On the Total Variation Distance of Semi-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. Semi-Markov chains (SMCs) are continuous-time probabilistic transition systems where the residence time on states is governed by generic distributions on the positive real line.
This paper shows the *tight* relation between the total variation distance on SMCs and their model checking problem over linear real-time specifications. Specifically, we prove that the total variation between two SMCs *coincides* with the maximal difference w.r.t. the likelihood of satisfying arbitrary MTL formulas or ω-languages recognized by timed automata. Computing this distance (i.e., solving its threshold problem) is NP-hard and its decidability is an open problem. Nevertheless, we propose an algorithm for approximating it with arbitrary precision.

1 Introduction

The growing interest in quantitative aspects in real world applications motivated the introduction of quantitative models and formal methods for studying their behaviors. Classically, the behavior of two models is compared by means of an equivalence (e.g., bisimilarity, trace equivalence, logical equivalence, etc.). However, when the models depend on numerical values that are subject to error estimates or obtained from statistical samplings, any notion of equivalence is too strong a concept. This motivated the study of *behavioral distances*. The idea is to generalize the concept of equivalence with that of *pseudometric*, aiming at measuring the behavioral dissimilarities between nonequivalent models.

Given a suitably large set of properties Φ, containing all the properties of interest, the behavioral dissimilarities of two states s, s' of a quantitative model are naturally measured by the pseudometric $d(s, s') = \sup_{\phi \in \Phi} |\phi(s) - \phi(s')|$, where $\phi(s)$ denotes the value of ϕ at s. This has been the leading idea for several proposals of behavioral distances, the first one given by Desharnais et al. [12] on probabilistic systems, and further developed by De Alfaro, van Breugel, Worrell, and others [10,11,19,16].

For probabilistic models $\phi(s)$ may represent the probability of satisfaction of a modal formula ϕ measured at s, hence relating the distance d to the *probabilistic*

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

model checking problem. In this context an immediate application is that the probability $\phi(s)$ of satisfying the formula ϕ at s can be approximated by $\phi(s')$ with an error bounded by $d(s, s')$, for any $\phi \in \Phi$. This may lead to savings in the overall cost of model checking.

In this paper we study the total variation distance of probabilistic systems, a popular distance used in a number of domains such as networks security and artificial intelligence, that measures the maximal difference in the probabilities of two systems of realizing the same event. We show that it is a genuine behavioral distance in the above sense by relating it to the probabilistic model checking problem over linear real-time specifications. Specifically, we prove that the total variation distance on semi-Markov chains coincides with the maximal difference in the probability of satisfying the same property, expressed either as an MTL formula [2,3] or an ω-language accepted by a timed automaton (TA) [1].

Semi-Markov chains (SMCs) are continuous-time probabilistic transition systems where the residence time on states is governed by generic distributions on the positive real line. SMCs subsume many probabilistic models, e.g., Markov chains (MCs) and continuous-time Markov Chains (CTMCs). Our attention on linear real-time properties is motivated by applications where the system to be modeled cannot be internally accessed but only tested via observations performed over a set of random executions. For instance, this is mostly common in domains such as systems biology, modeling/testing and machine learning, where real-time features are important e.g. for performance evaluation of cyber-physical systems or dependability analysis.

The total variation distance was already known to be a bound for the maximal difference w.r.t. the probability of satisfying linear-time formulas; our result guarantees that it is the tightest one. Since SMCs and MTL subsume MCs and LTL, respectively, the result holds also in the discrete-time case.

This further motivates the study of efficient methods for computing the total variation. Unfortunately, in [15,9] the threshold problem for the total variation distance is proven to be NP-hard in the case of MCs, and to the best of our knowledge, its decidability is still an open problem. Nevertheless, we prove that the problem of approximating the total variation distance with arbitrary precision is computable. This is done providing two effective sequences that converge from below and above to the total variation distance. This result generalizes that of [9] to the real-time setting. Our approach, however, is different, as it is based on a duality that characterizes the total variation between two measures as the minimal discrepancy associated with their couplings.

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

1. We solved the open problem of how tight is the upper-bound given by the total variation distance w.r.t. the variational distance ranging over MTL formulas and TA specifications, respectively. This has been made possible due to a more general result (Theorem 6) that entails many other nontrivial characterizations of the total variation distance on SMCs.

2. We provided sufficient conditions to construct sequences that converge, from below and above, to the total variation distance. Differently from [9], these

conditions are not specific to the probabilistic transition system at hand, but the results hold for probability measures on an arbitrary measurable space.

3. Lastly, we proved the computability of the converging sequences of the previous point. This yields a decidable procedure to approximate the total variation distance with arbitrary precision.

An extended version of the paper containing all the proofs is available at [5].

2 Preliminaries

The set of functions from X to Y is denoted by Y^X and for $f \in Y^X$, let $\equiv_f = \{(x, x') \mid f(x) = f(x')\}$. Given an equivalence relation $R \subseteq X \times X$, $X/_R$ denotes the set of R-equivalence classes and $[x]_R$ the equivalence class of $x \in X$.

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*. The σ-algebra generated by $\Sigma \subseteq 2^X$, denoted by $\sigma(\Sigma)$, is the smallest σ-algebra containing Σ. Hereafter $(\mathbb{R}_+, \mathbb{B})$ denotes the measurable space of positive real numbers (including zero) with Borel algebra.

Given two measurable spaces (X, Σ) and (Y, Θ), a function $f \colon X \to Y$ is *measurable* if for all $E \in \Theta$, $f^{-1}(E) = \{x \mid f(x) \in E\} \in \Sigma$. The *product space*, $(X, \Sigma) \otimes (Y, \Theta)$, is the measurable space $(X \times Y, \Sigma \otimes \Theta)$, where $\Sigma \otimes \Theta$ is the σ-algebra generated by the *rectangles* $E \times F$ for $E \in \Sigma$ and $F \in \Theta$.

A *measure* on (X, Σ) is a function $\mu \colon \Sigma \to \mathbb{R}_+$ s.t. $\mu(\bigcup_{E \in \mathcal{F}} E) = \sum_{E \in \mathcal{F}} \mu(E)$ for all countable families $\mathcal{F}$ of pairwise disjoint measurable sets (σ-*additive*); it is a *probability measure* if, in addition, $\mu(X) = 1$. In what follows $\Delta(X, \Sigma)$ denotes the set of probability measures on (X, Σ) and let $\mathcal{D}(X) = \Delta(X, 2^X)$.

Given a measurable function $f \colon (X, \Sigma) \to (Y, \Theta)$, 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*.

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

A measure ω on $(X, \Sigma) \otimes (Y, \Theta)$ is a *coupling* for (μ, ν) if for all $E \in \Sigma$ and $F \in \Theta$, $\omega(E \times Y) = \mu(E)$ and $\omega(X \times F) = \nu(F)$ (μ is the *left* and ν the *right marginals* of ω). We denote by $\Omega(\mu, \nu)$ the set of couplings for (μ, ν).

Metric spaces. Given a set X, $d \colon X \times X \to \mathbb{R}_+$ is a *pseudometric* on X if for arbitrary $x, y, z \in X$, $d(x, x) = 0$, $d(x, y) = d(y, x)$ 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*.

Given a measurable space (X, Σ), the set of measures $\Delta(X, \Sigma)$ is metrized by the *total variation distance*, defined by $\|\mu - \nu\| = \sup_{E \in \Sigma} |\mu(E) - \nu(E)|$.

The space of timed paths. A *timed path* over a set X is an infinite sequence $\pi = x_0, t_0, x_1, t_1 \ldots$, where $x_i \in X$ and $t_i \in \mathbb{R}_+$; t_i are called *time delays*. For any $i \in \mathbb{N}$, let $\pi[i] = x_i$, $\pi\langle i \rangle = t_i$, $\pi|^i = x_0, t_0, .., t_{i-1}, x_i$, and $\pi|_i = x_i, t_i, x_{i+1}, t_{i+1}, \ldots$. Let $\Pi(X)$ denote the set of timed paths on X.

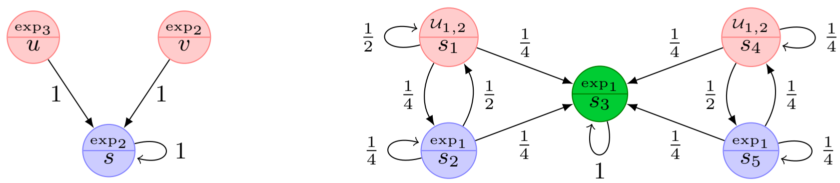

Fig. 1. Two SMCs. (left) the differences are only in the residence time distributions; (right) the behavioral differences arise only from their transition distributions.

The *cylinder set* (of rank n) for $X_i \subseteq X$ and $R_i \subseteq \mathbb{R}_+$, $i = 0..n$ is the set $\mathfrak{C}(X_0, R_0, .., R_{n-1}, X_n) = \{\pi \in \Pi(X) \mid \pi|^n \in X_0 \times R_0 \times \cdots \times R_{n-1} \times X_n\}$. For $\mathcal{F} \subseteq 2^X$ and $\mathcal{I} \subseteq 2^{\mathbb{R}_+}$, let $\mathfrak{C}_n(\mathcal{F}, \mathcal{I}) = \{\mathfrak{C}(X_0, R_0, .., R_{n-1}, X_n) \mid X_i \in \mathcal{F}, R_i \in \mathcal{I}\}$, for $n \in \mathbb{N}$, and $\mathfrak{C}(\mathcal{F}, \mathcal{I}) = \bigcup_{n \in N} \mathfrak{C}_n(\mathcal{F}, \mathcal{I})$.

If (X, Σ) is a measurable space, $\Pi(X, \Sigma)$ denotes the measurable space of timed paths with σ-algebra generated by $\mathfrak{C}(\Sigma, \mathbb{B})$. If $\Sigma = \sigma(\mathcal{F})$ and $\mathbb{B} = \sigma(\mathcal{I})$, then $\sigma(\mathfrak{C}(\Sigma, \mathbb{B})) = \sigma(\mathfrak{C}(\mathcal{F}, \mathcal{I}))$. Moreover, if both $\mathcal{F}$ and $\mathcal{I}$ are fields, so is $\mathfrak{C}(\mathcal{F}, \mathcal{I})$.

Any function $f \colon X \to Y$ can be stepwise extended to $f^\omega \colon \Pi(X) \to \Pi(Y)$. Note that if f is measurable, so is f^ω.

3 Semi-Markov Chains and Trace Distance

In this section we recall labelled *semi-Markov chains* (SMCs), models that subsume most of the space-finite Markovian models including Markov chains (MCs) and continuous-time Markov chains (CTMCs). We define the total variation distance between SMCs, called *trace distance*, which measures the difference between two SMCs w.r.t. their probabilities of generating labelled timed traces.

In what follows we fix a countable set $\mathbb{A}$ of atomic properties.

Definition 1 (Semi-Markov Chains). *A labelled* semi-Markov chain *is a tuple* $\mathcal{M} = (S, \tau, \rho, \ell)$ *consisting of a finite set* S *of* states*, a* transition probability function $\tau \colon S \to \mathcal{D}(S)$*, a* residence-time probability function $\rho \colon S \to \Delta(\mathbb{R}_+)$*, and a* labelling function $\ell \colon S \to 2^{\mathbb{A}}$.

In what follows we use $\mathcal{M} = (S, \tau, \rho, \ell)$ to range over the class of SMCs.

Intuitively, if $\mathcal{M}$ is in the state s, it moves to an arbitrary $s' \in S$ within time $t \in \mathbb{R}_+$ with probability $\rho(s)([0, t]) \cdot \tau(s)(s')$. For example, in Fig. 1(right) the SMC moves from s_1 to s_2 before time $t > 0$ with probability $\frac{1}{4} \cdot U[1, 2]([0, t))$, where $U[i, j]$ is the uniform distribution on $[i, j]$. An atomic proposition $p \in \mathbb{A}$ is said to hold in s if $p \in l(s)$.

Notice that MCs are the SMCs s.t. for all $s \in S$, $\rho(s)$ is the Dirac measure at 0 (transitions happen instantaneously); while CTMCs are the SMCs s.t. for all $s \in S$, $\rho(s) = \mathit{Exp}(\lambda)$ —the exponential distribution with rate $\lambda > 0$.

An SMC in an initial state is a stochastic process generating timed paths. They are distributed as in the next definition.

Definition 2. *Given $s \in S$ state in $\mathcal{M}$, let $\mathbb{P}_s$ be the unique probability measure*[1] *on $\Pi(S)$ such that for all $s_i \in S$ and $R_i \in \mathbb{B}$, $i = 0..n$,*

$$\mathbb{P}_s(\mathfrak{C}(\{s_0\}, R_0, \dots, R_{n-1}, \{s_n\})) = \mathbb{1}_{\{s\}}(s_0) \cdot \prod_{i=0}^{n-1} P(s_i, R_i, s_{i+1}) \,,$$

where $\mathbb{1}_A$ is the indicator function of A and $P(u, R, v) = \rho(u)(R) \cdot \tau(u)(v)$.

Since the only things that we observe in a state are the atomic properties (labels), timed paths are considered up to label equivalence. This leads to the definition of *trace cylinders*, which are elements in $\mathfrak{C}(S/_{\equiv_\ell}, \mathbb{B})$, and to the following equivalence between states.

Definition 3 (Trace Equivalence). *For arbitrary $\mathcal{M} = (S, \tau, \rho, \ell)$, $s, s' \in S$ are* trace equivalent*, written $s \approx s'$, if for all $T \in \mathfrak{C}(S/_{\equiv_\ell}, \mathbb{B})$, $\mathbb{P}_s(T) = \mathbb{P}_{s'}(T)$.*

Hereafter, we use $\mathcal{T}$ to denote the set $\mathfrak{C}(S/_{\equiv_\ell}, \mathbb{B})$ of trace cylinders.

If two states of an SMCs are *not* trace equivalent, then their difference is usually measured by the total variation distance between their corresponding probabilities restricted to events generated by labelled traces.

Definition 4 (Trace Pseudometric). *Given $\mathcal{M} = (S, \tau, \rho, \ell)$, the* trace pseudometric $\delta \colon S \times S \to [0, 1]$ *is defined, for arbitrary $s, s' \in S$, by*

$$\delta(s, s') = \sup_{E \in \sigma(\mathcal{T})} |\mathbb{P}_s(E) - \mathbb{P}_{s'}(E)| \,.$$

It is not difficult to observe that two states $s, s' \in S$ are trace equivalent if and only if $\delta(s, s') = 0$. This demonstrates that the trace equivalence is a *behavioural distance.*

4 Trace Distance and Probabilistic Model Checking

In this section we investigate the connections between the trace distance and model checking SMCs over linear real-time specifications. We show that the variational distance over measurable sets expressed either as Metric Temporal Logic (MTL) formulas or as languages accepted by Timed Automata (TAs) coincides with the trace distance introduced in the previous section. Both these results are instances of a more general result (Theorem 6), which also entails other similar nontrivial characterizations of the trace distance.

A measure μ on (X, Σ) induces the so-called *Fréchet-Nikodym pseudometric* on Σ, $d_\mu \colon \Sigma \times \Sigma \to \mathbb{R}_+$ defined for arbitrary $E, F \in \Sigma$, by $d_\mu(E, F) = \mu(E \triangle F)$, where $E \triangle F := (E \setminus F) \cup (F \setminus E)$ is the symmetric difference between sets.

Recall that in a (pseudo)metric space a subset D is dense if its closure $\overline{D}$ (i.e., the set of all the points arbitrarily close to D) coincides with the entire space. In order to prove the aforementioned general result, we need firstly to provide a sufficient condition for a family of measurable sets to be dense w.r.t. the Fréchet-Nikodym pseudometric for some finite measure.

[1] Existence and uniqueness of $\mathbb{P}_s$ is guaranteed by the Hahn-Kolmogorov extension theorem and by the fact that, for all $s \in S$, $\tau(s)$ and $\rho(s)$ are finite measures.

Lemma 5. *Let (X, Σ) be a measurable space and μ be a measure on it. If Σ is generated by a field $\mathcal{F}$, then $\mathcal{F}$ is dense in the pseudometric space (Σ, d_μ).*

Proof (sketch). We show that $\overline{\mathcal{F}} := \{E \in \Sigma \mid \forall \varepsilon > 0.\ \exists F \in \mathcal{F}.\ d_\mu(E, F) < \varepsilon\} = \Sigma$. To prove $\Sigma \subseteq \overline{\mathcal{F}}$, it is sufficient to show that $\overline{\mathcal{F}}$ is a σ-algebra. The closure under complement follows from $E \triangle F = (X \setminus E) \triangle (X \setminus F)$. The closure under countable union follows from monotonicity, additivity and ω-continuity from below of μ given that for any $\{E_i \mid i \in \mathbb{N}\} \subseteq \overline{\mathcal{F}}$ and $\varepsilon > 0$ the following hold:

a) there exists $k \in \mathbb{N}$, such that $d_\mu(\bigcup_{i \in \mathbb{N}} E_i, \bigcup_{i=0}^k E_i) < \frac{\varepsilon}{2}$;
b) for all $n \in \mathbb{N}$, there exist $F_0, \dots, F_n \in \mathcal{F}$, such that $d_\mu(\bigcup_{i=0}^n E_i, \bigcup_{i=0}^n F_i) < \frac{\varepsilon}{2}$.

Indeed, by triangular inequality, for arbitrary $F_0, \dots, F_k \in \mathcal{F}$,

$$d_\mu(\textstyle\bigcup_{i \in \mathbb{N}} E_i, \bigcup_{i=0}^k F_i) \leq d_\mu(\bigcup_{i \in \mathbb{N}} E_i, \bigcup_{i=0}^k E_i) + d_\mu(\bigcup_{i=0}^k E_i, \bigcup_{i=0}^k F_i) < \varepsilon \,.$$

Then, the lemma follows since $\mathcal{F}$ is a field. □

With this result in hands we can state the main theorem of this section.

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

Proof. For $Y \neq \emptyset$ and $f \colon Y \to \mathbb{R}$ bounded and continuous, if $D \subseteq Y$ is dense then $\sup f(D) = \sup f(Y)$. By Lemma 5, $\mathcal{F}$ is dense in $(\Sigma, d_{\mu+\nu})$. We show that $|\mu - \nu| \colon \Sigma \to \mathbb{R}$ is bounded and continuous. Boundedness follows since μ and ν are finite. By monotonicity, positivity, and additivity of the measures one can show that μ and ν are 1-Lipschitz continuous, so $|\mu - \nu|$ is continuous. □

4.1 Model Checking for MTL Formulas

Metric Temporal Logic [2] has been introduced as a formalism for reasoning on sequences of events in a real-time setting. The grammar of formulas is as follows

$$\varphi ::= p \mid \bot \mid \varphi \to \varphi \mid \mathsf{X}^{[t,t']}\varphi \mid \varphi \, \mathsf{U}^{[t,t']} \, \varphi \,,$$

where $p \in \mathbb{A}$ and $[t, t']$ are positive-reals intervals with rational boundaries.

The formal semantics[2] of MTL is given by means of a satisfiability relation defined, for an arbitrary SMC $\mathcal{M}$ and a timed path $\pi \in \Pi(S)$, as follows [17].

$$
\begin{array}{ll}
\mathcal{M}, \pi \models p & \text{if } p \in \ell(\pi[0]) \,, \\
\mathcal{M}, \pi \models \bot & \text{never} \,, \\
\mathcal{M}, \pi \models \varphi \to \psi & \text{if } \mathcal{M}, \pi \models \psi \text{ whenever } \mathcal{M}, \pi \models \varphi \,, \\
\mathcal{M}, \pi \models \mathsf{X}^{[t,t']}\varphi & \text{if } \pi\langle 0 \rangle \in [t, t'] \text{, and } \mathcal{M}, \pi|_1 \models \varphi \,, \\
\mathcal{M}, \pi \models \varphi \, \mathsf{U}^{[t,t']} \, \psi & \text{if } \exists i > 0 \text{ such that } \sum_{k=0}^{i-1} \pi\langle k \rangle \in [t, t'],\ \mathcal{M}, \pi|_i \models \psi, \\
& \quad \text{and } \mathcal{M}, \pi|_j \models \varphi \text{ whenever } 0 \leq j < i \,.
\end{array}
$$

[2] This is known as the *point-based* semantics, since the connectives quantify over a countable set of positions in the path; it differs from the *interval-based* semantics, adopted in [7,18], which associates a state with each point in the real line, and let the temporal connectives quantify over intervals with uncountable many points.

Having fixed an SMC $\mathcal{M}$, let $[\![\varphi]\!] = \{\pi \mid \mathcal{M}, \pi \models \varphi\}$ and $[\![\mathcal{L}]\!] = \{[\![\varphi]\!] \mid \varphi \in \mathcal{L}\}$, for any $\mathcal{L} \subseteq$ MTL. Let MTL$^-$ be the fragment of MTL without until operator.

Lemma 7. *(i)* $[\![\text{MTL}]\!] \subseteq \sigma(\mathcal{T})$ *and (ii)* $\mathcal{T} \subseteq \sigma([\![\text{MTL}^-]\!])$.

Lemma 7 states that (i) MTL formulas describe events in the σ-algebra generated by the trace cylinders; and (ii) the trace cylinders are measurable sets generated by MTL formulas without until operator. Consequently, the probabilistic model checking problem for SMC, which is to determine the probability $\mathbb{P}_s([\![\varphi]\!])$ given the initial state s of $\mathcal{M}$, is well defined. Moreover, for any $\mathcal{L} \subseteq$ MTL,

$$\delta_{\mathcal{L}}(s, s') = \sup_{\varphi \in \mathcal{L}} |\mathbb{P}_s([\![\varphi]\!]) - \mathbb{P}_{s'}([\![\varphi]\!])|$$

is a well-defined pseudometric that distinguishes states w.r.t. their maximal difference in the likelihood of satisfying formulas in $\mathcal{L}$.

Obviously, the trace distance δ is an upper bound of $\delta_{\mathcal{L}}$; however, Theorem 6 reveals a set of conditions on $\mathcal{L}$ guaranteeing that the two actually coincide.

Corollary 8 (Logical Characterization). *Let $\mathcal{L}$ be a Boolean-closed fragment of* MTL *s.t.* $\mathcal{T} \subseteq \sigma([\![\mathcal{L}]\!])$. *Then,* $\delta_{\mathcal{L}} = \delta$. *In particular,* $\delta_{\text{MTL}} = \delta_{\text{MTL}^-} = \delta$.

Remark 9. The supremum in the definition of δ_{MTL} is not a maximum. Fig.1 shows two examples. The SMC on the right is taken from [9, Example 1][3], where it is proven that $\delta(s_1, s_4)$ has a maximizing event that is not an ω-regular language, hence not describable by an LTL formula. As for the SMC on the left, the maximizing event corresponding to $\delta(u, v)$ should have the form $\mathsf{X}^I \top$ for $I = [0, \log(3) - \log(2)]$. However the previous is not an MTL formula since I has an irrational endpoint. ∎

4.2 Model Checking for Timed Automata

Timed Automata (TAs) [1] have been introduced to model the behavior of real-time systems over time. Here we consider TAs without location invariants.

Let $\mathcal{X}$ be a finite set of variables (*clocks*) and $\mathcal{V}(\mathcal{X})$ the set of *valuations* $v \colon \mathcal{X} \to \mathbb{R}_+$. As usual, for $v \in \mathcal{V}(\mathcal{X})$, $t \in \mathbb{R}_+$ and $X \subseteq \mathcal{X}$, we denote by $\mathbf{0}$ the null valuation, by $v + t$ the t-delay of v and by $v[X := t]$ the update of X in v.

A *clock guard* $g \in \mathcal{G}(\mathcal{X})$ over $\mathcal{X}$ is a finite set of expressions of the form $x \bowtie q$, for $x \in \mathcal{X}$, $q \in \mathbb{Q}_+$ and $\bowtie \in \{<, \leq, >, \geq\}$. We say that a valuation $v \in \mathcal{V}(\mathcal{X})$ *satisfies* a clock guard $g \in \mathcal{G}(\mathcal{X})$, written $v \models g$, if $v(x) \bowtie n$ holds, for all $x \bowtie q \in g$. Two clock guards $g, g' \in \mathcal{G}(\mathcal{X})$ are *orthogonal* (or *non-overlapping*), written $g \perp g'$, if there is no $v \in \mathcal{V}(\mathcal{X})$ such that $v \models g$ and $v \models g'$.

Definition 10 (Timed Automaton). *A* timed (Muller) automaton *over a set of clocks $\mathcal{X}$ is a tuple $\mathcal{A} = (Q, L, q_0, F, \rightarrow)$ consisting of a finite set Q of* locations*, a set L of* input symbols*, an* initial location *$q_0 \in Q$, a family $F \subseteq 2^Q$ of* final sets of locations*, and a* transition relation *$\rightarrow \subseteq Q \times L \times \mathcal{G}(\mathcal{X}) \times 2^{\mathcal{X}} \times Q$.*

$\mathcal{A}$ is deterministic *if $(q, a, g, X, q'), (q, a, g', X', q'') \in \rightarrow$ and $g \neq g'$ implies $g \perp g'$; it is* resetting *if $(q, a, g, X, q') \in \rightarrow$ implies $X = \mathcal{X}$.*

[3] The SMC has been adapted to the current setting where the labels are in the state, instead of in the transitions.

A *run* of $\mathcal{A} = (Q, L, q_0, F, \rightarrow)$ over a timed path $\pi = a_0, t_0, a_1, t_1, \dots$ is an infinite sequence

$$(q_0, v_0) \xrightarrow{a_0, t_0} (q_1, v_1) \xrightarrow{a_1, t_1} (q_2, v_2) \xrightarrow{a_2, t_2} \cdots$$

with $q_i \in Q$ and $v_i \in \mathcal{V}(\mathcal{X})$ for all $i \geq 0$, satisfying the following requirements: (*initialization*) $v_0 = \mathbf{0}$; (*consecution*) for all $i \geq 0$, exists $(q_i, a_i, g_i, X_i, q_{i+1}) \in \rightarrow$ such that $v_{i+1} = (v_i + t_i)[X_i := 0]$ and $v_i + t_i \models g_i$.

A run over π is *accepting* (π is *accepted* by $\mathcal{A}$) if the set of locations visited infinitely often is in F. Let $\mathcal{L}(\mathcal{A})$ be the set of timed paths accepted by $\mathcal{A}$.

A deterministic TA (DTA) has at most one accepting run over a given timed path in $\Pi(L)$. With respect to TAs, which are only closed under finite union and intersection, DTAs are also closed under complement [1].

To relate TAs and SMCs, consider $\mathcal{M} = (S, \tau, \rho, \ell)$ and a TA $\mathcal{A}$ that uses the labels of $\mathcal{M}$ as input symbols. Let $[\![\mathcal{A}]\!] = \{\pi \mid \ell^\omega(\pi) \in \mathcal{L}(\mathcal{A})\}$ be the set of timed paths in $\mathcal{M}$ accepted by $\mathcal{A}$ and $[\![\mathcal{F}]\!] = \{[\![\mathcal{A}]\!] \mid \mathcal{A} \in \mathcal{F}\}$ for any set $\mathcal{F} \in \mathrm{TA}$.

Lemma 11. *(i)* $[\![TA]\!] \subseteq \sigma(\mathcal{T})$ *and (ii)* $\mathcal{T} \subseteq \sigma([\![\mathrm{DTA}]\!])$.

Lemma 11 states that the model checking problem of an SMC $\mathcal{M}$ against a TA $\mathcal{A}$, which is to determine the probability $\mathbb{P}_s([\![\mathcal{A}]\!])$ given the initial state s of $\mathcal{M}$, is well defined and for any $\Phi \subseteq \mathrm{TA}$ we can define the pseudometric

$$\delta_\Phi(s, s') = \sup_{\mathcal{A} \in \Phi} |\mathbb{P}_s([\![\mathcal{A}]\!]) - \mathbb{P}_{s'}([\![\mathcal{A}]\!])|$$

that distinguishes states looking at a specific subclass Φ of TA specifications. For a generic $\Phi \subseteq \mathrm{TA}$, the trace distance is an upper bound of δ_Φ. However, Theorem 6 provides conditions that guarantee the equality of the two distances.

Corollary 12. *Let* $\Phi \subseteq \mathrm{TA}$ *be closed under Boolean operations and such that* $\mathcal{T} \subseteq \sigma([\![\Phi]\!])$. *Then,* $\delta_\Phi = \delta$. *In particular,* $\delta_{\mathrm{TA}} = \delta_{\mathrm{DTA}} = \delta$.

Single-clock Resetting DTAs. The decidability of model checking CTMCs against TA specifications is open, even for the subclass of DTAs. Recently, Chen et al. [8] provided a decidable algorithm for the case of *single-clock* DTAs (1-DTAs). In this context, an alternative characterization of the trace distance in terms of 1-DTAs is appealing. Notice however that Corollary 12 cannot be applied, since 1-DTAs are not closed under union. We show that the *resetting* 1-DTAs (1-RDTA) satisfy the requirements, hence $\delta_{\text{1-DTA}} = \delta_{\text{1-RDTA}} = \delta$.

Lemma 13. *(i)* $[\![\text{1-RDTA}]\!]$ *is a field and (ii)* $\mathcal{T} \subseteq \sigma([\![\text{1-RDTA}]\!])$.

5 General Convergence Criteria

In this section we provide sufficient conditions to construct sequences that converge, from below and from above, to the total variation distance between a generic pair of probability measures. Eventually, we instantiate these results to the specific case of the trace distance on SMCs.

Convergence from Below. To define a converging sequence of under-approximations of the total variation distance we exploit Theorem 6 as follows.

Theorem 14. *Let (X, Σ) be a measurable space and μ, ν be probability measures on it. Let $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots$ be a sequence s.t. $\mathcal{F} = \bigcup_{i \in \mathbb{N}} \mathcal{F}_i$ is a field that generates Σ and*

$$l_i = \sup \{ |\mu(E) - \nu(E)| \mid E \in \mathcal{F}_i \} \ .$$

Then, $l_i \leq l_{i+1}$ and $\sup_{i \in \mathbb{N}} l_i = \|\mu - \nu\|$, for all $i \in \mathbb{N}$.

Proof. $l_i \leq l_{i+1}$ follows from $\mathcal{F}_i \subseteq \mathcal{F}_{i+1}$. Because $\mathcal{F}$ is a field s.t. $\sigma(\mathcal{F}) = \Sigma$, μ and ν are finite measures and $\sup_{i \in \mathbb{N}} l_i = \sup_{E \in \mathcal{F}} |\mu(E) - \nu(E)|$, Theorem 6 concludes our proof. □

According to Theorem 14, to approximate the trace distance δ from below, we just need to find an increasing sequence of collections of measurable sets of timed paths whose union is a field generating $\sigma(\mathcal{T})$. We define it as follows.

For $k \in \mathbb{N}$, let $\mathcal{E}_k$ be the set of all finite unions of cylinders in $\mathfrak{C}_k(S/_{\equiv_\ell}, \mathfrak{R}_k)$, where $\mathfrak{R}_k = \left\{ \left[\frac{n}{2^k}, \frac{n+1}{2^k} \right) \mid 0 \leq n < k2^k \right\} \cup \{[k, \infty)\}$. Note that, these cylinders are pairwise disjoint and, in particular, they form a $\sigma(\mathcal{T})$-measurable partition of $\Pi(S)$. The choice is justified by the following result.

Lemma 15. *For all $k \in \mathbb{N}$, $\mathcal{E}_k \subseteq \mathcal{E}_{k+1}$ and $\bigcup_{k \in \mathbb{N}} \mathcal{E}_k$ is a field generating $\sigma(\mathcal{T})$.*

Given an SMC $\mathcal{M}$, a sequence of under-approximations of the trace distance δ is given, for $k \in \mathbb{N}$, by $\delta\!\uparrow_k \colon S \times S \to [0,1]$ defined by

$$\delta\!\uparrow_k(s, s') = \sup \{ |\mathbb{P}_s(E) - \mathbb{P}_{s'}(E)| \mid E \in \mathcal{E}_k \} \ . \tag{1}$$

The next result is an immediate consequence of Lemma 15 and Theorem 14.

Corollary 16. *For all $k \in \mathbb{N}$, $\delta\!\uparrow_k \leq \delta\!\uparrow_{k+1}$ and $\delta = \sup_{k \in \mathbb{N}} \delta\!\uparrow_k$.*

Remark 17 (A logical convergence). Note that Theorem 14 suggests alternative constructions of convergent sequences. For example, as lower-approximations of δ one can use the pseudometrics $\delta_{\mathrm{MTL}_k^-}$, where MTL_k^- is the set of MTL^- formulas with *modal depth* at most $k \in \mathbb{N}$. ■

Convergence from Above. The construction of the converging sequence of over-approximations of the total variation is based on a classic duality result asserting that *the total variation of two measures corresponds to the minimal discrepancy measured among all their possible couplings* [14].

Recall that a coupling $\omega \in \Omega(\mu, \nu)$ for two probability measures μ, ν on (X, Σ) is a measure in the product space $(X, \Sigma) \otimes (X, \Sigma)$ whose left and right marginals are μ and ν, respectively. The *discrepancy* associated with ω is the value $\omega(\ncong)$, where $\cong = \bigcap_{E \in \Sigma} \{(x, y) \mid x \in E \text{ iff } y \in E\}$ is the *inseparability relation* w.r.t. measurable sets in Σ. Then, the following duality holds.

Lemma 18 ([14, Th.5.2]). *Let μ, ν be probability measures on (X, Σ). Then, provided that $\ncong$ is measurable in $\Sigma \otimes \Sigma$, $\|\mu - \nu\| = \min \{ \omega(\ncong) \mid \omega \in \Omega(\mu, \nu) \}$.*

Given the above result, we can state a second general converging criterion to approach the total variation distance from above.

Theorem 19. *Let (X, Σ) be a measurable space s.t. $\cong \in \Sigma \otimes \Sigma$ and μ, ν be probability measures on it. Let $\Omega_0 \subseteq \Omega_1 \subseteq \Omega_2 \dots$ be an increasing sequence s.t. $\bigcup_{i \in \mathbb{N}} \Omega_i$ is dense in $\Omega(\mu, \nu)$ w.r.t. the total variation distance and define*

$$u_i = \inf \{\omega(\ncong) \mid \omega \in \Omega_i\} \ .$$

Then, $u_i \geq u_{i+1}$ and $\inf_{i \in \mathbb{N}} u_i = \|\mu - \nu\|$, for all $i \in \mathbb{N}$.

Proof. $u_i \geq u_{i+1}$ follows from $\Omega_i \subseteq \Omega_{i+1}$. To prove $\inf_{i \in \mathbb{N}} u_i = \|\mu - \nu\|$, recall that for $Y \neq \emptyset$ and $f \colon Y \to \mathbb{R}$ bounded and continuous, if $D \subseteq Y$ is dense then $\inf f(D) = \inf f(Y)$. By hypothesis $\bigcup_{i \in \mathbb{N}} \Omega_i \subseteq \Omega(\mu, \nu)$ is dense; moreover, $\mu \times \nu \in \Omega(\mu, \nu) \neq \emptyset$. We show that $ev_{\ncong} \colon \Omega(\mu, \nu) \to \mathbb{R}$, defined by $ev_{\ncong}(\omega) = \omega(\ncong)$ is bounded and continuous. It is bounded since all $\omega \in \Omega(\mu, \nu)$ are finite measures. It is continuous because $\|\omega - \omega'\| \geq |\omega(\ncong) - \omega'(\ncong)| = |ev_{\ncong}(\omega) - ev_{\ncong}(\omega')|$ (1-Lipschitz continuity). Now, applying Lemma 18, we derive our result. □

To conclude this section, we define a sequence of sets of couplings that, according to Theorem 19, characterizes the trace distance δ on SMCs.

Observe that the inseparability relation w.r.t. the σ-algebra generated by trace cylinders is measurable and it can be characterized as follows.

Lemma 20. $\equiv_{\ell^\omega} = \bigcap_{E \in \sigma(\mathcal{T})} \{(\pi, \pi') \mid \pi \in E \textit{ iff } \pi' \in E\} \in \sigma(\mathcal{T}) \otimes \sigma(\mathcal{T})$.

Next we introduce the notion of *coupling structure* for an SMC. Let $\Pi^k(S) = \{s_0, t_0, .., t_{k-1}, s_k \mid s_i \in S,\, t_i \in \mathbb{R}_+\}$ be the measurable space with σ-algebra generated by $\mathcal{R}_k = \{\{s_0\} \times R_0 \times .. \times R_{k-1} \times \{s_k\} \mid s_i \in S,\, R_i \in \mathbb{B}\}$. Note that, the prefix function $(\cdot)|^k \colon \Pi(S) \to \Pi^k(S)$ is measurable, hence, the push forward w.r.t. it on $\mu \in \Delta(\Pi(S))$, denoted by $\mu|^k$, is a measure in $\Pi^k(S)$.

Definition 21 (Coupling Structure). *A* coupling structure of rank $k \in \mathbb{N}$ *for an SMC $\mathcal{M}$ is a function $\mathcal{C} \colon S \times S \to \Delta(\Pi^k(S) \times \Pi^k(S))$ such that, for all states $s, s' \in S$, $\mathcal{C}(s, s') \in \Omega(\mathbb{P}_s|^k, \mathbb{P}_{s'}|^k)$.*

The set of coupling structures of rank k for $\mathcal{M}$ is denoted by $\mathbb{C}_k(\mathcal{M})$.

A coupling structure of rank k together with a distinguished initial pair of states, can be intuitively seen as a stochastic process generating pairs of timed paths divided in multi-steps of length k and distributed according to the following probability.

Definition 22. *For $k \in \mathbb{N}$, $s, s' \in S$ states in $\mathcal{M}$ and $\mathcal{C} \in \mathbb{C}_k(\mathcal{M})$, let $\mathbb{P}^{\mathcal{C}}_{s,s'}$ be the unique probability measure*[4] *on $\Pi(S) \otimes \Pi(S)$ such that, for all $n \in \mathbb{N}$ and $E = \{u_0\} \times R_0 \times .. \times R_{nk-1} \times \{u_{nk}\}, F = \{v_0\} \times H_0 \times .. \times H_{nk-1} \times \{v_{nk}\} \in \mathcal{R}_{nk}$*

$$\mathbb{P}^{\mathcal{C}}_{s,s'}(\mathfrak{C}(E) \times \mathfrak{C}(F)) = \mathbb{1}_{\{(s,s')\}}(u_0, v_0) \cdot \textstyle\prod_{h=0}^{n-1} \mathcal{C}(u_{hk}, v_{hk})(E_h \times F_h) \,,$$

where $\mathfrak{C}(E)$ denotes the cylinder obtained as the pre-image under $(\cdot)|^{nk}$ of E and $E_h = \{u_{hk}\} \times R_{hk} \times .. \times R_{(h+1)k-1} \times \{u_{(h+1)k}\}$ (similarly for F).

[4] The existence and the uniqueness of this measure follow by Hahn-Kolmogorov extension theorem and the fact that any cylinder of rank k can always be represented as a disjoint union of cylinders of rank $k' \geq k$ (see e.g., [6, pp.29–32]).

The name "coupling structure" is justified by the following result.

Lemma 23. *Let $\mathcal{C}$ be a coupling structure for $\mathcal{M}$, then $\mathbb{P}^{\mathcal{C}}_{s,s'} \in \Omega(\mathbb{P}_s, \mathbb{P}_{s'})$.*

We are finally ready to describe a decreasing sequence that converges to the trace distance on SMCs. Given $\mathcal{M}$, let $\delta{\downarrow_k} \colon S \times S \to [0,1]$ for $k \in \mathbb{N}$, be

$$\delta{\downarrow_k}(s,s') = \min \left\{ \mathbb{P}^{\mathcal{C}}_{s,s'}(\not\equiv_{\ell^\omega}) \mid \mathcal{C} \in \mathbb{C}_{2^k}(\mathcal{M}) \right\} . \tag{2}$$

According to Theorem 14 the following suffices to prove the convergence.

Lemma 24. *Let $s, s' \in S$ be a pair of states of an SMC $\mathcal{M}$. Then,*
(i) for all $k \in \mathbb{N}$, $\left\{ \mathbb{P}^{\mathcal{C}}_{s,s'} \mid \mathcal{C} \in \mathbb{C}_k(\mathcal{M}) \right\} \subseteq \left\{ \mathbb{P}^{\mathcal{C}}_{s,s'} \mid \mathcal{C} \in \mathbb{C}_{2k}(\mathcal{M}) \right\}$;
(ii) $\bigcup_{k\in\mathbb{N}} \left\{ \mathbb{P}^{\mathcal{C}}_{s,s'} \mid \mathcal{C} \in \mathbb{C}_{2^k}(\mathcal{M}) \right\}$ is dense in $\Omega(\mathbb{P}_s, \mathbb{P}_{s'})$ w.r.t. the total variation.

Proof (sketch). (i) Let $k > 0$ and $\mathcal{C} \in \mathbb{C}_k(\mathcal{M})$. Define $\mathcal{D}(s,s')$ as the unique measure on $\Pi^{2k}(S) \otimes \Pi^{2k}(S)$ s.t., for all $E = \{u_0\} \times R_0 \times .. \times R_{2k-1} \times \{u_{2k}\}$ and $F = \{v_0\} \times H_0 \times .. \times H_{2k-1} \times \{v_{2k}\}$ 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 $E'{=}\{u_0\} \times R_0 \times .. \times R_{k-1} \times \{u_k\}$ and $E''{=}\{u_k\} \times R_k \times .. \times R_{2k-1} \times \{u_{2k}\}$ (similarly for F). One can check that $\mathcal{D} \in \mathbb{C}_{2k}(\mathcal{M})$ and $\mathbb{P}^{\mathcal{C}}_{s,s'} = \mathbb{P}^{\mathcal{D}}_{s,s'}$.

(ii) Let $\Omega = \bigcup_{k\in\mathbb{N}} \left\{ \mathbb{P}^{\mathcal{C}}_{s,s'} \mid \mathcal{C} \in \mathbb{C}_{2^k}(\mathcal{M}) \right\}$. Let $\mathcal{F}_k$ be the collection of all finite union of sets of the form $\mathfrak{C}(E) \times \mathfrak{C}(F)$, for $E, F \in \mathcal{R}_k$. Note that $\mathcal{F} = \bigcup_{k\in\mathbb{N}} \mathcal{F}_k$ is a field generating the σ-algebra of $\Pi(S) \otimes \Pi(S)$. By Lemma 5 and Definition 22, to prove that Ω is dense it suffices that for all $\mu \in \Omega(\mathbb{P}_s, \mathbb{P}_{s'})$, $k \in \mathbb{N}$ and $F \in \mathcal{F}_k$, there exists $\omega \in \Omega$ s.t. $\omega(F) = \mu(F)$. One can check that $\omega = \mathbb{P}^{\mathcal{C}}_{s,s'}$, where $\mathcal{C} \in \mathbb{C}_{2^k}(\mathcal{M})$ is s.t. $\mathcal{C}(s,s') = \mu[(\cdot)|^{2^k} \times (\cdot)|^{2^k}]$ (i.e., the push forward of μ along the function $(\pi, \pi') \mapsto (\pi|^{2^k}, \pi'|^{2^k}))$ has the desired property. □

The following corollary derives from Lemma 24 and Theorem 19.

Corollary 25. *For all $k \in \mathbb{N}$, $\delta{\downarrow_k} \geq \delta{\downarrow_{k+1}}$ and $\delta = \inf_{k\in\mathbb{N}} \delta{\downarrow_k}$.*

6 An Approximation Algorithm

This section exploits the aforementioned results to propose a decidable procedure for approximating the trace distance δ on SMCs with arbitrary precision.

Let $\varepsilon > 0$ and consider the sequences $\{\delta{\uparrow_k}\}_{k\in\mathbb{N}}$ and $\{\delta{\downarrow_k}\}_{k\in\mathbb{N}}$ from Section 5. The procedure proceeds step-wise (increasing $k \geq 0$) by computing the point-wise difference $\delta{\downarrow_k} - \delta{\uparrow_k}$ until is smaller then ε. Termination and correctness is ensured by the convergence of the sequences from above and below to δ.

Theorem 26. *Let $\mathcal{M}$ be a SMC. There exists an algorithm that, given a rational number $\varepsilon > 0$, computes a function $d \colon S \times S \to [0,1] \cap \mathbb{Q}_+$ such that $|d - \delta| < \varepsilon$.*

We prove this theorem under two reasonable assumptions regarding SMCs:

A1. For all $s \in S$ and $q, q' \in \mathbb{Q}_+$, $\rho(s)([q, q'))$ is computable;
A2. For all $s, s' \in S$, $\|\rho(s) - \rho(s')\|$ is computable.

In the above $\rho(s)([q, q'))$ and $\|\rho(s) - \rho(s')\|$ may assume real values, and with the term "compute" we mean that there exists an effective Cauchy sequence of rationals that converges to the value.

Lemma 27. *Assuming A1, $\delta{\uparrow}_k$ is computable for all $k \in \mathbb{N}$.*

Proof (sketch). For each $k \in \mathbb{N}$, the set $\mathcal{E}_k$ is finite. Moreover, for each $s \in S$ and $E \in \mathcal{E}_k$, $\mathbb{P}_s(E)$ is computable thanks to its additivity and the hypothesis A1. □

The computability of the sequence $\{\delta{\downarrow}_k\}_{k \in \mathbb{N}}$ is less trivial. Equation (2) suggests to look for a coupling structure $\mathcal{C} \in \mathbb{C}_{2^k}(\mathcal{M})$ that minimizes the discrepancy $\mathbb{P}^{\mathcal{C}}_{s,s'}(\not\equiv_{\ell^\omega})$. This is done by following a searching strategy similar to the one in [4] and structured as follows: (i) we provide an alternative characterization of the discrepancy associated with a coupling structure (Section 6.1); (ii) we describe how to construct an optimal coupling structure and show that its associated discrepancy is computable (Section 6.2).

6.1 Fixed Point Characterization of the Discrepancy

We characterize the discrepancy associated with a coupling structure $\mathcal{C}$ by means of the least fixed point of a suitable operator parametric in $\mathcal{C}$. To define the fixed point operator it is convenient to split a coupling structure into two "projections": on discrete state transitions (regardless of time delays); and on residence times (given that a sequence of transitions has occurred). To this end define $\mathbb{S}^k \colon S \to \mathcal{D}(S^{k+1})$ and $\mathbb{T}^k \colon S^k \to \Delta(\mathbb{R}^k_+)$ as follows

$$\mathbb{S}^k(s)(u_0..u_k) = \mathbb{1}_s(u_0) \cdot \textstyle\prod_{i=0}^{k-1} \tau(u_i)(u_{i+1}), \quad \mathbb{T}^k(v_1..v_k) = \rho(v_1) \times \cdots \times \rho(v_k).$$

Lemma 28. *The set $\mathbb{C}_k(\mathcal{M})$ is in bijection with the set of pairs of functions $\tau_{\mathcal{C}} \colon S \times S \to \mathcal{D}(S^{k+1} \times S^{k+1})$ and $\rho_{\mathcal{C}} \colon S^k \times S^k \to \Delta(\mathbb{R}^k_+ \times \mathbb{R}^k_+)$ such that*

$$\tau_{\mathcal{C}}(u,v) \in \Omega(\mathbb{S}^k(u), \mathbb{S}^k(v)) \quad \textit{and} \quad \rho_{\mathcal{C}}(u_1..u_k, v_1..v_k) \in \Omega(\mathbb{T}^k(u_1..u_k), \mathbb{T}^k(v_1..v_k)).$$

Hereafter we identify the coupling structure $\mathcal{C}$ with its bijective image $(\tau_{\mathcal{C}}, \rho_{\mathcal{C}})$.

Intuitively, $\tau_{\mathcal{C}}(u,v)(u_0..u_k, v_0..v_k)$ is the probability that two copies of $\mathcal{M}$, scheduled according to $\mathcal{C}$, have respectively generated the sequences of states $u_0..u_k$ and $v_0..v_k$ starting from u and v; while $\rho(u_0..u_{k-1}, v_0..v_{k-1})(R \times R')$ is the probability that, having observed $u_0..u_{k-1}$ and $v_0..v_{k-1}$, the generated sequence of time delays are in $R, R' \subseteq \mathbb{R}^k_+$, respectively.

For a coupling structure $\mathcal{C} = (\tau_{\mathcal{C}}, \rho_{\mathcal{C}}) \in \mathbb{C}_k(\mathcal{M})$, define the self-map $\Gamma^{\mathcal{C}}$ over $[0,1]$-valued functions on $S^{k+1} \times S^{k+1}$ as follows[5]

$$\Gamma^{\mathcal{C}}(d)(u_0..u_k, v_0..v_k) = \begin{cases} 0 & \text{if } \alpha = 0 \\ 1 & \text{if } \alpha \neq 0,\ \exists i.\, u_i \not\equiv_\ell v_i \\ \beta + (1-\beta) \cdot \int d \, \mathrm{d}\tau_{\mathcal{C}}(u_k, v_k) & \text{otherwise} \end{cases}$$

[5] Since, for all $u, v \in S$, $\tau_{\mathcal{C}}(u,v)$ is a discrete measure on a finite space, the Lebesgue integral $\int d \, \mathrm{d}\tau_{\mathcal{C}}(u,v)$ in the definition of $\Gamma^{\mathcal{C}}$ is $\sum_{x,y \in S^{k+1}} d(x,y) \cdot \tau_{\mathcal{C}}(u,v)(x,y)$.

where $\beta = \rho_{\mathcal{C}}(u_0..u_{k-1}, v_0..v_{k-1})(\neq)$ and $\alpha = \tau_{\mathcal{C}}(u_0, v_0)(u_0..u_k, v_0..v_k)$.

The operator $\Gamma^{\mathcal{C}}$ is monotonic w.r.t. the point-wise order on $[0,1]$-valued functions. Hence, by Tarski's fixed point theorem, $\Gamma^{\mathcal{C}}$ has a least fixed point, which we denote by $\gamma^{\mathcal{C}}$. The next result shows that $\gamma^{\mathcal{C}}$ is closely related to the discrepancy associated with the coupling structure $\mathcal{C}$, and this will eventually be used to compute it.

Lemma 29. *For any coupling structure $\mathcal{C}$, $\mathbb{P}^{\mathcal{C}}_{s,s'}(\not\equiv_{\ell^\omega}) = \int \gamma^{\mathcal{C}} \,\mathrm{d}\tau_{\mathcal{C}}(s, s')$.*

6.2 Construction of an Optimal Coupling Structure

In this subsection we construct an optimal coupling structure by iterating successive updates of a given coupling structure. We provide necessary and sufficient conditions for a coupling structure $\mathcal{C}$ to ensure that $\delta{\downarrow_k}$ is obtained from $\gamma^{\mathcal{C}}$.

To this end, we first introduce the notion of update for a coupling structure.

Definition 30 (Update). *Let $\mathcal{C} = (\tau_{\mathcal{C}}, \rho_{\mathcal{C}}) \in \mathbb{C}_k(\mathcal{M})$. For $\mu \in \Omega(\mathbb{S}^k(u), \mathbb{S}^k(v))$ and $\nu \in \Omega(\mathbb{T}^k(u_1..u_k), \mathbb{T}^k(v_1..v_k))$, define*

- *transition update: $\mathcal{C}[(u, v)/\mu] = (\tau_{\mathcal{C}}[(u, v) \mapsto \mu], \rho_{\mathcal{C}})$;*
- *delay update: $\mathcal{C}\langle(u_1..u_k, v_1..v_k)/\nu\rangle = (\tau_{\mathcal{C}}, \rho_{\mathcal{C}}[(u_1..u_k, v_1..v_k) \mapsto \nu])$.*

where, for a function $f\colon X \to Y$, $f[x \mapsto y]$ denotes the update of f at x with y.

Our update strategy relies on the following result.

Lemma 31 (Update criteria). *Let $\mathcal{C} = (\tau_{\mathcal{C}}, \rho_{\mathcal{C}}) \in \mathbb{C}_k(\mathcal{M})$ 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 $i \leq k$, $u_i \equiv_\ell v_i$. Then, for $\mu \in \Omega(\mathbb{S}^k(u_k), \mathbb{S}^k(v_k))$, $\nu \in \Omega(\mathbb{T}^k(u_0..u_{k-1}), \mathbb{T}^k(v_0..v_{k-1}))$ and $\mathcal{D} = \mathcal{C}[(u_k, v_k)/\mu]\langle(u_0..u_{k-1}, v_1..v_{k-1})/\nu\rangle$, it holds $\gamma^{\mathcal{D}} < \gamma^{\mathcal{C}}$ whenever*

(i) $\nu(\neq) < \rho_{\mathcal{C}}(u_0..u_{k-1}, v_1..v_{k-1})(\neq)$ and $\int \gamma^{\mathcal{C}} \,\mathrm{d}\mu \leq \int \gamma^{\mathcal{C}} \,\mathrm{d}\tau_{\mathcal{C}}(u_k, v_k)$, or
(ii) $\nu(\neq) \leq \rho_{\mathcal{C}}(u_0..u_{k-1}, v_1..v_{k-1})(\neq)$ and $\int \gamma^{\mathcal{C}} \,\mathrm{d}\mu < \int \gamma^{\mathcal{C}} \,\mathrm{d}\tau_{\mathcal{C}}(u_k, v_k)$.

Condition (i) in Lemma 31 ensures that any $\mathcal{C} = (\tau_{\mathcal{C}}, \rho_{\mathcal{C}}) \in \mathbb{C}_k(\mathcal{M})$ is improved by replacing $\rho_{\mathcal{C}}$ with the function $\rho^*\colon S^k \times S^k \to \Delta(\mathbb{R}^k_+ \times \mathbb{R}^k_+)$ defined as

$$\begin{aligned}
\rho^*(u_0..u_{k-1}, v_1..v_{k-1}) &= \min\left\{\nu(\neq) \mid \nu \in \Omega(\mathbb{T}^k(u_0..u_{k-1}), \mathbb{T}^k(v_0..v_{k-1}))\right\} \\
&= \|\mathbb{T}^k(u_0..u_{k-1}) - \mathbb{T}^k(v_0..v_{k-1})\| \qquad \text{(Lemma 18)} \\
&= 1 - \textstyle\prod_{i=0}^{k-1}(1 - \|\rho(u_i) - \rho(v_i)\|) = \beta^*\,,
\end{aligned}$$

where the last equality follows by the definition of $\mathbb{T}^k(u_0..u_{k-1})$ and $\mathbb{T}^k(v_0..v_{k-1})$ as product measures. Notice that, assuming A2, the above is computable. By replacing β in the definition of $\Gamma^{\mathcal{C}}$ with β^*, $\gamma^{\mathcal{C}}$ can be computed as the least solution of the linear equation system induced by the definition of $\Gamma^{\mathcal{C}}$.

Condition (ii) of Lemma 31 suggests to improve $\mathcal{C}$ with $\mathcal{C}[(u_k, v_k)/\mu^*]$ where

$$\begin{aligned}
\mu^* &= \arg\min\left\{\textstyle\int \gamma^{\mathcal{C}} \,\mathrm{d}\mu \mid \mu \in \Omega(\mathbb{S}^k(u_k), \mathbb{S}^k(v_k))\right\} \\
&= \arg\min\left\{\textstyle\sum_{x,y \in S^{k+1}} \gamma^{\mathcal{C}}(x, y) \cdot \mu(x, y) \mid \mu \in \Omega(\mathbb{S}^k(u_k), \mathbb{S}^k(v_k))\right\}\,.
\end{aligned}$$

The above is a linear program (a.k.a. *transportation problem*), hence computable.

The sufficient conditions for termination is provided by the following lemma.

Lemma 32. *Let* $\mathcal{C} = (\tau_{\mathcal{C}}, \rho^*) \in \mathbb{C}_{2^k}(\mathcal{M})$ *be such that* $\delta{\downarrow_k}(u, v) \neq \int \gamma^{\mathcal{C}} \, \mathrm{d}\tau_{\mathcal{C}}(u, v)$ *for some* $u, v \in S$. *Then there exist* $u', v' \in S$ *and* $\mu \in \Omega(\mathbb{S}^{2^k}(u'), \mathbb{S}^{2^k}(v'))$ *such that* $\int \gamma^{\mathcal{C}} \, \mathrm{d}\mu < \int \gamma^{\mathcal{C}} \, \mathrm{d}\tau_{\mathcal{C}}(u', v')$.

Intuitively, the above ensures that, unless $\mathcal{C}$ is an optimal coupling structure, (ii) in Lemma 31 is satisfied, so that, we can further improve $\mathcal{C}$ as aforesaid.

Proposition 33. *Assuming A2,* $\delta{\downarrow_k}$ *is computable for all* $k \in \mathbb{N}$.

Proof (sketch). The aforementioned strategy ensures that the updated couplings are chosen from the vertices of the polytopes $\Omega(\mathbb{S}^k(u), \mathbb{S}^k(v))$, for $u, v \in S$. Since these polytopes have finitely many vertexes, the procedure eventually terminates. By Lemma 32, the last coupling describes $\delta{\downarrow_k}$. □

7 Conclusions and Future Work

In this paper we showed that the total variation distance of SMCs (i.e., the trace distance) is the appropriate behavioral distance to reason about linear real-time properties. This has been done by giving characterizations in terms of MTL formulas or timed ω-regular languages that arise naturally in the context of linear real-time probabilistic model checking. Notably, the technique that has been proposed to prove this result is more general and allows for many more interesting characterizations. We showed, for instance, that the distance can be characterized by considering strictly less expressive fragments of MTL, namely MTL$^-$; analogously, it suffices to consider only the subclass of ω-languages recognized by single-clock always resetting DTAs.

Moreover, we studied the problem of approximating the trace distance within any absolute error. We showed that the problem is computable by approximating the total variation distance both from above and below by means of the sequences $\{\delta{\downarrow_k}\}_k$ and $\{\delta{\uparrow_k}\}_k$, that are proved to be effective. This both extends the result of [9] to the real-time setting and gives an alternative way to approximate the total variation distance on MCs.

As a future work we consider to further explore the potentiality of the presented results by studying how fast the sequences converge to the total variation distance. Moreover, we would like to see if similar results can be used to link different behavioral distances, such as the Kantorovich-based bisimilarity distance and the total variation (for which the former is know to be an upper bound of the latter), opening for the possibility of “bridging the gap” between trace and branching-based behavioral distances.

From a computational perspective, also motivated by our previous work [4] on MCs, we would like to implement an on-the-fly algorithm for computing tight over-approximations of the trace distance.

References

1. R. Alur and D. Dill. Automata for Modeling real-time Systems. In M. S. Paterson, editor, *Automata, Languages and Programming*, volume 443 of *Lecture Notes in Computer Science*, pages 322–335. Springer Berlin Heidelberg, 1990.
2. R. Alur and T. A. Henzinger. Real-Time Logics: Complexity and Expressiveness. *Information and Computation*, 104(1):35–77, 1993.
3. R. Alur and T. A. Henzinger. A Really Temporal Logic. *Journal of the ACM*, 41(1):181–204, 1994.
4. G. Bacci, G. Bacci, K. G. Larsen, and R. Mardare. On-the-Fly Exact Computation of Bisimilarity Distances. In *TACAS*, volume 7795 of *Lecture Notes in Computer Science*, pages 1–15, 2013.
5. G. Bacci, G. Bacci, K. G. Larsen, and R. Mardare. On the Total Variation Distance of Semi-Markov Chains. Technical report, AAU, DK, 2014. `http://people.cs.aau.dk/~grbacci/Papers/tvsmc_ext.pdf`.
6. P. Billingsley. *Probability and Measure*. Wiley, New York, 3rd edition, 1995.
7. T. Chen, M. Diciolla, M. Z. Kwiatkowska, and A. Mereacre. Time-Bounded Verification of CTMCs against Real-Time Specifications. In *FORMATS*, volume 6919 of *Lecture Notes in Computer Science*, pages 26–42, 2011.
8. T. Chen, T. Han, J.-P. Katoen, and A. Mereacre. Model checking of continuous-time markov chains against timed automata specifications. *Logical Methods in Computer Science*, 7(1), 2011.
9. T. Chen and S. Kiefer. On the Total Variation Distance of Labelled Markov Chains. In *Proc. of CSL-LICS '14*, pages 33:1–33:10, New York, NY, USA, 2014. ACM.
10. L. de Alfaro, M. Faella, and M. Stoelinga. Linear and branching metrics for quantitative transition systems. In *ICALP*, volume 3142 of *Lecture Notes in Computer Science*, pages 97–109, 2004.
11. L. de Alfaro, R. Majumdar, V. Raman, and M. Stoelinga. Game Relations and Metrics. In *LICS*, pages 99–108, July 2007.
12. J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for labelled Markov processes. *Theoretical Compututer Science*, 318(3):323–354, 2004.
13. N. Dunford and J. T. Schwartz. *Linear Operators, Part 1, General Theory*. Wiley Classic Library. John Wiley, New York, 1988.
14. T. Lindvall. *Lectures on the Coupling Method*. Wiley Series in Probability and Mathematical Statistics. John Wiley, New York, 1992.
15. R. B. Lyngsø and C. N. Pedersen. The consensus string problem and the complexity of comparing hidden Markov models. *Journal of Computer and System Sciences*, 65(3):545–569, 2002. Special Issue on Computational Biology 2002.
16. M. Mio. Upper-Expectation Bisimilarity and Łukasiewicz μ-Calculus. In *FoSSaCS*, volume 8412 of *Lecture Notes in Computer Science*, pages 335–350, 2014.
17. J. Ouaknine and J. Worrell. On the decidability and complexity of Metric Temporal Logic over finite words. *Logical Methods in Computer Science*, 3(8), 2007.
18. A. Sharma and J.-P. Katoen. Weighted Lumpability on Markov Chains. In *Ershov Memorial Conference*, volume 7162 of *Lecture Notes in Computer Science*, pages 322–339, 2011.
19. F. van Breugel and J. Worrell. Approximating and computing behavioural distances in probabilistic transition systems. *Theoretical Computer Science*, 360(3):373–385, 2006.

A Technical proofs

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

Proof (of Lemma 5). The closure of $\mathcal{F}$ under d_μ is given by

$$\overline{\mathcal{F}} = \{E \in \Sigma \mid \forall \varepsilon > 0.\ \exists F \in \mathcal{F}.\ d_\mu(E, F) < \varepsilon\} \ .$$

We show that $\overline{\mathcal{F}} = \Sigma$. Clearly $\overline{\mathcal{F}} \subseteq \Sigma$. The converse inclusion follows by $\mathcal{F} \subseteq \overline{\mathcal{F}}$ and $\Sigma = \sigma(\mathcal{F})$, showing that $\overline{\mathcal{F}}$ is a σ-algebra:

Complement. Let $E \in \overline{\mathcal{F}}$. We want to show that $E^c \in \overline{\mathcal{F}}$, where $E^c := \Sigma \setminus E$ denotes the complement of E in Σ. Let $\varepsilon > 0$. By $E \in \overline{\mathcal{F}}$, there exists $F \in \mathcal{F}$ such that $d_\mu(E, F) < \varepsilon$. Moreover, note that $E \triangle F = E^c \triangle F^c$, so

$$d_\mu(E^c, F^c) = \mu(E^c \triangle F^c) = \mu(E \triangle F) = d_\mu(E, F) \,,$$

and $d_\mu(E^c, F^c) < \varepsilon$. By hypothesis, $\mathcal{F}$ is a field, hence $F^c \in \mathcal{F}$. Due to the generality of $\varepsilon > 0$, this proves $E^c \in \overline{\mathcal{F}}$.

Countable Union. Let $\{E_i \mid i \in \mathbb{N}\} \subseteq \overline{\mathcal{F}}$. We want to show that $\bigcup_{i\in\mathbb{N}} E_i \in \overline{\mathcal{F}}$. Let $\varepsilon > 0$. To prove the thesis it suffices to show that following statements hold:

a) there exists $k \in \mathbb{N}$, such that $d_\mu(\bigcup_{i\in\mathbb{N}} E_i, \bigcup_{i=0}^k E_i) < \frac{\varepsilon}{2}$;

b) for all $n \in \mathbb{N}$, there exist $F_0, \dots, F_n \in \mathcal{F}$, such that $d_\mu(\bigcup_{i=0}^n E_i, \bigcup_{i=0}^n F_i) < \frac{\varepsilon}{2}$.

Indeed, by applying the triangular inequality on (a) and (b), we have that there exist $k \in \mathbb{N}$ and $F_0, \dots, F_k \in \mathcal{F}$ such that

$$d_\mu(\textstyle\bigcup_{i\in\mathbb{N}} E_i, \bigcup_{i=0}^k F_i) \leq d_\mu(\bigcup_{i\in\mathbb{N}} E_i, \bigcup_{i=0}^k E_i) + d_\mu(\bigcup_{i=0}^k E_i, \bigcup_{i=0}^k F_i) < \varepsilon \,.$$

Since by hypothesis $\mathcal{F}$ is a field, we also have $\bigcup_{i=0}^k F_i \in \mathcal{F}$. Therefore, due to the generality of $\varepsilon > 0$, we obtain $\bigcup_{i\in\mathbb{N}} E_i \in \overline{\mathcal{F}}$.

(a). Since $\{\bigcup_{i=0}^n E_i\}_{n\in\mathbb{N}}$ is an increasing sequence converging to $\bigcup_{i\in\mathbb{N}} E_i$, by ω-continuity from below of μ, we have that $\{\mu(\bigcup_{i=0}^n E_i)\}_{n\in\mathbb{N}}$ converges in $\mathbb{R}$ to $\mu(\bigcup_{i\in\mathbb{N}} E_i)$, hence there exists and index $k \in \mathbb{N}$ such that

$$|\mu(\textstyle\bigcup_{i\in\mathbb{N}} E_i) - \mu(\bigcup_{i=0}^k E_i)| < \frac{\varepsilon}{2} \,.$$

By $\bigcup_{i\in\mathbb{N}} E_i \subseteq \bigcup_{i=0}^k E_i$ and monotonicity, additivity and finiteness of μ,

$$\begin{aligned} d_\mu(\textstyle\bigcup_{i\in\mathbb{N}} E_i, \bigcup_{i=0}^k E_i) &= \mu(\textstyle\bigcup_{i\in\mathbb{N}} E_i \triangle \bigcup_{i=0}^k E_i) \\ &= \mu(\textstyle\bigcup_{i\in\mathbb{N}} E_i \setminus \bigcup_{i=0}^k E_i) \\ &= \mu(\textstyle\bigcup_{i\in\mathbb{N}} E_i) - \mu(\bigcup_{i=0}^k E_i) < \frac{\varepsilon}{2} \,. \end{aligned}$$

(b). Let $n \in \mathbb{N}$. By $E_0, \dots, E_n \in \overline{\mathcal{F}}$, there exists $F_0, \dots, F_n \in \mathcal{F}$ such that $d_\mu(E_i, F_i) < \frac{\varepsilon}{2n}$. Moreover, note that $\bigcup_{i=0}^n E_i \triangle \bigcup_{i=0}^n F_i \subseteq \bigcup_{i=0}^n (E_i \triangle F_i)$, so that by monotonicity and sub-additivity of μ we have

$$\begin{aligned} d_\mu(\textstyle\bigcup_{i=0}^n E_i, \bigcup_{i=0}^k F_i) &= \mu(\textstyle\bigcup_{i=0}^n E_i \triangle \bigcup_{i=0}^n F_i) \\ &\leq \mu(\textstyle\bigcup_{i=0}^n (E_i \triangle F_i)) \\ &\leq \textstyle\sum_{i=0}^n \mu(E_i \triangle F_i) < \sum_{i=0}^n \frac{\varepsilon}{2n} = \frac{\varepsilon}{2} \,. \end{aligned}$$

□

Proof (of Theorem 6). For $Y \neq \emptyset$ and $f\colon Y \to \mathbb{R}$ bounded and continuous, if $D \subseteq Y$ is dense then $\sup f(D) = \sup f(Y)$. By Lemma 5, $\mathcal{F}$ is dense in $(\Sigma, d_{\mu+\nu})$. We show that $|\mu - \nu|\colon \Sigma \to \mathbb{R}$ is bounded and continuous. Let E and F be arbitrary measurable sets in Σ, then

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

This implies that, for all $E, F \in \Sigma$, $d_{\mu+\nu}(E, F) \geq |\mu(E) - \mu(F)|$, hence $\mu\colon \Sigma \to \mathbb{R}$ is 1-Lipschitz continuous. Analogously, also $\nu\colon \Sigma \to \mathbb{R}$ is 1-Lipschitz continuous. Then, continuity of $|\mu - \nu|\colon \Sigma \to \mathbb{R}$ follows by composition of continuous functions. Moreover, $|\mu - \nu|$ is bounded because, by hypothesis, μ and ν are finite. □

Proof (of Lemma 7). We prove the two statements separately.

(i) By structural induction on the syntax of $\varphi \in \text{MTL}$ we prove that $[\![\varphi]\!] \in \sigma(\mathcal{T})$.
Atomic prop. $[\![a]\!] = \{\pi \mid a \in \ell(\pi[0])\} = \bigcup \{\mathfrak{C}([s]_{\equiv_\ell}) \mid s \in \ell^{-1}(\{a\})\}$. Since S is finite and $\mathfrak{C}([s]_{\equiv_\ell}) \in \mathcal{T}$ for all $s \in S$, then $[\![a]\!] \in \sigma(\mathcal{T})$.
False. $[\![\bot]\!] = \emptyset \in \sigma(\mathcal{T})$.
Implication. $[\![\phi \to \psi]\!] = [\![\neg\phi \vee \psi]\!] = [\![\phi]\!]^c \cup [\![\psi]\!]$. By inductive hypothesis, $[\![\phi]\!], [\![\psi]\!] \in \sigma(\mathcal{T})$, therefore $[\![\phi \to \psi]\!] \in \sigma(\mathcal{T})$.
Next. Consider $\mathsf{X}^I \phi$. The following hold

$$\begin{aligned}
[\![\mathsf{X}^I \phi]\!] &= \{\pi \mid \pi\langle 0\rangle \in I, \text{ and } \mathcal{M}, \pi|_1 \models \phi\} && (\text{by def. of } \mathsf{X})\\
&= \{\pi \mid \pi\langle 0\rangle \in I, \text{ and } \pi|_1 \in [\![\phi]\!]\} && (\text{by def. of } [\![\cdot]\!])\\
&= (\cdot)\langle 0\rangle^{-1}(I) \cap (\cdot)|_1^{-1}([\![\phi]\!]) && (\text{by def. of } (\cdot)\langle 0\rangle \text{ and } (\cdot)|_1)
\end{aligned}$$

By inductive hypothesis and the fact that both $(\cdot)\langle 0\rangle$ and $(\cdot)|_1$ are measurable functions, it follows that $[\![\mathsf{X}^I \phi]\!] \in \sigma(\mathcal{T})$.
Until. Consider $[\![\phi\, \mathsf{U}^{[a,b]}\, \psi]\!]$. For $k > 0$ we define the set $\mathit{OnTime@k}$ as

$$\mathit{OnTime@k} = \bigcup \left\{ \mathfrak{C}(X) \;\middle|\; \begin{array}{l} C_i \in S/_{\equiv_\ell},\ t_i^-, t_i^+ \in \mathbb{Q}_+, \text{ for } 0 \leq i \leq k, \\ \sum_{i=0}^{k-1} t_i^- \geq a,\ \sum_{i=0}^{k-1} t_i^+ \leq b,\ t_i^- \leq t_i^+, \\ X = C_0, [t_0^-, t_0^+], \ldots, [t_{k-1}^-, t_{k-1}^+], C_n \end{array} \right\}$$

Notice that $\mathit{OnTime@k}$ is a countable union of cylinders in $\mathcal{T}$ (the number of unions is bounded by $|(S \times \mathbb{Q}_+^2)^{k+1}|$), hence it is a measurable set in $\sigma(\mathcal{T})$.

$$\mathit{OnTime@k} = \{\pi \mid \forall i < k.\ \textstyle\sum_{i=0}^{k-1} \pi\langle i\rangle \in [a, b]\} \tag{3}$$

The inclusion from left to right trivially holds by definition of $\mathit{OnTime@k}$. As for the reverse inclusion, let π be a timed path over S, such that $\pi\langle i\rangle = t_i$

$(i = 0..k-1)$ and $\sum_{i=0}^{k-1} t_i \in [a,b]$. We have to prove that there exist $t_i^-, t_i^+ \in \mathbb{Q}_+$ such that $t_i^- \leq t_i \leq t_i^+$, $\sum_{i=0}^{k-1} t_i^- \geq a$, and $\sum_{i=0}^{k-1} t_i^+ \leq b$. When $k = 1$ it suffices to take $t_1^- = a$ and $t_1^+ = b$. Assume $k > 1$. Let $\Delta = 2h/10^h$ for some $h \in \mathbb{N}$ large enough to satisfy the following two inequalities $\sum_{i=0}^{k-1} t_i - \Delta > a$ and $\sum_{i=0}^{k-2} t_i + \Delta < b$. Let $t_i^- = t_i = t_i^+$ if $t_i \in \mathbb{Q}_+$, otherwise we choose some $t_i^-, t_i^+ \in \mathbb{Q}_+$ that satisfy

$$t_i^- < t_i < t_i^+, \qquad t_i^- > t_i - \Delta/2k, \text{ and} \qquad t_i^+ < t_i + \Delta/2k\,. \tag{4}$$

We proceed by showing that the constraints (4) are sufficient to prove that $\sum_{i=0}^{k-1} t_i^- \geq a$ and $\sum_{i=0}^{k-1} t_i^+ \leq b$, then we show how to pick $t_i^-, t_i^+ \in \mathbb{Q}_+$ in order to satisfy (4). The following hold

$$\begin{aligned} \textstyle\sum_{i=0}^{k-1} t_i - \Delta &< \textstyle\sum_{i=0}^{k-1} (t_i^- + \Delta/2k) - \Delta = && \text{(by (4))} \\ &= \textstyle\sum_{i=0}^{k-1} t_i^- - \Delta/2 \leq \sum_{i=0}^{k-1} t_i^- \,. && \text{(by } \Delta \geq 0) \end{aligned}$$

By construction, $\sum_{i=0}^{k-1} t_i - \Delta > a$, hence $\sum_{i=0}^{k-1} t_i^- > a$. Analogously,

$$\begin{aligned} \textstyle\sum_{i=0}^{k-2} t_i + \Delta &> \textstyle\sum_{i=0}^{k-1} (t_i^+ - \Delta/2k) - \Delta = && \text{(by (4))} \\ &= \textstyle\sum_{i=0}^{k-2} t_i^+ + (k+1)\Delta/2k \geq \sum_{i=0}^{k-2} t_i^+ \,. && \text{(by } \Delta \geq 0) \end{aligned}$$

By construction, $\sum_{i=0}^{k-2} t_i + \Delta < b$, hence $\sum_{i=0}^{k-2} t_i^+ < b$.
One can check that the constraints (4) are easily satisfied if we pick

$$t_i^- = \lfloor t_i \rfloor + \frac{\lfloor 10^h \cdot \{t_i\} \rfloor}{10^h}\,, \qquad t_i^+ = \lfloor t_i \rfloor + \frac{\lfloor 10^h \cdot \{t_i\} \rfloor + 1}{10^h}\,,$$

for some large enough $h \in \mathbb{N}$, where $\{t_i\}$ denotes the fractional part of $t_i \notin \mathbb{Q}_+$. This proves (3).

$$\begin{aligned} &[\![\phi \,\mathsf{U}^{[a,b]}\, \psi]\!] \\ &= \left\{ \pi \;\middle|\; \begin{array}{l} \exists i > 0.\ \sum_{k=0}^{i-1} \pi\langle i \rangle \in [a,b],\ \text{and } \mathcal{M}, \pi|_i \models \psi, \\ \forall 0 \leq j < i.\, \mathcal{M}, \pi|_j \models \phi \end{array} \right\} && \text{(by def. } \mathsf{U}) \\ &= \left\{ \pi \;\middle|\; \begin{array}{l} \exists i > 0.\ \sum_{k=0}^{i-1} \pi\langle i \rangle \in [a,b],\ \text{and } \pi|_i \in [\![\psi]\!], \\ \forall 0 \leq j < i.\, \pi|_j \in [\![\phi]\!] \end{array} \right\} && \text{(by def. } [\![\cdot]\!]) \\ &= \bigcup_{i>0} \bigcap_{0 \leq j < i} ((\cdot)|_j^{-1}([\![\phi]\!]) \cap (\cdot)|_i^{-1}([\![\psi]\!]) \cap \mathit{OnTime@i})\,. && \text{(by def. } (\cdot)|_k \text{ and (3))} \end{aligned}$$

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

(ii) We show $\sigma(\mathcal{T}) \subseteq \sigma([\![\mathrm{MTL}^-]\!])$. Let $\mathcal{I}$ be the family of closed intervals in $\mathbb{R}_+$ with rational endpoints. It is standard that $\sigma(\mathcal{I}) = \mathbb{B}$, and from it one can easily

verify $\sigma(\mathfrak{C}(S/_{\equiv_\ell}, \mathcal{I})) = \sigma(\mathcal{T})$. Therefore, to prove $\sigma(\mathcal{T}) \subseteq \sigma([\![\mathrm{MTL}^-]\!])$, it suffices to show $\mathfrak{C}(S/_{\equiv_\ell}, \mathcal{I}) \subseteq \sigma([\![\mathrm{MTL}^-]\!])$. Let define $Ap \colon \mathbb{A} \times \mathfrak{C}(S/_{\equiv_\ell}, \mathcal{I}) \to \mathrm{MTL}^-$ as

$$Ap(a, \mathfrak{C}(C)) = \begin{cases} a & \text{if } C \subseteq \ell^{-1}(a) \\ \neg a & \text{othewise} \end{cases}$$

$$Ap(a, \mathfrak{C}(C, I, X)) = \begin{cases} a \wedge \mathsf{X}^I Ap(a, \mathfrak{C}(X)) & \text{if } C \subseteq \ell^{-1}(a) \\ \neg a \wedge \mathsf{X}^I Ap(a, \mathfrak{C}(X)) & \text{otherwise}\,, \end{cases}$$

Let $C = \mathfrak{C}(C_0, I_0, \dots, I_{n-1}, C_n) \in \mathfrak{C}(S/_{\equiv_\ell}, \mathcal{I})$, one can prove by induction on n that $\bigcap_{a \in \mathbb{A}} [\![Ap(a, C)]\!] = C$. Since $\sigma([\![\mathrm{MTL}^-]\!])$ is closed under countable intersection, we conclude that $C \in \sigma([\![\mathrm{MTL}^-]\!])$. $\square$

Proof (of Corollary 8). $\mathcal{L}$ is closed under all Boolean operators, therefore $[\![\mathcal{L}]\!]$ is a field. By Lemma 7(i) and $\mathcal{L} \subseteq \mathrm{MTL}$, $[\![\mathcal{L}]\!] \subseteq \sigma(\mathcal{T})$. Since $\mathcal{T} \subseteq \sigma([\![\mathcal{L}]\!])$, it follows that $\sigma([\![\mathcal{L}]\!]) = \sigma(\mathcal{T})$. The equality $\delta_{\mathcal{L}} = \delta$ now follows by Theorem 6. In particular $\delta_{\mathrm{MTL}} = \delta_{\mathrm{MTL}^-} = \delta$ follows by Lemma 7(ii). $\square$

Proof (Lemma 11). We prove the two statements separately.

(i) This is proven in [8, Theorem 3.2] and the proof can be left unchanged.
(ii) We show $\sigma(\mathcal{T}) \subseteq \sigma([\![\mathrm{DTA}]\!])$. Let $\mathcal{I}$ be the family of closed intervals in $\mathbb{R}_+$ with rational endpoints. It is standard that $\sigma(\mathcal{I}) = \mathbb{B}$, and from it one can easily verify $\sigma(\mathfrak{C}(S/_{\equiv_\ell}, \mathcal{I})) = \sigma(\mathcal{T})$. Therefore, to show $\sigma(\mathcal{T}) \subseteq \sigma([\![\mathrm{DTA}]\!])$ it suffices to prove $\mathfrak{C}(S/_{\equiv_\ell}, \mathcal{I}) \subseteq \sigma([\![\mathrm{DTA}]\!])$.

Let $C = \mathfrak{C}([s_0]_{\equiv_\ell}, I_0, \dots, I_{n-1}, [s_n]_{\equiv_\ell}) \in \mathfrak{C}(S/_{\equiv_\ell}, \mathcal{I})$. We define a DTA $\mathcal{A} = (Q, 2^{\mathbb{A}}, q_0, F, \to)$ such that $[\![\mathcal{A}]\!] = C$. Let $Q = \{q_0, \dots, q_n\}$, $F = \{\{q_n\}\}$, and, for a (shared) clock $x \in \mathcal{X}$ in each guard,

$$\begin{aligned} \to = & \{(q_i, \ell(s_i), g_i, \mathcal{X}, q_{i+1}) \mid g_i = a \leq x \leq b \text{ for } I_i = [a, b], 0 \leq i \leq n\} \\ & \cup \{(q_n, l, \emptyset, \mathcal{X}, q_n) \mid l \subseteq \mathbb{A}\}\ . \end{aligned}$$

It is easy to see that the only accepted timed paths $\pi \in \mathcal{L}(\mathcal{A})$ are such that $\pi|^n = \ell(s_0), t_0, \dots, t_{n-1}, \ell(s_n)$, and $t_i \in I_i$ $(0 \leq i \leq n-1)$, because clocks are always resetting. So the thesis. $\square$

Proof (of Corollary 12). Φ is closed under all Boolean operators, therefore $[\![\Phi]\!]$ is a field. By Lemma 11(i) and $\Phi \subseteq \mathrm{TA}$, $[\![\Phi]\!] \subseteq \sigma(\mathcal{T})$. Since $\mathcal{T} \subseteq \sigma([\![\Phi]\!])$, it follows that $\sigma([\![\Phi]\!]) = \sigma(\mathcal{T})$. The equality $\delta_\Phi = \delta$ follows by Theorem 6. In particular $\delta_{\mathrm{DTA}} = \delta$ follows by Lemma 11(ii) and the fact that DTAs are closed under all Boolean operators [1]. The equality $\delta_{\mathrm{TA}} = \delta$, follows by $\delta \geq \delta_{\mathrm{TA}} \geq \delta_{\mathrm{DTA}}$. $\square$

Proof (of Lemma 13).

(i) It suffices to prove that 1-RDTAs are closed under union and complement. As for the latter, one need only to complement the set of final locations (similar to [1]). Closure under union is proven via a product construction similar to [1, p.334], by noticing that the resetting condition on the automata allows one to use a single clock in the product.

(ii) The proof of Lemma 11(ii) actually uses 1-RDTAs. □

Proof (of Lemma 15). 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 $C' = \mathfrak{C}(C_0, R_0, .., R_{k-1}, C_k) \in \mathfrak{C}_k(S/_{\equiv_\ell}, \mathfrak{R}_{k+1})$ can be represented as

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

Since $\mathfrak{R}_{k+1}$ and $S/_{\equiv_\ell}$ 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 $\Pi(S)$, 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

$$[q, q') = \bigcup \left\{ \left[\frac{n}{2^k}, \frac{n+1}{2^k} \right) \middle|\ q \leq \frac{n}{2^k} < \frac{n+1}{2^k} \leq q',\ \text{for } k \in \mathbb{N},\, 0 \leq n < k2^k \right\} ,$$
$$[q, \infty) = \bigcup \left\{ \left[\frac{n}{2^k}, \frac{n+1}{2^k} \right) \middle|\ q \leq \frac{n}{2^k},\ \text{for } k \in \mathbb{N},\, 0 \leq n < k2^k \right\} .$$

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 $\sigma(\mathcal{E}) \subseteq \sigma(\mathcal{T})$. □

Proof (of Lemma 18 —restated from [14, Th.5.2]). We prove that $\|\mu - \nu\|$ is a lower bound for $\{\omega(\ncong) \mid \omega \in \Omega(\mu, \nu)\}$. Let $\omega \in \Omega(\mu, \nu)$ and $E \in \Sigma$, then

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

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

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

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

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

The above is a well defined measure (a.k.a. the meet of μ and ν, see [13, Corr.6 pp.163]). Now define the following derived measures

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

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

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

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

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

□

Proof (of Lemma 20). We first show $\equiv_{\ell^\omega} = \cap_{E \in \sigma(\mathcal{T})} \{(\pi, \pi') \mid \pi \in E \text{ iff } \pi' \in E\}$. ($\subseteq$) It suffices to prove inseparability w.r.t trace cylinders. Let $\pi \equiv_{\ell^\omega} \pi'$ and $\pi \in C = \mathfrak{C}(C_0, R_0, \dots, R_{n-1}, C_n) \in \mathcal{T}$, for $C_i \in S/_{\equiv_\ell}$ and $R_i \in \mathbb{B}$, $i = 0..n$. Then, for all $j \in \mathbb{N}$, $\ell(\pi[j]) = \ell(\pi'[j])$ (hence, $\pi[j] \equiv_\ell \pi'[j]$) and $\pi\langle j \rangle = \pi'\langle j \rangle$, so that $\pi' \in C$. ($\supseteq$) By contraposition. Let $\pi \not\equiv_{\ell^\omega} \pi'$, then there exist $j \in \mathbb{N}$ such that $\pi[j] \not\equiv_\ell \pi[j]$ or $\pi\langle j \rangle \neq \pi'\langle j \rangle$. Let $E = (\cdot)|_j^{-1}(\mathfrak{C}([\pi[j]]_{\equiv_\ell}, \{\pi\langle j \rangle\}, S))$, then $\pi \in E$ but $\pi' \notin E$. The inclusion follows since the function $(\cdot)|_j$ is measurable.

As for the measurability of $\equiv_{\ell^\omega}$, it suffices to show that its complement $\not\equiv_{\ell^\omega} \in \sigma(\mathcal{T}) \otimes \sigma(\mathcal{T})$. Define $\mathit{DiffS}(k)$ and $\mathit{DiffT}(k)$, for $k \geq 0$, as

$$\mathit{DiffS}(k) := \bigcup_{C \in S/_{\equiv_\ell}} (\cdot)|_k^{-1}(\mathfrak{C}(C)) \times (\cdot)|_k^{-1}(\mathfrak{C}(S \setminus C))\,,$$
$$\mathit{DiffT}(k) := \bigcup_{t<t' \in \mathbb{Q}_+} (\cdot)|_k^{-1}(\mathfrak{C}(S,(t,t'),S)) \times (\cdot)|_k^{-1}(\mathfrak{C}(S,\mathbb{R}_+ \setminus (t,t'),S))\,.$$

Measurability of $\mathit{DiffS}(k)$ and $\mathit{DiffT}(k)$ follows immediately by measurability of $(\cdot)|_k$. Now we show that $\not\equiv_{\ell^\omega} = \bigcup_{k\in\mathbb{N}} \big(\mathit{DiffS}(k) \cup \mathit{DiffT}(k)\big)$.

($\subseteq$) Let $\pi \not\equiv_{\ell^\omega} \pi'$. Then, $\pi[j] \not\equiv_\ell \pi'[j]$ or $\pi\langle j\rangle \neq \pi'\langle j\rangle$, for some $j \in \mathbb{N}$. Assume $\pi[j] \not\equiv_\ell \pi'[j]$, then $(\pi,\pi') \in (\cdot)|_j^{-1}(\mathfrak{C}(C)) \times (\cdot)|_j^{-1}(\mathfrak{C}(S \setminus C))$. Assume $\pi\langle j\rangle \neq \pi'\langle j\rangle$. Let $\varepsilon = |\pi\langle j\rangle - \pi'\langle j\rangle|$. Since $\mathbb{Q}_+$ is dense in $\mathbb{R}_+$, every nonempty open set has nonempty intersection with $\mathbb{Q}_+$, so that there exist $t \in \mathbb{Q}_+ \cap (\pi\langle j\rangle - \varepsilon, \pi\langle j\rangle)$ and $t' \in \mathbb{Q}_+ \cap (\pi\langle j\rangle, \pi\langle j\rangle + \varepsilon)$. Clearly, $\pi\langle j\rangle \in (t,t')$ and $\pi'\langle j\rangle \notin (t,t')$, therefore, $(\pi,\pi') \in (\cdot)|_j^{-1}(\mathfrak{C}(S,(t,t'),S)) \times (\cdot)|_j^{-1}(\mathfrak{C}(S,\mathbb{R}_+ \setminus (t,t'),S))$.

($\supseteq$) Let $(\pi,\pi') \in \mathit{DiffS}(k) \cup \mathit{DiffT}(k)$, for some $k \in \mathbb{N}$. Then, since

$$\begin{aligned}\mathit{DiffS}(k) &= \bigcup_{C \in S/_{\equiv_\ell}} \{(\pi,\pi') \mid \pi|_k \in \mathfrak{C}(C) \text{ and } \pi'|_k \in \mathfrak{C}(S \setminus C)\} \\ &= \{(\pi,\pi') \mid \pi[k] \not\equiv_\ell \pi'[k]\}\ ,\end{aligned}$$

$$\begin{aligned}\mathit{DiffT}(k) &= \bigcup_{t<t' \in \mathbb{Q}_+} \{(\pi,\pi') \mid \pi|_k \in \mathfrak{C}(S,(t,t'),S),\ \pi'|_k \in \mathfrak{C}(S,\mathbb{R}_+ \setminus (t,t'),S)\} \\ &= \bigcup_{t<t' \in \mathbb{Q}_+} \{(\pi,\pi') \mid \pi\langle k\rangle \in (t,t') \text{ and } \pi'\langle k\rangle \notin (t,t')\} \\ &\subseteq \{(\pi,\pi') \mid \pi\langle k\rangle \neq \pi'\langle k\rangle\}\ ,\end{aligned}$$

there exists $k \in \mathbb{N}$ such that $\pi[k] \not\equiv_\ell \pi'[k]$ or $\pi\langle k\rangle \neq \pi'\langle k\rangle$. Thus, $\pi \not\equiv_{\ell^\omega} \pi'$. □

Proof (of Lemma 23). Let $\mathcal{C} \in \mathbb{C}_k(\mathcal{M})$. To prove $\mathbb{P}^{\mathcal{C}}_{s,s'} \in \Omega(\mathbb{P}_s, \mathbb{P}_{s'})$ it suffices to show that, for all $n \in \mathbb{N}$ and $E = \{u_0\} \times R_0 \times .. \times R_{nk-1} \times \{u_{nk}\} \in \mathcal{R}_{nk}$

$$\mathbb{P}^{\mathcal{C}}_{s,s'}(\mathfrak{C}(E) \times \Pi(S)) \overset{\text{(i)}}{=} \mathbb{P}_s(\mathfrak{C}(E))\,, \qquad \mathbb{P}^{\mathcal{C}}_{s,s'}(\Pi(S) \times \mathfrak{C}(E)) \overset{\text{(ii)}}{=} \mathbb{P}_{s'}(\mathfrak{C}(E))\,.$$

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

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

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

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

Proof (of Lemma 24). We prove the two items separately.

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

$$
\mathcal{D}(s, s')(E \times \Pi^{2k}(S)) \overset{\text{(i)}}{=} \mathbb{P}_s|^{2k}(E) \,, \qquad \mathcal{D}(s, s')(\Pi^{2k}(S) \times E) \overset{\text{(ii)}}{=} \mathbb{P}_{s'}|^{2k}(E) \,.
$$

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

$$
\begin{aligned}
&\mathcal{D}(s, s')(E \times \Pi^{2k}(S)) = \\
&\quad = \textstyle\sum_{\boldsymbol{v} \in S^{2k+1}} \mathcal{D}(s, s')(E \times F^{\boldsymbol{v}}) && \text{(additivity)} \\
&\quad = \textstyle\sum_{\boldsymbol{v} \in S^{2k+1}} \mathcal{C}(s, s')(E' \times F^{\boldsymbol{v}}_0) \cdot \mathcal{C}(u_k, v_k)(E'' \times F^{\boldsymbol{v}}_1) && (\text{def. } \mathcal{D}) \\
&\quad = \textstyle\sum_{\boldsymbol{v} \in S^{k+1}} \mathcal{C}(s, s')(E' \times F^{\boldsymbol{v}}_0) \cdot \mathcal{C}(u_k, v_k)(E'' \times \Pi^k(S)) && \text{(additivity)} \\
&\quad = \textstyle\sum_{\boldsymbol{v} \in S^{k+1}} \mathcal{C}(s, s')(E' \times F^{\boldsymbol{v}}_0) \cdot \mathbb{P}_{u_k}|^k(E'') && (\mathcal{C} \in \mathbb{C}_k(\mathcal{M})) \\
&\quad = \mathcal{C}(s, s')(E' \times \Pi^k(S)) \cdot \mathbb{P}_{u_k}|^k(E'') && \text{(additivity)} \\
&\quad = \mathbb{P}_s|^k(E') \cdot \mathbb{P}_{u_k}|^k(E'') && (\mathcal{C} \in \mathbb{C}_k(\mathcal{M})) \\
&\quad = \mathbb{P}_s(\mathfrak{C}(E')) \cdot \mathbb{P}_{u_k}(\mathfrak{C}(E'')) && \text{(preimage)} \\
&\quad = \mathbb{P}_s(\mathfrak{C}(E)) && (\text{def. } \mathbb{P}_s) \\
&\quad = \mathbb{P}_s|^{2k}(E) \,. && \text{(preimage)}
\end{aligned}
$$

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 $E = \{u_0\} \times R_0 \times .. \times R_{2nk-1} \times \{u_{2nk}\}$, $F = \{v_0\} \times H_0 \times .. \times H_{2nk-1} \times \{v_{2nk}\}$ 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 $E^i_h = \{u_{hi}\} \times R_{hi} \times .. \times R_{(h+1)i-1} \times \{u_{(h+1)i}\}$ (similarly for F). Then the following holds:

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

From the above it immediately follows that $\mathbb{C}_k(\mathcal{M}) \subseteq \mathbb{C}_{2k}(\mathcal{M})$.
(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 $d_E \colon \Delta(X) \times \Delta(X) \to \mathbb{R}_+$ be the pseudometric defined as $d_E(\mu, \nu) = |\mu(E) - \nu(E)|$, for $\mu, \nu \in \Delta(X)$. Since $\|\mu - \nu\| = \sup_{E \in \Sigma} d_E(\mu, \nu)$, to prove that D is dense w.r.t. the total variation distance it suffices to show that D is dense w.r.t. d_E, for any $E \in \Sigma$ (see Proposition 37). 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 [13, Corr.6 pp.163] (note that $\tilde{\mu}$ is not necessarily finite). By Lemma 5, $\mathcal{F} \subseteq \Sigma$ is dense in $(\Sigma, d_{\tilde{\mu}})$, where $d_{\tilde{\mu}}$ is the Fréchet-Nikodym pseudometric[6], 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

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

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

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

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

Let $s, s' \in S$, $\Omega = \bigcup_{i \in \mathbb{N}} \left\{ \mathbb{P}^{\mathcal{C}}_{s,s'} \mid \mathcal{C} \in \mathbb{C}_{2^i}(\mathcal{M}) \right\}$. Given the general result above, to prove (ii) it is sufficient to provide a field $\mathcal{F}$ that generates the σ-algebra of $\Pi(S) \otimes \Pi(S)$ 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 $\Pi(S) \otimes \Pi(S)$.

Let $\mu \in \Omega(\mathbb{P}_s, \mathbb{P}_{s'})$, $k \in \mathbb{N}$ and $\mathcal{D} \in \mathbb{C}_k(\mathcal{M})$. We define $\omega_k = \mathbb{P}^{\mathcal{C}_k}_{s,s'}$, where $\mathcal{C}_k \colon S \times S \to \Delta(\Pi^k(S) \times \Pi^k(S))$ 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 $\mathbb{C}_k(\mathcal{M})$ is nonempty, $\mathcal{C}_k$ is well defined. We show $\mathcal{C}_k \in \mathbb{C}_k(\mathcal{M})$. 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 $\mu[(\cdot)|^k \times (\cdot)|^k](E \times \Pi^k(S)) = \mathbb{P}_s|^k(E)$ and $\mu[(\cdot)|^k \times (\cdot)|^k](\Pi^k(S) \times E) = \mathbb{P}_{s'}|^k(E)$ for arbitrary $E \in \mathcal{R}_k$ (we check one equality, the other follows similarly):

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

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).

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

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 $\omega_{2^i} \in \Omega$. □

Proof (of Lemma 28). Consider the functions p_1 and p_2 defined as

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

Note that p_1, p_2 are measurable. For $\mathcal{C} \in \mathbb{C}_k(\mathcal{M})$ 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

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

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

Proof (of Lemma 29). Let $k \in \mathbb{N}$ and $\mathcal{C} = (\tau_\mathcal{C}, \rho_\mathcal{C}) \in \mathbb{C}_k(\mathcal{M})$ be a coupling structure for $\mathcal{M} = (S, \tau, \rho, \ell)$. Define $g \colon S^{k+1} \times S^{k+1} \to [0, 1]$, for $x, y \in S^{k+1}$, as

$$g(x, y) = \begin{cases} 0 & \text{if } \tau_\mathcal{C}(x_0, y_0)(x, y) = 0 \\ \mathbb{P}^\mathcal{C}_{x_0,y_0}(\not\equiv_{\ell^\omega} | \{(\pi_1, \pi_2)[0..k] = (x, y)\}) & \text{otherwise} \end{cases} \tag{5}$$

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 $\mathbb{P}^\mathcal{C}_{s,s'}(\not\equiv_{\ell^\omega}) = \int \gamma^\mathcal{C} \, \mathrm{d}\tau_\mathcal{C}(s, s')$ it suffices to show that $g = \gamma^\mathcal{C}$. Indeed

$$\begin{aligned} \mathbb{P}^\mathcal{C}_{s,s'}(\not\equiv_{\ell^\omega}) &= \int \mathbb{P}^\mathcal{C}_{s,s'}(\not\equiv_{\ell^\omega} | \{(\pi_1, \pi_2)[0..k] = (\cdot, \cdot)\}) \, \mathrm{d}\mathbb{P}^\mathcal{C}_{s,s'}[(\cdot, \cdot)[0..k]] && \text{(cond. pr.)} \\ &= \int \mathbb{P}^\mathcal{C}_{s,s'}(\not\equiv_{\ell^\omega} | \{(\pi_1, \pi_2)[0..k] = (\cdot, \cdot)\}) \, \mathrm{d}\tau_\mathcal{C}(s, s') && \text{(def. } \mathbb{P}^\mathcal{C}) \\ &= \int g \, \mathrm{d}\tau_\mathcal{C}(s, s') = \int \gamma^\mathcal{C} \, \mathrm{d}\tau_\mathcal{C}(s, s')\,. && \text{(by (5) and } g = \gamma^\mathcal{C}) \end{aligned}$$

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 (5), $\Gamma^\mathcal{C}(g)(x, y) = 0 = g(x, y)$.
Case $\tau_\mathcal{C}(x_0, y_0)(x, y) > 0$ and $\exists i \leq k.\, x_i \not\equiv_\ell y_i$. The following hold

$$\begin{aligned} g(x, y) &= \mathbb{P}^\mathcal{C}_{x_0,y_0}(\not\equiv_{\ell^\omega} | \{(\pi_1, \pi_2)[0..k] = (x, y)\}) && \text{(by (5))} \\ &= \frac{\mathbb{P}^\mathcal{C}_{x_0,y_0}(\not\equiv_{\ell^\omega} \cap \{(\pi_1, \pi_2)[0..k] = (x, y)\})}{\mathbb{P}^\mathcal{C}_{x_0,y_0}(\{(\pi_1, \pi_2)[0..k] = (x, y)\})} && \text{(cond. pr.)} \\ &= \frac{\mathbb{P}^\mathcal{C}_{x_0,y_0}(\{(\pi_1, \pi_2)[0..k] = (x, y)\})}{\mathbb{P}^\mathcal{C}_{x_0,y_0}(\{(\pi_1, \pi_2)[0..k] = (x, y)\})} = 1 = \Gamma^\mathcal{C}(g)(x, y)\,, \end{aligned}$$

where the last equalities follow by $\{(\pi_1, \pi_2)[0..k] = (x, y)\} \subseteq \not\equiv_{\ell^\omega}$ (because by hypothesis $\exists i.\, x_i \not\equiv_\ell y_i$) 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) $\mathbb{P}^{\mathcal{C}}_{x_0,y_0}(\not\equiv_{\ell^\omega} \cap B|A) = \beta$;
(ii) $\mathbb{P}^{\mathcal{C}}_{x_0,y_0}(\not\equiv_{\ell^\omega} \cap B^c|A) = (1-\beta) \cdot \int g \, \mathrm{d}\tau_{\mathcal{C}}(x_k, y_k)$.
Note that once we have shown (i–ii), $g(x, y) = \Gamma^{\mathcal{C}}(g)(x, y)$ follows immediately:

$$\begin{aligned}
g(x,y) &= \mathbb{P}^{\mathcal{C}}_{x_0,y_0}(\not\equiv_{\ell^\omega} |A) && \text{(by (5))} \\
&= \mathbb{P}^{\mathcal{C}}_{x_0,y_0}(\not\equiv_{\ell^\omega} \cap B|A) + \mathbb{P}^{\mathcal{C}}_{x_0,y_0}(\not\equiv_{\ell^\omega} \cap B^c|A) && \text{(by additivity)} \\
&= \beta + (1-\beta) \cdot \int g \, \mathrm{d}\tau_{\mathcal{C}}(x_k, y_k) && \text{(by (i) and (ii))} \\
&= \Gamma^{\mathcal{C}}(g)(x,y)\,. && \text{(by def. } \Gamma^{\mathcal{C}})
\end{aligned}$$

We show (i):

$$\begin{aligned}
\mathbb{P}^{\mathcal{C}}_{x_0,y_0}(\not\equiv_{\ell^\omega} \cap B|A) &= \mathbb{P}^{\mathcal{C}}_{x_0,y_0}(B|A) && \text{(by } B \subseteq \not\equiv_{\ell^\omega}) \\
&= \rho_{\mathcal{C}}(x_0..x_{k-1}, y_0..y_{k-1})(\neq) && \text{(by def. } \mathbb{P}^{\mathcal{C}}) \\
&= \beta && \text{(by def. } \beta)
\end{aligned}$$

We show (ii):

$$\begin{aligned}
&\mathbb{P}^{\mathcal{C}}_{x_0,y_0}(\not\equiv_{\ell^\omega} \cap B^c|A) = \\
&= \frac{\mathbb{P}^{\mathcal{C}}_{x_0,y_0}(\not\equiv_{\ell^\omega} \cap B^c \cap A)}{\mathbb{P}^{\mathcal{C}}_{x_0,y_0}(A)} && \text{(by cond. pr.)} \\
&= \frac{\tau_{\mathcal{C}}(x_0,y_0)(x,y) \cdot \rho_{\mathcal{C}}(x_0..x_{k-1}, y_0..y_{k-1})(=) \cdot \mathbb{P}^{\mathcal{C}}_{x_k,y_k}(\not\equiv_{\ell^\omega})}{\tau_{\mathcal{C}}(x_0,y_0)(x,y)} && \text{(by def. } \mathbb{P}^{\mathcal{C}}) \\
&= (1-\beta) \cdot \mathbb{P}^{\mathcal{C}}_{x_k,y_k}(\not\equiv_{\ell^\omega}) && \text{(by def. } \beta \text{ and compl.)} \\
&= (1-\beta) \cdot \int \mathbb{P}^{\mathcal{C}}_{x_k,y_k}(\not\equiv_{\ell^\omega} | \{(\pi_1,\pi_2)[0..k] = (\cdot,\cdot)\}) \, \mathrm{d}\mathbb{P}^{\mathcal{C}}_{x_k,y_k}[(\cdot,\cdot)[0..k]] \\
& && \text{(cond. pr.)} \\
&= (1-\beta) \cdot \int \mathbb{P}^{\mathcal{C}}_{x_k,y_k}(\not\equiv_{\ell^\omega} | \{(\pi_1,\pi_2)[0..k] = (\cdot\,,\cdot)\}) \, \mathrm{d}\tau_{\mathcal{C}}(x_k, y_k) && \text{(def. } \mathbb{P}^{\mathcal{C}}) \\
&= (1-\beta) \cdot \int g \, \mathrm{d}\tau_{\mathcal{C}}(x_k, y_k)\,. && \text{(by (5))}
\end{aligned}$$

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

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

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

$$
\begin{aligned}
m &= g(x, y) - \gamma^{\mathcal{C}}(x, y) && \text{(by } x\ M\ y\text{)} \\
&= \Gamma^{\mathcal{C}}(g)(x, y) - \Gamma^{\mathcal{C}}(\gamma^{\mathcal{C}})(x, y) && (g \text{ and } \gamma^{\mathcal{C}} \text{ fixed points}) \\
&= (1 - \beta) \cdot \textstyle\int (g - \gamma^{\mathcal{C}}) \, \mathrm{d}\tau_{\mathcal{C}}(x_k, y_k) && \text{(by def. } \Gamma^{\mathcal{C}}\text{)} \\
&= (1 - \beta) \cdot \textstyle\sum_{u,v \in S^{k+1}} \big(g(u, v) - \gamma^{\mathcal{C}}(u, v)\big) \cdot \tau_{\mathcal{C}}(x_k, y_k)(u, v) \, . && (6)
\end{aligned}
$$

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

$$
(1 - \beta) \cdot \textstyle\sum_{u,v \in S^{k+1}} \big(g(u, v) - \gamma^{\mathcal{C}}(u, v)\big) \cdot \tau_{\mathcal{C}}(x_k, y_k)(u, v) \leq (1 - \beta)m \, . \quad (7)
$$

We distinguish two cases:

– if $\beta > 0$, then $1 - \beta < 1$. By (6) and (7) we have that $m \leq (1 - \beta)m$. By the assumption on β this holds only for $m = 0$;
– if $\beta = 0$, by (7) and (7) 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

$$
g(x, y) \overset{(5)}{=} \mathbb{P}^{\mathcal{C}}_{x_0,y_0}(\not\equiv_{\ell^\omega} | \, \{\pi_1[0..k] = x, \pi_2[0..k] = y\}) = 0 \, .
$$

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

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

This proves the thesis. □

Proof (of Lemma 31). Let $\mathcal{C} = (\tau_{\mathcal{C}}, \rho_{\mathcal{C}}) \in \mathbb{C}_k(\mathcal{M})$ 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 $\mathcal{D} = \mathcal{C}[(u_k, v_k)/\mu]\langle(u_0..u_{k-1}, v_1..v_{k-1})/\nu\rangle$ 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

$$
\begin{aligned}
\alpha &= \textstyle\int \gamma^{\mathcal{C}} \, \mathrm{d}\mu & \alpha' &= \textstyle\int \gamma^{\mathcal{C}} \, \mathrm{d}\tau_{\mathcal{C}}(u_k, v_k) \, , \\
\beta &= \nu(\neq) & \beta' &= \rho_{\mathcal{C}}(u_0..u_{k-1}, v_1..v_{k-1})(\neq) \, .
\end{aligned}
$$

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 $\gamma^{\mathcal{D}} < \gamma^{\mathcal{C}}$. □

Proof (of Lemma 32). By contradiction. Assume $\delta{\downarrow_k}(u,v) \neq \int \gamma^{\mathcal{C}} \, \mathrm{d}\tau_{\mathcal{C}}(u,v)$ 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 $\int \gamma^{\mathcal{C}} \, \mathrm{d}\mu \geq \int \gamma^{\mathcal{C}} \, \mathrm{d}\tau_{\mathcal{C}}(u', v')$. By hypothesis and Lemma 31, we have that $\int \gamma^{\mathcal{C}} \, \mathrm{d}\tau_{\mathcal{C}}(u,v) = \min\left\{\int \gamma^{\mathcal{D}} \, \mathrm{d}\tau_{\mathcal{D}}(u,v) \mid \mathcal{D} \in \mathbb{C}_{2^k}(\mathcal{M})\right\}$. But at the same time

$$
\begin{aligned}
\delta{\downarrow_k}(u,v) &= \min\left\{\mathbb{P}^{\mathcal{D}}_{u,v}(\not\equiv_{\ell^\omega}) \mid \mathcal{D} \in \mathbb{C}_{2^k}(\mathcal{M})\right\} && (\text{by } (2)) \\
&= \min\left\{\int \gamma^{\mathcal{D}} \, \mathrm{d}\tau_{\mathcal{D}}(u,v) \mid \mathcal{D} \in \mathbb{C}_{2^k}(\mathcal{M})\right\} . && (\text{by Lemma 29})
\end{aligned}
$$

This contradicts hypothesis that $\int \gamma^{\mathcal{C}} \, \mathrm{d}\tau_{\mathcal{C}}(u,v) \neq \delta{\downarrow_k}(u,v)$. □

B Folklore Results about Metric Spaces

Proposition 34. *Let $A \subseteq \mathbb{R}$ be a bounded nonempty set. Then,*

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

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

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

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

Let $\alpha = \sup A$. (i) We prove that $\alpha \in \overline{A}$. Let $\varepsilon > 0$, then $\alpha - \varepsilon$ is not an upper bound for A. This means that there exists $x \in A$ such that $\alpha - \varepsilon < x \leq \alpha$ and, in particular, that $x \in (\alpha - \varepsilon, \alpha + \varepsilon) \cap A$. Therefore $\alpha \in \overline{A}$. (ii) Let $\beta = \sup \overline{A}$.

By $A \subseteq \overline{A} = \overline{\overline{A}}$ and (i), we have $\alpha \leq \beta \in \overline{A}$. We prove that $\alpha = \beta$. Assume by contradiction that $\alpha \neq \beta$ and let $\varepsilon := \beta - \alpha$. Clearly $\varepsilon > 0$, so that, by $\beta \in \overline{A}$, we have that $(\beta - \varepsilon, \beta + \varepsilon) \cap A \neq \emptyset$. This means that there exists $x \in A$ such that $\alpha = \beta - \varepsilon < x$, in contradiction with the hypothesis that $\alpha = \sup A$. □

Proposition 35. *Let $f \colon X \to Y$ be continuous and $A \subseteq X$, then $\overline{f(A)} = \overline{f(\overline{A})}$.*

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

Proposition 36. *Let X be nonempty, $f \colon X \to \mathbb{R}$ be a bounded continuous real-valued function, and $D \subseteq X$ be dense in X. Then $\sup f(D) = \sup f(X)$.*

Proof. Notice that, since $X \neq \emptyset$ and f is bounded, by Dedekind axiom, both $\sup f(D)$ and $\sup f(X)$ exist. By Propositions 34, 35, and $\overline{D} = X$, we have

$$\sup f(D) \overset{\text{(Prop.34)}}{=} \sup \overline{f(D)} \overset{\text{(Prop.35)}}{=} \sup \overline{f(\overline{D})} = \sup \overline{f(X)} \overset{\text{(Prop.34)}}{=} \sup f(X)\,,$$

which proves the thesis. □

Proposition 37.

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

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

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

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

(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

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

which holds since all the pseudometrics d_i are positive. □