNCS*, pages 185–199, 2015.

Christel Baier, Bettina Engelen, and Mila E. Majster-Cederbaum. Deciding bisimilarity and similarity for probabilistic processes. *J. Comput. Syst. Sci.*, 60(1):187–231, 2000.

Christel Baier and Joost-Pieter Katoen. *Principles of Model Checking.* MIT Press, 2008.

Taolue Chen and Stefan Kiefer. On the Total Variation Distance of Labelled Markov Chains. In *CSL-LICS'14*, pages 33:1–33:10. ACM, 2014.

Corinna Cortes, Mehryar Mohri, and Ashish Rastogi. Lp Distance and Equivalence of Probabilistic Automata. *Int. Journal of Foundations of Computer Science*, 18(04):761–779, 2007.

Di Chen, Franck van Breugel, and James Worrell. On the Complexity of Computing Probabilistic Bisimilarity. In *FoSSaCS*, volume 7213 of *LNCS*, pages 437–451. Springer, 2012.

Costas Courcoubetis and Mihalis Yannakakis. Markov decision processes and regular events. In *ICALP*, volume 443 of *Lecture Notes in Computer Science*, pages 336–349. Springer, 1990.

Luca de Alfaro, Marco Faella, and Mariëlle Stoelinga. Linear and branching metrics for quantitative transition systems. In *ICALP*, volume 3142 of *LNCS*, pages 97–109, 2004.

Josee Desharnais, Vineet Gupta, Radha Jagadeesan, and Prakash Panangaden. Metrics for Labeled Markov Systems. In *CONCUR*, volume 1664 of *LNCS*, pages 258–273. Springer, 1999.

Josee Desharnais, Vineet Gupta, Radha Jagadeesan, and Prakash Panangaden. Metrics for labelled Markov processes. *Theoretical Compututer Science*, 318(3):323–354, 2004.

Josee Desharnais, Radha Jagadeesan, Vineet Gupta, and Prakash Panangaden. The Metric Analogue of Weak Bisimulation for Probabilistic Processes. In *LICS*, pages 413–422. IEEE Computer Society, 2002.

Nelson Dunford and Jacob T. Schwartz. *Linear Operators, Part 1, General Theory.* Wiley Classic Library. John Wiley, New York, 1988.

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

Norm Ferns, Prakash Panangaden, and Doina Precup. Metrics for finite Markov Decision Processes. In *UAI*, pages 162–169. AUAI Press, 2004.

Hongfei Fu. Computing game metrics on markov decision processes. In *ICALP*, volume 7392 of *LNCS*, pages 227–238, 2012.

Leslie Lamport. What good is temporal logic? In *IFIP*, pages 657–668, 1983.

Torgny Lindvall. *Lectures on the Coupling Method*. Wiley Series in Probability and Mathematical Statistics. John Wiley, New York, 1992.

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

Kim Guldstrand Larsen and Arne Skou. Bisimulation through probabilistic testing. *Information and Computation*, 94(1):1–28, 1991.

Matteo Mio. Upper-Expectation Bisimilarity and Łukasiewicz μ-Calculus. In *FoSSaCS*, volume 8412 of *LNCS*, pages 335–350, 2014.

Amir Pnueli. The temporal logic of programs. In *SFCS*, pages 46–57. IEEE Computer Society, 1977.

Robert Paige and Robert Endre Tarjan. Three partition refinement algorithms. *SIAM J. Comput.*, 16(6):973–989, 1987.

Moshe Y. Vardi. Automatic verification of probabilistic concurrent finite-state programs. In *26th Annual Symposium on Foundations of Computer Science, Portland, Oregon, USA, 21-23 October 1985*, pages 327–338. IEEE Computer Society, 1985.

Moshe Y. Vardi. Probabilistic linear-time model checking: An overview of the automata-theoretic approach. In *Formal Methods for Real-Time and Probabilistic Systems, 5th International AMAST Workshop, ARTS'99, Bamberg, Germany, May 26-28, 1999. Proceedings*, volume 1601 of *Lecture Notes in Computer Science*, pages 265–276. Springer, 1999.

Appendix A. Technical proofs

This section contains all the technical proofs that have been omitted in the paper.

Let recall the definition of the function $q\colon S^\omega \to S^\omega$ given in Proposition 2. For $\pi \in S^\omega$,

$$q(\pi) = \begin{cases} \pi[0]q(\pi|_k) & \text{if } \exists k \text{ s.t. } \pi[0] \not\equiv_\ell \pi[k] \text{ and } \forall j < k,\, \pi[0] \equiv_\ell \pi[j] \\ \pi & \text{otherwise (i.e., } \pi \text{ is } \equiv_{\ell^\omega}\text{-constant)} \end{cases}$$

Proof of Proposition 2. To prove the $\sigma(\mathcal{ST})$-$\sigma(\mathcal{T})$ measurability of q it suffices to show that for all cylinders $\mathfrak{C}(C_1 \cdots C_n) \in \mathcal{T}$, $q^{-1}(\mathfrak{C}(C_1 \cdots C_n)) \in \sigma(\mathcal{ST})$. By definition of q we have that

$$\begin{aligned} q^{-1}(\mathfrak{C}(C_1 \cdots C_n)) &= \{\pi \mid q(\pi) \in \mathfrak{C}(C_1 \cdots C_n)\} && \text{(preimage)} \\ &= \{\pi \mid \exists j_1 \leq \cdots \leq j_n, \text{ such that } \forall k \leq n\, .\pi[j_k] \in C_{j_k}\} && \text{(def. } q) \\ &= \mathfrak{C}_{\equiv_\ell}(C_1 \cdots C_n)\,. && \text{(stutter cylinder)} \end{aligned}$$

Now we show that $R = \{(\pi, \rho) \mid q(\pi) \equiv_{\ell^\omega} q(\rho)\} \subseteq S^\omega \times S^\omega$ is a stutter relation. Assume that $q(\pi) \equiv_{\ell^\omega} q(\rho)$. We check that the three conditions of Definition 4 hold.

(i) By definition of q, $q(\pi)[0] = \pi[0]$ and $q(\rho)[0] = \rho[0]$, and by $\equiv_{\ell^\omega}$, we get $\pi[0] \equiv_\ell \rho[0]$.

(ii) It suffices to prove that, for arbitrary $\pi, \rho \in S^\omega$, the following hold:

(a) π is $\equiv_{\ell^\omega}$-constant iff π is R-constant;

(b) if $\pi\ R\ \rho$ and π is $\equiv_{\ell^\omega}$-constant, then ρ is $\equiv_{\ell^\omega}$-constant.

((a) : $\Leftarrow$) Assume by contradiction that π is R-constant but not $\equiv_{\ell^\omega}$-constant. Then, there exists $k > 0$ such that $\pi[0] \not\equiv_\ell \pi[k]$. By definition of q, $q(\pi)[0] = \pi[0]$ and $q(\pi|_k)[0] = \pi[k]$, therefore $q(\pi) \equiv_{\ell^\omega} q(\pi|_k)$. In particular, this means that $\pi\ R\ \pi|_k$, so that π is not R-constant. This contradicts the hypothesis on π. ((a) : $\Rightarrow$) Assume that π is $\equiv_{\ell^\omega}$-constant. This implies also that, for all $i \in \mathbb{N}$, $\pi|_i$ is $\equiv_{\ell^\omega}$-constant. By definition of q, we have $q(\pi) = \pi$ and, for all $i \in \mathbb{N}$, $q(\pi|_i) = \pi|_i$. Hence, $q(\pi)$ is R-constant.

(b) By $\pi\ R\ \rho$, we have $q(\pi) \equiv_{\ell^\omega} q(\rho)$. Since π is $\equiv_{\ell^\omega}$-constant, by definition of q, $q(\pi) = \pi$, so that $\pi \ell^\omega q(\rho)$. In particular, $q(\rho)$ is $\equiv_{\ell^\omega}$-constant and, by definition of q, this is the case only when ρ is $\equiv_{\ell^\omega}$-constant.

(iii) We show that $q(\pi|_1) \not\equiv_{\ell^\omega} q(\rho)$ and $\pi \not\equiv_{\ell^\omega} q(\rho|_1)$ implies $q(\pi|_1) \equiv_{\ell^\omega} q(\rho|_1)$. By $q(\pi|_1) \not\equiv_{\ell^\omega} q(\rho)$, we have that $\pi[0] \not\equiv_\ell \pi[1]$. Indeed, if $\pi[0] \equiv_\ell \pi[1]$, by definition of q, $q(\pi) = q(\pi|_1)$, and this contradicts the hypothesis $q(\pi) \equiv_{\ell^\omega} q(\rho)$. Similarly, $\pi \not\equiv_{\ell^\omega} q(\rho|_1)$ implies $\rho[0] \not\equiv_\ell \rho[1]$. By definition of q, $\pi[0] \not\equiv_\ell \pi[1]$ and $\rho[0] \not\equiv_\ell \rho[1]$, we have $q(\pi) = \pi[0]q(\pi|_1)$ and $q(\rho) = \rho[0]q(\rho|_1)$. By $q(\pi) \equiv_{\ell^\omega} q(\rho)$ and definition of $\equiv_{\ell^\omega}$, we obtain $q(\pi|_1) \equiv_{\ell^\omega} q(\rho|_1)$.

□

Proof of Lemma 2 —restated from (Lin92, Th.5.2) We prove that $\|\mu - \nu\|$ is a lower bound for $\{\omega(\ncong) \mid \omega \in \Omega(\mu,\nu)\}$. Let $\omega \in \Omega(\mu,\nu)$ and $E \in \Sigma$, then

$$\begin{aligned}
\mu(E) &= \omega(E \times X) && (\omega \in \Omega(\mu,\nu)) \\
&\geq \omega((X \times E) \cap \cong) && (\text{def. } \cong) \\
&= 1 - \omega((X \times E)^c \cup \ncong) && (\text{complement}) \\
&\geq 1 - \omega((X \times E)^c) - \omega(\ncong) && (\text{sub additivity}) \\
&= \omega(X \times E) - \omega(\ncong) && (\text{complement}) \\
&= \nu(E) - \omega(\ncong)\,. && (\omega \in \Omega(\mu,\nu))
\end{aligned}$$

Thus, by the generality of $\omega \in \Omega(\mu,\nu)$ and $E \in \Sigma$, it immediately follows that $\|\mu - \nu\| = \sup_{E\in\Sigma} |\mu(E) - \nu(E)| \leq \min \{\omega(\ncong) \mid \omega \in \Omega(\mu,\nu)\}$.

Now we prove that there exists a coupling $\omega^* \in \Omega(\mu,\nu)$ such that $\omega^*(\ncong) = \|\mu - \nu\|$. Define $\psi \colon X \to X \times X$ by $\psi(x) = (x,x)$ (it is measurable because $\psi^{-1}(E \times E') = E \cap E'$, for all $E, E' \in \Sigma$). Note that $\psi^{-1}(\cong) = X$, since $\psi(x) = (x,x) \in \cong$.

If $\mu = \nu$, just define $\omega^* = \mu[\psi]$ (to check that this is a coupling and that it is such that $\omega^*(\ncong) = \|\mu - \nu\|$ is trivial). Let $\mu \neq \nu$. Define $\mu \wedge \nu \colon \Sigma \to \mathbb{R}_+$ as follows, for $E \in \Sigma$

$$(\mu \wedge \nu)(E) = \inf \{\mu(F) + \nu(E \setminus F) \mid F \in \Sigma \text{ and } F \subseteq E\}\ .$$

The above is a well defined measure (a.k.a. the meet of μ and ν, see (DS88, Corr.6 pp.163)). Now define the following derived measures

$$\eta = \mu - (\mu \wedge \nu)\,, \qquad \eta' = \nu - (\mu \wedge \nu)\,, \qquad \omega^* = \frac{\eta \times \eta'}{1-\gamma} + (\mu \wedge \nu)[\psi]\,.$$

where $\gamma = (\mu \wedge \nu)[\psi](\cong)$. Note that, since $\psi^{-1}(\cong) = X$, $(\mu \wedge \nu)[\psi]$ puts all its mass in $\cong$. Moreover, since $\mu \neq \nu$, we get $\gamma < 1$, so ω^* is well defined and, in particular, $\omega^*(\cong) = \gamma$. Now we show that $\omega^* \in \Omega(\mu,\nu)$. Let $E \in \Sigma$,

$$\begin{aligned}
\omega^*(E \times X) &= \frac{\eta(E) \cdot \eta'(X)}{1-\gamma} + (\mu \wedge \nu)[\psi](E \times S^\omega) && (\text{def. } \omega^*) \\
&= \frac{\eta(E) \cdot (\nu(X) - (\mu \wedge \nu)(X))}{1-\gamma} + (\mu \wedge \nu)[\psi](E \times X) && (\text{def. } \eta') \\
&= \frac{\eta(E) \cdot (1-\gamma)}{1-\gamma} + (\mu \wedge \nu)[\psi](E \times X) && (\text{def. } \mu \wedge \nu) \\
&= \mu(E) - (\mu \wedge \nu)(E) + (\mu \wedge \nu)[\psi](E \times X) && (\text{def. } \eta) \\
&= \mu(E) - (\mu \wedge \nu)(E) + (\mu \wedge \nu)(E) && (\text{def. } (\mu \wedge \nu)[\psi]) \\
&= \mu(E)\,.
\end{aligned}$$

Similarly $\omega^*(X \times E) = \nu(E)$. The following shows that ω^* is optimal

$$\begin{aligned}
\|\mu - \nu\| &= 1 - (\mu \wedge \nu)(X) && (\text{def. } \mu \wedge \nu \text{ and compl.}) \\
&= 1 - (\mu \wedge \nu)[\psi](\cong) && (\text{def. } \psi) \\
&= 1 - \gamma && (\text{def. } \gamma) \\
&= 1 - \omega^*(\cong) && (\text{def. } \omega^*) \\
&= \omega^*(\ncong) && (\text{compl.})
\end{aligned}$$

□

Proposition 4. Let Σ be a σ-algebra on X generated by $\mathcal{F} \subseteq 2^X$. Then the *separability relations* w.r.t. Σ and $\mathcal{F}$ coincide:

$$\ncong_\Sigma := \bigcup \{E \times E^c \mid E \in \Sigma\} = \bigcup \{F \times F^c \mid F \in \mathcal{F}\} =: \ncong_\mathcal{F} .$$

Proof. ($\supseteq$) It immediately follows by $\mathcal{F} \subseteq \Sigma$. ($\subseteq$) Let $\mathcal{U}$ be the smallest family of subsets of X that contains $\mathcal{F}$ and is closed under complement and (generic) union. Define $\ncong_\mathcal{U} := \bigcup \{E \times E^c \mid E \in \mathcal{U}\}$. Clearly $\Sigma \subseteq \mathcal{U}$, hence $\ncong_\Sigma \subseteq \ncong_\mathcal{U}$. This means that to prove the inclusion it suffices to prove $\ncong_\mathcal{F} \supseteq \ncong_\mathcal{U}$, which is equivalent to $\cong_\mathcal{F} \subseteq \cong_\mathcal{U}$. We proceed by contradiction. Assume that $(x, y) \in \cong_\mathcal{F}$ but $(x, y) \notin \cong_\mathcal{U}$. By definition of $\cong_\mathcal{U}$, there exists a set $E \in \mathcal{U}$ such that $x \in E$ and $y \in E^c$. By definition of $\mathcal{U}$, there exist $\mathcal{P}, \mathcal{N} \subseteq \mathcal{F}$ such that $E = \bigcup \mathcal{P} \cup \bigcap \mathcal{N}$. This means that, either $x \in P$ for some $P \in \mathcal{P}$ or $x \in \bigcap \mathcal{N}$. If $x \in P$, by $x \cong_\mathcal{F} y$ and $P \in \mathcal{F}$ we have $y \in P \subseteq E$, hence a contradiction. If, $x \in \bigcap \mathcal{N}$, then by $x \cong_\mathcal{F} y$ and $\mathcal{N} \subseteq \mathcal{F}$, we have that $x \in \bigcap \mathcal{N} \subseteq E$, hence another contradiction. □

Proposition 5. $\not\equiv_{\ell^\omega}$ is the separability relation w.r.t. $\sigma(\mathcal{T})$ and it is a measurable set in $\sigma(\mathcal{T}) \otimes \sigma(\mathcal{T})$, i.e., $\not\equiv_{\ell^\omega} = \ncong_{\sigma(\mathcal{T})} \in \sigma(\mathcal{T}) \otimes \sigma(\mathcal{T})$.

Proof. We first show $\not\equiv_{\ell^\omega} = \ncong_{\sigma(\mathcal{T})} = \bigcup \{E \times E^c \mid E \in \sigma(\mathcal{T})\}$. ($\supseteq$) By , it suffices to prove separability w.r.t trace cylinders. Let $\pi \ncong_{\sigma(\mathcal{T})} \rho$, then, by Proposition 4, there must be a trace cylinder $C \in \mathcal{T}$ such that $\pi \in C$ and $\rho \notin C$. Let $C = \mathfrak{C}(C_1 \cdots C_n)$, for some $C_i \in S/_{\equiv_\ell}$ $(i = 1..n)$. By $\pi \in C$ and $\rho \notin C$, there must be an index $1 \leq j \leq n$ such that $\pi[j] \not\equiv_\ell \rho[j]$, so that $\pi \not\equiv_{\ell^\omega} \rho$. ($\subseteq$) Let $\pi \not\equiv_{\ell^\omega} \rho$, then there exist $k \in \mathbb{N}$ such that $\pi[k] \not\equiv_\ell \rho[k]$. Let $E = (\cdot)|_k^{-1}(\mathfrak{C}([\pi[k]]_{\equiv_\ell}))$. Clearly $\pi \in E$ but $\rho \notin E$. The function $(\cdot)|_k$ is measurable, hence $E \in \sigma(\mathcal{T})$.

Since $\mathcal{T}$ is a countable family, the measurability of $\not\equiv_{\ell^\omega}$ follows by Proposition 4 and $\not\equiv_{\ell^\omega} = \ncong_{\sigma(\mathcal{T})}$, since $\not\equiv_{\ell^\omega} = \bigcup \{E \times E^c \mid E \in \mathcal{T}\}$. □

Appendix B. Folklore Results about Metric Spaces

Proposition 6. Let $A \subseteq \mathbb{R}$ be a bounded nonempty set. Then,

(i) $\sup A \in \overline{A}$;

(ii) $\sup A = \sup \overline{A}$.

Proof. First, notice that since $A \neq \emptyset$ and is bounded, by Dedekind axiom, the supremum of A (and $\overline{A}$) in $\mathbb{R}$ exists. Moreover, recall that, for any $B \subseteq \mathbb{R}$,

$$\overline{B} = ad(B) := \{x \in \mathbb{R} \mid \forall \varepsilon > 0.\ (x - \varepsilon, x + \varepsilon) \cap B \neq \emptyset\} ,$$

where $ad(B)$ denotes the set of points *adherent* to B.

Let $\alpha = \sup A$. (i) We prove that $\alpha \in \overline{A}$. Let $\varepsilon > 0$, then $\alpha - \varepsilon$ is not an upper bound for A. This means that there exists $x \in A$ such that $\alpha - \varepsilon < x \leq \alpha$ and, in particular, that $x \in (\alpha - \varepsilon, \alpha + \varepsilon) \cap A$. Therefore $\alpha \in \overline{A}$. (ii) Let $\beta = \sup \overline{A}$. By $A \subseteq \overline{A} = \overline{\overline{A}}$ and (i), we have $\alpha \leq \beta \in \overline{A}$. We prove that $\alpha = \beta$. Assume by contradiction that $\alpha \neq \beta$ and let $\varepsilon := \beta - \alpha$. Clearly $\varepsilon > 0$, so that, by $\beta \in \overline{A}$, we have that $(\beta - \varepsilon, \beta + \varepsilon) \cap A \neq \emptyset$.

This means that there exists $x \in A$ such that $\alpha = \beta - \varepsilon < x$, in contradiction with the hypothesis that $\alpha = \sup A$. □

Proposition 7. Let $f\colon X \to Y$ be continuous and $A \subseteq X$, then $\overline{f(A)} = \overline{f(\overline{A})}$.

Proof. ($\supseteq$) A function $f\colon X \to Y$ is continuous iff for all $B \subseteq X$, $f(\overline{B}) \subseteq \overline{f(B)}$. Therefore $f(\overline{A}) \subseteq \overline{f(A)}$. Since $\overline{f(A)}$ is closed, we have $\overline{f(\overline{A})} \subseteq \overline{f(A)}$. ($\subseteq$) The result follows by $A \subseteq \overline{A}$ and monotonicity of $f(\cdot)$ and $\overline{(\cdot)}$. □

Proposition 8. Let X be nonempty, $f\colon X \to \mathbb{R}$ be a bounded continuous real-valued function, and $D \subseteq X$ be dense in X. Then $\sup f(D) = \sup f(X)$.

Proof. Notice that, since $X \neq \emptyset$ and f is bounded, by Dedekind axiom, both $\sup f(D)$ and $\sup f(X)$ exist. By Propositions 6, 7, and $\overline{D} = X$, we have

$$\sup f(D) \stackrel{\text{(Prop.6)}}{=} \sup \overline{f(D)} \stackrel{\text{(Prop.7)}}{=} \sup \overline{f(\overline{D})} = \sup \overline{f(X)} \stackrel{\text{(Prop.6)}}{=} \sup f(X)\,,$$

which proves the thesis. □

Proposition 9.

(i) The set of 1-bounded pseudometrics over a set X is a complete lattice w.r.t. the point-wise order $d \sqsubseteq d'$ iff for all $x, y \in X$, $d(x,y) \leq d'(x,y)$;

(ii) $D \subseteq X$ is dense in all 1-bounded pseudometric spaces $\{(X, d_i) \mid i \in I\}$ iff is dense in $(X, \bigsqcup_{i \in I} d_i)$.

Proof. (i) Bottom and top elements are respectively given by the constant function $\mathbf{0}$ and the indiscrete metric $\mathbf{1}(x,y) = 0$ if $x = y$ and $\mathbf{1}(x,y) = 1$ otherwise. To complete the proof it suffices to show that the set of 1-bounded pseudometrics is closed under supremum. Let P be a set of 1-bounded pseudometrics over X. We define $(\bigsqcup P)(x,y) = \sup_{d \in P} d(x,y)$. It is easy to see that $\bigsqcup P$ is the least upper bound of P w.r.t. $\sqsubseteq$ and that is 1-bounded. We only have to check that $\bigsqcup P$ is a pseudometric. Reflexivity and symmetry are trivial. The only nontrivial part is to prove the triangular inequality:

$$\begin{aligned}
(\bigsqcup P)(x,y) + (\bigsqcup P)(y,z) &\leq \sup_{d \in P} d(x,y) + \sup_{d \in P} d(y,z) && \text{(def. and upper bound)} \\
&\leq \sup_{d \in P} d(x,y) + d(y,z)\,. && \text{(triang. ineq. } d \in P\text{)}
\end{aligned}$$

(ii) Recall that a subset $K \subseteq Y$ is dense in a pseudometric space (Y, d) iff $\overline{K} = \{y \in Y \mid d(y, K) = 0\} = X$, where $d(y, K) = \inf_{y' \in K} d(y, y')$. Then, both directions immediately follow by the equality below

$$\big\{x \in X \mid (\textstyle\bigsqcup_{i \in I} d_i)(x, D) = 0\big\} = \bigcap \{x \in X \mid d_i(x, D) = 0\}\ ,$$

which holds since all the pseudometrics d_i are positive. □