



AALBORG UNIVERSITY
DENMARK

Aalborg Universitet

Converging from branching to linear metrics on Markov chains

Bacci, Giorgio; Bacci, Giovanni; Larsen, Kim G.; Mardare, Radu

Published in:
Mathematical Structures in Computer Science

DOI (link to publication from Publisher):
[10.1017/S0960129517000160](https://doi.org/10.1017/S0960129517000160)

Publication date:
2019

Document Version
Accepted author manuscript, peer reviewed version

[Link to publication from Aalborg University](#)

Citation for published version (APA):
Bacci, G., Bacci, G., Larsen, K. G., & Mardare, R. (2019). Converging from branching to linear metrics on Markov chains. *Mathematical Structures in Computer Science*, 29(Special Issue 1), 3-37.
<https://doi.org/10.1017/S0960129517000160>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- ? Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- ? You may not further distribute the material or use it for any profit-making activity or commercial gain
- ? You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

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, Denmark
{grbacci,giovbacci,kgl,mardare}@cs.aau.dk

Abstract. We study the strong and strutter trace distances on Markov chains (MCs). 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 LTL^{-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 Kantorovich-like pseudometrics, i.e. branching-time distances, that converge point-wise to the linear-time metrics. This convergence is interesting in itself, since it reveals a nontrivial relation between branching and linear-time metric-based semantics that does not hold in the case of 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 their behaviour, metrics are preferred to equivalences since the latter are not robust with respect to small variations of the numerical values. *Behavioral metrics* generalize the concept of equivalence by measuring the behavioral dissimilarities of two states.

Several proposals of behavioral distances [10, 12, 8, 20, 13] measure the difference according to this general schema: $d(u, v) = \sup_{\phi \in \Phi} |\phi(u) - \phi(v)|$, where Φ is a suitable set of properties of interest and $\phi(u)$ denotes the value of the property ϕ evaluated at state u . A logical characterization as above is desirable in particular when the distances are defined in a different way (e.g., as a fixed-point [10, 8, 13], a Hausdorff lifting [8] or games [9]) because it relates them in terms of a set Φ of expressible properties. Many logical characterizations in the literature use quantitative logics, whose semantics is given in terms of real-valued functions.

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

Such real-valued logics are not supported by quantitative model checking tools (e.g., PRISM [15] and UPPAAL [4]). Therefore, it is desirable to also have logical characterizations relating the distances to the logics adopted by these tools.

In this work we are interested in the relation with the probabilistic model checking problem for LTL [21] against Markov chains (MCs). In particular we provide two logical characterizations. The first relates the trace distance δ_t , which generalizes trace equivalence, to the probabilistic LTL-model checking problem as $\delta_t(u, v) = \sup_{\varphi \in \text{LTL}} |\mathbb{P}(u)(\llbracket \varphi \rrbracket) - \mathbb{P}(v)(\llbracket \varphi \rrbracket)|$, where $\mathbb{P}(u)(\llbracket \varphi \rrbracket)$ is the probability of executing a run from u satisfying the formula φ . The second relates the stutter trace distance δ_{st} , which generalizes stutter trace equivalence, to $\text{LTL}^{-\times}$ (LTL without next operator) as $\delta_{st}(u, v) = \sup_{\varphi \in \text{LTL}^{-\times}} |\mathbb{P}(u)(\llbracket \varphi \rrbracket) - \mathbb{P}(v)(\llbracket \varphi \rrbracket)|$. An immediate application is that $\mathbb{P}(u)(\llbracket \varphi \rrbracket)$ (i.e., probabilistically model checking φ at u) can be approximated by $\mathbb{P}(v)(\llbracket \varphi \rrbracket)$ with an error bounded by $\delta_t(u, v)$, for any $\varphi \in \text{LTL}$. This may lead to savings in the overall cost of model checking.

This further motivates the study of efficient methods for computing these distances. Unfortunately, in [19, 6] 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 [6] it is shown that the problem of approximating this distance with arbitrary precision is decidable. This is done by providing two effective sequences that converge from below and above to the trace distance. In this paper we provide an alternative approximation schema that, differently from [6], 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 [6], our approach is more general with the nice consequence that the same result is obtained for the problem of approximating the stutter trace distance.

Notably, in our construction the upper-approximants are Kantorovich-like pseudometrics, i.e., branching-time distances. These metrics form a net — a concept used in topology that generalizes infinite sequences — that converges pointwise to the linear-time metrics. The result is interesting in itself, since it reveals a nontrivial link (by means of a converging net) between branching and linear-time metric-based semantics that does not hold when a more standard equivalence-based semantics on MCs is used instead. This opens new perspectives in the study of the operational behavior of quantitative systems, and suggests relating behavioral distances by means of converging nets rather than the standard ‘greater than or equal to’ relation, commonly used in the literature (e.g., in [8]).

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

1. We provide a logical characterization of the trace distance terms of LTL. This result, differently from previous proposals (e.g. [8, 10]), 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.
2. We construct two nets of bisimilarity-like distances that converge to the strong and stutter trace distance. This construction leverages on 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 two important results in [5], namely Theorem 8 and Corollary 11.

3. We demonstrate that each element of the proposed converging nets is computable in polynomial time in the size of the MC. Moreover, we provide other two sequences of pseudometrics that, respectively, converges from below to the two linear distances. Also the 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. The approximation schema for the trace distance improves the one proposed in [6].

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 $R \subseteq X \times X$ an equivalence relation, X/R is the quotient set, $[x]_R$ the R -equivalence class of x , and for $A \subseteq X$, $[A]_R = \bigcup_{x \in A} [x]_R$.

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 $\mathcal{F} \subseteq 2^X$, $\sigma(\Sigma)$ denotes the smallest σ -algebra containing \mathcal{F} . For $(X, \Sigma), (Y, \Theta)$ measurable spaces, $f: X \rightarrow Y$ 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 $\mu: \Sigma \rightarrow \mathbb{R}_+$, i.e., $\mu(\bigcup_{i \in \mathbb{N}} E_i) = \sum_{i \in \mathbb{N}} \mu(E_i)$ 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) . Given a measurable function $f: (X, \Sigma) \rightarrow (Y, \Theta)$, any measure μ on (X, Σ) defines a measure $\mu[f]$ on (Y, Θ) by $\mu[f](E) = \mu(f^{-1}(E))$, for all $E \in \Theta$; it is called the *push forward of μ under f* . A measure ω on $(X, \Sigma) \otimes (Y, \Theta)$ is a *coupling* for (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., μ is the *left* and ν the *right marginal* of ω). $\Omega(\mu, \nu)$ denotes the set of couplings for (μ, ν) .

Metric spaces. For a set X , $d: X \times X \rightarrow \mathbb{R}_+$ 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*. We define $\ker(d) = \{(u, v) \mid d(u, v) = 0\}$. For (X, Σ) a measurable space, $\Delta(X, \Sigma)$ can be metrized by the *total variation distance* $\|\mu - \nu\| = \sup_{E \in \Sigma} |\mu(E) - \nu(E)|$. A (pseudo-)metric $d: X \times X \rightarrow \mathbb{R}_+$ is lifted to $\Delta(X, \Sigma)$ by means of the *Kantorovich (pseudo-)metric*, defined as $\mathcal{K}(d)(\mu, \nu) = \min \left\{ \int d \, d\omega \mid \omega \in \Omega(\mu, \nu) \right\}$.

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 $\pi = x_0x_1\dots$ 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 $\pi|_i = x_ix_{i+1}\dots \in X^\omega$. For $A \subseteq X^n$, the

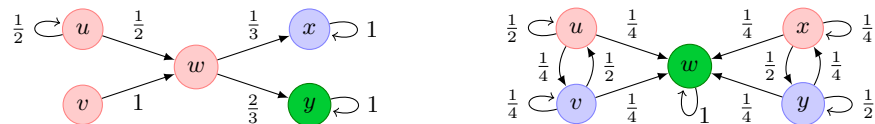


Fig. 1. (Left) u and v are stutter trace equivalent but neither bisimilar nor trace equivalent; (Right) $\delta_t(u, v) = \sqrt{2}/4$ (see [6]) and $\delta_b(u, v) = 1/2$. States are labeled by colors.

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 s.t., for all $n \in \mathbb{N}$, the prefix $(\cdot)|^n$ and tail $(\cdot)|_n$ functions are measurable). Note that, the stepwise extension $f^\omega: X^\omega \rightarrow Y^\omega$ of the function $f: X \rightarrow Y$ 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 discrete-time Markov chains and the notions of strong and stutter probabilistic trace equivalences on them.

In what follows we fix a finite set \mathbb{A} of atomic propositions.

Definition 1. A Markov chain is a tuple $\mathcal{M} = (S, \tau, \ell)$ consisting of a countable set S of states, a transition probability function $\tau: S \rightarrow \Delta(S)$ and a labeling function $\ell: S \rightarrow 2^{\mathbb{A}}$.

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 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 2. Let $\mathbb{P}: S \rightarrow \Delta(S^\omega)$ be such that, for all $u \in S$, $\mathbb{P}(u)$ is the unique probability measure¹ on S^ω such that, for all $n \geq 1$ and $U_i \subseteq S$ ($i = 0..n$)

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

where $\mathbf{1}_A$ 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 $E \subseteq S^\omega$. For example, $\mathbb{P}(u)(\mathfrak{C}(u_0..u_n)) = \mathbf{1}_{u_0}(u) \cdot \prod_{i=0}^{n-1} \tau(u_i)(u_{i+1})$.

¹ Existence and uniqueness follows by the Hahn-Kolmogorov extension theorem.

Remark 3. In Definition 2, since $\mathfrak{C}(U_0) = \mathfrak{C}(U_0S)$, the case $\mathbb{P}(u)(\mathfrak{C}(U_0))$ is covered implicitly. Indeed, $\mathbb{P}(u)(\mathfrak{C}(U_0S)) = \mathbf{1}_{U_0}(u) \cdot \int \mathbb{P}(\cdot)(\mathfrak{C}(S)) \, d\tau(u) = \mathbf{1}_{U_0}(u) \cdot \int 1 \, d\tau(u) = \mathbf{1}_{U_0}(u)$, since for all $v \in S$, $\mathbb{P}(v)$ is a probability measure. \square

Two states of an MC are considered equivalent if they exhibit the same “observable behaviour”. In this work we focus on linear-time properties. In this respect, we recall the most used linear-time equivalences on MCs: *strong* and *stutter probabilistic trace equivalences*.

Definition 4. *Two states $u, v \in S$ are probabilistic trace equivalent, written $u \sim_t v$, if for all $T \in \mathfrak{C}(S/\equiv_\ell)$, $\mathbb{P}(u)(T) = \mathbb{P}(v)(T)$.*

Intuitively, \sim_t 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 $\text{tl}_R^n: X^\omega \rightarrow X^\omega$, by induction on n , as follows

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

$$\text{tl}_R^{n+1}(\pi) = \text{tl}_R^1(\text{tl}_R^n(\pi)).$$

Intuitively, tl_R^1 seeks for the first tail whose head is not R -equivalent to $\pi[0]$ (if it exists!) and $\text{tl}_R^n(\pi)$ is the n -th composition of it. For example, let $\pi = aaabbbcc^\omega$, then $\text{tl}_R^1(\pi) = bbbcc^\omega$ and, for all $n > 1$, $\text{tl}_R^n(\pi) = c^\omega$. The n -th R -stutter prefix function $\text{pf}_R^n: X^\omega \rightarrow X^n$ is defined, by induction on $n \geq 1$, as $\text{pf}_R^1(\pi) = \pi[0]$ and $\text{pf}_R^{n+1}(\pi) = \pi[0]\text{pf}_R^n(\text{tl}_R^1(\pi))$.

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 $\mathfrak{C}_R(A) = \{\pi \in X^\omega \mid \text{pf}_R^n(\pi) \in A\}$. For a family $\mathcal{F} \subseteq 2^X$, denote by $\mathfrak{C}_R^n(\mathcal{F}) = \{\mathfrak{C}_R(E_1 \cdots E_n) \mid E_i \in \mathcal{F}\}$ the set of all R -stutter cylinders of rank n over \mathcal{F} and $\mathfrak{C}_R(\mathcal{F}) = \bigcup_{n \geq 1} \mathfrak{C}_R^n(\mathcal{F})$. If (X, Σ) a measurable space, we denote by $(X, \Sigma)_R^\omega$ 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 5. *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_ℓ -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(left) 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 model checking problem.

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

$$\begin{aligned} \delta_t(u, v) &= \sup_{E \in \sigma(\mathcal{T})} |\mathbb{P}(u)(E) - \mathbb{P}(v)(E)|, & (\text{STRONG TRACE DISTANCE}) \\ \delta_{st}(u, v) &= \sup_{E \in \sigma(\mathcal{ST})} |\mathbb{P}(u)(E) - \mathbb{P}(v)(E)|. & (\text{STUTTER TRACE DISTANCE}) \end{aligned}$$

Observe that two states $u, v \in S$ are strong (resp. stutter) trace equivalent iff $\delta_t(u, v) = 0$ (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 [21]. The LTL formulas are generated by the following grammar

$$\varphi ::= p \mid \perp \mid \varphi \rightarrow \varphi \mid \mathbf{X}\varphi \mid \varphi \mathbf{U} \varphi, \quad \text{where } p \in \mathbb{A}.$$

Let LTL^u and LTL^x be the fragments, respectively, built without until (U) and next (X) operators. 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

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

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

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

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

In the rest of the section we characterize δ_t and δ_{st} respectively as δ_{LTL} (or δ_{LTL^u}) and δ_{LTL^x} . We do this by exploiting the following result.

Lemma 6 ([2]). *Let μ and ν be two finite measures on a measurable space (X, Σ) . If Σ is generated by a field \mathcal{F} , then $\|\mu - \nu\| = \sup_{E \in \mathcal{F}} |\mu(E) - \nu(E)|$.*

By Lemma 6, to provide a logical characterization for δ_t it suffices to show that the σ -algebra $\sigma(\mathcal{T})$ is generated by $\llbracket \text{LTL} \rrbracket$ (or $\llbracket \text{LTL}^u \rrbracket$).

Theorem 7. (i) $\sigma(\mathcal{T}) = \sigma(\llbracket \text{LTL} \rrbracket) = \sigma(\llbracket \text{LTL}^u \rrbracket)$, (ii) $\delta_t = \delta_{\text{LTL}} = \delta_{\text{LTL}^u}$.

Remark 8. $\delta_t = \delta_{\text{LTL}}$ is not trivial. Fig. 1(right) shows an MC from [6, Ex. 1]² where it is proven that $\delta_t(u, x)$ is obtained on a maximizing event in $\sigma(\mathcal{T})$ that is not ω -regular, hence it cannot be expressed by a single LTL formula. \square

In Theorem 7, the proof of $\sigma(\mathcal{T}) \subseteq \sigma(\llbracket \text{LTL} \rrbracket)$ uses the measurability of the n -th tail function $(\cdot)|_n$ w.r.t. $\sigma(\mathcal{T})$. However, $(\cdot)|_n$ 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* [16] (for a standard definition see e.g. [3, §7.7.1]). For a relation $R \subseteq S^\omega \times S^\omega$, $\pi \in S^\omega$ is said *R-constant* if, for all $i \in \mathbb{N}$, $\pi R \pi|_i$.

Definition 9. 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.

Stutter relations are closed under union and reflexive/symmetric/transitive closure, therefore \simeq is an equivalence and a stutter relation.

Proposition 10. $\pi \simeq \rho$ iff $\forall \varphi \in \text{LTL}^\times. (\mathcal{M}, \pi \models \varphi \Leftrightarrow \mathcal{M}, \rho \models \varphi)$.

The above states that \simeq characterizes the logical equivalence w.r.t. LTL^\times . Definition 9 and Proposition 10 are essential to prove the next result.

Theorem 11. (i) $\sigma(\mathcal{ST}) = \sigma(\llbracket \text{LTL}^\times \rrbracket)$, (ii) $\delta_{st} = \delta_{\text{LTL}^\times}$.

Proof. We prove (i), then (ii) follows by Lemma 6. (\supseteq) We prove $\llbracket \varphi \rrbracket \in \sigma(\mathcal{ST})$ by induction on φ . We show the case $\varphi = \phi \cup \psi$. Define $q: S^\omega \rightarrow S^\omega$, as $q(\pi) = \text{pf}_{\equiv_\ell}^1(\pi)q(\text{tl}_{\equiv_\ell}^1(\pi))$ ³. The function q is idempotent, moreover, it is $\sigma(\mathcal{ST})$ - $\sigma(\mathcal{T})$ measurable, i.e., for all $E \in \sigma(\mathcal{T})$, $q^{-1}(E) \in \sigma(\mathcal{ST})$. It can be shown that $R = \{(\pi, \rho) \mid q(\pi) \equiv_{\ell^\omega} q(\rho)\}$ is a stutter relation. Therefore, by $q(\pi) \equiv_{\ell^\omega} q(q(\pi))$, we get $\pi R q(\pi)$, hence $\pi \simeq q(\pi)$. Then, the following hold:

$$\begin{aligned} \llbracket \phi \cup \psi \rrbracket &= \{\pi \mid \exists i \geq 0. q(\pi)|_i \in \llbracket \psi \rrbracket, \forall 0 \leq j < i. q(\pi)|_j \in \llbracket \phi \rrbracket\} \quad (\llbracket \cdot \rrbracket \text{ \& Prop. 10}) \\ &= \bigcup_{i \geq 0} \bigcap_{0 \leq j < i} ((\cdot)|_i \circ q)^{-1}(\llbracket \psi \rrbracket) \cap ((\cdot)|_j \circ q)^{-1}(\llbracket \phi \rrbracket). \quad (\text{preimage}) \end{aligned}$$

By inductive hypothesis on ϕ, ψ and $\sigma(\mathcal{ST})$ -measurability of $(\cdot)|_k \circ q$, for any $k \in \mathbb{N}$, it follows that $\llbracket \phi \cup \psi \rrbracket \in \sigma(\mathcal{ST})$. (\subseteq) The σ -algebra $\sigma(\mathcal{ST})$ is alternatively generated by the family $\mathcal{F} = \{\mathfrak{C}_{\equiv_\ell}(C_1 \cdots C_n) \in \mathcal{ST} \mid C_i \neq C_{i+1}\}$. Hence, it suffices to show $\mathcal{F} \subseteq \sigma(\llbracket \text{LTL}^\times \rrbracket)$. Define $B: \mathcal{F} \rightarrow \text{LTL}^\times$ by induction as follows,

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

where $A(p, C) = p$ if there exists $s \in C$ s.t. $p \in \ell(s)$, otherwise $A(p, C) = \neg p$. For $T \in \mathcal{F}$ one can prove that $\llbracket B(T) \rrbracket = T$. \square

² The MC has been adapted to the case of labeled states, instead of labeled transitions.

³ Note that $q = \lim_{n \geq 1} \text{pf}_{\equiv_\ell}^n$, i.e., it is the unique map s.t., for all $n \geq 1$, $\text{pf}_{\equiv_\ell}^n = (\cdot)|^n \circ q$.

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 [5, Cor. 11], where the bisimilarity pseudometric δ_b of Desharnais et al. [11] 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” [5, Th. 8]. The proof of $\delta_t \leq \delta_b$ uses a classic duality result asserting that *the total variation of two measures coincides to the minimal discrepancy measured among all their couplings* (Lemma 12). Formally, given $\mu, \nu \in \Delta(X, \Sigma)$, the *discrepancy* of $\omega \in \Omega(\mu, \nu)$ is the value $\omega(\not\cong_\Sigma)$, where $\cong_\Sigma = \bigcap \{E \times E \mid E \in \Sigma\}$ is the *inseparability relation* w.r.t. Σ .

Lemma 12 ([18, Th.5.2]). *Let μ, ν be probability measures on (X, Σ) . Then, provided that \cong_Σ is measurable in $\Sigma \otimes \Sigma$, $\|\mu - \nu\| = \min \{\omega(\not\cong_\Sigma) \mid \omega \in \Omega(\mu, \nu)\}$.*

Along the way to obtain our construction, we nontrivially extend (and improve the proofs of) both Corollary 11 and Theorem 8 in [5]. 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 generalization of probabilistic bisimulation.

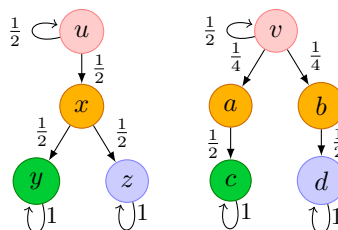
Definition 13. *Let $k \geq 1$. An equivalence relation $R \subseteq S \times S$ is a k -probabilistic bisimulation on \mathcal{M} if whenever $u R v$, then, for all $E_i \in S/\equiv_i$ and $C \in S/R$,*

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

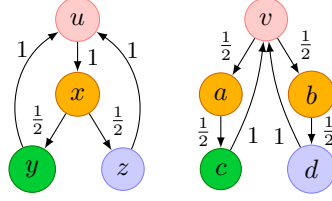
Two states $u, v \in S$ are k -probabilistic bisimilar, written $u \sim_b^k v$, if they are related by some k -probabilistic bisimulation.

The notion of k -bisimulation weakens that of probabilistic bisimulation of Larsen and Skou [17] by equating states that have the same probability to move to the same k -bisimilarity class after having observed the same labels within k -steps. Note that \sim_b^1 coincides with Larsen and Skou bisimilarity. Moreover, for all $k \geq 1$, \sim_b^k is a k -bisimulation and, by σ -additivity of the measures, $\sim_b^1 \subseteq \sim_b^k \subseteq \sim_t$.

Remark 14. Clearly, $\bigcup_{k \geq 1} \sim_b^k \subseteq \sim_t$. However, the converse inclusion does not hold. A counterexample is shown in the picture aside, where states are labeled by colors. It is easy to see that u and v are probabilistic trace equivalent, but they are not probabilistic k -bisimilar for any $k \geq 1$. \square



Remark 15. 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 aside where $u \sim_b^4 v$ but $u \not\sim_b^5 v$, hence $\sim_b^4 \not\subseteq \sim_b^5$. \square



Next we show how to “lift” the above equivalences to behavioral pseudometrics. A pseudometric that lifts bisimilarity is δ_b [11], defined as the least fixed point of the following operator on 1-bounded pseudometrics $d: S \times S \rightarrow [0, 1]$

$$\Theta(d)(u, v) = \begin{cases} 1 & \text{if } u \neq_\ell v \\ \mathcal{K}(d)(\tau(u), \tau(v)) & \text{otherwise.} \end{cases} \quad (\text{KANTOROVICH OPERATOR})$$

Intuitively, two states are incomparable 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 $\tau^k: S \rightarrow \Delta(S^k)$ 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

$$\Theta^k(d)(u, v) \begin{cases} 1 & \text{if } u \neq_\ell v \\ \mathcal{K}(\Lambda^k(d))(\tau^k(u), \tau^k(v)) & \text{otherwise.} \end{cases}$$

where $\Lambda^k(d)(u_1..u_k, v_1..v_k) = 1$ if $u_i \neq_\ell v_i$ 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 fixed point theorem, it has least fixed point, hereafter denoted by δ_b^k . Note that $\delta_b^1 = \delta_b$, moreover the following hold.

Lemma 16 (k -Bisimilarity Distance). $u \sim_b^k v$ iff $\delta_b^k(u, v) = 0$.

Due to the above result we call δ_b^k the k -bisimilarity pseudometric.

Next we 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 of 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 17 (Coupling Structure). A function $\mathcal{C}: S \times S \rightarrow \Delta(S^k \otimes S^k)$ 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 18. For $k \geq 1$ and $\mathcal{C} \in \mathbb{C}_k$, let $\mathbb{P}_{\mathcal{C}}: S \times S \rightarrow \Delta(S^\omega \otimes S^\omega)$ 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 $U_i, V_i \subseteq S$ ($i = 0..nk$)

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

where, $U_{i,j} = U_i \cdots U_j$ (similarly for V)⁴ 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 [5, Th. 8] to k -bisimilarity pseudometrics and provides the alternative characterization of δ_b^k in terms of coupling structures.

Lemma 19 (Coupling Lemma). $\delta_b^k(u, v) = \inf \{ \mathbb{P}_{\mathcal{C}}(u, v)(\neq_{\ell\omega}) \mid \mathcal{C} \in \mathbb{C}_k \}$.

Thanks to Lemma 19 and the next result we can show that the k -bisimilarity pseudometrics δ_b^k form a net that converges point-wise to the trace distance δ_t .

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 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 $(x_i)_{i \in D} \rightarrow x$, if for every open subset $A \subseteq X$ such that $x \in A$, there exists $h \in D$ such that, for all $j \succeq h$, $x_j \in A$.

Theorem 20. *Let (X, Σ) be a measurable space s.t. $\cong_{\Sigma} \in \Sigma \otimes \Sigma$, μ, ν be probability measures on it, (D, \preceq) be a directed poset and $\Omega: D \rightarrow 2^{\Omega(\mu, \nu)}$ 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 $u_i = \inf \{ \omega(\neq_{\Sigma}) \mid \omega \in \Omega(i) \}$, converges to $\|\mu - \nu\|$.*

Proof. By Lemma 12, for all $i \in D$, $u_i \geq \|\mu - \nu\|$. Moreover, by monotonicity of Ω , $i \preceq j$ implies $u_i \leq u_j$. Therefore, to prove $(u_i)_{i \in D} \rightarrow \|\mu - \nu\|$, it suffices to show $\inf_{i \in D} u_i = \|\mu - \nu\|$. Recall that for $Y \neq \emptyset$ and $f: Y \rightarrow \mathbb{R}$ 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 $ev_{\neq}: \Omega(\mu, \nu) \rightarrow \mathbb{R}$, defined by $ev_{\neq}(\omega) = \omega(\neq)$ is bounded and continuous. It is bounded since all $\omega \in \Omega(\mu, \nu)$ are probability measures. It is continuous because $\|\omega - \omega'\| \geq |\omega(\neq) - \omega'(\neq)| = |ev_{\neq}(\omega) - ev_{\neq}(\omega')|$ (1-Lipschitz continuity). Now, applying Lemma 12, we derive our result. \square

Recall that, $\delta_t(u, v)$ is the total variation distance between $\mathbb{P}(u)$ and $\mathbb{P}(v)$ restricted on $\sigma(\mathcal{T})$. Observe that the inseparability relation w.r.t. $\sigma(\mathcal{T})$ is $\equiv_{\ell\omega}$, which is easily seen to be measurable in $\sigma(\mathcal{T}) \otimes \sigma(\mathcal{T})$. Therefore, by Lemma 12,

$$\delta_t(u, v) = \min \{ \omega(\neq_{\ell\omega}) \mid \omega \in \Omega(\mathbb{P}(u), \mathbb{P}(v)) \}.$$

The next lemma shows that (i) a coupling structure \mathcal{C} induces a 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 21. *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.*

⁴ We assume that $U_{i,j} = \{\epsilon\}$ whenever $i > j$.

Proof. (sketch) (i) It follows directly by definition of $\mathbb{P}_{\mathcal{C}}$ and the definitional conditions of coupling structures. (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 $E, F \subseteq S^{hk}$,

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

Then, $\mathcal{D} \in \mathbb{C}_{hk}$ and $\mathbb{P}_{\mathcal{C}}(u, v) = \mathbb{P}_{\mathcal{D}}(u, v)$. (iii) Let $\Omega = \bigcup_{k \geq 1} \{\mathbb{P}_{\mathcal{C}}(u, v) \mid \mathcal{C} \in \mathbb{C}_k\}$. Note that $\bigcup_{n \in \mathbb{N}} \{\mathfrak{C}(E) \times \mathfrak{C}(F) \mid E, F \subseteq S^n\}$ is a field generating the σ -algebra of $S^\omega \otimes S^\omega$. To prove that Ω is dense w.r.t. the total variation it suffices to show that, for all $\mu \in \Omega(\mathbb{P}(u), \mathbb{P}(v))$, $n \in \mathbb{N}$ and $E, F \subseteq S^n$, there exists $\omega \in \Omega$ s.t. $\omega(\mathfrak{C}(E) \times \mathfrak{C}(F)) = \mu(\mathfrak{C}(E) \times \mathfrak{C}(F))$ (consequence of [2, Lemma 5]). One can check that this equality holds for $\omega = \mathbb{P}_{\mathcal{C}}(u, v)$ and $\mathcal{C} \in \mathbb{C}_n$ s.t. $\mathcal{C}(u, v) = \mu[f]$ is the push forward of μ along $f: S^\omega \rightarrow S^n$, defined as $f(\pi, \rho) = (\pi|_1^n, \rho|_1^n)$. \square

Note that Lemmas 19 and 21(i) imply that, for all $k \geq 1$, $\delta_b^k \geq \delta_t$. This generalizes [5, Cor. 11] to arbitrary k -bisimilarity distances.

Denote by \mathbb{K} the poset over $\mathbb{N} \setminus \{0\}$ with partial order $n \preceq m$ iff there exists $k \in \mathbb{N}$ such that $m = nk$. It is easy to see that \mathbb{K} is directed. According to Theorem 20, Lemmas 19 and 21 suffice to prove the following net-convergence.

Theorem 22 (Convergence). *The net $(\delta_b^k)_{k \in \mathbb{K}}$ converges point-wise to δ_t .*

Remark 23. The use of the preorder \preceq in the definition of the directed poset \mathbb{K} is essential in Theorem 22. Indeed, if \preceq is replaced by the standard total order \leq over natural numbers, the net-convergence does not hold (by Lemma 16, the MC shown in Remark 15 provides a counterexample). \square

Remark 24 (Equivalence vs Metric-based semantics). Although $\bigcup_{k \geq 1} \sim_b^k \neq \sim_b$ (see Remark 14), by Theorem 22, we have that $\inf_{k \geq 1} \delta_b^k = \delta_t$. Note that this is not in contradiction with Lemma 16. 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*. \square

5.2 The Stutter Case

We show how the construction that led to Theorem 22 can be easily adapted to obtain a net that converges to the stutter trace distance δ_{st} . This proves that the method is general enough to accommodate nontrivial convergence results.

Definition 25. *Let $k \geq 1$. An equivalence relation $R \subseteq S \times S$ is a \equiv_ℓ -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_ℓ -stutter k -probabilistic bisimilar, written $u \sim_{sb}^k v$, if they are related by some \equiv_ℓ -stutter k -probabilistic bisimulation.

The above definition weakens that of k -probabilistic bisimulation by restricting the events to be tested only to those that are \equiv_ℓ -stutter invariant.

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

Now we lift these equivalences to pseudometrics by means of a Kantorovich-like operator. For $k \geq 1$, define the \equiv_ℓ -stuttered k -steps transition probability function $\tau_s^k: S \rightarrow \Delta(S^k)$ 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., $\tau_s^k(u) = \mathbb{P}(u)[\text{pf}_{\equiv_\ell}^k \circ \text{tl}_{\equiv_\ell}^1]$). Define, for $d: S \times S \rightarrow [0, 1]$ pseudometric,

$$\Psi^k(d)(u, v) = \begin{cases} 1 & \text{if } u \neq_\ell v \\ \mathcal{K}(\Lambda^k(d))(\tau_s^k(u), \tau_s^k(v)) & \text{otherwise.} \end{cases}$$

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_ℓ -stutter k -bisimilarity distance.

Lemma 26 (Stutter k -Bisimilarity Distance). $u \sim_{sb}^k v$ iff $\delta_{sb}^k(u, v) = 0$.

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_ℓ -stutter step.

Definition 27. A function $\mathcal{C}: S \times S \rightarrow \Delta(S^k \otimes S^k)$ 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 28. For $k \geq 1$ and $\mathcal{C} \in \mathbb{C}_k^s$, let $\mathbb{P}_\mathcal{C}: S \times S \rightarrow \Delta(st(S^\omega) \otimes st(S^\omega))$ 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 $U_i, V_i \subseteq S$ ($i = 0..nk$)

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

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 10, \simeq is the inseparability relation w.r.t. $\sigma(\mathcal{ST})$ and, since LTL^\times is countable, it holds $\simeq \in \sigma(\mathcal{ST}) \otimes \sigma(\mathcal{ST})$.

Lemma 29 (Coupling Lemma). $\delta_{sb}^k(u, v) = \inf \{\mathbb{P}_\mathcal{C}(u, v)(\not\sim) \mid \mathcal{C} \in \mathbb{C}_k^s\}$.

According to Theorem 20 what follows suffices to prove the convergence.

Lemma 30. *Let $u, v \in S$ be a pair of states of an MC \mathcal{M} . Then,*

- i. for $k \geq 1$ and $C \in \mathbb{C}_k^s$, $\mathbb{P}_C(u, v) \in \Omega(\mathbb{P}(u), \mathbb{P}(v))$;*
- ii. for $k, h \geq 1$, $\{\mathbb{P}_C(u, v) \mid C \in \mathbb{C}_k^s\} \subseteq \{\mathbb{P}_C(u, v) \mid C \in \mathbb{C}_{hk}^s\}$;*
- iii. $\bigcup_{k \geq 1} \{\mathbb{P}_C(u, v) \mid C \in \mathbb{C}_k^s\}$ is dense in $\Omega(\mathbb{P}(u), \mathbb{P}(v))$ w.r.t. the total variation, where $\mathbb{P}(u)$ is assumed to be restricted on the sub- σ -algebra $\sigma(\mathfrak{C}_{\equiv_\ell}(2^S))$.*

The next result is a direct consequence of Theorem 20, Lemmas 29, and 30.

Theorem 31 (Convergence). *The net $(\delta_{sb}^k)_{k \in \mathbb{K}}$ 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 converges 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 its transition probabilities are rational (i.e., $\tau(u)(v) \in \mathbb{Q} \cap [0, 1]$). 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

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 $l^k: S \times S \rightarrow [0, 1]$ 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 converges point-wise to the trace distance δ_t .

Lemma 32. *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 [2, Theorem 6] and the fact that $\bigcup_{k \geq 1} \mathcal{E}_k$ is a field generating $\sigma(\mathcal{T})$. \square

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 thin cylinders of rank k .

Proposition 33. $l^k(u, v) = \frac{1}{2} \sum_{C \in \mathfrak{C}^k(S)} |\mathbb{P}(u)(C) - \mathbb{P}(v)(C)|$.

Proof. Note that \mathcal{E}_k is finite and closed under complement. Let \mathcal{F} be the family of cylinders $C \in \mathfrak{C}^k(S)$ s.t. $\mathbb{P}(u)(C) \geq \mathbb{P}(v)(C)$. By Hahn 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

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

where the second equality holds since $\mathbb{P}(v)(F^c) = 1 - \mathbb{P}(v)(F)$. \square

Note that the cylinders in $\mathfrak{C}^k(S)$ are all those of the form $\mathfrak{C}(u_1..u_k)$, for some $u_i \in S$ ($i = 1..k$), and $\mathbb{P}(u)(\mathfrak{C}(u_1..u_k)) = \mathbf{1}_{u_1}(u) \cdot \prod_{i=1}^{k-1} \tau(u_i)(u_{i+1})$. Then, by Proposition 33, to compute $l^k(u, v)$ we need only $2kS^k$ multiplications, S^k subtractions and $S^k - 1$ summations. Hence l^k can be computed in $O(kS^{2+k})$.

Theorem 34. l^k can be computed in polynomial time in the size of \mathcal{M} .

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, $d: X \times X \rightarrow [0, 1]$ a pseudometric and $\mu, \nu \in \Delta(X)$, the value of $\mathcal{K}(d)(\mu, \nu)$ coincides with the optimal value of the following linear programs.

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

Consider the linear program in Figure 2, 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 $\Upsilon^k: [0, 1]^{S \times S} \rightarrow [0, 1]^{S \times S}$ defined by $\Upsilon^k(d)(u, v) = 0$ if $u \sim_b^k v$, otherwise $\Upsilon^k(d)(u, v) = \Theta^k(d)(u, v)$.

Lemma 35. Υ^k has a unique fixed point that coincides with δ_b^k .

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

Fig. 2. Linear program characterization of the k -bisimilarity distance δ_b^k .

This 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 36. \sim_b^k can be computed in polynomial time in the size of \mathcal{M} .

Theorem 37. δ_b^k can be computed in polynomial time in the size of \mathcal{M} .

Proof. (sketch) By Lemma 36, D can be constructed in polynomial time. Since the number of constraints and unknowns in D are bounded by a polynomial in the size of \mathcal{M} , D can be solved in polynomial time with the ellipsoid method. \square

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 $k \geq 1$

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

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 32:

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

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

- (i) $U_i \subseteq S$, $\mathbb{P}_{\mathcal{M}}(u)(\mathfrak{C}_{\equiv_\ell}(U_1 \cdots U_k)) = \mathbb{P}_{\mathcal{N}}(u)(\mathfrak{C}(U_1 \cdots U_k))$;
- (ii) $\Psi_{\mathcal{M}}^k = \Theta_{\mathcal{N}}^k$.

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 $\tau_s^1(u)(v) = \mathbb{P}_{\mathcal{M}}(u)(\mathfrak{C}([u]_{\equiv_\ell}^\pm v))$. 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 $\mathbb{P}_{\mathcal{M}}(u)([u]_{\equiv_\ell} \mathbf{U} \{v\})$. 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. [3, §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_ℓ -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 $1 - \sum_{x \not\equiv_\ell v} \tau_s^1(v)(x)$, reusing the probabilities computed in the previous case.

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

Lemma 39. *$\mathcal{N} = (S, \tau_s^1, \ell)$ 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 $\tau_s^1(u)(v) = \mathbb{P}_{\mathcal{M}}(u)([u]_{\equiv_\ell} \mathbf{U} \{v\})$. 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 $u \equiv_\ell v$ follows by the previous one since $\tau_s^1(u)(v) = \tau(u)(v) \cdot (1 - \sum_{x \not\equiv_\ell v} \tau_s^1(v)(x))$.

By Lemmas 38 and 39, and Theorems 34 and 37, the following holds.

Theorem 40. *l_{sb}^k and δ_{sb}^k can be computed in polynomial time in the size of \mathcal{M} .*

Remark 41. Theorems 37 and 40 do not contradict the fact that the problem of approximating the trace distances up to a given precision $\epsilon > 0$ is NP-hard [7]. 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 . \square

7 Conclusions and Future Work

In this paper we provided the strong and stutter trace distances with a logical characterization in terms of LTL and LTL^{-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 [5], 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 14 is shown not 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 proposed generalizes and improves the linear program characterization given in [5, 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 [6] both for the generality of its applicability and in terms of computational complexity.

Natural questions are (i) to see if the on-the-fly algorithm for the computation of bisimilarity distance in [1] 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 recent result by Fu [14] gives new insight on how to obtain minimal information in case the distance is not a bisimilarity metric, and where the PSPACE-complexity results is sharpened to $\text{NP} \cap \text{coNP}$); (iii) whether a similar construction can be applied to stochastic models with continuous time, such as CTMCs.

References

1. 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.

2. 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.
3. Christel Baier and Joost-Pieter Katoen. *Principles of Model Checking*. MIT Press, 2008.
4. Gerd Behrmann, Alexandre David, Kim G. Larsen, John Håkansson, Paul Pettersson, Wang Yi, and Martijn Hendriks. UPPAAL 4.0. In *QEST*, pages 125–126. IEEE Computer Society, 2006.
5. 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.
6. 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.
7. 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.
8. 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.
9. Luca de Alfaro, Rupak Majumdar, Vishwanath Raman, and Mariëlle Stoelinga. Game Relations and Metrics. In *LICS*, pages 99–108, July 2007.
10. 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.
11. Josee Desharnais, Vineet Gupta, Radha Jagadeesan, and Prakash Panangaden. Metrics for labelled Markov processes. *Theoretical Computer Science*, 318(3):323–354, 2004.
12. 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.
13. 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.
14. Hongfei Fu. Computing game metrics on markov decision processes. In *ICALP*, volume 7392 of *LNCS*, pages 227–238, 2012.
15. M. Kwiatkowska, G. Norman, and D. Parker. PRISM 4.0: Verification of probabilistic real-time systems. In *CAV*, volume 6806 of *LNCS*, pages 585–591. Springer, 2011.
16. Leslie Lamport. What good is temporal logic? In *IFIP*, pages 657–668, 1983.
17. Kim Guldstrand Larsen and Arne Skou. Bisimulation through probabilistic testing. *Information and Computation*, 94(1):1–28, 1991.
18. Torgny Lindvall. *Lectures on the Coupling Method*. Wiley Series in Probability and Mathematical Statistics. John Wiley, New York, 1992.
19. 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.
20. Matteo Mio. Upper-Expectation Bisimilarity and Łukasiewicz μ -Calculus. In *FoSSaCS*, volume 8412 of *LNCS*, pages 335–350, 2014.
21. Amir Pnueli. The temporal logic of programs. In *SFCS*, pages 46–57. IEEE Computer Society, 1977.

A Trashed

This result opens for the study of effective methods for approximating both the strong and the stutter trace distances.

It is known that computing the trace distance is NP-hard and its decidability is still an open problem [6]. In this section we show how the above convergence sequences can be useful for approximating the trace and the stutter-trace distances.

Here we show that the problem of computing the stutter-trace distance is as hard as that of computing the trace distance. To this end we will reduce the problem of computing the stutter trace distance on an MC \mathcal{M} to the problem of computing the stutter-trace distance on another MC $\bar{\mathcal{M}}$ constructed from \mathcal{M} .

Given $\mathcal{M} = (S, \tau, \ell)$, the MC $\bar{\mathcal{M}}$ is constructed as $(S \uplus \bar{S}, \theta, l)$ where $\bar{S} = \{\bar{u} \mid u \in S\}$, $\theta(u) = \delta_{\bar{u}}$ (the dirac measure at \bar{u}) if $u \in S$, and $\theta(\bar{u}) = \tau(u)$ if $\bar{u} \in \bar{S}$, and $l(u) = \ell(u)$ if $u \in S$, otherwise \square , where \square is a label that differs from all the labels in \mathcal{M} (see Figure 3).



Fig. 3. (Left) A state in the original MC \mathcal{M} ; (Right) The MC $\bar{\mathcal{M}}$ constructed from \mathcal{M} . In $\bar{\mathcal{M}}$, each state $w \in S$ moves to its counterpart $\bar{w} \in \bar{S}$ that proceeds as w does in \mathcal{M} .

Intuitively, each run $u_0 u_1 \dots$ generated by \mathcal{M} is emulated (with same probability) by a run $u_0 \bar{u}_0 u_1 \bar{u}_1 \dots$ in $\bar{\mathcal{M}}$ where each state $u_i \in S$ is followed by its counterpart $\bar{u}_i \in \bar{S}$ (for all $i \in \mathbb{N}$). It is worth noting that the underlying graph of $\bar{\mathcal{M}}$ is bipartite and the labeling function l induces a 2-coloring for this graph.

Proposition 42. *Let \mathcal{M} be an MC. Then, for all $u, v \in S$, $\delta_t^{\mathcal{M}}(u, v) = \delta_{st}^{\bar{\mathcal{M}}}(u, v)$.*

Proof. Let $f: S^\omega \rightarrow (S \uplus \bar{S})^\omega$ be defined as $f(\pi) = u\bar{u}f(\pi|_i)$, where $u = \pi[0]$.

(i) The following equality holds

$$\sup_{E \in \sigma(\mathcal{T}_{\bar{\mathcal{M}}})} |\mathbb{P}_{\bar{\mathcal{M}}}(u)(E) - \mathbb{P}_{\bar{\mathcal{M}}}(v)(E)| = \sup_{E \in \sigma(\mathcal{T}_{\mathcal{M}})} |\mathbb{P}_{\mathcal{M}}(u)(f^{-1}(E)) - \mathbb{P}_{\mathcal{M}}(v)(f^{-1}(E))| \quad (\text{i})$$

(\geq) follows by the fact that f is $\sigma(\mathcal{T}_{\mathcal{M}})$ - $\sigma(\mathcal{T}_{\bar{\mathcal{M}}})$ measurable. (\leq) since for all $T \in \mathcal{T}_{\mathcal{M}}$ there exists $\bar{T} \in \mathcal{T}_{\bar{\mathcal{M}}}$ such that $f^{-1}(\bar{T}) = T$. Thus $\sigma(\mathcal{T}_{\mathcal{M}}) = \{f^{-1}(\bar{E}) \mid \bar{E} \in \sigma(\mathcal{T}_{\bar{\mathcal{M}}})\}$.

(ii) The following equality holds

$$\sup_{E \in \sigma(\mathcal{T}_{\bar{\mathcal{M}}})} |\mathbb{P}_{\bar{\mathcal{M}}}(u)(E) - \mathbb{P}_{\bar{\mathcal{M}}}(v)(E)| = \sup_{E \in \sigma(\mathcal{T}_{\mathcal{M}})} |\mathbb{P}_{\mathcal{M}}(u)(f^{-1}(E)) - \mathbb{P}_{\mathcal{M}}(v)(f^{-1}(E))| \quad (\text{ii})$$

The above is proven showing that, for any $u \in S$, $\mathbb{P}_{\bar{\mathcal{M}}}(u) = \mathbb{P}_{\mathcal{M}}(u)[f]$. This is done by showing that for any cylinder set $\mathfrak{C}(u_0..u_n) \in \mathfrak{C}(2^{S \uplus \bar{S}})$ (for $n \in 2\mathbb{N} + 1$) it holds $\mathbb{P}_{\bar{\mathcal{M}}}(u)(\mathfrak{C}(u_0..u_n)) = \mathbb{P}_{\mathcal{M}}(u)(f^{-1}(\mathfrak{C}(u_0..u_n)))$. This can be done by cases.

(iii) The following equality holds

$$\sup_{E \in \sigma(\mathcal{T}_{\bar{\mathcal{M}}})} |\mathbb{P}_{\bar{\mathcal{M}}}(u)(E) - \mathbb{P}_{\bar{\mathcal{M}}}(v)(E)| = \sup_{E \in \sigma(\mathcal{S}\mathcal{T}_{\bar{\mathcal{M}}})} |\mathbb{P}_{\bar{\mathcal{M}}}(u)(E) - \mathbb{P}_{\bar{\mathcal{M}}}(v)(E)| \quad (\text{iii})$$

(\geq) immediate by $\sigma(\mathcal{S}\mathcal{T}_{\bar{\mathcal{M}}}) \subseteq \sigma(\mathcal{T}_{\bar{\mathcal{M}}})$. (\leq) We first show that for all $u \in S$ and for all $E \in \sigma(\mathcal{T}_{\bar{\mathcal{M}}})$ there exists $F \in \sigma(\mathcal{S}\mathcal{T}_{\bar{\mathcal{M}}})$ such that $E \subseteq F$ and $\mathbb{P}_{\bar{\mathcal{M}}}(u)(E) = \mathbb{P}_{\bar{\mathcal{M}}}(u)(F)$. To this end it suffices to show that for an arbitrary cylinder set $\mathfrak{C}(u_0..u_n) \subseteq \mathfrak{C}(2^{S \uplus \bar{S}})$, it holds $\mathbb{P}_{\bar{\mathcal{M}}}(u)(\mathfrak{C}(u_0..u_n)) = \mathbb{P}_{\bar{\mathcal{M}}}(u)(\mathfrak{C}_{\equiv_\varepsilon}(u_0..u_n))$: (\leq) follows by monotonicity of $\mathbb{P}_{\bar{\mathcal{M}}}(u)$ and $\mathfrak{C}(u_0..u_n) \subseteq \mathfrak{C}_{\equiv_\varepsilon}(u_0..u_n)$; (\geq) follows from the fact that, for any $v \in S$ and $\bar{v} \in \bar{S}$, $l(v) \neq l(\bar{v})$ and $\theta(v)(S) = \theta(\bar{v})(S) = 0$.

Thus, coming back to (iii) we have that for arbitrary $E \in \sigma(\mathcal{T}_{\bar{\mathcal{M}}})$ the following hold

$$\begin{aligned} & |\mathbb{P}_{\bar{\mathcal{M}}}(u)(E) - \mathbb{P}_{\bar{\mathcal{M}}}(v)(E)| \quad (\text{w.l.o.g. we assume } \mathbb{P}_{\bar{\mathcal{M}}}(u)(E) \geq \mathbb{P}_{\bar{\mathcal{M}}}(v)(E)) \\ &= |\mathbb{P}_{\bar{\mathcal{M}}}(u)(E) - \mathbb{P}_{\bar{\mathcal{M}}}(v)(F)| \quad (\text{for some } F \in \sigma(\mathcal{S}\mathcal{T}_{\bar{\mathcal{M}}}) \text{ s.t. } E \subseteq F) \\ &\leq |\mathbb{P}_{\bar{\mathcal{M}}}(u)(F) - \mathbb{P}_{\bar{\mathcal{M}}}(v)(F)| \quad (\text{monotonicity of } \mathbb{P}_{\bar{\mathcal{M}}}(u)) \end{aligned}$$

This proves the inequality (\leq) in (iii).

The thesis follows from (i), (ii) and (iii).

A direct consequence of Proposition 42 and [6, Proposition 9] is the following result.

Theorem 43. *Approximating the stutter-trace distance up to any ε whose size polynomial in the given MC is NP-hard.*

B Technical proofs

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

Proposition 44. *Let $R \subseteq S \times S$ be an equivalence such that whenever $u R v$, then $\tau(u)(C) = \tau(v)(C)$, for all $C \in S/R$. Then $u R v$ iff $\mathbb{P}(u)(E) = \mathbb{P}(v)(E)$, for all $E \in \mathfrak{C}(S/R)$.*

Proof. (\Leftarrow) By contraposition. Let $(u, v) \notin R$ and $C = [u]_R \in S/R$. Then $u \in C$ and $v \notin C$. By def. of \mathbb{P} , $\mathbb{P}(u)(\mathfrak{C}(C)) = \mathbb{1}_C(u) = 1$ and $\mathbb{P}(v)(\mathfrak{C}(C)) = \mathbb{1}_C(v) = 0$. (\Rightarrow) By induction on the rank $n \geq 1$ of the cylinders in $\mathfrak{C}(S/R)$.

Base case ($n = 1$). $\mathbb{P}(u)(\mathfrak{C}(C_1)) = \mathbb{1}_{C_1}(u) = \mathbb{1}_{C_1}(v) = \mathbb{P}(v)(\mathfrak{C}(C_1))$.

Inductive step ($n \geq 1$). Let $c_2 \in C_2$.

$$\begin{aligned}
 \mathbb{P}(u)(\mathfrak{C}(C_1 \dots C_n)) &= \mathbb{1}_{C_1}(u) \cdot \int_{C_2} \mathbb{P}(\cdot)(\mathfrak{C}(C_2 \dots C_n)) \, d\tau(u) && \text{(def. } \mathbb{P} \text{)} \\
 &= \mathbb{1}_{C_1}(u) \cdot \mathbb{P}(c_2)(\mathfrak{C}(C_2 \dots C_n)) \cdot \int \mathbb{1}_{C_2} \, d\tau(u) && \text{(ind. hp.)} \\
 &= \mathbb{1}_{C_1}(u) \cdot \mathbb{P}(c_2)(\mathfrak{C}(C_2 \dots C_n)) \cdot \tau(u)(C_2) && \text{(def. } \int \text{)} \\
 &= \mathbb{1}_{C_1}(v) \cdot \mathbb{P}(c_2)(\mathfrak{C}(C_2 \dots C_n)) \cdot \tau(v)(C_2) && (u R v) \\
 &= \mathbb{P}(v)(\mathfrak{C}(C_1 \dots C_n)). && \text{(def. } \int \text{ + ind. hp. + def. } \mathbb{P} \text{)}
 \end{aligned}$$

□

Proposition 45. $u \sim_b v$ iff for all $E \in \mathfrak{C}(S/\sim_b)$, $\mathbb{P}(u)(E) = \mathbb{P}(v)(E)$.

Proof. (\Leftarrow) By contraposition. Assume $u \not\sim_b v$. Let $C = [u]_{\sim_b} \in S/\sim_b$, then $u \in C$ and $v \notin C$. By def. of \mathbb{P} , $\mathbb{P}(u)(\mathfrak{C}(C)) = \mathbb{1}_C(u) = 1$ and $\mathbb{P}(v)(\mathfrak{C}(C)) = \mathbb{1}_C(v) = 0$. (\Rightarrow) We proceed by induction on the rank $n \geq 1$ of the cylinders in $\mathfrak{C}(S/\sim_b)$.

Base case ($n = 1$). $\mathbb{P}(u)(\mathfrak{C}(C_1)) = \mathbb{1}_{C_1}(u) = \mathbb{1}_{C_1}(v) = \mathbb{P}(v)(\mathfrak{C}(C_1))$.

Inductive step ($n \geq 1$). Let $c_2 \in C_2$.

$$\begin{aligned}
 \mathbb{P}(u)(\mathfrak{C}(C_1 \dots C_n)) &= \mathbb{1}_{C_1}(u) \cdot \int_{C_2} \mathbb{P}(\cdot)(\mathfrak{C}(C_2 \dots C_n)) \, d\tau(u) && \text{(def. } \mathbb{P} \text{)} \\
 &= \mathbb{1}_{C_1}(u) \cdot \mathbb{P}(c_2)(\mathfrak{C}(C_2 \dots C_n)) \cdot \int \mathbb{1}_{C_2} \, d\tau(u) && \text{(ind. hp.)} \\
 &= \mathbb{1}_{C_1}(u) \cdot \mathbb{P}(c_2)(\mathfrak{C}(C_2 \dots C_n)) \cdot \tau(u)(C_2) && \text{(def. } \int \text{)} \\
 &= \mathbb{1}_{C_1}(v) \cdot \mathbb{P}(c_2)(\mathfrak{C}(C_2 \dots C_n)) \cdot \tau(v)(C_2) && (u \sim_b v) \\
 &= \mathbb{P}(v)(\mathfrak{C}(C_1 \dots C_n)). && \text{(def. } \int \text{ + ind. hp. + def. } \mathbb{P} \text{)}
 \end{aligned}$$

□

Proposition 46. (i) \sim_{sb} is a stutter bisimulation and (ii) $\sim_b \subseteq \sim_{sb}$.

Proof. Let $\mathcal{M} = (S, \tau, \ell)$ be an MC. (i) Let \mathcal{B} be the smallest equivalence that contains all the stutter bisimulations on \mathcal{M} . Explicitly, $\mathcal{B} = \bigcup_{n \in \mathbb{N}} \mathcal{B}_n$ for

$$\mathcal{B}_0 = \bigcup \{R \mid R \text{ stutter bisimulation on } \mathcal{M}\} \quad \mathcal{B}_{n+1} = \mathcal{B}_n; \mathcal{B}_n$$

where “;” denotes the relational composition. Clearly $\mathcal{B} \subseteq \equiv_\ell$. We prove, by induction on $n \geq 0$ that whenever $u \mathcal{B}_n v$, then $\mathbb{P}(u)(\mathfrak{C}_{\mathcal{B}}(CD)) = \mathbb{P}(v)(\mathfrak{C}_{\mathcal{B}}(CD))$, for all $C, D \in S/\mathcal{B}$. This suffices to show that \mathcal{B} is a stutter bisimulation.

Base case ($n = 0$). If $u \mathcal{B}_0 v$, then $u R v$ for some stutter bisimulation R . Let $C, D \in S/\mathcal{B}$. By R and \mathcal{B} being equivalence relations and $R \subseteq \mathcal{B}$, we have that C and D can be represented as disjoint unions of R -equivalence classes: $C = \bigcup \mathcal{C}$

and $D = \bigcup \mathcal{D}$, where $\mathcal{C} = \{C' \in S/R \mid C' \subseteq C\}$ and $\mathcal{D} = \{D' \in S/R \mid D' \subseteq D\}$. In particular $\mathfrak{C}_{\mathcal{B}}(CD) = \bigcup \mathcal{A}$, for $\mathcal{A} = \{\mathfrak{C}_{\mathcal{B}}(C'D') \mid C' \in \mathcal{C}, D' \in \mathcal{D}\}$. Define $\mathcal{A}' = \{A \in \mathcal{A} \mid \mathbb{P}(u)(A) > 0\}$. By $u R v$ and R being a stutter bisimulation, we also have $\mathcal{A}' = \{A \in \mathcal{A} \mid \mathbb{P}(v)(A) > 0\}$. By [?, Lemma 2.2], \mathcal{A}' is at most countable, moreover its elements are all pairwise disjoint. Hence

$$\begin{aligned} \mathbb{P}(u)(\mathfrak{C}_{\mathcal{B}}(CD)) &= \mathbb{P}(u)(\bigcup \mathcal{A}') \\ &= \sum_{A \in \mathcal{A}'} \mathbb{P}(u)(A) && (\sigma\text{-additivity}) \\ &= \sum_{A \in \mathcal{A}'} \mathbb{P}(v)(A) && (u R v) \\ &= \mathbb{P}(v)(\bigcup \mathcal{A}') = \mathbb{P}(v)(\mathfrak{C}_{\mathcal{B}}(CD)). \end{aligned}$$

Inductive step ($n > 0$). If $u \mathcal{B}_{n+1} v$, then there exists some $w \in S$ such that $u \mathcal{B}_n w$ and $w \mathcal{B}_n v$. Then, by applying the inductive hypothesis twice, we have $\mathbb{P}(u)(\mathfrak{C}_{\mathcal{B}}(CD)) = \mathbb{P}(w)(\mathfrak{C}_{\mathcal{B}}(CD)) = \mathbb{P}(v)(\mathfrak{C}_{\mathcal{B}}(CD))$, for all $C, D \in S/\mathcal{B}$.

By definition, $\sim_{sb} = \mathcal{B}_0 \subseteq \mathcal{B}$. Since \mathcal{B} is a stutter bisimulation we also have $\mathcal{B} \subseteq \sim_{sb}$. This proves that \sim_{sb} is a stutter bisimulation.

(ii) It suffices to show that \sim_b is a stutter bisimulation. Clearly \sim_b is an equivalence relation, moreover $\sim_b \subseteq \equiv_{\ell}$. Let $u \sim_b v$ and $C, D \in S/\sim_b$, then

$$\begin{aligned} \mathbb{P}(u)(\mathfrak{C}(CD)) &= \mathbb{P}(u)(\mathfrak{C}(C^+D^+)) && (\text{stutter cylinder}) \\ &= \mathbb{P}(u)(\mathfrak{C}(C^+D)) && (\mathfrak{C}(C^+D^+) = \mathfrak{C}(C^+D)) \\ &= \mathbb{P}(u)(\bigcup_{n \geq 1} \mathfrak{C}(C^n D)) && (\text{Kleene plus}) \\ &= \sum_{n \geq 1} \mathbb{P}(u)(\mathfrak{C}(C^n D)) && (\sigma\text{-additivity}) \\ &= \sum_{n \geq 1} \mathbb{P}(v)(\mathfrak{C}(C^n D)) && (\text{Prop. 45}) \\ &= \mathbb{P}(v)(\mathfrak{C}(CD)). && (\sigma\text{-additivity} + \text{Kleene plus}) \end{aligned}$$

Proof (of Theorem 7). (ii) Is a direct consequence of (i) and Lemma 6, since both $\llbracket \text{LTL} \rrbracket$ and $\llbracket \text{LTL}^{-u} \rrbracket$ are fields.

To prove (i) it suffices to show (a) $\llbracket \text{LTL} \rrbracket \subseteq \sigma(\mathcal{T})$ and (b) $\mathcal{T} \subseteq \sigma(\llbracket \text{LTL}^{-u} \rrbracket)$.

(a) By structural induction on the syntax of $\varphi \in \text{LTL}$ we prove that $\llbracket \varphi \rrbracket \in \sigma(\mathcal{T})$.

Atomic prop. $\llbracket p \rrbracket = \bigcup \{\mathfrak{C}([u]_{\equiv_{\ell}}) \mid u \in S, p \in \ell(u)\}$. Since S is countable and $\mathfrak{C}([u]_{\equiv_{\ell}}) \in \mathcal{T}$ for all $u \in S$, then $\llbracket p \rrbracket \in \sigma(\mathcal{T})$.

False. $\llbracket \perp \rrbracket = \emptyset \in \sigma(\mathcal{T})$.

Implication. $\llbracket \phi \rightarrow \psi \rrbracket = \llbracket \neg\phi \vee \psi \rrbracket = \llbracket \phi \rrbracket^c \cup \llbracket \psi \rrbracket$. By inductive hypothesis, $\llbracket \phi \rrbracket, \llbracket \psi \rrbracket \in \sigma(\mathcal{T})$, therefore $\llbracket \phi \rightarrow \psi \rrbracket \in \sigma(\mathcal{T})$.

Next. Consider $X\phi$. The following hold

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

By inductive hypothesis and $(\cdot)|_1$ being measurable, it follows $\llbracket X^I \phi \rrbracket \in \sigma(\mathcal{T})$.

Until. Consider $\llbracket \phi \text{ U } \psi \rrbracket$. The following hold

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

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

(b) To prove $\sigma(\mathcal{T}) \subseteq \sigma(\llbracket \text{LTL}^{\text{u}} \rrbracket)$ it suffices to show $\mathcal{T} \subseteq \sigma(\llbracket \text{LTL}^{\text{u}} \rrbracket)$. Define $A: \mathbb{A} \times 2^S \rightarrow \text{LTL}^{\text{u}}$ and $B: \mathcal{T} \rightarrow \text{LTL}^{\text{u}}$ as follows

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

$$B(\mathfrak{C}(C_0)) = \bigwedge_{p \in \mathbb{A}} A(p, C_0)$$

$$B(\mathfrak{C}(C_0 \cdots C_{n+1})) = B(\mathfrak{C}(C_0)) \wedge \text{X}B(\mathfrak{C}(C_1 \cdots C_{n+1})).$$

For $T \in \mathcal{T}$ one can prove that $\llbracket B(T) \rrbracket = T$. □

Proof (of Proposition 10). We prove the two implications separately.

(\Rightarrow) Let R be a stutter relation such that $\pi R \rho$. Without loss of generality assume R to be symmetric (indeed $R \cup R^{-1}$ is a stutter bisimulation if R is so). It suffices to prove that, for any $\varphi \in \text{LTL}^{\times}$, $\mathcal{M}, \pi \models \varphi$ implies $\mathcal{M}, \rho \models \varphi$. We proceed by induction on the structure of φ .

(Case $\varphi = \perp$). Immediate, by the semantics of \perp .

(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 $\varphi = \phi \rightarrow \psi$). Assume $\mathcal{M}, \rho \models \phi$. By symmetry of R , $\rho R \pi$. By inductive hypothesis, $\mathcal{M}, \pi \models \phi$. By hypothesis $\mathcal{M}, \pi \models \phi \rightarrow \psi$, hence $\mathcal{M}, \pi \models \psi$. By inductive hypothesis, $\mathcal{M}, \rho \models \psi$, hence $\mathcal{M}, \rho \models \phi \rightarrow \psi$.

(Case $\varphi = \phi \text{ U } \psi$). For $\pi' \in \llbracket \phi \text{ U } \psi \rrbracket$, 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, $\mathcal{M}, \rho \models \phi \text{ U } \psi$. [Inductive step: $i^*(\pi) > 0$] By definition of i^* , $\mathcal{M}, \pi|_1 \models \phi \text{ U } \psi$. Since $i^*(\pi|_1) = i^*(\pi) - 1$, by inductive hypothesis on $\pi|_1$, for every ρ' such that $\pi|_1 R \rho'$, it holds $\mathcal{M}, \rho' \models \phi \text{ U } \psi$. By $\pi R \rho$, one of the following cases holds:

- Case $\pi|_1 \simeq \rho$: then $\mathcal{M}, \rho \models \phi \text{ U } \psi$.
- Case $\pi|_1 \simeq \rho|_1$: then $\mathcal{M}, \rho|_1 \models \phi \text{ U } \psi$. By $i^*(\pi) > 0$ and $\mathcal{M}, \pi \models \phi \text{ U } \psi$ we have $\mathcal{M}, \pi \models \phi$. By inductive hypothesis on the formula, $\mathcal{M}, \rho \models \phi$. From this and $\mathcal{M}, \rho|_1 \models \phi \text{ U } \psi$ we conclude that $\mathcal{M}, \rho \models \phi \text{ U } \psi$.

(\Leftarrow) We show that $\equiv_{\text{LTL}^{\times}} = \{(\pi, \rho) \mid \forall \varphi \in \text{LTL}^{\times}, \mathcal{M}, \pi \models \varphi \text{ iff } \mathcal{M}, \rho \models \varphi\}$ is a stutter relation. Assume $\pi \equiv_{\text{LTL}^{\times}} \rho$.

(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_{LTL^*} -constant. Then, there exists an index $i > 0$ such that $\pi \not\equiv_{\text{LTL}^*} \pi|_i$. Let k be the least such index. 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 $\mathcal{M}, \pi \models \varphi \text{ U } \neg\varphi$. By $\pi \equiv_{\text{LTL}^*} \rho$, we have $\mathcal{M}, \rho \models \phi \text{ U } \neg\varphi$ 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_{LTL^*} -constant.

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

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

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

Let $\varphi \in \text{LTL}^*$ 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 $\pi \equiv_{\text{LTL}^*} \rho$ 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 $\mathcal{M}, \rho \models \alpha \text{ U } \neg\varphi$, hence, by $\pi \equiv_{\text{LTL}^*} \rho$, we have $\mathcal{M}, \pi \models \alpha \text{ U } \neg\varphi$. 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 $\pi \equiv_{\text{LTL}^*} \rho$, $\mathcal{M}, \pi \models \beta$. By this and hypothesis on $\pi|_1$, we have $\mathcal{M}, \pi \models \beta \text{ U } \varphi$. Then, by $\pi \equiv_{\text{LTL}^*} \rho$, we have $\mathcal{M}, \rho \models \beta \text{ U } \varphi$ and similarly, by the hypothesis made on π , we also have $\mathcal{M}, \rho \models \neg\varphi$. This means that $\mathcal{M}, \rho|_1 \models \varphi$. \square

Proposition 47. *Define $q: S^\omega \rightarrow S^\omega$ 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 $\equiv_{\ell^\omega \circ q}$ is a stutter relation.

Proof. To prove the measurability of q it suffices to show that for all cylinders $T \in \mathcal{T}$, $q^{-1}(T) \in \sigma(\mathcal{ST})$. We proceed by induction on the rank $n \geq 1$ of T .

(Base case, $n = 1$). Let $C \in S/\equiv_\ell$, then the following holds:

$$\begin{aligned} q^{-1}(\mathfrak{C}(C)) &= \{\pi \mid q(\pi) \in \mathfrak{C}(C)\} && \text{(preimage)} \\ &= \{\pi \mid \pi[0] \in C\} && \text{(def. } q) \\ &= \mathfrak{C}(C) && \text{(cylinder set)} \\ &= \mathfrak{C}(C) && (*) \end{aligned}$$

The equality (*) follows by $\mathfrak{C}(C) \subseteq \bigcup_{k \geq 1} \mathfrak{C}(C^k) = \mathfrak{C}(C)$ and the fact that, for all $k \geq 1$, $\mathfrak{C}(C^k) = \mathfrak{C}(C) \cap ((\cdot)|_1)^{-1}(\mathfrak{C}(C^{k-1})) \subseteq \mathfrak{C}(C)$.

(Inductive step, $n > 1$). Let $C_i \in S/\equiv_\ell$, for $i = 1..n$. We consider two cases. If for all $i = 1..n$, $C_1 = C_i$, then the following holds:

$$\begin{aligned} q^{-1}(\mathfrak{C}(C_1 \cdots C_n)) &= \{\pi \mid q(\pi) \in \mathfrak{C}(C_1 \cdots C_n)\} && \text{(preimage)} \\ &= \{\pi \mid \pi[0] \in C_1\} && \text{(def. } q) \\ &= \mathfrak{C}(C_1) && \text{(by } (*)) \end{aligned}$$

Assume that for some $1 < j \leq n$, $C_1 \not\equiv_\ell C_j$. Then the following holds:

$$\begin{aligned}
 q^{-1}(\mathfrak{C}(C_1 \cdots C_n)) &= \{\pi \mid q(\pi) \in \mathfrak{C}(C)\} && \text{(preimage)} \\
 &= \left\{ \pi \mid \begin{array}{l} \exists k. \pi[0] \not\equiv_\ell \pi[k], \forall j < k. \pi[0] \equiv_\ell \pi[j], \\ \pi[0] \in C_1 \text{ and } q(\pi|_k) \in \mathfrak{C}(C_2 \cdots C_n) \end{array} \right\} && \text{(def. } q) \\
 &= \bigcup_{k \geq 1} (\mathfrak{C}(C_1^k) \cap ((\cdot)|_k)^{-1}(q^{-1}(\mathfrak{C}(C_2 \cdots C_n)))) && \text{(def. } (\cdot)|_k)
 \end{aligned}$$

By inductive hypothesis $q^{-1}(\mathfrak{C}(C_2 \cdots C_n)) \in \sigma(\mathcal{ST})$. Now we show that for all $E \in \sigma(\mathcal{ST})$, $\bigcup_{k \geq 1} (\mathfrak{C}(C_1^k) \cap ((\cdot)|_k)^{-1}(E)) \in \sigma(\mathcal{ST})$. To do this it suffices to prove that, for all stutter cylinders $W \in \mathcal{ST}$, $\bigcup_{k \geq 1} (\mathfrak{C}(C_1^k) \cap ((\cdot)|_k)^{-1}(W)) \in \sigma(\mathcal{ST})$. Let $W = \mathfrak{C}(D_1^+ \cdots D_m^+)$, for some $D_i \in S/\equiv_\ell$, s.t. $D_i \neq D_{i+1}$, for all $i = 1..m$. We consider two cases. If $C_1 = D_1$, then

$$\bigcup_{k \geq 1} (\mathfrak{C}(C_1^k) \cap ((\cdot)|_k)^{-1}(\mathfrak{C}(D_1^+ \cdots D_m^+))) = \mathfrak{C}(D_1^+ \cdots D_m^+) \in \sigma(\mathcal{ST}).$$

If $C_1 \neq D_1$, then

$$\begin{aligned}
 \bigcup_{k \geq 1} (\mathfrak{C}(C_1^k) \cap ((\cdot)|_k)^{-1}(\mathfrak{C}(D_1^+ \cdots D_m^+))) &= \bigcup_{k \geq 1} (\mathfrak{C}(C_1^k D_1^+ \cdots D_m^+)) \\
 &= \mathfrak{C}(C_1^+ D_1^+ \cdots D_m^+) \in \sigma(\mathcal{ST}).
 \end{aligned}$$

This concludes the proof for the $\sigma(\mathcal{ST})$ - $\sigma(\mathcal{T})$ measurability of q .

We show that $\equiv_{\ell^\omega \circ q} = \{(\pi, \rho) \mid q(\pi) \equiv_{\ell^\omega} q(\rho)\}$ is a stutter relation. Assume that $q(\pi) \equiv_{\ell^\omega} q(\rho)$, then

(i) By definition of q , $q(\pi)[0] = \pi[0]$ and $q(\rho)[0] = \rho[0]$. By definition of \equiv_{ℓ^ω} , 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_{ℓ^ω} -constant iff π is $\equiv_{\ell^\omega \circ q}$ -constant;
- (b) if $\pi \equiv_{\ell^\omega \circ q} \rho$ and π is \equiv_{ℓ^ω} -constant, then ρ is \equiv_{ℓ^ω} -constant.

(1: \Leftarrow) Assume by contradiction that π is $\equiv_{\ell^\omega \circ q}$ -constant but not \equiv_{ℓ^ω} -constant. Then, there exists $k > 0$ such that $\pi[0] \not\equiv_\ell \pi[k]$. By definition of q , we have that $q(\pi)[0] = \pi[0]$, $q(\pi|_k)[0] = \pi[k]$, therefore $q(\pi) \equiv_{\ell^\omega} q(\pi|_k)$. In particular, this means that $\pi \equiv_{\ell^\omega \circ q} \pi|_k$, hence π is not $\equiv_{\ell^\omega \circ q}$ -constant. This is in contradiction with the hypothesis on π . (1: \Rightarrow) Assume that π is \equiv_{ℓ^ω} -constant. This implies also that, for all $i \in \mathbb{N}$, $\pi|_i$ is \equiv_{ℓ^ω} -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 $\equiv_{\ell^\omega \circ q}$ -constant.

(2) By $\pi \equiv_{\ell^\omega \circ q} \rho$, we have $q(\pi) \equiv_{\ell^\omega} q(\rho)$. By π is \equiv_{ℓ^ω} -constant and definition of q , $q(\pi) = \pi$, therefore $\pi \equiv_{\ell^\omega} q(\rho)$. In particular, $q(\rho)$ is \equiv_{ℓ^ω} -constant and, by definition of q , this is the case only when ρ is \equiv_{ℓ^ω} -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_{ℓ^ω} , we obtain $q(\pi|_1) \equiv_{\ell^\omega} q(\rho|_1)$. \square

Proof (of Theorem 11). (ii) Is a direct consequence of (i) and Lemma 6, 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 structural induction on $\varphi \in \text{LTL}^{\text{x}}$ we prove that $\llbracket \varphi \rrbracket \in \sigma(\mathcal{ST})$.

Atomic prop. $\llbracket p \rrbracket = \bigcup \{ \mathfrak{C}([u]_{\equiv_\ell}) \mid u \in S, p \in \ell(u) \}$. Since S is countable and $\mathfrak{C}([u]_{\equiv_\ell}) = \mathfrak{C}([u]_{\equiv_\ell}) \in \mathcal{ST}$ for all $u \in S$, then $\llbracket p \rrbracket \in \sigma(\mathcal{ST})$.

False. $\llbracket \perp \rrbracket = \emptyset \in \sigma(\mathcal{ST})$.

Implication. $\llbracket \phi \rightarrow \psi \rrbracket = \llbracket \neg\phi \vee \psi \rrbracket = \llbracket \phi \rrbracket^c \cup \llbracket \psi \rrbracket$. By inductive hypothesis, $\llbracket \phi \rrbracket, \llbracket \psi \rrbracket \in \sigma(\mathcal{ST})$, therefore $\llbracket \phi \rightarrow \psi \rrbracket \in \sigma(\mathcal{ST})$.

Until. Consider the function $q: S^\omega \rightarrow S^\omega$ of Proposition 47. By Proposition 47, $\equiv_{\ell^\omega \circ q} = \{ (\pi, \rho) \mid q(\pi) \equiv_{\ell^\omega} q(\rho) \}$ is a stutter relation. One can easily prove that $q(\pi) \equiv_{\ell^\omega} q(q(\pi))$ for all $\pi \in S^\omega$, hence $\pi \simeq q(\pi)$ $\pi \in S^\omega$. Then,

$$\begin{aligned} \llbracket \phi \text{ U } \psi \rrbracket &= \{ \pi \mid \exists i \geq 0. \mathcal{M}, \pi|_i \models \psi \text{ and } \forall 0 \leq j < i. \mathcal{M}, \pi|_j \models \phi \} \quad (\text{by def. U}) \\ &= \{ \pi \mid \exists i \geq 0. \mathcal{M}, q(\pi)|_i \models \psi \text{ and } \forall 0 \leq j < i. \mathcal{M}, q(\pi)|_j \models \phi \} \\ &\quad (\text{Prop.10}) \\ &= \{ \pi \mid \exists i \geq 0. q(\pi)|_i \in \llbracket \psi \rrbracket, \forall 0 \leq j < i. q(\pi)|_j \in \llbracket \phi \rrbracket \} \quad (\text{by def. } \llbracket \cdot \rrbracket) \\ &= \bigcup_{i \geq 0} \bigcap_{0 \leq j < i} ((\cdot)|_i \circ q)^{-1}(\llbracket \psi \rrbracket) \cap ((\cdot)|_j \circ q)^{-1}(\llbracket \phi \rrbracket). \quad (\text{preimage}) \end{aligned}$$

By Proposition 47, 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 $\llbracket \phi \text{ U } \psi \rrbracket \in \sigma(\mathcal{ST})$.

(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 $A: \mathbb{A} \times 2^S \rightarrow \text{LTL}^{\text{x}}$ and $B: \mathcal{ST} \rightarrow \text{LTL}^{\text{x}}$ as follows, for $i = 1..n$ and $C_i \in S/\equiv_\ell$ s.t. $C_i \neq C_{i+1}$,

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

$$B(\mathfrak{C}(C_1)) = \bigwedge_{p \in \mathbb{A}} A(p, C_0)$$

$$B(\mathfrak{C}(C_1 \cdots C_{n+1})) = (B(\mathfrak{C}(C_1)) \wedge \neg B(\mathfrak{C}(C_2)) \text{ U } B(\mathfrak{C}(C_2 \cdots C_{n+1}))),$$

For $T \in \mathcal{ST}$ one can prove that $\llbracket B(T) \rrbracket = T$. \square

Proof (of Lemma 12 —restated from [18, Th.5.2]). We prove that $\|\mu - \nu\|$ is a lower bound for $\{\omega(\cong) \mid \omega \in \Omega(\mu, \nu)\}$. Let $\omega \in \Omega(\mu, \nu)$ and $E \in \Sigma$, then

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

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

Now we prove that there exists an optimal coupling $\omega^* \in \Omega(\mu, \nu)$ such that $\omega^*(\cong) = \|\mu - \nu\|$. Define $\psi: X \rightarrow X \times X$ 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 $\omega^*(\cong) = \|\mu - \nu\|$ is trivial). Let $\mu \neq \nu$. Define $\mu \wedge \nu: \Sigma \rightarrow \mathbb{R}_+$ 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 [?, Corr.6 pp.163]). Now define the following derived measures

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

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 $E \in \Sigma$, then

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

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

$$\begin{aligned}
\|\mu - \nu\| &= 1 - (\mu \wedge \nu)(X) && \text{(def. } \mu \wedge \nu \text{ and compl.)} \\
&= 1 - (\mu \wedge \nu)[\psi](\cong) && \text{(def. } \psi) \\
&= 1 - \gamma && \text{(def. } \gamma) \\
&= 1 - \omega^*(\cong) && \text{(def. } \omega^*) \\
&= \omega^*(\not\cong) && \text{(compl.)}
\end{aligned}$$

□

Proof (of Lemma ??). For $\mathcal{E}_k \subseteq \mathcal{E}_{k+1}$, it suffices to prove $\mathfrak{C}^k(S/\equiv_\ell, \mathfrak{R}_k) \subseteq \mathcal{E}_{k+1}$. We proceed by induction on $k \geq 0$. The base case is trivial. Assume $k > 0$ and let $C \in \mathfrak{C}^k(S/\equiv_\ell, \mathfrak{R}_k)$. Note that, for any $n \in \mathbb{N}$ such that $0 \leq n < k2^k$, $\frac{n}{2^k} = \frac{2n}{2^{k+1}}$ and $2n < (k+1)2^{k+1}$. From this is immediate to prove that there exists $\mathcal{F} \subseteq \mathfrak{C}^k(S/\equiv_\ell, \mathfrak{R}_{k+1})$ such that $C = \bigcup \mathcal{F}$. Note that \mathfrak{R}_{k+1} is a partition of \mathbb{R}_+ (i.e., a family of pairwise disjoint subsets of \mathbb{R}_+ whose union is \mathbb{R}_+). So, any $C' = \mathfrak{C}(C_0, R_0, \dots, R_{k-1}, C_k) \in \mathfrak{C}^k(S/\equiv_\ell, \mathfrak{R}_{k+1})$ can be represented as

$$C' = \bigcup \{ \mathfrak{C}(C_0, R_0, \dots, R_{k-1}, C_k, R'', C'') \mid R'' \in \mathfrak{R}_{k+1}, C'' \in S/\equiv_\ell \}.$$

Since \mathfrak{R}_{k+1} and S/\equiv_ℓ are finite, from the above we get that C can be represented as a finite union of cylinders in $\mathfrak{C}^{k+1}(S/\equiv_\ell, \mathfrak{R}_{k+1})$, hence $C \in \mathcal{E}_{k+1}$.

Let $\mathcal{E} = \bigcup_{k \in \mathbb{N}} \mathcal{E}_k$. Since each $\mathfrak{C}^k(S/\equiv_\ell, \mathfrak{R}_k)$ forms a finite partition of S^ω , it is immediate to prove that \mathcal{E}_k is a field. Since the limit of an increasing sequence of fields is a field, then \mathcal{E} is a field.

It remains to show $\sigma(\mathcal{E}) = \sigma(\mathcal{T})$. Clearly $\mathcal{E} \subseteq \sigma(\mathcal{T})$, hence $\sigma(\mathcal{E}) \subseteq \sigma(\mathcal{T})$. As for the converse inclusion, let $\mathcal{R} = \bigcup_{k \in \mathbb{N}} \mathfrak{R}_k$ and recall that $\mathbb{B} = \sigma(\mathcal{CO})$, where $\mathcal{CO} = \{[q, q'] \mid q < q' \in \mathbb{Q}_+\} \cup \{[q, \infty) \mid q \in \mathbb{Q}_+\}$ is the family of left-closed right-open intervals with rational endpoints (or ∞). Let $q < q' \in \mathbb{Q}_+$, then the following hold

$$\begin{aligned}
[q, q'] &= \bigcup \left\{ \left[\frac{n}{2^k}, \frac{n+1}{2^k} \right) \mid q \leq \frac{n}{2^k} < \frac{n+1}{2^k} \leq q', \text{ for } k \in \mathbb{N}, 0 \leq n < k2^k \right\}, \\
[q, \infty) &= \bigcup \left\{ \left[\frac{n}{2^k}, \frac{n+1}{2^k} \right) \mid q \leq \frac{n}{2^k}, \text{ for } k \in \mathbb{N}, 0 \leq n < k2^k \right\}.
\end{aligned}$$

The above suffices to prove $\mathcal{CO} \subseteq \sigma(\mathcal{R})$, hence $\mathbb{B} = \sigma(\mathcal{CO}) \subseteq \sigma(\mathcal{R})$. This proves $\sigma(\mathfrak{C}(S/\equiv_\ell, \mathcal{R})) \subseteq \sigma(\mathcal{T})$. Clearly, $\mathcal{E} \subseteq \sigma(\mathfrak{C}(S/\equiv_\ell, \mathcal{R}))$, therefore $\sigma(\mathcal{E}) \subseteq \sigma(\mathcal{T})$. □

Proposition 48. *Let Σ be a σ -algebra on X generated by $\mathcal{F} \subseteq 2^X$. Then the inseparability relations w.r.t. Σ and \mathcal{F} coincide:*

$$\cong_\Sigma := \bigcap \{E \times E \mid E \in \Sigma\} = \bigcap \{F \times F \mid F \in \mathcal{F}\} =: \cong_{\mathcal{F}}.$$

Proof. (\subseteq) It immediately follows by $\mathcal{F} \subseteq \Sigma$. (\supseteq) Let \mathcal{U} be the smallest family of subsets of X that contains \mathcal{F} and is closed under complement and (generic) union. Define $\cong_{\mathcal{U}} := \bigcap \{E \times E \mid E \in \mathcal{U}\}$. Clearly $\Sigma \subseteq \mathcal{U}$. This means that to prove the inclusion it suffices to prove $\cong_{\mathcal{F}} \subseteq \cong_{\mathcal{U}}$. 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. \square

Proposition 49. \equiv_{ℓ^ω} is the inseparability relation w.r.t. $\sigma(\mathcal{T})$ and it is a measurable set in $\sigma(\mathcal{T}) \otimes \sigma(\mathcal{T})$, i.e., $\equiv_{\ell^\omega} = \cong_{\sigma(\mathcal{T})} \in \sigma(\mathcal{T}) \otimes \sigma(\mathcal{T})$.

Proof. We first show $\equiv_{\ell^\omega} = \bigcap \{E \times E \mid E \in \sigma(\mathcal{T})\}$. (\subseteq) By Proposition 48, it suffices to prove inseparability w.r.t trace cylinders. Let $\pi \equiv_{\ell^\omega} \rho$ and $\pi \in C = \mathfrak{C}(C_1 \cdots C_n)$, for some $C_i \in S/\equiv_{\ell}$ ($i = 1..n$). By definition of \equiv_{ℓ^ω} , for all $j \in \mathbb{N}$, $\pi[j] \equiv_{\ell} \rho[j]$, so that $\rho \in \mathfrak{C}(C_1 \cdots C_n)$. (\supseteq) By contraposition. 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 \equiv_{ℓ^ω} follows by Proposition 48 and $\equiv_{\ell^\omega} = \cong_{\sigma(\mathcal{T})}$. \square

Proof (of Lemma 19). Let $\mathcal{C} \in \mathbb{C}_k$. To prove $\mathbb{P}_{\mathcal{C}}(u, v) \in \Omega(\mathbb{P}(u), \mathbb{P}(v))$ it suffices to show that, for all $n \in \mathbb{N}$ and $U_i \subseteq S$ ($i = 0..nk$) the following hold

- (i) $\mathbb{P}_{\mathcal{C}}(u, v)(\mathfrak{C}(U_0 \cdots U_{nk}) \times S^\omega) = \mathbb{P}(u)(\mathfrak{C}(U_0 \cdots U_{nk}))$;
- (ii) $\mathbb{P}_{\mathcal{C}}(u, v)(S^\omega \times \mathfrak{C}(U_0 \cdots U_{nk})) = \mathbb{P}(v)(\mathfrak{C}(U_0 \cdots U_{nk}))$.

We prove (i) by induction on $n \geq 0$. The base case is trivial. Let $n > 0$. For any $\mathbf{v} \in S^{nk+1}$ define $F^{\mathbf{v}} = \{v_0\} \times \mathbb{R}_+ \times \dots \times \mathbb{R}_+ \times \{v_{nk}\}$ and, for $h < n$, let $F_h^{\mathbf{v}} = \{v_{hk}\} \times \mathbb{R}_+ \times \dots \times \mathbb{R}_+ \times \{v_{(h+1)k}\}$. Then the following hold

$$\begin{aligned}
 & \mathbb{P}_{\mathcal{C}}(s, s')(\mathfrak{C}(E) \times S^\omega) = \\
 &= \sum_{\mathbf{v} \in S^{nk+1}} \mathbb{P}_{\mathcal{C}}(s, s')(\mathfrak{C}(E) \times \mathfrak{C}(F^{\mathbf{v}})) \quad (\text{additivity}) \\
 &= \sum_{\mathbf{v} \in S^{nk+1}} \mathbb{1}_{\{(s, s')\}}(u_0, v_0) \cdot \prod_{h=0}^{n-1} \mathcal{C}(u_{hk}, v_{hk})(E_h \times F_h^{\mathbf{v}}) \quad (\text{def. } \mathbb{P}_{\mathcal{C}}(s, s')) \\
 &= \sum_{\mathbf{v} \in S^{(n-1)k+1}} \mathbb{P}_{\mathcal{C}}(s, s')(\mathfrak{C}(E') \times \mathfrak{C}(F^{\mathbf{v}})) \cdot \mathcal{C}(s_{(n-1)k}, v_{(n-1)k})(E_{(n-1)} \times S^k) \\
 & \quad (\text{def. } \mathbb{P}_{\mathcal{C}}(s, s'))
 \end{aligned}$$

$$\begin{aligned}
&= \sum_{\mathbf{v} \in S^{(n-1)k+1}} \mathbb{P}_{\mathcal{C}}(s, s')(\mathfrak{C}(E') \times \mathfrak{C}(F^{\mathbf{v}})) \cdot \mathbb{P}(s_{(n-1)k})(E_{(n-1)}) && (\mathcal{C} \in \mathbb{C}_k) \\
&= \mathbb{P}_{\mathcal{C}}(s, s')(\mathfrak{C}(E') \times S^{\omega}) \cdot \mathbb{P}(s_{(n-1)k})(E_{(n-1)}) && (\text{additivity}) \\
&= \mathbb{P}(s)(\mathfrak{C}(E')) \cdot \mathbb{P}(s_{(n-1)k})(E_{(n-1)}) && (\text{inductive hp.}) \\
&= \mathbb{P}(s)(\mathfrak{C}(E)) && (\text{def. } \mathbb{P}_{\mathcal{C}}(s, s'))
\end{aligned}$$

where $E' = \{u_0\} \times R_0 \times \dots \times R_{(n-1)k-1} \times \{u_{(n-1)k}\}$. (ii) follows similarly. \square

Proof (of Lemma 21). We prove the two items separately.

(i) Let $k > 0$ and $\mathcal{C} \in \mathbb{C}_k$. Define, for all $s, s' \in S$, $\mathcal{D}(s, s')$ as the unique measure on $S^{2k} \otimes S^{2k}$ s.t., for all $E = \{u_0\} \times R_0 \times \dots \times R_{2k-1} \times \{u_{2k}\}$ and $F = \{v_0\} \times H_0 \times \dots \times H_{2k-1} \times \{v_{2k}\}$ in \mathcal{R}_{2k}

$$\mathcal{D}(s, s')(E \times F) = \mathcal{C}(s, s')(E' \times F') \cdot \mathcal{C}(u_k, v_k)(E'' \times F''),$$

where $E' = \{u_0\} \times R_0 \times \dots \times R_{k-1} \times \{u_k\}$ and $E'' = \{u_k\} \times R_k \times \dots \times R_{2k-1} \times \{u_{2k}\}$ (similarly for F). To show $\mathcal{D} \in \mathbb{C}_{2k}$ we need to prove that for all $s, s' \in S$, $\mathcal{D}(s, s') \in \Omega(\mathbb{P}(s)|^{2k}, \mathbb{P}(s')|^{2k})$. To this end it is sufficient that, for all measurable sets $E = \{u_0\} \times R_0 \times \dots \times R_{2k-1} \times \{u_{2k}\} \in \mathcal{R}_{2k}$, the following hold

$$\mathcal{D}(s, s')(E \times S^{2k}) \stackrel{(i)}{=} \mathbb{P}(s)|^{2k}(E), \quad \mathcal{D}(s, s')(S^{2k} \times E) \stackrel{(ii)}{=} \mathbb{P}(s')|^{2k}(E).$$

We prove only (i). For any $\mathbf{v} \in S^{2k+1}$ define $F^{\mathbf{v}} = \{v_0\} \times \mathbb{R}_+ \times \dots \times \mathbb{R}_+ \times \{v_{2k}\}$ and, for $h = 0..1$, let $F_h^{\mathbf{v}} = \{v_{hk}\} \times \mathbb{R}_+ \times \dots \times \mathbb{R}_+ \times \{v_{(h+1)k}\}$. Then we have

$$\begin{aligned}
\mathcal{D}(s, s')(E \times S^{2k}) &= \\
&= \sum_{\mathbf{v} \in S^{2k+1}} \mathcal{D}(s, s')(E \times F^{\mathbf{v}}) && (\text{additivity}) \\
&= \sum_{\mathbf{v} \in S^{2k+1}} \mathcal{C}(s, s')(E' \times F_0^{\mathbf{v}}) \cdot \mathcal{C}(u_k, v_k)(E'' \times F_1^{\mathbf{v}}) && (\text{def. } \mathcal{D}) \\
&= \sum_{\mathbf{v} \in S^{k+1}} \mathcal{C}(s, s')(E' \times F_0^{\mathbf{v}}) \cdot \mathcal{C}(u_k, v_k)(E'' \times S^k) && (\text{additivity}) \\
&= \sum_{\mathbf{v} \in S^{k+1}} \mathcal{C}(s, s')(E' \times F_0^{\mathbf{v}}) \cdot \mathbb{P}(u_k)|^k(E'') && (\mathcal{C} \in \mathbb{C}_k) \\
&= \mathcal{C}(s, s')(E' \times S^k) \cdot \mathbb{P}(u_k)|^k(E'') && (\text{additivity}) \\
&= \mathbb{P}(s)|^k(E') \cdot \mathbb{P}(u_k)|^k(E'') && (\mathcal{C} \in \mathbb{C}_k) \\
&= \mathbb{P}(s)(\mathfrak{C}(E')) \cdot \mathbb{P}(u_k)(\mathfrak{C}(E'')) && (\text{preimage}) \\
&= \mathbb{P}(s)(\mathfrak{C}(E)) && (\text{def. } \mathbb{P}(s)) \\
&= \mathbb{P}(s)|^{2k}(E). && (\text{preimage})
\end{aligned}$$

We show that, for arbitrary $s, s' \in S$, $\mathbb{P}_{\mathcal{C}}(s, s') = \mathbb{P}_{\mathcal{D}}(s, s')$. To this end it suffices to check the following for all $n \in \mathbb{N}$ and $E = \{u_0\} \times R_0 \times \dots \times R_{2nk-1} \times \{u_{2nk}\}$, $F = \{v_0\} \times H_0 \times \dots \times H_{2nk-1} \times \{v_{2nk}\}$ in \mathcal{R}_{2nk} :

$$\mathbb{P}_{\mathcal{C}}(s, s')(\mathfrak{C}(E) \times \mathfrak{C}(F)) = \mathbb{P}_{\mathcal{D}}(s, s')(\mathfrak{C}(E) \times \mathfrak{C}(F))$$

We proceed by induction on $n \geq 0$. The base case is trivial. Assume $n > 0$ and, for $i \in \{k, 2k\}$, define $E_h^i = \{u_{hi}\} \times R_{hi} \times \dots \times R_{(h+1)i-1} \times \{u_{(h+1)i}\}$ (similarly for F). Then the following holds:

$$\begin{aligned} \mathbb{P}_{\mathcal{C}}(s, s')(\mathfrak{C}(E) \times \mathfrak{C}(F)) &= \mathbb{1}_{\{(s, s')\}}(u_0, v_0) \cdot \prod_{h=0}^{2n-1} \mathcal{C}(u_{hk}, v_{hk})(E_h^k \times F_h^k) && (\text{def. } \mathbb{P}_{\mathcal{C}}) \\ &= \mathbb{1}_{\{(s, s')\}}(u_0, v_0) \cdot \prod_{h=0}^{n-1} \mathcal{D}(u_{2hk}, v_{2hk})(E_h^{2k} \times F_h^{2k}) && (\text{def. } \mathcal{D}) \\ &= \mathbb{P}_{\mathcal{D}}(s, s')(\mathfrak{C}(E) \times \mathfrak{C}(F)). && (\text{def. } \mathbb{P}_{\mathcal{D}}) \end{aligned}$$

From the above it immediately follows that $\mathbb{C}_k \subseteq \mathbb{C}_{2k}$.

(ii) We prove the following more general result from which we will obtain (ii).

Let (X, Σ) be a measurable space such that \mathcal{F} is a field that generates 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 $d_E: \Delta(X) \times \Delta(X) \rightarrow \mathbb{R}_+$ 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 53). 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 [?, Corr.6 pp.163] (note that $\tilde{\mu}$ is not necessarily finite). By Lemma ??, $\mathcal{F} \subseteq \Sigma$ is dense in $(\Sigma, d_{\tilde{\mu}})$, where $d_{\tilde{\mu}}$ is the Fréchet-Nikodym pseudometric⁵, hence 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

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

This implies $|\omega(E) - \omega(F)| \leq d_{\tilde{\mu}}(E, F)$, and in particular that $|\mu(E) - \mu(F)| < \frac{\varepsilon}{2}$ and $|\nu(E) - \nu(F)| < \frac{\varepsilon}{2}$. Then, the density of D follows by

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

⁵ Notice that Lemma ?? does not assume the measure to be finite, hence it can be safely applied to $\tilde{\mu}$.

Let $s, s' \in S$, $\Omega = \bigcup_{i \in \mathbb{N}} \{\mathbb{P}_{\mathcal{C}}(s, s') \mid \mathcal{C} \in \mathbb{C}_{2^i}\}$. Given the general result above, to prove (ii) it is sufficient to provide a field \mathcal{F} that generates the σ -algebra of $S^\omega \otimes S^\omega$ and to 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 $\mathcal{F} = \bigcup_{k \in \mathbb{N}} \mathcal{F}_k$, 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 $E, F \in \mathcal{R}_k$. It holds that $\mathcal{F}_k \subseteq \mathcal{F}_{k+1}$ and \mathcal{F}_k is a field, for all $k \in \mathbb{N}$. Therefore \mathcal{F} is a field that generates the σ -algebra of $S^\omega \otimes S^\omega$.

Let $\mu \in \Omega(\mathbb{P}(s), \mathbb{P}(s'))$, $k \in \mathbb{N}$ and $\mathcal{D} \in \mathbb{C}_k$. We define $\omega_k = \mathbb{P}_{\mathcal{C}_k}(s, s')$, where $\mathcal{C}_k: S \times S \rightarrow \Delta(S^k \times S^k)$ is defined by

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

where $\mu[(\cdot)^k \times (\cdot)^k]$ denotes the push forward of μ along $(\pi, \pi') \mapsto (\pi|^k, \pi'|^k)$. Note that, since \mathbb{C}_k is nonempty, \mathcal{C}_k is well defined. We show $\mathcal{C}_k \in \mathbb{C}_k$. We just need to prove $\mu[(\cdot)^k \times (\cdot)^k] \in \Omega(\mathbb{P}(s)^k, \mathbb{P}(s')^k)$ that corresponds to check $\mu[(\cdot)^k \times (\cdot)^k](E \times S^k) = \mathbb{P}(s)^k(E)$ and $\mu[(\cdot)^k \times (\cdot)^k](S^k \times E) = \mathbb{P}(s')^k(E)$ for arbitrary $E \in \mathcal{R}_k$ (we check one equality, the other follows similarly):

$$\begin{aligned} \mu[(\cdot)^k \times (\cdot)^k](E \times S^k) &= \mu(\mathfrak{C}(E) \times S^\omega) && \text{(preimage)} \\ &= \mathbb{P}(s)(\mathfrak{C}(E)) && (\mu \in \Omega(\mathbb{P}(s), \mathbb{P}(s'))) \\ &= \mathbb{P}(s)^k(E). && \text{(preimage)} \end{aligned}$$

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 $j \leq k$. Let $A = \bigcup_{i=0}^n \mathfrak{C}(E_i) \times \mathfrak{C}(F_i) \in \mathcal{F}_k$, for some $n \in \mathbb{N}$ and $E_i, F_i \in \mathcal{R}_k$ ($i = 0..n$). 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).

$$\begin{aligned} \omega_k(A) &= \mathbb{P}_{\mathcal{C}_k}(s, s')(A) && \text{(def. } \omega_k) \\ &= \sum_{i=0}^n \mathbb{P}_{\mathcal{C}_k}(s, s')(\mathfrak{C}(E_i) \times \mathfrak{C}(F_i)) && \text{(additivity)} \\ &= \sum_{i=0}^n \mathcal{C}_k(s, s')(E_i \times F_i) && \text{(def. } \mathbb{P}_{\mathcal{C}_k}(s, s')) \\ &= \sum_{i=0}^n \mu(\mathfrak{C}(E_i) \times \mathfrak{C}(F_i)) && \text{(def. } \mathcal{C}_k) \\ &= \mu(A). && \text{(additivity)} \end{aligned}$$

To conclude the proof, observe that, given $\mu \in \Omega(\mathbb{P}(s), \mathbb{P}(s'))$ and $F \in \mathcal{F}$, there exists $i \in \mathbb{N}$ such that $F \in \mathcal{F}_i$, and that for ω_{2^i} defined as above (w.r.t. μ) is such that $\omega_{2^i}(F) = \mu(F)$ and $\omega_{2^i} \in \Omega$. \square

Proof (of Lemma ??). Consider the functions p_1 and p_2 defined as

$$\begin{aligned} p_1: S^k &\rightarrow S^{k+1} & p_2: S^k &\rightarrow \mathbb{R}_+^k \\ p_1(s_0, t_0, \dots, t_{k-1}, s_k) &= (s_0, \dots, s_k) & p_2(s_0, t_0, \dots, t_{k-1}, s_k) &= (t_0, \dots, t_{k-1}). \end{aligned}$$

Note that p_1, p_2 are measurable. For $\mathcal{C} \in \mathbb{C}_k$ and (ϑ, η) satisfying the conditions of the statement, the bijection is given by $\mathcal{C} \mapsto (\tau_{\mathcal{C}}, \rho_{\mathcal{C}})$ and $(\vartheta, \eta) \mapsto \mathcal{D}$, where

$$\begin{aligned} \tau_{\mathcal{C}}(u, v) &= \mathcal{C}(u, v)[p_1 \times p_1], \\ \rho_{\mathcal{C}}(u_1..u_k, v_1..v_k) &= \begin{cases} \frac{\mathcal{C}(u_1, v_1)[p_2 \times p_2]}{\alpha} & \text{if } \alpha = \tau_{\mathcal{C}}(u_1, v_1)(u_1..u_k S \times v_1..v_k S) \neq 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

and

$$\mathcal{D}(s, s')(E \times F) = \vartheta(u_0, v_0)(u_0..u_k, v_0..v_k) \cdot \eta(u_0..u_{k-1}, v_0..v_{k-1})(R, H).$$

for $E = \{u_0\} \times R_0 \times \dots \times R_{k-1} \times \{u_k\}$, $F = \{v_0\} \times H_0 \times \dots \times H_{k-1} \times \{v_k\} \in \mathcal{R}_k$, and $R = R_0 \times \dots \times R_{k-1}$, $H = H_0 \times \dots \times H_{k-1}$. \square

Proof (of Lemma ??). Let $k \in \mathbb{N}$ and $\mathcal{C} = (\tau_{\mathcal{C}}, \rho_{\mathcal{C}}) \in \mathbb{C}_k$ be a coupling structure for $\mathcal{M} = (S, \tau, \rho, \ell)$. Define $g: S^{k+1} \times S^{k+1} \rightarrow [0, 1]$, for $x, y \in S^{k+1}$, as

$$g(x, y) = \begin{cases} 0 & \text{if } \tau_{\mathcal{C}}(x_0, y_0)(x, y) = 0 \\ \mathbb{P}_{\mathcal{C}}(x_0, y_0)(\neq_{\ell\omega} \{(\pi_1, \pi_2)[0..k] = (x, y)\}) & \text{otherwise} \end{cases} \quad (3)$$

where $\mathbb{P}(A|B)$ denotes the conditional probability of A given B w.r.t. \mathbb{P} (defined as $\mathbb{P}(A|B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$, when $\mathbb{P}(B) > 0$) and $\{(\pi_1, \pi_2)[0..k] = (x, y)\}$ stands for the event $(\cdot, \cdot)[0..k]^{-1}(\{(x, y)\})$, where the function $(\cdot, \cdot)[0..k]$ is defined by $(\pi_1, \pi_2) \mapsto (\pi_1[0].. \pi_1[k], \pi_2[0].. \pi_2[k])$ (easily checked to be measurable). Note that g is well defined since $\mathbb{P}_{\mathcal{C}}(x_0, y_0)(\{(\pi_1, \pi_2)[0..k] = (x, y)\}) = \tau_{\mathcal{C}}(x_0, y_0)(x, y)$.

To prove $\mathbb{P}_{\mathcal{C}}(s, s')(\neq_{\ell\omega}) = \int \gamma^{\mathcal{C}} d\tau_{\mathcal{C}}(s, s')$ it suffices to show that $g = \gamma^{\mathcal{C}}$. Indeed

$$\begin{aligned} \mathbb{P}_{\mathcal{C}}(s, s')(\neq_{\ell\omega}) &= \int \mathbb{P}_{\mathcal{C}}(s, s')(\neq_{\ell\omega} | \{(\pi_1, \pi_2)[0..k] = (\cdot, \cdot)\}) d\mathbb{P}_{\mathcal{C}}(s, s')[(\cdot, \cdot)[0..k]] \\ &\quad \text{(cond. pr.)} \\ &= \int \mathbb{P}_{\mathcal{C}}(s, s')(\neq_{\ell\omega} | \{(\pi_1, \pi_2)[0..k] = (\cdot, \cdot)\}) d\tau_{\mathcal{C}}(s, s') \quad \text{(def. } \mathbb{P}_{\mathcal{C}}) \\ &= \int g d\tau_{\mathcal{C}}(s, s') = \int \gamma^{\mathcal{C}} d\tau_{\mathcal{C}}(s, s'). \quad \text{(by (3) and } g = \gamma^{\mathcal{C}}) \end{aligned}$$

First we prove that g is a fixed point of $\Gamma^{\mathcal{C}}$. We proceed by cases

Case $\tau_{\mathcal{C}}(x_0, y_0)(x, y) = 0$. By definition of $\Gamma^{\mathcal{C}}$ and (3), $\Gamma^{\mathcal{C}}(g)(x, y) = 0 = g(x, y)$.

Case $\tau_{\mathcal{C}}(x_0, y_0)(x, y) > 0$ and $\exists i \leq k. x_i \neq_{\ell} y_i$. The following hold

$$\begin{aligned} g(x, y) &= \mathbb{P}_{\mathcal{C}}(x_0, y_0)(\neq_{\ell\omega} | \{(\pi_1, \pi_2)[0..k] = (x, y)\}) \quad \text{(by (3))} \\ &= \frac{\mathbb{P}_{\mathcal{C}}(x_0, y_0)(\neq_{\ell\omega} \cap \{(\pi_1, \pi_2)[0..k] = (x, y)\})}{\mathbb{P}_{\mathcal{C}}(x_0, y_0)(\{(\pi_1, \pi_2)[0..k] = (x, y)\})} \quad \text{(cond. pr.)} \\ &= \frac{\mathbb{P}_{\mathcal{C}}(x_0, y_0)(\{(\pi_1, \pi_2)[0..k] = (x, y)\})}{\mathbb{P}_{\mathcal{C}}(x_0, y_0)(\{(\pi_1, \pi_2)[0..k] = (x, y)\})} = 1 = \Gamma^{\mathcal{C}}(g)(x, y), \end{aligned}$$

where the last equalities follow by $\{(\pi_1, \pi_2)[0..k] = (x, y)\} \subseteq \neq_{\ell\omega}$ (because by hypothesis $\exists i. x_i \neq_{\ell} y_i$) and definition of $\Gamma^{\mathcal{C}}$.

Case $\tau_{\mathcal{C}}(x_0, y_0)(x, y) > 0$ and $\forall i \leq k. x_i \equiv_{\ell} y_i$. Let $A = \{(\pi_1, \pi_2)[0..k] = (x, y)\}$ and $B = \{(\pi_1, \pi_2)\langle 0..k-1 \rangle \in \neq\}$ (i.e., the event $(\cdot, \cdot)\langle 0..k-1 \rangle^{-1}(\neq)$, where the function $(\cdot, \cdot)\langle 0..k-1 \rangle$ is defined by $(\pi_1, \pi_2) \mapsto (\pi_1\langle 0 \rangle.. \pi_1\langle k-1 \rangle, \pi_2\langle 0 \rangle.. \pi_2\langle k-1 \rangle)$ and it is easy to see that it is measurable).

Let $\beta = \rho_{\mathcal{C}}(x_0..x_{k-1}, y_0..y_{k-1})(\neq)$. We show that the following hold

- (i) $\mathbb{P}_{\mathcal{C}}(x_0, y_0)(\neq_{\ell\omega} \cap B|A) = \beta$;
- (ii) $\mathbb{P}_{\mathcal{C}}(x_0, y_0)(\neq_{\ell\omega} \cap B^c|A) = (1 - \beta) \cdot \int g \, d\tau_{\mathcal{C}}(x_k, y_k)$.

Note that once we have shown (i–ii), $g(x, y) = \Gamma^{\mathcal{C}}(g)(x, y)$ follows immediately:

$$\begin{aligned}
g(x, y) &= \mathbb{P}_{\mathcal{C}}(x_0, y_0)(\neq_{\ell\omega} |A) && \text{(by (3))} \\
&= \mathbb{P}_{\mathcal{C}}(x_0, y_0)(\neq_{\ell\omega} \cap B|A) + \mathbb{P}_{\mathcal{C}}(x_0, y_0)(\neq_{\ell\omega} \cap B^c|A) && \text{(by additivity)} \\
&= \beta + (1 - \beta) \cdot \int g \, d\tau_{\mathcal{C}}(x_k, y_k) && \text{(by (i) and (ii))} \\
&= \Gamma^{\mathcal{C}}(g)(x, y). && \text{(by def. } \Gamma^{\mathcal{C}})
\end{aligned}$$

We show (i):

$$\begin{aligned}
\mathbb{P}_{\mathcal{C}}(x_0, y_0)(\neq_{\ell\omega} \cap B|A) &= \mathbb{P}_{\mathcal{C}}(x_0, y_0)(B|A) && \text{(by } B \subseteq \neq_{\ell\omega}) \\
&= \rho_{\mathcal{C}}(x_0..x_{k-1}, y_0..y_{k-1})(\neq) && \text{(by def. } \mathbb{P}_{\mathcal{C}}) \\
&= \beta && \text{(by def. } \beta)
\end{aligned}$$

We show (ii):

$$\begin{aligned}
&\mathbb{P}_{\mathcal{C}}(x_0, y_0)(\neq_{\ell\omega} \cap B^c|A) = \\
&= \frac{\mathbb{P}_{\mathcal{C}}(x_0, y_0)(\neq_{\ell\omega} \cap B^c \cap A)}{\mathbb{P}_{\mathcal{C}}(x_0, y_0)(A)} && \text{(by cond. pr.)} \\
&= \frac{\tau_{\mathcal{C}}(x_0, y_0)(x, y) \cdot \rho_{\mathcal{C}}(x_0..x_{k-1}, y_0..y_{k-1})(=) \cdot \mathbb{P}_{\mathcal{C}}(x_k, y_k)(\neq_{\ell\omega})}{\tau_{\mathcal{C}}(x_0, y_0)(x, y)} && \text{(by def. } \mathbb{P}_{\mathcal{C}}) \\
&= (1 - \beta) \cdot \mathbb{P}_{\mathcal{C}}(x_k, y_k)(\neq_{\ell\omega}) && \text{(by def. } \beta \text{ and compl.)} \\
&= (1 - \beta) \cdot \int \mathbb{P}_{\mathcal{C}}(x_k, y_k)(\neq_{\ell\omega} | \{(\pi_1, \pi_2)[0..k] = (\cdot, \cdot)\}) \, d\mathbb{P}_{\mathcal{C}}(x_k, y_k)[(\cdot, \cdot)[0..k]] && \\
& && \text{(cond. pr.)} \\
&= (1 - \beta) \cdot \int \mathbb{P}_{\mathcal{C}}(x_k, y_k)(\neq_{\ell\omega} | \{(\pi_1, \pi_2)[0..k] = (\cdot, \cdot)\}) \, d\tau_{\mathcal{C}}(x_k, y_k) && \text{(def. } \mathbb{P}_{\mathcal{C}}) \\
&= (1 - \beta) \cdot \int g \, d\tau_{\mathcal{C}}(x_k, y_k). && \text{(by (3))}
\end{aligned}$$

Now we prove by contradiction that g is actually the least fixed point of $\Gamma^{\mathcal{C}}$ (i.e., $\gamma^{\mathcal{C}} = g$). Assume that $\gamma^{\mathcal{C}} < g$ and let

$$m = \max_{x, y \in S^{k+1}} \{g(x, y) - \gamma^{\mathcal{C}}(x, y)\}, \quad x M y \iff g(x, y) - \gamma^{\mathcal{C}}(x, y) = m.$$

We show that $m = 0$, that is $\gamma^{\mathcal{C}} = g$. Assume $x M y$, we distinguish 3 cases

1. If $\tau_{\mathcal{C}}(x_0, y_0)(x, y) = 0$, then by definition of $\Gamma^{\mathcal{C}}$ and the fact that g and $\gamma^{\mathcal{C}}$ are fixed points of it, we have that $m = g(x, y) - \gamma^{\mathcal{C}}(x, y) = 0 - 0 = 0$.
2. If $\tau_{\mathcal{C}}(x_0, y_0)(x, y) > 0$ and $x_i \not\equiv_{\ell} y_i$ for some $0 \leq i \leq k$. Analogously, we have that $m = g(x, y) - \gamma^{\mathcal{C}}(x, y) = 1 - 1 = 0$.
3. If $\tau_{\mathcal{C}}(x_0, y_0)(x, y) > 0$ and $x_i \equiv_{\ell} y_i$ for all $0 \leq i \leq k$. Let $\beta = \rho_{\mathcal{C}}(x, y)(\neq)$, then the following equalities hold

$$\begin{aligned}
 m &= g(x, y) - \gamma^{\mathcal{C}}(x, y) && \text{(by } x M y\text{)} \\
 &= \Gamma^{\mathcal{C}}(g)(x, y) - \Gamma^{\mathcal{C}}(\gamma^{\mathcal{C}})(x, y) && \text{(} g \text{ and } \gamma^{\mathcal{C}} \text{ fixed points)} \\
 &= (1 - \beta) \cdot \int (g - \gamma^{\mathcal{C}}) d\tau_{\mathcal{C}}(x_k, y_k) && \text{(by def. } \Gamma^{\mathcal{C}}\text{)} \\
 &= (1 - \beta) \cdot \sum_{u, v \in S^{k+1}} (g(u, v) - \gamma^{\mathcal{C}}(u, v)) \cdot \tau_{\mathcal{C}}(x_k, y_k)(u, v). && (4)
 \end{aligned}$$

By hypothesis on m and $\tau_{\mathcal{C}}$ we have respectively that $g(u, v) - \gamma^{\mathcal{C}}(u, v) \leq m$ for all $u, v \in S^{k+1}$ and $\sum_{u, v \in S^{k+1}} \tau_{\mathcal{C}}(x_k, y_k)(u, v) = 1$, therefore it holds that

$$(1 - \beta) \cdot \sum_{u, v \in S^{k+1}} (g(u, v) - \gamma^{\mathcal{C}}(u, v)) \cdot \tau_{\mathcal{C}}(x_k, y_k)(u, v) \leq (1 - \beta)m. \quad (5)$$

We distinguish two cases:

- if $\beta > 0$, then $1 - \beta < 1$. By (4) and (5) we have that $m \leq (1 - \beta)m$. By the assumption on β this holds only for $m = 0$;
- if $\beta = 0$, by (5) and (5) we have that $g(u, v) - \gamma^{\mathcal{C}}(u, v) = m$ whenever $\tau_{\mathcal{C}}(x_k, y_k)(u, v) > 0$. Thus $\tau_{\mathcal{C}}(x_k, y_k)$ has support contained in M . By the generality of x and y one can prove that

$$g(x, y) \stackrel{(3)}{=} \mathbb{P}_{\mathcal{C}}(x_0, y_0)(\neq_{\ell\omega} | \{\pi_1[0..k] = x, \pi_2[0..k] = y\}) = 0.$$

Therefore $\gamma^{\mathcal{C}}(x, y) \not\leq g(x, y) = 0$, hence $m = 0$.

This proves that $\gamma^{\mathcal{C}} = g$.

This proves the thesis. \square

Proof (of Lemma ??). Let $\mathcal{C} = (\tau_{\mathcal{C}}, \rho_{\mathcal{C}}) \in \mathbb{C}_k$ be a coupling structure and $u_0..u_k, v_0..v_k \in S$ such that $\tau_{\mathcal{C}}(u_0..u_k, v_0..v_k) > 0$ and, for all $i \leq k$, $u_i \equiv_{\ell} v_i$. Consider $\mu \in \Omega(\mathbb{S}^k(u_k), \mathbb{S}^k(v_k))$, $\nu \in \Omega(\mathbb{T}^k(u_0..u_{k-1}), \mathbb{T}^k(v_0..v_{k-1}))$ and let $\mathcal{D} = \mathcal{C}[(u_k, v_k)/\mu] \langle (u_0..u_{k-1}, v_0..v_{k-1})/\nu \rangle$ be an update of \mathcal{C} .

We will prove that if (i) or (ii) holds then $\gamma^{\mathcal{C}}$ is a proper prefix point of $\Gamma^{\mathcal{D}}$, that is, $\Gamma^{\mathcal{D}}(\gamma^{\mathcal{C}}) < \gamma^{\mathcal{C}}$. Then, the thesis follows by Tarski's fixed point theorem.

To this end, define α, α' and β, β' as

$$\begin{aligned}
 \alpha &= \int \gamma^{\mathcal{C}} d\mu && \alpha' = \int \gamma^{\mathcal{C}} d\tau_{\mathcal{C}}(u_k, v_k), \\
 \beta &= \nu(\neq) && \beta' = \rho_{\mathcal{C}}(u_0..u_{k-1}, v_0..v_{k-1})(\neq).
 \end{aligned}$$

Then, the following inequalities hold

$$\begin{aligned}
\Gamma^{\mathcal{D}}(\gamma^{\mathcal{C}})(u_0..u_k, v_0..v_k) &= \\
&= \beta + (1 - \beta)\alpha && \text{(def. } \Gamma^{\mathcal{D}}) \\
&\leq \beta + (1 - \beta)\alpha' && (\alpha \leq \alpha') \\
&= \alpha' - \alpha' + \beta + (1 - \beta)\alpha' \\
&= \alpha' - \beta\alpha' - (1 - \beta)\alpha' + \beta + (1 - \beta)\alpha' && (0 \leq \beta \leq 1) \\
&= \alpha' - \beta\alpha' + \beta = \alpha' + (1 - \alpha')\beta \\
&\leq \alpha' + (1 - \alpha')\beta' && (\beta \leq \beta') \\
&= \beta' + (1 - \beta')\alpha' && \text{(same as for } \beta + (1 - \beta)\alpha' = \alpha' + (1 - \alpha')\beta) \\
&= \Gamma^{\mathcal{C}}(\gamma^{\mathcal{C}})(u_0..u_k, v_0..v_k) && \text{(def. } \Gamma^{\mathcal{C}}) \\
&= \gamma^{\mathcal{C}}. && \text{(def. } \gamma^{\mathcal{C}})
\end{aligned}$$

In particular, for (i) $\beta < \beta'$ or (ii) $\alpha < \alpha'$, the above inequality is strict.

By construction of \mathcal{D} and definition of Γ , it is immediate to prove that, for arbitrary $u, v \in S^{k+1}$, $\Gamma^{\mathcal{D}}(\gamma^{\mathcal{C}})(u, v) \leq \gamma^{\mathcal{C}}(u, v)$. This proves that if (i) or (ii) hold, then $\gamma^{\mathcal{D}} < \gamma^{\mathcal{C}}$. \square

Proof (of Lemma ??). By contradiction. Assume $\delta_{\downarrow k}(u, v) \neq \int \gamma^{\mathcal{C}} d\tau_{\mathcal{C}}(u, v)$ for some $u, v \in S$ and that for all $u', v' \in S$ and all $\mu \in \Omega(\mathbb{S}^{2^k}(u'), \mathbb{S}^{2^k}(v'))$ it holds that $\int \gamma^{\mathcal{C}} d\mu \geq \int \gamma^{\mathcal{C}} d\tau_{\mathcal{C}}(u', v')$. By hypothesis and Lemma ??, we have that $\int \gamma^{\mathcal{C}} d\tau_{\mathcal{C}}(u, v) = \min \{ \int \gamma^{\mathcal{D}} d\tau_{\mathcal{D}}(u, v) \mid \mathcal{D} \in \mathbb{C}_{2^k} \}$. But at the same time

$$\begin{aligned}
\delta_{\downarrow k}(u, v) &= \min \{ \mathbb{P}_{\mathcal{D}}(u, v)(\neq_{\ell^{\omega}}) \mid \mathcal{D} \in \mathbb{C}_{2^k} \} && \text{(by (??))} \\
&= \min \{ \int \gamma^{\mathcal{D}} d\tau_{\mathcal{D}}(u, v) \mid \mathcal{D} \in \mathbb{C}_{2^k} \}. && \text{(by Lemma ??)}
\end{aligned}$$

This contradicts hypothesis that $\int \gamma^{\mathcal{C}} d\tau_{\mathcal{C}}(u, v) \neq \delta_{\downarrow k}(u, v)$. \square

C Folklore Results about Metric Spaces

Proposition 50. *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 $\alpha = \sup A$. \square

Proposition 51. *Let $f: X \rightarrow Y$ be continuous and $A \subseteq X$, then $\overline{f(A)} = \overline{f(\overline{A})}$.*

Proof. (\supseteq) A function $f: X \rightarrow Y$ 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 $\overline{(\cdot)}$. \square

Proposition 52. *Let X be nonempty, $f: X \rightarrow \mathbb{R}$ 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 50, 51, and $\overline{D} = X$, we have

$$\sup f(D) \stackrel{(\text{Prop.50})}{=} \sup \overline{f(D)} \stackrel{(\text{Prop.51})}{=} \sup \overline{f(\overline{D})} = \sup \overline{f(X)} \stackrel{(\text{Prop.50})}{=} \sup f(X),$$

which proves the thesis. \square

Proposition 53.

- (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:

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

(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

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

which holds since all the pseudometrics d_i are positive. \square