A behavioural pseudometric for continuous-time Markov processes

Linan Chen (McGill University) *
Florence Clerc (Heriot-Watt University)†
Prakash Panangaden (McGill University)‡

Abstract

In this work, we generalize the concept of bisimulation metric in order to metrize the behaviour of continuous-time processes. Similarly to what is done for discrete-time systems, we follow two approaches and show that they coincide: as a fixpoint of a functional and through a real-valued logic.

The whole discrete-time approach relies entirely on the step-based dynamics: the process jumps from state to state. We define a behavioural pseudometric for processes that evolve continuously through time, such as Brownian motion or involve jumps or both.

1 Introduction

Bisimulation is a concept that captures behavioural equivalence of states in a variety of types of transition systems. It has been widely studied in a discrete-time setting where the notion of a step is fundamental. An important and especially useful further notion is that of bisimulation metric which quantifies "how similar two states are".

Most of the theoretical work that exists is on discrete time but a growing part of what computer science allows us to do is in real-time: robotics, self-driving cars, online machine-learning etc. A common solution is to discretize

*Linan Chen was supported by an NSERC discovery grant

†Florence Clerc was supported by the EPSRC-UKRI grant EP/Y000455/1, by a CREATE grant for the project INTER-MATH-AI, by NSERC and by IVADO through the DEEL Project CRDPJ 537462-18

‡Prakash Panangaden was supported by an NSERC discovery grant

time, however it is well-known that this can lead to errors that are hopefully small but that may accumulate over time and lead to vastly different outcomes. For that reason, it is important to have a continuous-time way of quantifying the error made.

Bisimulation [23, 26, 29] is a fundamental concept in the theory of transition systems capturing a strong notion of behavioural equivalence. The extension to probabilistic systems is due to Larsen and Skou [22]; henceforth we will simply say "bisimulation" instead of "probabilistic bisimulation". Bisimulation has been studied for discrete-time systems where transitions happen as steps, both on discrete [22] and continuous state spaces [4, 13, 14]. In all these types of systems, a crucial ingredient of the definition of bisimulation is the ability to talk about *the next step*. This notion of bisimulation is characterized by a modal logic [22] even when the state space is continuous [13].

Some work had previously been done in what are called continuous-time systems, see for example [2], but even in so-called continuous-time Markov chains there is a discrete notion of time *step*; it is only that there is a real-valued duration associated with each state that leads to people calling such systems continuous-time. They are often called "jump processes" in the mathematical literature (see, for example, [27, 32]), a phrase that better captures the true nature of such processes. Metrics and equivalences for such processes were studied by Gupta et al. [19, 20].

The processes we consider have continuous state spaces and are governed by a continuous-time evolution, a paradigmatic example is Brownian motion. When approximating such processes by discrete-time processes, entirely new phenomena and difficulties manifest themselves in this procedure. For example, even the basic properties of trajectories of Brownian motion are vastly more complicated than the counterparts of a random walk. Basic concepts like "the time at which a process exits a given subset of the state space" becomes intricate to define. Notions like "matching transition steps" are no longer applicable as the notion of "step" does not make sense.

In [9, 10, 11], we proposed different notions of behavioural equivalences on continuous-time processes. We showed that there were several possible extensions of the notion of bisimulation to continuous time and that the continuous-time notions needed to involve trajectories in order to be meaningful. There were significant mathematical challenges in even proving that an equivalence relation existed. For example, obstacles occurred in establishing measurability of various functions and sets, due to the inability to countably generate the relevant σ-algebras. Those papers left completely open the question of defining a suitable pseudometric analogue, a concept that would be more useful in practice than an equivalence relation.

Previous work on discrete-time Markov processes by Desharnais et al. [15, 17] extended the modal logic characterizing bisimulation to a real-valued logic that allowed to not only state if two states were "behaviourally equivalent" but, more interestingly, how similarly they behaved. This shifts the notion from a qualitative notion (an equivalence) to a quantitative one (a pseudometric).

Other work also on discrete-time Markov processes by van Breugel et al. [30] introduced a slightly different real-valued logic and compared the corresponding pseudometric to another pseudometric obtained as a terminal coalgebra of a carefully crafted functor. We also mention in this connexion the work by Ferns et al. on Markov Decision Processes and the connexion between bisimulation and optimal value functions [18].

In this work, we are looking to extend the notion of bisimulation metric to a behavioural pseudometric on continuous-time processes. Very broadly speaking, we are following a familiar path from equivalences to logics to metrics. However, it is necessary for us to redevelop the framework and the mathematical techniques from scratch. Indeed, a very important aspect in discrete-time is the fact that the process is a jump process, "hopping" from state to state. This limitation also applies to continuous-time Markov chains. In our case, we want to cover processes that evolve through time. A standard example would be Brownian motion or other diffusion processes (often described by stochastic differential equations). As one will see throughout this work, there are new mathematical challenges that need to be overcome. This means that the similarity between the pre-existing work on discrete-time and our generalization to continuous-time is only at the highest level of abstraction.

Outline of the paper: TODO change

The first two sections after the introduction are background. We will start by recalling some mathematical notions in Section 2, introducing the continuous-time processes that we will be studying in Section 3. A very brief overview of bisimulation metrics in discrete time can also be found in Appendix 4. In Section 5, we will introduce a functional $\mathcal{F}$ and define a pseudometric $\overline{\delta}$ using this functional. We will also show that the pseudometric $\overline{\delta}$ is a fixpoint of $\mathcal{F}$. In Section 6, we will show that this pseudometric is characterized by a real-valued logic. We will further emphasize the novelty of this work wrt discrete time and summarize the obstacles that we had to overcome in Section 7. We will provide some examples in Section 8. Finally we will discuss the limitations of our approach and how it relates to previous works in Section 10.

2 Mathematical background

We assume the reader to be familiar with basic measure theory and topology. Nevertheless we provide a brief review of the relevant notions and theorems. Let us start with clarifying a few notations on integrals: Given a measurable space X equipped with a measure μ and a measurable function $f : X \longrightarrow \mathbb{R}$, we can write either $\int f \, \mathrm{d}\mu$ or $\int f(x) \, \mu(\mathrm{d}x)$ interchangeably. The second notation will be especially useful when considering a Markov kernel P_t for some $t \geq 0$ and $x \in X$: $\int f(y) \, P_t(x, \mathrm{d}y) = \int f \, \mathrm{d}P_t(x)$.

All the proofs for this Section can be found in Appendix A.

2.1 Lower semi-continuity

Definition 2.1. Given a topological space X, a function $f : X \longrightarrow \mathbb{R}$ is *lower semi-continuous* if for every $x_0 \in X$, $\liminf_{x \to x_0} f(x) \geq f(x_0)$. This condition is equivalent to the following one: for any $y \in \mathbb{R}$, the set $f^{-1}((y, +\infty)) = \{x \mid f(x) > y\}$ is open in X.

Let X be a metric space. A function $f : X \longrightarrow \mathbb{R}$ is lower semi-continuous, if and only if f is the limit of an increasing sequence of real-valued continuous functions on X. Details can be found in the Appendix A.1.

2.2 Couplings

Definition 2.2. Let (X, Σ_X, P) and (Y, Σ_Y, Q) be two probability spaces. Then a *coupling* γ of P and Q is a probability distribution on $(X \times Y, \Sigma_X \otimes \Sigma_Y)$ such that for every $B_X \in \Sigma_X$, $\gamma(B_X \times Y) = P(B_X)$ and for every $B_Y \in \Sigma_Y$, $\gamma(X \times B_Y) = Q(B_Y)$ (P, Q are called the *marginals* of γ). We write $\Gamma(P, Q)$ for the set of couplings of P and Q.

Lemma 2.3. Given two probability measures P and Q on Polish spaces X and Y respectively, the set of couplings $\Gamma(P, Q)$ is compact under the topology of weak convergence.

2.3 Optimal transport theory

A lot of this work is based on optimal transport theory. This whole subsection is based on [31] and will be adapted to our framework.

Consider a Polish space $\mathcal{X}$ and a lower semi-continuous cost function $c : \mathcal{X} \times \mathcal{X} \longrightarrow [0, 1]$ such that for every $x \in \mathcal{X}$, $c(x, x) = 0$.

For every two probability distributions μ and ν on $\mathcal{X}$, we write $W(c)(\mu,\nu)$ for the optimal transport cost from μ to ν. Adapting Theorem 5.10(iii) of [31] to our framework (see Remark A.8 in Appendix A.3), we get the following statement for the Kantorovich duality:

$$W(c)(\mu,\nu) = \min_{\gamma\in\Gamma(\mu,\nu)} \int c \, \mathrm{d}\gamma = \max_{h\in\mathcal{H}(c)} \left| \int h \, \mathrm{d}\mu - \int h \, \mathrm{d}\nu \right|$$

where $\mathcal{H}(c) = \{h : \mathcal{X} \longrightarrow [0,1] \mid \forall x, y \ \ |h(x) - h(y)| \leq c(x,y)\}$.

Lemma 2.4. If the cost function c is a 1-bounded pseudometric on $\mathcal{X}$, then $W(c)$ is a 1-bounded pseudometric on the space of probability distributions on $\mathcal{X}$.

We will later need the following technical lemma. Theorem 5.20 of [31] states that a sequence $W(c_k)(P_k, Q_k)$ converges to $W(c)(P,Q)$ if c_k uniformly converges to c and P_k and Q_k converge weakly to P and Q respectively. Uniform convergence in the cost function may be too strong a condition for us, but the following lemma is enough for what we need.

Lemma 2.5. Consider a Polish space $\mathcal{X}$ and a cost function $c : \mathcal{X} \times \mathcal{X} \longrightarrow [0,1]$ such that there exists an increasing ($c_{k+1} \geq c_k$ for every k) sequence of continuous cost functions $c_k : \mathcal{X}\times\mathcal{X} \longrightarrow [0,1]$ that converges to c pointwise. Then, given two probability distributions P and Q on $\mathcal{X}$,

$$\lim_{k\to\infty} W(c_k)(P,Q) = W(c)(P,Q).$$

3 Background on continuous-time Markov processes

This work focuses on continuous-time processes that are honest (without loss of mass over time) and with additional regularity conditions. In order to define what we mean by continuous-time Markov processes here, we first define Feller-Dynkin processes. Much of this material is adapted from [27] and we use their notations. Another useful source is [5].

Let E be a locally compact, Hausdorff space with a countable base. We also equip the set E with its Borel σ-algebra $\mathcal{B}(E)$, denoted $\mathcal{E}$. The previous topological hypotheses also imply that E is σ-compact and Polish (see corollary IX.57 in [7]). We will denote Δ for the 1-bounded metric that generates the topology making E Polish.

Definition 3.1. A *semigroup* of operators on any Banach space X is a family of linear continuous (bounded) operators $\mathcal{P}_t : X \longrightarrow X$ indexed by $t \in \mathbb{R}_{\geq 0}$ such that

$$\forall s, t \geq 0, \mathcal{P}_s \circ \mathcal{P}_t = \mathcal{P}_{s+t} \qquad \text{(semigroup property)}$$

and

$$\mathcal{P}_0 = I \qquad \text{(the identity).}$$

Definition 3.2. For X a Banach space, we say that a semigroup $\mathcal{P}_t : X \longrightarrow X$ is *strongly continuous* if

$$\forall x \in X, \lim_{t \downarrow 0} \|\mathcal{P}_t x - x\| \longrightarrow 0.$$

What the semigroup property expresses is that we do not need to understand the past (what happens before time t) in order to compute the future (what happens after some additional time s, so at time $t+s$) as long as we know the present (at time t).

We say that a continuous real-valued function f on E "vanishes at infinity" if for every $\varepsilon > 0$ there is a compact subset $K \subseteq E$ such that for every $x \in E \setminus K$, we have $|f(x)| \leq \varepsilon$. To give an intuition, if E is the real line, this means that $\lim_{x \to \pm\infty} f(x) = 0$. The space $C_0(E)$ of continuous real-valued functions that vanish at infinity is a Banach space with the sup norm.

Definition 3.3. A *Feller-Dynkin (FD) semigroup* is a strongly continuous semigroup $(\hat{P}_t)_{t \geq 0}$ of linear operators on $C_0(E)$ satisfying the additional condition:

$$\forall t \geq 0 \quad \forall f \in C_0(E), \text{ if } \quad 0 \leq f \leq 1, \text{ then } \quad 0 \leq \hat{P}_t f \leq 1$$

The Riesz representation theorem can be found as Theorem II.80.3 of [27]. From it, we can derive the following important proposition which relates these FD-semigroups with Markov kernels (see Appendix B for the details). This allows one to see the connection with familiar probabilistic transition systems.

Proposition 3.4. Given an FD-semigroup $(\hat{P}_t)_{t \geq 0}$ on $C_0(E)$, it is possible to define a unique family of sub-Markov kernels $(P_t)_{t \geq 0} : E \times \mathcal{E} \longrightarrow [0, 1]$ such that for all $t \geq 0$ and $f \in C_0(E)$,

$$\hat{P}_t f(x) = \int f(y) P_t(x, \mathrm{d}y).$$

Given a time t and a state x, we will often write $P_t(x)$ for the measure $P_t(x, \cdot)$ on E. Note that since E is Polish, then $P_t(x)$ is tight.

Definition 3.5. A process described by the FD-semigroup $(\hat{P}_t)_{t\geq 0}$ is *honest* if for every $x \in E$ and every time $t \geq 0$, $P_t(x, E) = 1$.

Worded differently, a process is honest if there is no loss of mass over time. A standard example of an honest process is Brownian motion. We refer the reader to Appendix C for an introduction to Brownian motion.

3.1 Observables

In previous sections, we defined Feller-Dynkin processes. In order to bring the processes more in line with the kind of transition systems that have hitherto been studied in the computer science literature, we also equip the state space E with an additional continuous function $obs : E \longrightarrow [0, 1]$. One should think of it as the interface between the process and the user (or an external observer): external observers won't see the exact state in which the process is at a given time, but they will see the associated *observables*. What could be a real-life example is the depth at which a diver goes: while the diver does not know precisely his location underwater, at least his watch is giving him the depth at which he is.

Note that the condition on the observable is a major difference from our previous work [9, 10] since we used a countable set of atomic propositions AP and obs was a discrete function $E \longrightarrow 2^{AP}$.

Definition 3.6. In this study, a *Continuous-time Markov process* (abbreviated CTMP) is an honest FD-semigroup on $C_0(E)$ equipped with a continuous function $obs : E \longrightarrow [0, 1]$ and that satisfies the following additional property: if a sequence $(x_n)_{n\in\mathbb{N}}$ converges to x in E, then for every t, the sequence of measures $(P_t(x_n))_{n\in\mathbb{N}}$ weakly converges to the measure $P_t(x)$.

Remark 3.7. Some properties could be relaxed. For instance, in some cases, a non honest process could be made into a CTMP by adding a state ∂. Another hypothesis that could be relaxed is the one on obs by imposing some stronger conditions on the FD-process.

4 Bisimulation metric for discrete-time Labelled Markov Processes

Let us recall the existing work on discrete-time Labelled Markov Processes adapted to our framework. For the detailed version, we refer the readers to the works by Desharnais et al. [16, 17, 25] and van Breugel et al. [30].

A labelled Markov Process is a family of Markov sub-kernels indexed by a set of what are called actions. Our work can easily be adapted to this framework by adding indices. For the sake of readability, we will not consider actions in this work and that is why we will only present the discrete-time work for a single Markov kernel.

Consider a Markov kernel τ on a Polish space $(E, \mathcal{E})$. Note that τ is a Markov kernel and in particular for every $x \in E$, $\tau(x, E) = 1$.

Define the functional F on 1-bounded pseudometrics on E by

$$F(m)(x, y) = cW(m)(\tau(x), \tau(y)),$$

where $0 < c < 1$ is a discount factor. This functional has a fixpoint which is a bisimulation metric and is denoted $d_{\mathcal{C}}$ in [30]. It is also characterized by the logic defined inductively:

$$\phi := 1 \mid \diamond \phi \mid \min(\phi, \phi) \mid 1 - \phi \mid \phi \ominus q,$$

which is interpreted as $(\diamond\phi)(x) = c \int \phi \, \mathrm{d}\tau(x)$ etc. Indeed, this logic defines a pseudometric on E by $d_{\mathcal{L}}(x, y) = \sup_\phi |\phi(x) - \phi(y)|$ and the two pseudometrics coincide: $d_{\mathcal{L}} = cd_{\mathcal{C}}$.

5 Generalizing to continuous-time through a functional

We start by defining a behavioural pseudometric on our CTMPs by defining a functional $\mathcal{F}$ on the lattice of 1-bounded pseudometrics. As we will see, unlike in the discrete-time case, it is not possible to apply the Banach fixpoint theorem and get a fixpoint metric a priori: instead we need to construct a candidate and then show that it is a fixpoint of our functional. More specifically, the idea is to iteratively apply our functional to a metric and then consider the supremum of the sequence of pseudometrics. Doing so requires to first restrict the scope of our functional $\mathcal{F}$.

5.1 Lattices

At the core of this construction is the definition of a functional on the lattice of 1-bounded pseudometrics.

Let $\mathcal{M}$ be the lattice of 1-bounded pseudometrics on the state space E equipped with the order $\leq$ defined as: $m_1 \leq m_2$ if and only if for every (x, y), $m_1(x, y) \leq m_2(x, y)$. We can define a sublattice $\mathcal{P}$ of $\mathcal{M}$ by restricting to pseudometrics that are lower semi-continuous (wrt the original topology

$\mathcal{O}$ on E generated by the metric Δ making the space E Polish). We will further require to define the sublattice $\mathcal{C}$ which is the set of pseudometrics $m \in \mathcal{M}$ on the state space E such that the topology generated by m on E is a subtopology of the original topology $\mathcal{O}$, *i.e.* m is a continuous function $E \times E \longrightarrow [0, 1]$.

We have the following inclusion: $\mathcal{C} \subset \mathcal{P} \subset \mathcal{M}$.

Remark 5.1. One has to be careful here. The topology $\mathcal{O}$ on E is generated by the 1-bounded metric Δ, and hence Δ is in $\mathcal{C}$. However, we can define many pseudometrics that are not related to $\mathcal{O}$. As an example, the discrete pseudometric[1] on the real line is not related to the usual topology on $\mathbb{R}$.

5.2 Defining our functional

Throughout the rest of the paper, $(P_t)_{t>=0}$ is the family of Markov kernels associated with a CTMP. Given a discount factor $0 < c < 1$, we define the functional $\mathcal{F}_c : \mathcal{P} \longrightarrow \mathcal{M}$ as follows: for every pseudometric $m \in \mathcal{P}$ and every two states x, y,

$$\mathcal{F}_c(m)(x, y) = \sup_{t \geq 0} c^t W(m)(P_t(x), P_t(y)).$$

$\mathcal{F}_c(m)(x, y)$ compares all the distributions $P_t(x)$ and $P_t(y)$ through transport theory and takes their supremum.

There are several remarks to make on this definition. First, we can only define $\mathcal{F}_c(m)$ if m is lower semi-continuous since we are using optimal transport theory which is why the domain of $\mathcal{F}_c$ is only $\mathcal{P}$.

Additionally, even if m is lower semi-continuous, $\mathcal{F}_c(m)$ may not even be measurable which means that the range of $\mathcal{F}_c$ is not the lattice $\mathcal{P}$. At least, Lemma 2.4 ensures that $\mathcal{F}_c(m)$ is indeed in $\mathcal{M}$, as a supremum of pseudometrics. This subtlety was not present in the work on continuous-time Markov chains in [19, 20].

Second, we will use the Kantorovich duality throughout this work. It only holds for probability measures, and that is why we restrict this work to honest processes.

As a direct consequence of the definition of $\mathcal{F}_c$, we have that $\mathcal{F}_c$ is monotone: if $m_1 \leq m_2$ in $\mathcal{P}$, then $\mathcal{F}_c(m_1) \leq \mathcal{F}_c(m_2)$.

Lemma 5.2. For every pseudometric m in $\mathcal{P}$, discount factor $0 < c < 1$ and pair of states x, y,

$$m(x, y) \leq \mathcal{F}_c(m)(x, y).$$

[1]The discrete pseudometric is defined as $m(x, y) = 1$ if $x \neq y$ and $m(x, x) = 0$

Proof. Consider a pair of states x, y. Then

$$\begin{aligned}\mathcal{F}_c(m)(x,y) &= \sup_{t\geq 0} c^t W(m)(P_t(x), P_t(y)) \\ &\geq W(m)(P_0(x), P_0(y)) \\ &= \inf_{\gamma\in\Gamma(P_0(x),P_0(y))} \int m \, \mathrm{d}\gamma.\end{aligned}$$

Since $P_0(x)$ is the dirac distribution at x and similarly for $P_0(y)$, the only coupling γ between $P_0(x)$ and $P_0(y)$ is the product measure $P_0(x) \times P_0(y)$ and thus $W(m)(P_0(x), P_0(y)) = m(x, y)$, which concludes the proof. ∎

5.3 When restricted to continuous pseudometrics

We wish to iteratively apply $\mathcal{F}_c$ in order to construct a fixpoint (in a similar fashion to the proof of the Knaster-Tarski theorem). While $\mathcal{F}_c(m)$ is a pseudometric (for $m \in \mathcal{P}$), there is no reason for it to be in $\mathcal{P}$. This means that we cannot hastily apply $\mathcal{F}_c$ iteratively to just any pseudometric in order to obtain a fixpoint.

However, if m is a pseudometric which is continuous wrt the original topology, then so is $\mathcal{F}_c(m)$.

Lemma 5.3. Consider a pseudometric $m \in \mathcal{C}$. Then the topology generated by $\mathcal{F}_c(m)$ is a subtopology of the original topology $\mathcal{O}$ for any discount factor $0 < c < 1$.

This is where we need that the discount factor $c < 1$. The condition that $c < 1$ enables us to maintain continuity by allowing to bound the time interval we consider. Indeed, given $T > 0$, for any time $t \geq T$ and any $x, y \in E$, we know that $c^t W(m)(P_t(x), P_t(y)) \leq c^T$.

Proof. Since c and m are fixed throughout the proof, we will omit noting them and for instance write $\mathcal{F}(x, y)$ instead of $\mathcal{F}_c(m)(x, y)$ and $W(P_t(x), P_t(y))$ instead of $W(m)(P_t(x), P_t(y))$. We will also write $\Phi(t, x, y) = c^t W(m)(P_t(x), P_t(y))$, i.e. $\mathcal{F}(x, y) = \sup_t \Phi(t, x, y)$.

It is enough to show that for a fixed state x, the map $y \mapsto \mathcal{F}(x, y)$ is continuous.

Pick $\epsilon > 0$ and a sequence of states $(y_n)_{n\in\mathbb{N}}$ converging to y. We want to show that there exists M such that for all $n \geq M$,

$$|\mathcal{F}(x, y) - \mathcal{F}(x, y_n)| \leq \epsilon. \quad (1)$$

Pick t such that $\mathcal{F}(x,y) = \sup_s \Phi(s,x,y) \leq \Phi(t,x,y) + \epsilon/4$, i.e.

$$|\Phi(t,x,y) - \mathcal{F}(x,y)| \leq \epsilon/4. \tag{2}$$

Recall that $P_t(y_n)$ converges weakly to $P_t(y)$ and hence we can apply Theorem 5.20 of [31] and get:

$$\lim_{n\to\infty} W(P_t(x), P_t(y_n)) = W(P_t(x), P_t(y)).$$

This means that there exists N' such that for all $n \geq N'$,

$$|W(P_t(x), P_t(y_n)) - W(P_t(x), P_t(y))| \leq \epsilon/4.$$

This further implies that for all $n \geq N'$,

$$|\Phi(t,x,y_n) - \Phi(t,x,y)| \leq c^t\epsilon/4 \leq \epsilon/4. \tag{3}$$

In order to show (1), it is enough to show that there exists N such that for every $n \geq N$,

$$|\Phi(t,x,y_n) - \mathcal{F}(x,y_n)| \leq \epsilon/2. \tag{4}$$

Indeed, in that case, $\forall n \geq \max\{N, N'\}$, using Equations (2) and (3),

$$\begin{aligned}
&|\mathcal{F}(x,y) - \mathcal{F}(x,y_n)| \\
&\leq |\mathcal{F}(x,y) - \Phi(t,x,y)| + |\Phi(t,x,y) - \Phi(t,x,y_n)| + |\Phi(t,x,y_n) - \mathcal{F}(x,y_n)| \\
&\leq \epsilon/4 + \epsilon/4 + \epsilon/2 = \epsilon.
\end{aligned}$$

So let us show (4). Assume it is not the case: for all N, there exists $n \geq N$ such that $|\Phi(t,x,y_n) - \mathcal{F}(x,y_n)| > \epsilon/2$, i.e.

$$\Phi(t,x,y_n) + \epsilon/2 < \mathcal{F}(x,y_n).$$

Define the sequence $(N_k)_{k\in\mathbb{N}}$ by: $N_{-1} = -1$ and if N_k is defined, define N_{k+1} to be the smallest $n \geq N_k + 1$ such that $\Phi(t,x,y_n) + \epsilon/2 < \mathcal{F}(x,y_n)$. In particular for every $k \in \mathbb{N}$, $\mathcal{F}(x,y_{N_k}) > \epsilon/2$. There exists T such that for every $s \geq T$, $c^s < \epsilon/2$. We thus have that

$$\forall k \in \mathbb{N} \quad \mathcal{F}(x,y_{N_k}) = \sup_{0\leq s\leq T} \Phi(s,x,y_{N_k}).$$

Therefore for every $k \in \mathbb{N}$, there exists $s_k \in [0,T]$ such that

$$\mathcal{F}(x,y_{N_k}) \leq \Phi(s_k,x,y_{N_k}) + \epsilon/8. \tag{5}$$

We get a sequence $(s_k)_{k\in\mathbb{N}} \subset [0,T]$, and there is thus a subsequence $(t_k)_{k\in\mathbb{N}}$ converging to some $t' \in [0,T]$. There is a corresponding subsequence $(z_k)_{k\in\mathbb{N}}$ of the original sequence $(y_{N_k})_{k\in\mathbb{N}}$. Since $\lim_{n\to\infty} y_n = y$, $\lim_{k\to\infty} z_k = y$.

We constructed the sequence $(N_k)_{k\in\mathbb{N}}$ such that $\Phi(t,x,y_{N_k}) + \epsilon/2 < \mathcal{F}(x, y_{N_k})$. Hence by Equation (5),

$$\Phi(t,x,z_k) + \epsilon/2 < \mathcal{F}(x,y_{N_k}) \leq \Phi(t_k,x,z_k) + \epsilon/8,$$

which means that by taking the limit $k \longrightarrow \infty$,

$$\Phi(t,x,y) + \epsilon/2 \leq \Phi(t',x,y) + \epsilon/8. \tag{6}$$

At the start of this proof, we picked t such that $\mathcal{F}(x,y) = \sup_s \Phi(s,x,y) \leq \Phi(t,x,y) + \epsilon/4$ which means that

$$\Phi(t',x,y) \leq \Phi(t,x,y) + \epsilon/4. \tag{7}$$

Equations (6) and (7) are incompatible which concludes the proof. ∎

5.4 Defining our family of pseudometrics

We are now finally able to iteratively apply our functional $\mathcal{F}_c$ on continuous pseudometrics and thus construct a sequence of increasing pseudometrics and its limit.

Since *obs* is a continuous function $E \longrightarrow \mathbb{R}_{\geq 0}$ and by Lemma 5.3, we can define the sequence of pseudometrics in $\mathcal{C}$ for each $0 < c < 1$:

$$\begin{aligned} \delta_0^c(x,y) &= |obs(x) - obs(y)|, \\ \delta_{n+1}^c &= \mathcal{F}_c(\delta_n^c). \end{aligned}$$

By Lemma 5.2, for every two states x and y, $\delta_{n+1}^c(x,y) \geq \delta_n^c(x,y)$. Define the pseudometric $\overline{\delta}^c = \sup_n \delta_n^c$ (which is also a limit since the sequence is non-decreasing).

As a direct consequence of Lemma A.1, the pseudometric $\overline{\delta}^c$ is lower semi-continuous and is thus in the lattice $\mathcal{P}$ for any $0 < c < 1$.

Remark 5.4. Note that the lattice $\mathcal{C}$ is not complete which means that, although the metrics δ_n^c all belong to $\mathcal{C}$, $\overline{\delta}^c$ does not need to be in $\mathcal{C}$. For that reason, we cannot directly use the Knaster-Tarski theorem in this work.

5.5 Fixpoint

Even though we are not able to define the metric $\overline{\delta}^c$ as a fixpoint directly, it is actually a fixpoint.

Theorem 5.5. The pseudometric $\overline{\delta}^c$ is a fixpoint for $\mathcal{F}_c$.

Proof. We will omit writing c as an index for the pseudometrics δ_n and $\overline{\delta}$ throughout this proof. Fix two states x, y and a time t.

The space of finite measures on $E \times E$ is a linear topological space. Using Lemma 2.3, we know that the set of couplings $\Gamma(P_t(x), P_t(y))$ is a compact subset. It is also convex.

The space of bounded pseudometrics on E is also a linear topological space. We have defined a sequence $\delta_0, \delta_1, ...$ on that space. We define Y to be the set of linear combinations of pseudometrics $\sum_{n\in\mathbb{N}} a_n\delta_n$ such that for every n, $a_n \geq 0$ (and finitely many are non-zero) and $\sum_{n\in\mathbb{N}} a_n = 1$. This set Y is convex.

Define the function

$$\begin{aligned}\Xi : \Gamma(P_t(x), P_t(y)) \times Y &\longrightarrow [0, 1]\\ (\gamma, m) &\mapsto \int m \ \mathrm{d}\gamma.\end{aligned}$$

For $\gamma \in \Gamma(P_t(x), P_t(y))$, the map $\Xi(\gamma, \cdot)$ is continuous by the dominated convergence theorem and it is monotone and hence quasiconcave. For a given $m \in Y$, by definition of the Lebesgue integral, $\Xi(\cdot, m)$ is continuous and linear.

We can therefore apply Sion's minimax theorem:

$$\min_{\gamma\in\Gamma(P_t(x),P_t(y))} \sup_{m\in Y} \int m \ \mathrm{d}\gamma = \sup_{m\in Y} \min_{\gamma\in\Gamma(P_t(x),P_t(y))} \int m \ \mathrm{d}\gamma. \tag{8}$$

Now note that for an arbitrary functional $\Psi : Y \longrightarrow \mathbb{R}$ such that for any two pseudometrics m and m', $m \leq m' \Rightarrow \Psi(m) \leq \Psi(m')$,

$$\sup_{m\in Y} \Psi(m) = \sup_{n\in\mathbb{N}} \Psi(\delta_n). \tag{9}$$

Indeed since $\delta_n \in Y$ for every n, we have that $\sup_{m\in Y} \Psi(m) \geq \sup_{n\in\mathbb{N}} \Psi(\delta_n)$. For the other direction, note that for every $m \in Y$, there exists n such that for all $k \geq n$, $a_k = 0$. Thus $m \leq \delta_n$ and therefore $\Psi(m) \leq \Psi(\delta_n)$ which is enough to prove the other direction.

The right-hand side of Equation (8) is

$$\sup_{m \in Y} \min_{\gamma \in \Gamma(P_t(x), P_t(y))} \int m \, \mathrm{d}\gamma = \sup_{m \in Y} W(m)(P_t(x), P_t(y))$$
$$= \sup_{n \in \mathbb{N}} W(\delta_n)(P_t(x), P_t(y)) \quad \text{(previous result in Equation (9))}.$$

The left-hand side of Equation (8) is

$$\begin{aligned}
&\min_{\gamma \in \Gamma(P_t(x), P_t(y))} \sup_{m \in Y} \int m \, \mathrm{d}\gamma \\
&= \min_{\gamma \in \Gamma(P_t(x), P_t(y))} \sup_{n \in \mathbb{N}} \int \delta_n \, \mathrm{d}\gamma \quad \text{(previous result in Equation (9))} \\
&= \min_{\gamma \in \Gamma(P_t(x), P_t(y))} \int \sup_{n \in \mathbb{N}} \delta_n \, \mathrm{d}\gamma \quad \text{(dominated convergence theorem)} \\
&= \min_{\gamma \in \Gamma(P_t(x), P_t(y))} \int \overline{\delta} \, \mathrm{d}\gamma \\
&= W(\overline{\delta})(P_t(x), P_t(y)).
\end{aligned}$$

We thus have that

$$\begin{aligned}
\mathcal{F}_c(\overline{\delta})(x, y) &= \sup_{t \geq 0} c^t W(\overline{\delta})(P_t(x), P_t(y)) \\
&= \sup_{t \geq 0} c^t \sup_{n \in \mathbb{N}} W(\delta_n)(P_t(x), P_t(y)) \quad \text{using Sion's minimax theorem} \\
&= \sup_{n \in \mathbb{N}} \sup_{t \geq 0} c^t W(\delta_n)(P_t(x), P_t(y)) \\
&= \sup_{n \in \mathbb{N}} \mathcal{F}_c(\delta_n)(x, y) \\
&= \sup_{n \in \mathbb{N}} \delta_{n+1}(x, y) \quad \text{by definition of } (\delta_n)_{n \in \mathbb{N}} \\
&= \overline{\delta}(x, y) \quad \text{by definition of } \overline{\delta}.
\end{aligned}$$

∎

Lemma 5.6. Consider a discount factor $0 < c < 1$ and a pseudometric m in $\mathcal{P}$ such that

1. m is a fixpoint for $\mathcal{F}_c$,
2. for every two states x and y, $m(x, y) \geq |obs(x) - obs(y)|$,

then $m \geq \overline{\delta}^c$.

Proof. We show by induction on n that $m \geq \delta_n^c$. First, we have that $m \geq \delta_0^c$. Then assume it is true for n, i.e. $m \geq \delta_n^c$ and let us show that it also holds for $n+1$. Since the functional $\mathcal{F}_c$ is monotone:

$$m = \mathcal{F}_c(m) \geq \mathcal{F}_c(\delta_n^c) = \delta_{n+1}^c.$$

Since for every n, $m \geq \delta_n^c$, $m \geq \sup_n \delta_n^c = \overline{\delta}^c$. ∎

We even have the following characterization of $\overline{\delta}^c$ using Lemma 5.6 and Theorem 5.5.

Theorem 5.7. The pseudometric $\overline{\delta}^c$ is the least fixpoint of $\mathcal{F}_c$ that is greater than the pseudometric $(x, y) \mapsto |obs(x) - obs(y)|$.

6 Corresponding real-valued logic

Similarly to what happens in the discrete-time setting, this behavioural pseudometric $\overline{\delta}^c$ can be characterized by a real-valued logic. This real-valued logic should be thought of as tests performed on the diffusion process, for instance "what is the expected value of *obs* after letting the process evolve for time t?" and it generates a pseudometric on the state space by looking at how different the process performs on those tests starting from different positions.

6.1 The logic

Definition of the logic: The logic is defined inductively and is denoted Λ:

$$f \in \Lambda := q \mid obs \mid \min\{f_1, f_2\} \mid 1 - f \mid f \ominus q \mid \langle t \rangle f$$

for all $f_1, f_2, f \in \Lambda$, $q \in [0,1] \cap \mathbb{Q}$ and $t \in \mathbb{Q}_{\geq 0}$.

This logic closely resembles the ones introduced for discrete-time systems by Desharnais et al. [15, 17] and by van Breugel et al. [30]. The key difference is the term $\langle t \rangle f$ which deals with continuous-time.

Interpretation of the logics: We fix a discount factor $0 < c < 1$. The expressions in Λ are interpreted inductively as functions $E \longrightarrow [0,1]$ as follows for a state $x \in E$:

$$\begin{aligned}
q(x) &= q, \\
obs(x) &= obs(x), \\
(\min\{f_1, f_2\})(x) &= \min\{f_1(x), f_2(x)\}, \\
(1 - f)(x) &= 1 - f(x), \\
(f \ominus q)(x) &= \max\{0, f(x) - q\}, \\
(\langle t \rangle f)(x) &= c^t \int f(y)\ P_t(x, \mathrm{d}y) = c^t\ \left(\hat{P}_t f\right)(x).
\end{aligned}$$

Whenever we want to emphasize the fact that the expressions are interpreted for a discount factor $0 < c < 1$, we will write Λ_c.

Remark 6.1. Let us clarify what the difference is between an expression in Λ and its interpretation. Expressions can be thought of as the notation $+, ^2, \times$ etc. They don't carry much meaning by themselves but one can then interpret them for a given set: $\mathbb{R}, C_0(E)$ (continuous functions $E \longrightarrow \mathbb{R}$ that vanish at infinity) for instance. Combining notations, one can write expressions that can then be interpreted on a given set.

From Λ, we can also define the expression $f \oplus q = 1 - ((1 - f) \ominus q)$ which is interpreted as a function $E \longrightarrow [0, 1]$ as $(f \oplus q)(x) = \min\{1, f(x) + q\}$.

6.2 Definition of the pseudometric

The pseudometric we derive from the logic Λ corresponds to how different the test results are when the process starts from x compared to the case when it starts from y.

Given a fixed discount factor $0 < c < 1$, we can define the pseudometric λ^c:

$$\lambda^c(x, y) = \sup_{f \in \Lambda_c} (f(x) - f(y)) = \sup_{f \in \Lambda_c} |f(x) - f(y)|.$$

The latter equality holds since for every $f \in \Lambda_c$, Λ_c also contains $1 - f$.

6.3 Comparison to the fixpoint metric

This real-valued logic Λ_c is especially interesting as the corresponding pseudometric λ^c matches the fixpoint pseudometric $\overline{\delta}^c$ for the functional $\mathcal{F}_c$ that we defined in Section 5.4. In order to show that $\lambda^c = \overline{\delta}^c$, we establish the inequalities in both directions.

Lemma 6.2. For every f in Λ_c, there exists n such that for every x, y, $f(x) - f(y) \leq \delta^c_n(x, y)$.

Proof. This proof is done by induction on the structure of f:

- If $f = q$, then $f(x) - f(y) = 0 \leq \delta^c_0(x, y)$.
- If $f = obs$, then $f(x) - f(y) = obs(x) - obs(y) \leq \delta^c_0(x, y)$.
- If $f = 1 - g$ and there exists n such that for every x, y, $g(x) - g(y) \leq \delta^c_n(x, y)$, then $f(x) - f(y) = g(y) - g(x) \leq \delta^c_n(y, x) = \delta^c_n(x, y)$.
- If $f = g \ominus q$ and there exists n such that for every x, y, $g(x) - g(y) \leq \delta^c_n(x, y)$, then it is enough to study the case when $f(x) = g(x) - q \geq 0$ and $f(y) = 0 \geq g(y) - q$. In that case,

  $$f(x) - f(y) = g(x) - q \leq g(x) - q - (g(y) - q) = g(x) - g(y) \leq \delta^c_n(x, y)$$
  $$f(y) - f(x) = q - g(x) \leq 0 \leq \delta^c_n(x, y).$$

- If $f = \min\{f_1, f_2\}$ and for $i = 1, 2$ there exists n_i such that for every x, y, $f_i(x) - f_i(y) \leq \delta^c_{n_i}(x, y)$. Then write $n = \max\{n_1, n_2\}$. There really is only one case to consider: $f(x) = f_1(x) \leq f_2(x)$ and $f(y) = f_2(y) \leq f_1(y)$. In that case

  $$f(x) - f(y) = f_1(x) - f_2(y) \leq f_2(x) - f_2(y) \leq \delta^c_{n_2}(x, y) \leq \delta^c_n(x, y).$$

- If $f = \langle t \rangle g$ and there exists n such that for every x, y, $g(x) - g(y) \leq \delta^c_n(x, y)$, then

  $$\begin{aligned} f(x) - f(y) &= c^t \hat{P}_t g(x) - c^t \hat{P}_t g(y) = c^t \left(\hat{P}_t g(x) - \hat{P}_t g(y) \right) \\ &\leq c^t W(\delta^c_n)(P_t(x), P_t(y)) \\ &\leq \delta^c_{n+1}(x, y). \end{aligned}$$

∎

Since $\lambda^c(x, y) = \sup_{f \in \Lambda_c}(f(x) - f(y))$, we get the following corollary as an immediate consequence of Lemma 6.2.

Corollary 6.3. For every discount factor $0 < c < 1$, $\overline{\delta}^c \geq \lambda^c$.

We now aim at proving the reverse inequality. We will use the following result (Lemma A.7.2) from [1] which is also used in [30] for discrete-time processes.

Lemma 6.4. Let X be a compact Hausdorff space. Let A be a subset of the set of continuous functions $X \longrightarrow \mathbb{R}$ such that if $f, g \in A$, then $\max\{f, g\}$ and $\min\{f, g\}$ are also in A. Consider a function h that can be approximated at each pair of points by functions in A, meaning that

$$\forall x, y \in X \ \forall \epsilon > 0 \ \exists g \in A \ |h(x) - g(x)| \leq \epsilon \text{ and } |h(y) - g(y)| \leq \epsilon$$

Then h can be approximated by functions in A, meaning that $\forall \epsilon > 0 \ \exists g \in A \ \forall x \in X \ |h(x) - g(x)| \leq \epsilon$.

In order to use this lemma, we need the following lemmas:

Lemma 6.5. For every $f \in \Lambda_c$, the function $x \mapsto f(x)$ is continuous.

Proof. This is done by induction on the structure of Λ_c. The only case that is not straightforward is when $f = \langle t \rangle g$. By induction hypothesis, g is continuous. Since the map $x \mapsto P_t(x)$ is continuous (onto the weak topology), $f = c^t \hat{P}_t g$ is continuous. ∎

Lemma 6.6. Consider a continuous function $h : E \longrightarrow [0, 1]$ and two states z, z' such that there exists f in the logic Λ_c such that $|h(z) - h(z')| \leq |f(z) - f(z')|$. Then for every $\epsilon > 0$, there exists $g \in \Lambda_c$ such that $|h(z) - g(z)| \leq 2\epsilon$ and $|h(z') - g(z')| \leq 2\epsilon$.

Proof. WLOG $h(z) \geq h(z')$ and $f(z) \geq f(z')$ (otherwise consider $1 - f$ instead of f).

Pick $p, q, r \in \mathbb{Q} \cap [0, 1]$ such that

$$\begin{aligned}
p &\in [f(z') - \epsilon, f(z')], \\
q &\in [h(z) - h(z') - \epsilon, h(z) - h(z')], \\
r &\in [h(z'), h(z') + \epsilon].
\end{aligned}$$

Define $g = (\min\{f \ominus p, q\}) \oplus r$. Then,

$$\begin{aligned}
(f \ominus p)(z) &\in [f(z) - f(z'), f(z) - f(z') + \epsilon] \ \text{ since } f(z) \geq f(z'), \\
(\min\{f \ominus p, q\})(z) &= q \ \text{ since } q \leq h(z) - h(z') \leq f(z) - f(z'), \\
g(z) &= \min\{1, q + r\} \in [h(z) - \epsilon, h(z) + \epsilon)] \ \text{ as } h(z) \leq 1,
\end{aligned}$$

which means that $|h(z) - g(z)| \leq \epsilon$.

$$\begin{aligned}
(f \ominus p)(z') &= \max\{0, f(z') - p\} \in [0, \epsilon], \\
(\min\{f \ominus p, q\})(z') &\in [0, \epsilon], \\
g(z') &\in [h(z'), h(z') + 2\epsilon],
\end{aligned}$$

which means that $|h(z') - g(z')| \leq 2\epsilon$. ∎

Corollary 6.7. Consider a continuous function $h : E \longrightarrow [0,1]$ such that for any two states z, z' there exists f in the logic Λ_c such that $|h(z) - h(z')| \leq |f(z) - f(z')|$. Then for every compact set K in E, there exists a sequence $(g_n)_{n\in\mathbb{N}}$ in Λ_c that approximates h on K.

Proof. We have proven in Lemma 6.6 that such a function h can be approximated at pairs of states by functions in Λ_c. Now recall that all the functions in Λ_c are continuous (Lemma 6.5).

We can thus apply Lemma 6.4 on the compact set K, and we get that the function h can be approximated by functions in Λ_c. ∎

Theorem 6.8. The pseudometric λ^c is a fixpoint of $\mathcal{F}_c$.

Proof. We will omit writing c as an index in this proof. We already have that $\lambda \leq \mathcal{F}(\lambda)$ (cf Lemma 5.2), so we only need to prove the reverse direction.

There are countably many expressions in Λ so we can number them: $\Lambda = \{f_0, f_1, ...\}$. Write $m_k(x,y) = \max_{j\leq k} |f_j(x) - f_j(y)|$. Since all the f_j are continuous (Lemma 6.5), the map m_k is also continuous. Furthermore, $m_k \leq m_{k+1}$ and for every two states x and y, $\lim_{k\to\infty} m_k(x,y) = \lambda(x,y)$.

Using Lemma 2.5, we know that for every states x, y and time t,

$$\sup_k W(m_k)(P_t(x), P_t(y)) = W(\lambda)(P_t(x), P_t(y)).$$

This implies that

$$\mathcal{F}(\lambda)(x,y) = \sup_{t\geq 0} c^t W(\lambda)(P_t(x), P_t(y)) = \sup_k \sup_{t\geq 0} c^t W(m_k)(P_t(x), P_t(y)).$$

It is therefore enough to show that for every k, every time $t \in \mathbb{Q}_{\geq 0}$ and every pair of states x, y, $c^t W(m_k)(P_t(x), P_t(y)) \leq \lambda(x,y)$. There exists $h \in \mathcal{H}(m_k)$ such that $W(m_k)(P_t(x), P_t(y)) = \left|\int h \, \mathrm{d}P_t(x) - \int h \, \mathrm{d}P_t(y)\right|$.

Since $P_t(x)$ and $P_t(y)$ are tight, for every $\epsilon > 0$, there exists a compact set $K \subset E$ such that $P_t(x, E\setminus K) \leq \epsilon/4$ and $P_t(y, E \setminus K) \leq \epsilon/4$.

By Corollary 6.7, there exists $(g_n)_{n\in\mathbb{N}}$ in Λ that pointwise converge to h on K. In particular, for n large enough,

$$\left|\int_K g_n \, \mathrm{d}P_t(x) - \int_K h \, \mathrm{d}P_t(x)\right| \leq \epsilon/4,$$

and similarly for $P_t(y)$. We get that for n large enough.

$$\begin{aligned}&\left|\int_E g_n \, \mathrm{d}P_t(x) - \int_E h \, \mathrm{d}P_t(x)\right| \\ &\leq \left|\int_K g_n \, \mathrm{d}P_t(x) - \int_K h \, \mathrm{d}P_t(x)\right| + \int_{E\setminus K} |g_n - h| \, \mathrm{d}P_t(x) \leq \epsilon/2,\end{aligned}$$

and similarly for $P_t(y)$. We can thus conclude that

$$
\begin{aligned}
&c^t W(m_k)(P_t(x), P_t(y)) = c^t \left| \int_E h \; \mathrm{d}P_t(x) - \int_E h \; \mathrm{d}P_t(y) \right| \\
&\leq c^t \left| \int_E g_n \; \mathrm{d}P_t(x) - \int_E h \; \mathrm{d}P_t(x) \right| + c^t \left| \int_E g_n \; \mathrm{d}P_t(y) - \int_E h \; \mathrm{d}P_t(y) \right| \\
&\quad + |(\langle t \rangle g_n)(x) - (\langle t \rangle g_n)(y)| \\
&\leq \epsilon + \lambda(x, y).
\end{aligned}
$$

Since ϵ is arbitrary, $c^t W(m_k)(P_t(x), P_t(y)) \leq \lambda(x, y)$. □ ■

As a consequence of Theorem 6.8 and using Lemma 5.6, we get that $\overline{\delta}^c \leq \lambda^c$. Then with Corollary 6.3, we finally get:

Theorem 6.9. The two pseudometrics are equal: $\overline{\delta}^c = \lambda^c$.

7 Obstacles in continuous time

As we have pointed out throughout this work, although the overall outline is similar to that employed in the discrete case, we are forced to develop new strategies to overcome technical challenges arising from the continuum setting, where the topological properties of the time/state space become crucial elements in the study.

For example, a key obstacle we face in this work is that the fixpoint pseudometric can't be derived directly from the Banach fixed point theorem. There is no notion of step and we therefore need to consider all times in $\mathbb{R}_{\geq 0}$. In discrete-time, the counterpart of our functional $\mathcal{F}_c$ is c-Lipschitz. However, in our case, this should be thought of as c^1. Since we are forced to consider all times and since $\sup_{t \geq 0} c^t = 1$, we cannot find a constant $k < 1$ such that $\mathcal{F}_c$ is k-Lipschitz. To overcome this issue, we construct a candidate pseudometric with brute force (we first define a sequence of pseudometrics δ^c_n and then define our candidate $\overline{\delta}^c$ as their supremum) and then prove that it is indeed the greatest fixpoint.

In addition, the scope of the functional $\mathcal{F}_c$ also requires careful treatment: for measurability reasons, we have to restrict to the lattice of pseudometrics which generate a subtopology of the pre-existing one. We cannot apply the Kleene fixed point theorem either since this lattice is not complete. This whole approach toward a fixpoint pseudometric differs substantially from the long-existing methodology.

Furthermore, for some of our key results (e.g., Theorem 6.8 and Lemma 5.3), the proofs rely on the compactness argument to establish certain convergence relations in temporal and/or spatial variables and to further achieve the goals. This type of procedure in general is not required in the discrete setting.

Although from time to time, we do restrict the time variable to rational values thanks to the continuity of the FD-semigroups, this is very different from treating discrete-time models, because rational time stamps cannot be "ordered" to represent the notion of "next step" in a continuous-time setting. Therefore, we are still working with a true continuous-time dynamics, and it cannot be reduced to a discrete-time problem.

8 Examples

8.1 A toy example

Let us consider a process defined on $\{0, x, y, z, \partial\}$. Let us first give an intuition for what we are trying to model. In the states x, y, z, the process is trying to learn a value. In the state 0, the correct value has been learnt, but in the state ∂, an incorrect value has been learnt. From the three " learning" states x, y and z, the process has very different learning strategies:

- from x, the process exponentially decays to the correct value represented by the state 0,
- from y, the process is not even attempting to learn the correct value and thus remains in a learning state, and
- from z, the process slowly learns but it may either learn the correct value (0) or an incorrect one (∂).

The word "learning" is here used only to give colour to the example.

Formally, this process is described by the time-indexed kernels:

$$\begin{array}{lll} P_t(x,\{0\}) = 1 - e^{-\lambda t} & P_t(z,\{0\}) = \frac{1}{2}(1 - e^{-\lambda t}) & P_t(0,\{0\}) = 1 \\ P_t(x,\{x\}) = e^{-\lambda t} & P_t(z,\{\partial\}) = \frac{1}{2}(1 - e^{-\lambda t}) & P_t(y,\{y\}) = 1 \\ & P_t(z,\{z\}) = e^{-\lambda t} & P_t(\partial,\{\partial\}) = 1 \end{array}$$

(where $\lambda \geq 0$) and by the observable function $obs(x) = obs(y) = obs(z) = r \in (0, 1)$, $obs(\partial) = 0$ and $obs(0) = 1$.

We will pick as a discount factor $c = e^{-\lambda}$ to simplify notations. We can compute the distance $\overline{\delta}$ (we omit adding c in the notation).

We will only detail the computations for $\overline{\delta}(0,z)$ in the main section of this paper. First of all, note that $\delta_0(0,z) = 1 - r$ and that

$$\begin{aligned}\delta_{n+1}(0,z) &= \sup_{t\geq 0} e^{-\lambda t}\left(e^{-\lambda t}\delta_n(0,z) + \frac{1}{2}(1-e^{-\lambda t})\delta_n(0,\partial)\right)\\ &= \sup_{0<\theta\leq 1} \theta\left(\theta\delta_n(0,z) + \frac{1}{2}(1-\theta)\right)\\ &= \sup_{0<\theta\leq 1} \theta\left[\frac{1}{2} + \theta\left(\delta_n(0,z) - \frac{1}{2}\right)\right]\end{aligned}$$

We are thus studying the function $\phi : \theta \mapsto \theta\left[\frac{1}{2} + \theta\left(\delta_n(0,z) - \frac{1}{2}\right)\right]$. Its derivative ϕ' has value 0 at $\theta_0 = \frac{1}{2(1-2\delta_n(0,z))}$. There are three distinct cases:

1. If $0 < \delta_n(0,z) \leq \frac{1}{4}$, in that case, $0 < \theta_0 \leq 1$ and $\sup_{1<\theta\leq 1}\phi(\theta)$ is attained in θ_0 and in this case, we have that $\delta_{n+1}(0,z) = \frac{1}{8(1-2\delta_n(0,z))} \leq \frac{1}{4}$. This means in particular that if $1 - r \leq \frac{1}{4}$, then $\overline{\delta}(0,z) = \frac{1}{4}$.
2. If $\frac{1}{4} \leq \delta_n(0,z) \leq \frac{1}{2}$, then $\theta_0 \geq 1$ and ϕ is increasing on $(-\infty, \theta_0]$. In that case, we therefore have that $\sup_{1<\theta\leq 1}\phi(\theta)$ is attained in 1 and therefore $\delta_{n+1}(0,z) = \delta_n(0,z)$.
3. If $\frac{1}{2} \leq \delta_n(0,z)$, then $\theta_0 \leq 0$ and ϕ is increasing on $[\theta_0, +\infty)$. In that case, we therefore have that $\sup_{1<\theta\leq 1}\phi(\theta)$ is attained in 1 and therefore $\delta_{n+1}(0,z) = \delta_n(0,z)$.

We therefore have that

$$\overline{\delta}(0,z) = \begin{cases}\frac{1}{4} & \text{if } r \geq \frac{3}{4}\\ 1-r & \text{otherwise.}\end{cases}$$

The other cases can be found in Appendix E but we summarize them in a table here. Note that the computation of $\overline{\delta}(x,z)$ is too involved and we therefore only provide an interval.

	x	y	z	∂	0
x	0	$\frac{1-r}{2}$	$\in \left[\frac{1}{8}, \frac{1}{4}\right]$	$\begin{cases} r & \text{if } r \geq \frac{1}{2} \\ \frac{1}{2} & \text{otherwise} \end{cases}$	$1-r$
y	$\frac{1-r}{2}$	0	$\frac{1}{4}$	r	$1-r$
z	$\in \left[\frac{1}{8}, \frac{1}{4}\right]$	$\frac{1}{4}$	0	$\begin{cases} r & \text{if } r \geq \frac{1}{4} \\ \frac{1}{4} & \text{otherwise} \end{cases}$	$\begin{cases} \frac{1}{4} & \text{if } r \geq \frac{3}{4} \\ 1-r & \text{otherwise} \end{cases}$
∂	$\begin{cases} r & \text{if } r \geq \frac{1}{2} \\ \frac{1}{2} & \text{otherwise} \end{cases}$	r	$\begin{cases} r & \text{if } r \geq \frac{1}{4} \\ \frac{1}{4} & \text{otherwise} \end{cases}$	0	1
0	$1-r$	$1-r$	$\begin{cases} \frac{1}{4} & \text{if } r \geq \frac{3}{4} \\ 1-r & \text{otherwise} \end{cases}$	1	0

Note that even though the process behaves vastly differently from x than from y, we have that $\overline{\delta}(x, 0) = \overline{\delta}(y, 0)$, even though for $t > 0$, we have that $\hat{P}_t obs(x) > \hat{P}_t obs(y)$. However note that x and y have different distances to other states.

This also happens in the discrete-time setting and for continuous-time Markov chains. A continuous-time Markov chain is a continuous-time type of processes but where the evolution still occurs as steps. They can be described as jump processes over continuous time. They have been studied in [20] by considering the whole trace starting from a single state. It is possible to adapt our work to traces (called trajectories in [27]) by adding some additional regularity conditions on the processes; this can be found in [12]. However one should consider the added complexity. For instance, for Brownian motion the kernels P_t are well-known and easy to describe with a density function but that is not the case for the probability measures on trajectories.

8.2 Two examples based on Brownian motion

Previous example is a very simple example which emphasizes some of the difficulties of computing our metric. It may then be tempting to think that our approach cannot yield any result when applied to the real world. The next examples show that we can still provide some meaningful results when looking at real-life processes such as Brownian motion. We refer the reader to Section C for some background on Brownian motion and hitting times.

First example: We denote $(B_t^x)_{t \geq 0}$ the standard Brownian motion on the real line starting from x. We can then define its first hitting time of 0 or 1: $\tau := \inf\{t \geq 0 : B_t^x = 0 \text{ or } B_t^x = 1\}$.

Our state space is the interval $[0,1]$. For every $x \in [0,1]$ and every $t \geq 0$, let $P_t(x,\cdot)$ be distribution of $B^x_{t\wedge\tau}$. In other words, $P_t(x,\cdot)$ is the distribution of Brownian motion starting from x running until hitting either 0 or 1 and getting trapped upon hitting a boundary. We equip the state space $[0,1]$ with *obs* defined as $obs(x) = x$. Then, $obs\left(B^x_{t\wedge\tau}\right) = B^x_{t\wedge\tau}$ and hence for every $x \in (0,1)$

$$\begin{aligned}\delta_1(0,x) &= \sup_{t\geq 0} c^t W(\delta_0)\left(P_t(0), P_t(x)\right) = \sup_{t\geq 0} c^t \mathbb{E}[B^x_{t\wedge\tau}] \\ &= \sup_{t\geq 0} c^t \cdot x \text{ (because } B^x_{t\wedge\tau} \text{ is a martingale)} \\ &= x.\end{aligned}$$

where the first equality follows from the fact that $P_t(0)$ is the dirac distribution $\mathfrak{d}_0$ at 0 and the second equality comes from the fact that any coupling $\gamma \in \Gamma(\mathfrak{d}_0, P_t(x))$ is reduced to the marginal $P_t(x)$. Since $\delta_0(0,x) = \delta_1(0,x)$, we then have $\bar{\delta}(0,x) = x$.

Similarly, $\bar{\delta}(1,y) = 1 - y$ for every $y \in [0,1]$. It is difficult to compute $\bar{\delta}(x,y)$ for general $x, y \in [0,1]$ though.

Second example: This example relies on stochastic differential equations (SDE). We refer the reader to [24] for a comprehensive introduction.

Let the state space and *obs* be the same as above. Let $Q_t(x,\cdot)$ be the distribution of the solution to the SDE

$$dX_t = X_t dB^0_t + \frac{1}{2} X_t dt \text{ with } X_0 = x.$$

It can be verified that the solution to this equation is the process $X_t = xe^{B^0_t}$. In this case, we also have $Q_t(0) = \mathfrak{d}_0$. Again, for every $x \in [0,1]$,

$$\begin{aligned}\delta_1(0,x) &= \sup_{t\geq 0} c^t W(\delta_0)\left(Q_t(0), Q_t(x)\right) = \sup_{t\geq 0} c^t \mathbb{E}[obs(xe^{B^0_t})] \\ &= \sup_{t\geq 0} c^t \left(x\mathbb{E}[e^{B^0_t} \mid t \leq \tau'] + \mathbb{P}(t > \tau') \right),\end{aligned}$$

where $\tau' := \inf\{s \geq 0 \mid B^0_s \geq -\ln x\}$ and $\mathbb{E}$ and $\mathbb{P}$ denote expected values and probabilities for the standard Brownian motion starting in 0. The distribution of τ', as well as the joint distribution of (B^0_t, τ'), has been determined explicitly (see Chapter 2, Section 8 in [21] for instance). Even if we only consider the second term in the expression above, we have

$$\delta_1(0,x) \geq \sup_{t\geq 0} c^t \mathbb{P}(t > \tau') = \sup_{t\geq 0} c^t \frac{2}{\sqrt{2\pi}} \int_{\frac{-\ln x}{\sqrt{t}}}^{\infty} e^{-\frac{y^2}{2}} dy.$$

It is possible for the right hand side above to be greater than x. For example, if $x = \frac{1}{e}$, then by choosing $t = 4$, we have that the right hand side above is no smaller than $c\frac{2}{\sqrt{2\pi}} \int_{\frac{1}{2}}^{\infty} e^{-\frac{y^2}{2}} dy$, which will be (strictly) greater than $\frac{1}{e}$ (i.e., x) provided that c is sufficiently close to 1. As a consequence, for this example, we have $\overline{\delta}(0, x) > x$.

These two examples demonstrate different behaviors of the two systems, while the first system "maintains" the mass at the starting point (expectation is constant x), the second system dissipates the mass to the right (which is the direction of larger values of *obs*). Therefore, processes starting from the same point x possess different distances to the static path between these two systems.

A very important observation from these examples based on Brownian motion is that even though we cannot explicitly compute values, we are still able to compare state behaviours.

9 Another approach through trajectories

BLABLA

In some previous work [9, 10], we showed that blabla need for trajectories, blabla turns out this work can be adapted to trajectories

9.1 Trajectories and diffusions

A very important ingredient in the theory is the space of trajectories of a FD-process (FD-semigroup) as a probability space. This space does not appear explicitly in the study of labelled Markov processes but one does see it in the study of continuous-time Markov chains and jump processes.

We will use trajectories in Section 9. We will impose further conditions (see Definition 9.5), but for now we follow [27] to introduce the concept of trajectories and the corresponding probabilities. We still require that our process be honest, however, this work could be adapted to non-honest processes with some additional continuity condition on *obs*.

Definition 9.1. A function $\omega : [0, \infty) \longrightarrow E$ is called a *trajectory on E* if it is *cadlag* [2] meaning that for every $t \geq 0$,

$$\lim_{s>t, s \longrightarrow t} \omega(s) = \omega(t) \quad \text{and} \quad \omega(t-) := \lim_{s<t, s \longrightarrow t} \omega(s) \text{ exists.}$$

[2]cadlag stands for the French "continu à droite, limite à gauche"

As an intuition, a cadlag function ω is an "almost continuous" function with jumping in a reasonable fashion.

It is possible to associate to such an FD-semigroup a *canonical FD-process*. Let Ω be the set of all trajectories $\omega : [0, \infty) \longrightarrow E$.

Definition 9.2. The *canonical FD-process* associated to the FD-semigroup $(\hat{P}_t)_{t\geq 0}$ is

$$(\Omega, \mathcal{A}, (\mathcal{A}_t)_{t\geq 0}, (X_t)_{0\leq t\leq\infty}, (\mathbb{P}^x)_{x\in E_\partial})$$

where

- the random variable $X_t : \Omega \longrightarrow E$ is defined as $X_t(\omega) = \omega(t)$ for every $0 \leq t$,
- $\mathcal{A} = \sigma(X_s \mid 0 \leq s)$ [3], $\mathcal{A}_t = \sigma(X_s \mid 0 \leq s \leq t)$ for every $0 \leq t$,
- given any probability measure μ on E_∂, by the Daniell-Kolmogorov theorem, there exists a unique probability measure $\mathbb{P}^\mu$ on $(\Omega, \mathcal{A})$ such that for all $n \in \mathbb{N}, 0 \leq t_1 \leq t_2 \leq ... \leq t_n$ and $x_0, x_1, ..., x_n$ in E_∂,

  $$\mathbb{P}^\mu(X_0 \in dx_0, X_{t_1} \in dx_1, ..., X_{t_n} \in dx_n) = \mu(dx_0)P_{t_1}(x_0, dx_1)...P_{t_n - t_{n-1}}(x_{n-1}, dx_n)$$

  The dx_i in this equation should be understood as infinitesimal volumes. This notation is standard in probability and should be understood by integrating it over measurable state sets C_i. We set $\mathbb{P}^x = \mathbb{P}^{\delta_x}$.

The distribution $\mathbb{P}^x$ is a measure on the space of trajectories for a system started at the point x.

Remark 9.3. Here, we have constructed one process which we called the canonical FD-process. Note that there could be other tuples

$$(\Omega', \mathcal{A}', (\mathcal{A}'_t)_{t\geq 0}, (Y_t)_{0\leq t\leq\infty}, (\mathbb{Q}^x)_{x\in E_\partial})$$

satisfying the same conditions except that the space Ω' would not be the space of trajectories.

Remark 9.4. Let us give an intuition for this new way of describing the dynamics of the system based on discrete-time. One standard way of describing such a system is through the Markov kernel τ. In continuous-time settings, the analogue is the time-indexed family $(P_t)_{t\geq 0}$. On the other hand, it is also possible to look at infinite runs of the process. For instance, for a coin toss,

[3] The σ-algebra $\mathcal{A}$ is the same as the one induced by the Skorohod metric, see theorem 16.6 of [3]

we can look at sequences of heads and tails. Such individual sequences correspond to trajectories and we can look at the probability of certain sets of trajectories (for instance, the set of trajectories that start with three heads). In continuous-time settings, this corresponds to the space-indexed family of distributions $(\mathbb{P}^x)_{x\in E}$.

Definition 9.5. A *diffusion* is an honest process satisfying some additional properties:

- the trajectories in Ω are continuous,
- the map $x \mapsto \mathbb{P}^x$ is weakly continuous meaning that if $y_n \underset{n\to\infty}{\longrightarrow} y$, then for every bounded and continuous function f on the set of trajectories, $\int f \mathrm{d}\mathbb{P}^{y_n} \underset{n\to\infty}{\longrightarrow} \int f \mathrm{d}\mathbb{P}^y$.

The set of continuous maps from a locally-compact space onto a Polish space is Polish which means that Ω is Polish.

In [27], FD-diffusions are also introduced. Our definition is different in the sense that we only define them for honest processes, but we require less conditions. While these conditions may appear restrictive, they cover classical examples of stochastic processes such as Brownian motion.

This work is restricted to diffusions and all the processes that we will consider later on in this section are diffusions. Note that compared to Definition 3.6, we replaced the weak continuity condition wrt space of the measures $P_t(x)$ by a weak continuity wrt space of the measures $\mathbb{P}^x$.

9.2 Through functionals

The work here is similar to that in Section 5: we define a functional $\mathcal{G}$ on the lattice $\mathcal{P}$ of lower semi-continuous 1-bounded pseudometrics on the state space E. We show that the image of a continuous pseudometric is also continuous, thus allowing us to iteratively apply $\mathcal{G}$ and to obtain a pseudometric $\overline{d}$ starting from $d_0(x,y) = \delta_0(x,y) = |obs(x) - obs(y)|$. We can then show that $\overline{d}$ is a fixpoint of $\mathcal{G}$. The key difference is that the cost function is no longer just the pseudometric on the state space.

Recall that Ω is Polish (see Section 9.1). This is important in order to use the machinery of optimal transport.

We also need to add a discount factor $0 < c < 1$ and thus obtain a family of such functionals $\mathcal{G}_c$ indexed by the discount factors and a corresponding family of pseudometrics obtained by iteratively applying $\mathcal{G}_c$.

9.2.1 Some more results on our lattices

Recall that we have defined three lattices $\mathcal{C} \subset \mathcal{P} \subset \mathcal{M}$ where $\mathcal{M}$ is the lattice of 1-bounded pseudometrics on E, $\mathcal{P}$ is the sublattice of lower semi-continuous pseudometrics and $\mathcal{C}$ of continuous pseudometrics. We refer the reader to Remark 5.1 for an important subtlety on those lattices: the pseudometrics in $\mathcal{M}$ should be thought of as a function $E \times E \longrightarrow [0, 1]$ and do not relate a priori to the original topology $\mathcal{O}$ on E.

As we have stated beforehand, we will be working with trajectories in this section. We will therefore need to define a cost function on the space of trajectories Ω. This is going to be done through the (discounted) uniform metric:

Definition 9.6. Given a pseudometric m on a set E and $0 < c \leq 1$, we define the *discounted uniform pseudometric* $U_c(m)$ on the set of trajectories as follows:

$$\text{for all } \omega, \omega' \in \Omega, \qquad U_c(m)(\omega, \omega') = \sup_t c^t m(\omega(t), \omega'(t)).$$

Remark 9.7. If m is a pseudometric, then $U_c(m)$ is also a pseudometric. Let us check quickly the triangular inequality: for ω_1, ω_2 and ω_3 in Ω,

$$U_c(m)(\omega_1, \omega_3) = \sup_t c^t m(\omega_1(t), \omega_3(t)) \leq \sup_t c^t (m(\omega_1(t), \omega_2(t)) + m(\omega_2(t), \omega_3(t)))$$

by triangular inequality for m. Since $c^t m(\omega_i(t), \omega_j(t)) \leq U_c(m)(\omega_i, \omega_j)$, we get the triangular inequality as desired.

Note that if $c = 1$, we get the uniform metric $U(m)$ wrt m. We also have that for any discount factor $c \leq 1$, $U_c(m) \leq U(m)$.

In Section 5, we restricted the pseudometrics, which were used as cost functions, to lower semi-continuous ones (lattice $\mathcal{P}$) or to continuous ones (lattice $\mathcal{C}$). When we are dealing with trajectories, our cost function is going to be a discounted uniform pseudometric. For that reason, we need to make sure that the regularity conditions on a pseudometric m on the state space are also passed on to $U_c(m)$.

Lemma 9.8. Let m be a pseudometric in $\mathcal{P}$. Then, $U_c(m) : \Omega \times \Omega \longrightarrow \mathbb{R}$ is lower semi-continuous for every $0 < c \leq 1$.

Proof. For $r \in [0, 1]$, let us consider the set $U_c(m)^{-1}((r, 1]) = \{(\omega, \omega') \mid U_c(m)(\omega, \omega') > r\}$. We intend to show that this set is open in $\Omega \times \Omega$. Since the trajectories are continuous, the topology on Ω is the topology generated by the uniform metric $U(\Delta)$ defined by $U(\Delta)(\omega, \omega') = \sup_t \Delta(\omega(t), \omega'(t))$. This topology then gives the product topology on $\Omega \times \Omega$.

Consider $(\omega, \omega') \in U_c(m)^{-1}((r, 1])$. This means that $\sup_t c^t m(\omega(t), \omega'(t)) > r$ which implies that there exists t such that $m(\omega(t), \omega'(t)) > rc^{-t}$.

Define the set $O = \{(x, y) \mid m(x, y) > rc^{-t}\} \subset E \times E$. This set O is open in $E \times E$ since the function m is lower semi-continuous.

The pair $(\omega(t), \omega'(t))$ is in the open set O which means that there exists $q > 0$ [4] such that $B_q \times B'_q \subset O$ where the two sets B_q and B'_q are defined as open balls in E:

$$B_q = \{z \mid \Delta(z, \omega(t)) < q\} \qquad \text{and} \qquad B'_q = \{z \mid \Delta(z, \omega'(t)) < q\}.$$

Define the two sets A and A' as

$$A = \{\theta \mid U(\Delta)(\omega, \theta) < q\} \qquad \text{and} \qquad A' = \{\theta \mid U(\Delta)(\omega', \theta) < q\}.$$

These are open balls in Ω and thus $A \times A'$ is open in $\Omega \times \Omega$.

The set $A \times A'$ is included in $U(m)^{-1}((r, 1])$. Indeed, consider $\theta \in A$ and $\theta' \in A'$. This means that $U(\Delta)(\omega, \theta) < q$, which implies that $\Delta(\omega(t), \theta(t)) < q$, i.e. $\theta(t) \in B_q$. Similarly, $\theta'(t) \in B'_q$, which means that $(\theta(t), \theta'(t)) \in B_q \times B'_q \subset O$, which means that $c^t m(\theta(t), \theta'(t)) > r$.

Furthermore, $\omega \in A$ and $\omega' \in A'$ which means that $A \times A'$ is an open neighbourhood of (ω, ω') in $U_c(m)^{-1}((r, 1])$ which concludes the proof that $U_c(m)^{-1}((r, 1])$ is open. ∎

Lemma 9.9. If $m \in \mathcal{C}$ and the discount factor $c < 1$, then $U_c(m)$ is continuous as a function $\Omega \times \Omega \longrightarrow \mathbb{R}$.

Proof. It is enough to show that if $(\omega_n)_{n \in \mathbb{N}}$ is a sequence of trajectories that converges to ω (in the topology generated by $U(\Delta)$), then $\lim_{n \longrightarrow \infty} U_c(m)(\omega, \omega_n) = 0$.

Assume it is not the case. This means that there exists $\epsilon > 0$ and an increasing sequence of natural numbers $(n_k)_{k \in \mathbb{N}}$ such that $U_c(m)(\omega, \omega_{n_k}) \geq 3\epsilon$. This means that for every k, there exists a time t_k such that $c^{t_k} m(\omega(t_k), \omega_{n_k}(t_k)) \geq 2\epsilon$. Since m is 1-bounded, those times t_k are bounded by $\ln(2\epsilon)/\ln c$.

For that reason we may assume that the sequence $(t_k)_{k \in \mathbb{N}}$ converges to some time t (otherwise there is a subsequence that converges, so we can consider that subsequence instead and the corresponding trajectories ω_{n_k}). We have the following inequalities for every k:

$$\begin{aligned} 2\epsilon &\leq c^{t_k} m(\omega(t_k), \omega_{n_k}(t_k)) \\ &\leq c^{t_k} m(\omega(t_k), \omega(t)) + c^{t_k} m(\omega(t), \omega_{n_k}(t_k)). \end{aligned}$$

[4] Note that this q also depends on how the topology on $E \times E$ is metrized, for instance p-distance or max distance

Since m is continuous and so are the trajectories, $\lim_{k \longrightarrow \infty} m(\omega(t_k), \omega(t)) = 0$. This means that there exists K such that for all $k \geq K$, $m(\omega(t_k), \omega(t)) < \epsilon$. Since $c < 1$, this implies that $c^{t_k} m(\omega(t_k), \omega(t)) < \epsilon$ and thus using previous inequality, $\epsilon < c^{t_k} m(\omega(t), \omega_{n_k}(t_k))$. Pushing this further, using that $c < 1$ once more, we get that for all $k \geq K$, $\epsilon < m(\omega(t), \omega_{n_k}(t_k))$.

Now recall that $(\omega_n)_{n \in \mathbb{N}}$ is a sequence of trajectories that converges to ω (in the topology generated by Δ), which means that $\lim_{n \longrightarrow \infty} U(\Delta)(\omega_n, \omega) = 0$. Using the triangular inequality, we have that

$$\Delta(\omega_{n_k}(t_k), \omega(t)) \leq \Delta(\omega_{n_k}(t_k), \omega(t_k)) + \Delta(\omega(t_k), \omega(t)) \leq U(\Delta)(\omega_{n_k}, \omega) + \Delta(\omega(t_k), \omega(t)).$$

The second term of the right-hand side converges to 0 as $k \longrightarrow \infty$ as $t_k \longrightarrow t$ and ω is continuous. This means that $\lim_{n_k \longrightarrow \infty} \Delta(\omega_{n_k}(t_k), \omega(t)) = 0$. This implies that $\lim_{n \longrightarrow \infty} m(\omega_{n_k}(t_k), \omega(t)) = 0$ which directly contradicts that $\epsilon < m(\omega(t), \omega_{n_k}(t_k))$ for every $k \geq K$, which concludes the proof. ∎

9.2.2 The family of functionals

A functional defined through trajectories: The (family of) functional(s) $\mathcal{F}_c$ that we defined in Section 5 compared the distributions on the state space $P_t(x)$. The new functional that we will define here compares the probability distributions $\mathbb{P}^x$ on the space of trajectories through transport theory. Similarly to that in Section 5, this functional is defined for diffusions as we need the map $x \mapsto \mathbb{P}^x$ to be continuous.

Just as the first functional $\mathcal{F}_c$ is defined with a discount factor c, the second functional that we will define also involves a discount factor c, it is denoted $\mathcal{G}_c$. The functional $\mathcal{G}_c$ can be defined for $0 < c \leq 1$ but we will later require $0 < c < 1$. The transport cost between two trajectories is given by the discounted uniform metric.

Given a discount factor $0 < c \leq 1$, we define the functional $\mathcal{G}_c : \mathcal{P} \longrightarrow \mathcal{M}$ as follows: for every pseudometric $m \in \mathcal{P}$ and every two states x, y,

$$\mathcal{G}_c(m)(x, y) = W(U_c(m))(\mathbb{P}^x, \mathbb{P}^y).$$

This is well-defined as we have proven in Lemma 9.8 that if m is lower semi-continuous, then so is $U_c(m)$.

Using the Kantorovich duality, we have:

$$\begin{aligned} \mathcal{G}_c(m)(x, y) &= \min_{\gamma \in \Gamma(\mathbb{P}^x, \mathbb{P}^y)} \int U_c(m) \, \mathrm{d}\gamma \\ &= \sup_{h \in \mathcal{H}(U_c(m))} \left(\int h \, \mathrm{d}\mathbb{P}^x - \int h \, \mathrm{d}\mathbb{P}^y \right). \end{aligned}$$

Lemma 2.4 ensures that $\mathcal{G}_c(m)$ is indeed in $\mathcal{M}$.

The remarks (5.1 and 5.4) that we made on $\mathcal{F}_c$ still apply here. In particular, we will need to restrict to the lattice $\mathcal{C}$ in order for $\mathcal{G}_c$ to be defined on a lattice since $\mathcal{G}_c(m)$ need not be in $\mathcal{P}$ when $m \in \mathcal{P}$. For the same reason as before (i.e. the lattice $\mathcal{C}$ is not complete), we cannot apply the Knaster-Tarski theorem so we will have to go through similar steps as for $\mathcal{F}_c$.

9.2.3 Some ordering results on our functionals

In this section, we order m, $\mathcal{G}_c(m)$ and $\mathcal{F}_c(m)$.

As a direct consequence from the definition of $\mathcal{G}_c$, we have that the functionals $\mathcal{G}_c$ are monotone (for any $0 < c \leq 1$) : if $m_1 \leq m_2$ in $\mathcal{P}$, then $\mathcal{G}_c(m_1) \leq \mathcal{G}_c(m_2)$.

Lemma 9.10. For every pseudometric m in $\mathcal{P}$, every discount factor $0 < c \leq 1$ and every pair of states x, y,

$$m(x,y) \leq \mathcal{F}_c(m)(x,y) \leq \mathcal{G}_c(m)(x,y).$$

Proof. The first inequality corresponds to Lemma 5.2.

For the second inequality: pick $t \geq 0$ and $h : E \longrightarrow [0,1] \in \mathcal{H}(m)$. Define $h_t : \Omega \longrightarrow [0,1]$ as $h_t(\omega) = c^t h(\omega(t))$. Then note that for any trajectories ω, ω',

$$\begin{aligned} |h_t(\omega) - h_t(\omega')| &= c^t |h(\omega(t)) - h(\omega'(t))| \\ &\leq c^t m(\omega(t), \omega'(t)) \\ &\leq U_c(m)(\omega, \omega'). \end{aligned}$$

This means that $h_t \in \mathcal{H}(U_c(m))$ and thus

$$\begin{aligned} \mathcal{G}_c(m)(x,y) &\geq \int h_t \, \mathrm{d}\mathbb{P}^x - \int h_t \, \mathrm{d}\mathbb{P}^y \\ &= c^t \left(\int h \, \mathrm{d}P_t(x) - \int h \, \mathrm{d}P_t(y) \right) \\ &= c^t \left(\hat{P}_t h(x) - \hat{P}_t h(y) \right), \end{aligned}$$

which allows us to conclude since it holds for every time $t \geq 0$. ∎

Corollary 9.11. A fixpoint for $\mathcal{G}_c$ is also a fixpoint for $\mathcal{F}_c$.

Proof. For a metric $m \in \mathcal{P}$ that is a fixpoint of $\mathcal{G}_c$ and two states x and y:

$$m(x,y) \leq \mathcal{F}_c(m)(x,y) \leq \mathcal{G}_c(m)(x,y) = m(x,y),$$

which means that $m = \mathcal{F}_c(m)$. ∎

9.2.4 When restricted to continuous metrics

As in Section 5, one does not have that $\mathcal{G}_c(m) \in \mathcal{P}$ when $m \in \mathcal{P}$ and we have to go through the lattice of continuous pseudometrics $\mathcal{C}$ as $\mathcal{G}_c(m) \in \mathcal{C}$ whenever $m \in \mathcal{C}$.

Lemma 9.12. Consider a pseudometric $m \in \mathcal{C}$. Then the topology on E generated by $\mathcal{G}_c(m)$ is a subtopology of the original topology for any $0 < c < 1$.

Proof. Using Lemma 9.9, we know that $U_c(m)$ is continuous wrt the topology on $\Omega \times \Omega$ generated by $U(\Delta)$.

In order to show that $\mathcal{G}_c(m)$ generates a topology on E which is a subtopology of $\mathcal{O}$ (generated by Δ), it is enough to show that for a fixed state $x \in E$, the map $y \mapsto \mathcal{G}_c(m)(x, y)$ is continuous on E.

Pick a sequence $(y_n)_{n\in\mathbb{N}}$ in E such that $y_n \underset{n\longrightarrow\infty}{\longrightarrow} y$ in the topology $\mathcal{O}$ with $y \in E$. Recall that $\mathbb{P}^{y_n}$ converges weakly to $\mathbb{P}^y$. Then,

$$\begin{aligned}\lim_{n\longrightarrow\infty} \mathcal{G}_c(m)(x, y_n) &= \lim_{n\longrightarrow\infty} W(U_c(m))(\mathbb{P}^x, \mathbb{P}^{y_n}) \\ &= \lim_{n\longrightarrow\infty} \min \left\{ \int U_c(m) \mathrm{d}\gamma \;\middle|\; \gamma \in \Gamma(\mathbb{P}^x, \mathbb{P}^{y_n}) \right\}.\end{aligned}$$

Using Theorem 5.20 of [31], we know that the optimal transport plan π_n between $\mathbb{P}^x$ and $\mathbb{P}^{y_n}$ converges, when the cost function is the continuous function $U_c(m)$, up to extraction of a subsequence, to an optimal transport plan π for $\mathbb{P}^x$ and $\mathbb{P}^y$. This means that $\mathcal{G}_c(m)(x, y_n)$ converges up to extraction of a subsequence, to $\mathcal{G}_c(m)(x, y)$.

However, the whole sequence $(\mathcal{G}_c(m)(x, y_n))_{n\in\mathbb{N}}$ converges to $\mathcal{G}_c(m)(x, y)$ for the following reason: assume it is not the case. Since $0 \leq \mathcal{G}_c(m)(x, y_n) \leq 1$ for every $n \in \mathbb{N}$, then there exists another subsequence $(z_k)_{k\in\mathbb{N}}$ such that the sequence $(\mathcal{G}_c(m)(x, z_k))_{k\in\mathbb{N}}$ converges to a limit $l \neq \mathcal{G}_c(m)(x, y)$. We also have that $\lim_{k\longrightarrow\infty} z_k = y$, hence using Theorem 5.20 of [31], there is a subsequence $(a_j)_{j\in\mathbb{N}}$ of $(z_k)_{k\longrightarrow\infty}$ such that $\lim_{j\longrightarrow\infty} \mathcal{G}_c(m)(x, a_j) = \mathcal{G}_c(m)(x, y) \neq l$ which yields a contradiction. ∎

9.2.5 Defining our family of pseudometrics

Lemma 9.12 enables us to define for each $0 < c < 1$ the following sequence of pseudometrics on E:

$$\begin{aligned}d_0^c(x, y) &= |obs(x) - obs(y)|, \\ d_{n+1}^c &= \mathcal{G}_c(d_n^c).\end{aligned}$$

And finally we can define $\overline{d}^c = \sup_{n\in\mathbb{N}} d_n^c$ (which is also a limit since the sequence is non-decreasing using Lemma 9.10). As a direct consequence of Lemma A.1, the pseudometric $\overline{d}^c$ is lower semi-continuous and is thus in the lattice $\mathcal{P}$ for any $0 < c < 1$.

9.2.6 Fixpoints

Similarly to $\mathcal{F}_c$ and $\overline{\delta}^c$, we can show that $\overline{d}^c$ is a fixpoint for $\mathcal{G}_c$.

Since $\mathcal{G}_c(\overline{d}^c)$ is defined as the optimal transport cost when the transport function is $U_c(\overline{d}^c)$, in order to study $\mathcal{G}_c(\overline{d}^c)$, we need to relate $U_c(\overline{d}^c)$ to $U_c(d_n^c)$ which the next lemma does.

Lemma 9.13. For two trajectories ω, ω',

$$U_c(\overline{d}^c)(\omega,\omega') = \lim_{n\longrightarrow\infty} U_c(d_n^c)(\omega,\omega') = \sup_n U_c(d_n^c)(\omega,\omega').$$

Proof. We will omit writing c as an index for the pseudometrics d_n and $\overline{d}$ throughout this proof.

First note that $U_c(d_{n+1}) \geq U_c(d_n)$ since U_c is monotone which shows the second equality. Besides,

$$\begin{aligned}
U_c(\overline{d})(\omega,\omega') &= \sup_{t\geq 0} c^t \overline{d}(\omega(t),\omega'(t)) \\
&= \sup_{t\geq 0} c^t \sup_n d_n(\omega(t),\omega'(t)) \\
&= \sup_n \sup_{t\geq 0} c^t d_n(\omega(t),\omega'(t)) \\
&= \sup_n U_c(d_n)(\omega,\omega').
\end{aligned}$$

∎

Theorem 9.14. The pseudometric $\overline{d}^c$ is a fixpoint of $\mathcal{G}_c$.

Proof. We will omit writing c as an index for the pseudometrics d_n and $\overline{d}$ throughout this proof. Fix two states x, y.

The space of finite measures on $\Omega\times\Omega$ is a linear topological space. Using Lemma 2.3, we know that the set of couplings $\Gamma(\mathbb{P}^x, \mathbb{P}^y)$ is a compact subset. Besides, it is also convex.

The space of bounded pseudometrics on E is also a linear topological space. We have defined a sequence $d_0, d_1, \ldots$ on that space. We define Y to be the set of linear combinations of pseudometrics $\sum_{n\in\mathbb{N}} a_n d_n$ such that for every n, $a_n \geq 0$ (and finitely many are non-zero) and $\sum_{n\in\mathbb{N}} a_n = 1$. This set Y is convex.

Define the function

$$\begin{aligned}\Xi : \Gamma(\mathbb{P}^x, \mathbb{P}^y) \times Y &\longrightarrow [0,1] \\ (\gamma, m) &\mapsto \int U_c(m) \ \mathrm{d}\gamma.\end{aligned}$$

For $\gamma \in \Gamma(\mathbb{P}^x, \mathbb{P}^y)$, the map $\Xi(\gamma, \cdot)$ is continuous by dominated convergence theorem and it is monotone and hence quasiconcave. For a given $m \in Y$, by definition of the Lebesgue integral, $\Xi(\cdot, m)$ is continuous and linear.

We can therefore apply Sion's minimax theorem:

$$\min_{\gamma\in\Gamma(\mathbb{P}^x,\mathbb{P}^y)} \sup_{m\in Y} \int U_c(m) \ \mathrm{d}\gamma = \sup_{m\in Y} \min_{\gamma\in\Gamma(\mathbb{P}^x,\mathbb{P}^y)} \int U_c(m) \ \mathrm{d}\gamma. \tag{10}$$

Similarly to what was shown in the proof of Theorem 5.5, for an arbitrary functional Ψ on Y such that $m \leq m' \Rightarrow \Psi(m) \leq \Psi(m')$,

$$\sup_{m\in Y} \Psi(m) = \sup_{n\in\mathbb{N}} \Psi(d_n). \tag{11}$$

We can now go back to Equation (10). The right-hand side of the equation is

$$\begin{aligned}\sup_{m\in Y} \min_{\gamma\in\Gamma(\mathbb{P}^x,\mathbb{P}^y)} \int U_c(m) \ \mathrm{d}\gamma &= \sup_{m\in Y} \mathcal{G}_c(m)(x,y) \\ &= \sup_{n\in\mathbb{N}} \mathcal{G}_c(d_n)(x,y) \quad \text{(previous result in Equation (11))} \\ &= \sup_{n\in\mathbb{N}} d_{n+1}(x,y) \quad \text{(definition of } (d_n)_{n\in\mathbb{N}}) \\ &= \overline{d}(x,y).\end{aligned}$$

The left-hand side of Equation (10) is

$$\begin{aligned}\min_{\gamma\in\Gamma(\mathbb{P}^x,\mathbb{P}^y)} \sup_{m\in Y} \int U_c(m) \ \mathrm{d}\gamma &= \min_{\gamma\in\Gamma(\mathbb{P}^x,\mathbb{P}^y)} \sup_{n\in\mathbb{N}} \int U_c(d_n) \ \mathrm{d}\gamma \quad \text{(previous result in Equation (11))} \\ &= \min_{\gamma\in\Gamma(\mathbb{P}^x,\mathbb{P}^y)} \int \sup_{n\in\mathbb{N}} U_c(d_n) \ \mathrm{d}\gamma \quad \text{(dominated convergence theorem)} \\ &= \min_{\gamma\in\Gamma(\mathbb{P}^x,\mathbb{P}^y)} \int U_c(\sup_{n\in\mathbb{N}} d_n) \ \mathrm{d}\gamma \quad \text{(Lemma 9.13)} \\ &= \min_{\gamma\in\Gamma(\mathbb{P}^x,\mathbb{P}^y)} \int U_c(\overline{d}) \ \mathrm{d}\gamma \\ &= \mathcal{G}_c(\overline{d})(x,y).\end{aligned}$$

This concludes the proof since we get that $\mathcal{G}_c(\overline{d})(x, y) = \overline{d}(x, y)$ for every two states x and y. ∎

Using Corollary 9.11 and Lemma 5.6, we directly obtain the following:

Corollary 9.15. The pseudometric $\overline{d}^c$ is a fixpoint for $\mathcal{F}_c$ for every $0 < c < 1$ and hence $\overline{\delta}^c \leq \overline{d}^c$.

The following lemma is similar to Lemma 5.6 and the proof is essentially the same.

Lemma 9.16. Consider a discount factor $0 < c < 1$ and a pseudometric m in $\mathcal{P}$ such that

1. m is a fixpoint for $\mathcal{G}_c$,
2. for every x, y, $m(x, y) \geq |obs(x) - obs(y)|$,

then $m \geq \overline{d}^c$.

What this lemma shows is the following important result:

Theorem 9.17. The pseudometric $\overline{d}^c$ is the least fixpoint of $\mathcal{G}_c$ that is greater than the pseudometric $(x, y) \mapsto |obs(x) - obs(y)|$.

9.3 A real-valued logic

Similarly to Section 6, the pseudometric $\overline{d}^c$ obtained from the functional $\mathcal{G}_c$ can also be described through a real-valued logic. It is worth noting that this logic needs to handle both states and trajectories and is thus defined in two parts: $\mathcal{L}_\sigma$ and $\mathcal{L}_\tau$.

9.3.1 Definition of the real-valued logic

Definition of the logics: This logic is defined inductively in two parts as follows:

$$f \in \mathcal{L}_\sigma = q \mid obs \mid 1 - f \mid \int g,$$
$$g \in \mathcal{L}_\tau = f \circ ev_t \mid \min\{g_1, g_2\} \mid \max\{g_1, g_2\} \mid g \ominus q \mid g \oplus q,$$

where $q \in \mathbb{Q} \cap [0, 1]$, $f \in \mathcal{L}_\sigma$, $g, g_1, g_2 \in \mathcal{L}_\tau$ and $t \in \mathbb{Q}_{\geq 0}$.

Interpretation of the logics: For a fixed discount factor $0 < c < 1$, the expressions in $\mathcal{L}_\sigma$ are interpreted as functions $E \longrightarrow [0,1]$ and expressions in $\mathcal{L}_\tau$ are interpreted as functions $\Omega \longrightarrow [0,1]$. The first terms of $\mathcal{L}_\sigma$ are interpreted as the corresponding ones in Λ: for a state $x \in E$,

$$\begin{aligned} q(x) &= q, \\ obs(x) &= obs(x), \\ (1-f)(x) &= 1 - f(x), \\ \left(\int g\right)(x) &= \int g \ \mathrm{d}\mathbb{P}^x. \end{aligned}$$

The remaining expressions in $\mathcal{L}_\tau$ are interpreted for a trajectory ω as

$$\begin{aligned} (f \circ ev_t)(\omega) &= c^t f(\omega(t)), \\ (\min\{g_1, g_2\})(\omega) &= \min\{g_1(\omega), g_2(\omega)\}, \\ (\max\{g_1, g_2\})(\omega) &= \max\{g_1(\omega), g_2(\omega)\}, \\ (g \ominus q)(\omega) &= \max\{0, g(\omega) - q\}, \\ (g \oplus q)(\omega) &= \min\{1, g(\omega) + q\}. \end{aligned}$$

Whenever we want to emphasize the fact that the expressions are interpreted for a certain discount factor $0 < c < 1$, we will write $\mathcal{L}_\sigma^c$ and $\mathcal{L}_\tau^c$.

Additional useful expressions From $\mathcal{L}_\tau$, we can define additional expressions in $\mathcal{L}_\sigma$ as follows:

$$\begin{aligned} f \ominus q &= \int [(f \circ ev_0) \ominus q], \\ \min\{f_1, f_2\} &= \int \min\{f_1 \circ ev_0, f_2 \circ ev_0\}, \\ \max\{f_1, f_2\} &= 1 - \min\{1 - f_1, 1 - f_2\}, \\ f \oplus q &= 1 - ((1 - f) \ominus q). \end{aligned}$$

Their interpretations as functions $E \longrightarrow [0,1]$ are the same as for Λ.

9.3.2 Definition of the pseudometric

The metric we derive from the logic $\mathcal{L}_\sigma$ corresponds to how different the test results are when the process starts from x compared to when it starts from y.

Given a fixed discount factor $0 < c < 1$, we can define the pseudometric ℓ^c:

$$\ell^c(x, y) = \sup_{f \in \mathcal{L}^c_\sigma} |f(x) - f(y)| = \sup_{f \in \mathcal{L}^c_\sigma} (f(x) - f(y)).$$

The last equality holds since for every $f \in \mathcal{L}_\sigma$, $\mathcal{L}_\sigma$ also contains $1 - f$.

9.3.3 Comparison to the fixpoint metric

In Theorem 6.9, we proved that the real-valued logic Λ generated a pseudometric that was equal to $\overline{\delta}^c$. Such a result also holds for $\mathcal{L}^c_\sigma$ and $\overline{d}^c$, but the proof is more complex as the trajectories play a bigger role. Indeed, the difficulty is now to approximate functions in $\mathcal{H}(U_c(\ell^c))$ by functions in $\mathcal{L}^c_\tau$.

Lemma 9.18. For any expression $f \in \mathcal{L}_\sigma$, the function $x \mapsto f(x)$ is continuous and for any expression $g \in \mathcal{L}_\tau$, the function $\omega \mapsto g(\omega)$ is continuous.

Proof. This is done by induction on the structure of f and g. There are two cases that are not straightforward.

- First, if $f = \int g$ with $g \in \mathcal{L}_\tau$, then, since $x \mapsto \mathbb{P}^x$ is continuous and g is continuous in Ω, we get that $x \mapsto \int g \, d\mathbb{P}^x$ is continuous in E.
- Second, if $g = f \circ ev_t$ with $f \in \mathcal{L}_\sigma$ and $t \in \mathbb{Q}_{\geq 0}$, then since trajectories are continuous and f is continuous on the state space, we get the desired result.

∎

Approximating functions on trajectories: We will later on use Lemma 6.4 to show that a function on trajectories in $\mathcal{H}(U_c(\ell^c))$ can be approximated by functions in $\mathcal{L}_\tau$. We first need to look at pairs of trajectories.

Lemma 9.19. Consider a function $h : \Omega \longrightarrow [0, 1]$ and two trajectories ω, ω' such that there exists f in the logic $\mathcal{L}_\sigma$ and a time $s \in \mathbb{Q}_{\geq 0}$ such that $|h(\omega) - h(\omega')| \leq c^s |f(\omega(s)) - f(\omega'(s))|$. Then for every $\delta > 0$, there exists g in the logic $\mathcal{L}_\tau$ such that $|h(\omega) - g(\omega)| \leq 2\delta$ and $|h(\omega') - g(\omega')| \leq 2\delta$.

Proof. WLOG $h(\omega) \geq h(\omega')$ and $f(\omega(s)) \geq f(\omega'(s))$ (otherwise consider $1 - f$ instead of f).

Pick $p, q, r \in \mathbb{Q} \cap [0, 1]$ such that

$$\begin{aligned}
&p \in [c^s f(\omega'(s)) - \delta, c^s f(\omega'(s))], \\
&q \in [(h(\omega) - h(\omega') - \delta, h(\omega) - h(\omega')], \\
&r \in [h(\omega'), h(\omega') + \delta].
\end{aligned}$$

Define $g = (\min\{(f \circ ev_s) \ominus p, q\}) \oplus r$. Then,

$$\begin{aligned}
(f \circ ev_s)(\omega) &= c^s f(\omega(s)), \\
((f \circ ev_s) \ominus p)(\omega) &= \max\{0, c^s f(\omega(s)) - p\} \\
&\in [c^s[f(\omega(s)) - f(\omega'(s))], c^s[f(\omega(s)) - f(\omega'(s))] + \delta] \text{ since } f(\omega(s)) \geq \\
(\min\{(f \circ ev_s) \ominus p, q\})(\omega) &= q \text{ since } q \leq h(\omega) - h(\omega') \leq c^s[f(\omega(s)) - f(\omega'(s))], \\
q + r &\in [h(\omega) - \delta, h(\omega) + \delta], \\
g(\omega) &= \min\{1, q + r\} \in [h(\omega) - \delta, h(\omega) + \delta] \text{ as } h(\omega) \leq 1.
\end{aligned}$$

Meanwhile,

$$\begin{aligned}
(f \circ ev_s)(\omega') &= c^s f(\omega'(s)), \\
((f \circ ev_s) \ominus p)(\omega') &= \max\{0, c^s f(\omega'(s)) - p\} \in [0, \delta], \\
(\min\{(f \circ ev_s) \ominus p, q\})(\omega') &\in [0, \delta], \\
g(\omega') &= \min\{1, (\min\{(f \circ ev_s) \ominus p, q\})(\omega') + r\} \\
&\in [h(\omega'), h(\omega') + 2\delta].
\end{aligned}$$

∎

Corollary 9.20. Consider a continuous function $h : \Omega \longrightarrow [0, 1]$ such that for any two trajectories ω, ω', there exists f in the logic and a time $s \in \mathbb{Q}$ such that $|h(\omega) - h(\omega')| \leq c^s|f(\omega(s)) - f(\omega'(s))|$. Then the function h can be approximated by functions (that are interpretations of expressions) in $\mathcal{L}_\tau$.

Proof. We have proven in Lemma 9.19 that such a function h can be approximated at pairs of trajectories by functions from $\mathcal{L}_\tau$ and $t \in \mathbb{Q}_{\geq 0}$.

Since the function h and all functions in $\mathcal{L}_\tau$ are continuous and since $\mathcal{L}_\tau$ is closed under min and max, we can apply Lemma 6.4, and we get that the function h can be approximated by functions in $\mathcal{L}_\tau$. ∎

A fixpoint of $\mathcal{G}$: Using previously proven lemmas, the proof is quite similar to the one in Section 6 in order to show that ℓ^c is a fixpoint of $\mathcal{G}_c$.

Theorem 9.21. The pseudometric ℓ^c is a fixpoint of $\mathcal{G}_c$: $\ell^c = \mathcal{G}_c(\ell^c)$.

Proof. We will omit writing the index c in this proof.

We already know that $\mathcal{G}(\ell) \geq \ell$ (Lemma 9.10) and we only need to prove the other inequality. Pick two states x, y and recall that

$$\mathcal{G}(\ell)(x, y) = W(U_c(\ell))(\mathbb{P}^x, \mathbb{P}^y) \text{ where } U_c(\ell)(\omega, \omega') = \sup_{t \geq 0, f \in \mathcal{L}_\sigma} c^t|f(\omega(t)) - f(\omega'(t))|.$$

Since for any $f \in \mathcal{L}_\sigma$ the map $z \mapsto f(z)$ is continuous (Lemma 9.18) and the trajectories are continuous, it is enough to consider rational times only and we have that

$$U_c(\ell)(\omega, \omega') = \sup_{t \in \mathbb{Q}_{\geq 0}, f \in \mathcal{L}_\sigma} c^t |f(\omega(t)) - f(\omega'(t))|.$$

Now, both the logic $\mathcal{L}_\sigma$ and the set of non-negative rationals $\mathbb{Q}_{\geq 0}$ are countable, so we can enumerate their elements as $\mathcal{L}_s = (f_k)_{k \in \mathbb{N}}$ and $\mathbb{Q}_{\geq 0} = (s_k)_{k \in \mathbb{N}}$. Define the map

$$c_k(\omega, \omega') = \max_{i=1,\dots,k \quad j=1,\dots,k} c^{s_i} |f_j(\omega(s_i)) - f_j(\omega'(s_i))|.$$

Note that $\lim_{k \longrightarrow \infty} c_k(\omega, \omega') = U_c(\ell)(\omega, \omega')$ for any two trajectories ω, ω'. By Lemma 2.5, we get that $\lim_{k \longrightarrow \infty} W(c_k)(\mathbb{P}^x, \mathbb{P}^y) = W(U_c(\ell))(\mathbb{P}^x, \mathbb{P}^y)$.

It is thus enough to study $W(c_k)(\mathbb{P}^x, \mathbb{P}^y)$. Pick $k \in \mathbb{N}$. There exists h such that

$$W(c_k)(\mathbb{P}^x, \mathbb{P}^y) = \int h \, \mathrm{d}\mathbb{P}^x - \int h \, \mathrm{d}\mathbb{P}^y \text{ and } \forall \omega, \omega' \ |h(\omega) - h(\omega')| \leq c_k(\omega, \omega').$$

Using Lemma 9.18, we know that c_k is continuous (as a maximum of finitely many continuous functions) which means that h is also continuous.Using the second condition about h and the definition of c_k, we know that for every pair of trajectories ω, ω' there exists i and $j \leq k$ such that $|h(\omega) - h(\omega')| \leq c^{s_i} |f_j(\omega(s_i)) - f_j(\omega'(s_i))|$. We can thus apply Theorem 9.21 and we thus know that h can be approximated by functions that are interpretations of expressions in $\mathcal{L}_\tau$. This means in particular that

$$\begin{aligned}
\int h \, \mathrm{d}\mathbb{P}^x - \int h \, \mathrm{d}\mathbb{P}^y &\leq \sup_{g \in \mathcal{L}_\tau} \left(\int g \, \mathrm{d}\mathbb{P}^x - \int g \, \mathrm{d}\mathbb{P}^y \right) \\
&= \sup_{g \in \mathcal{L}_\tau} \left[\left(\int g \right)(x) - \left(\int g \right)(y) \right] \\
&\leq \sup_{f \in \mathcal{L}_\sigma} |f(x) - f(y)| = \ell(x, y).
\end{aligned}$$

Moving back to $W(U(\ell))$, we thus get that

$$W(U(\ell))(\mathbb{P}^x, \mathbb{P}^y) = \lim_{k \longrightarrow \infty} W(c_k)(\mathbb{P}^x, \mathbb{P}^y) \leq \ell(x, y)$$

which concludes the proof. ∎

Comparison to the fixpoint metric: We are finally ready to compare the metric ℓ^c obtained through the logic $\mathcal{L}_\sigma$ and $\mathcal{L}_\tau$ to the metric $\overline{d}^c$ obtained through the functional $\mathcal{G}_c$.

Lemma 9.22. For every $f \in \mathcal{L}_\sigma^c$, there exists $n \in \mathbb{N}$ such that $f \in \mathcal{H}(d_n^c)$, i.e. for every states x, y,

$$|f(x) - f(y)| \leq d_n^c(x, y).$$

For every $g \in \mathcal{L}_\tau^c$, there exists $n \in \mathbb{N}$ such that $g \in \mathcal{H}(U_c(d_n^c))$, i.e. for every pair of trajectories ω, ω',

$$|g(\omega) - g(\omega')| \leq U_c(d_n^c)(\omega, \omega').$$

Proof. The proof is done by induction on the structure of the logic $\mathcal{L}_\sigma$ and $\mathcal{L}_\tau$ and quite similar to that of Lemma 6.2. We will only mention the different cases here.

- If $f = \int g$ with $g \in \mathcal{H}(U_c(d_n^c))$, then

$$\begin{aligned}|f(x) - f(y)| &= \left| \int g \ \mathrm{d}\mathbb{P}^x - \int g \ \mathrm{d}\mathbb{P}^y \right| \\ &\leq W(U_c(d_n^c))(\mathbb{P}^x, \mathbb{P}^y) \\ &= \mathcal{G}_c(d_n^c)(x, y) = d_{n+1}^c(x, y).\end{aligned}$$

- If $g = \min\{g_1, g_2\}$ with $g_i \in \mathcal{H}(U_c(d_{n_i}^c))$, then for any two trajectories ω, ω', there really is only one case to consider: $g_1(\omega) \leq g_2(\omega)$ and $g_2(\omega') \leq g_1(\omega')$. In that case,

$$\begin{aligned}|g(\omega) - g(\omega')| &= |g_1(\omega) - g_2(\omega')| \\ &= \max\{g_1(\omega) - g_2(\omega'), g_2(\omega') - g_1(\omega)\} \\ &\leq \max\{g_2(\omega) - g_2(\omega'), g_1(\omega') - g_1(\omega)\} \\ &\leq \max\{U_c(d_{n_2}^c)(\omega, \omega'), U_c(d_{n_1}^c)(\omega, \omega')\} \\ &\leq U_c(d_n^c)(\omega, \omega')\end{aligned}$$

  with $n = \max\{n_1, n_2\}$, which means that $g \in \mathcal{H}(U_c(d_n^c))$.

- If $g = f \circ ev_t$ with $f \in \mathcal{H}(d_n^c)$, then for any two trajectories ω, ω',

$$\begin{aligned}|g(\omega) - g(\omega')| &= \left|c^t f(\omega(t)) - c^t f(\omega'(t))\right| \\ &\leq c^t d_n^c(\omega(t), \omega'(t)) \\ &\leq U_c(d_n^c)(\omega, \omega').\end{aligned}$$

■

Finally, we obtain the equality of the pseudometrics defined through the fixpoint $\mathcal{G}_c$ and the logics $\mathcal{L}^c_\sigma$ and $\mathcal{L}^c_\tau$.

Theorem 9.23. The two pseudometrics ℓ^c and $\overline{d}^c$ are equal.

Proof. We know from Lemma 9.16 and Theorem 9.21 that $\ell^c \geq \overline{d}^c$.

The other direction is a direct consequence of Lemma 9.22. ■

10 Conclusion

In our previous work [9, 10], we showed that we needed to use trajectories in order to define a meaningful notion of behavioural equivalence. However working with trajectories is extremely complex as notions do not translate easily from states to trajectories; for instance, we said that a measurable set B of trajectories is time-*obs*-closed if for every two trajectories ω, ω' such that for every time t, $obs(\omega(t)) = obs(\omega'(t))$, then $\omega \in B$ if, and only if $\omega' \in B$. The σ-algebra of all time-*obs*-closed sets cannot be simply described.

To explain why this present work does not deal with trajectories, we need to first discuss the example that lead us to trajectories in [9]: consider Brownian motion on the real line equipped with the function $obs = \mathbb{1}_{\{0\}}$. There are four *obs*-closed sets: $\emptyset, \{0\}, \mathbb{R} \setminus \{0\}, \mathbb{R}$ and for any $x \neq 0$ and any time t, $P_t(x, \{0\}) = 0$. This meant that one could not distinguish between the states 1 and 1000 in this specific case. Using trajectories enabled us to consider intervals of time. In this current work, we have decided to instead "smooth" the function *obs* so as to prevent singling points out without needing to deal with trajectories.

Let us go back to our examples. As shown in those examples, computing $\overline{\delta}$ is quite an involved process. It would have been interesting to adapt our example in Section 8.1 to the real-line with $obs(x) = e^{-x^2}$ and consider other processes such as Brownian motion which stops once it hits 0 or the Ornstein–Uhlenbeck process (a variation of Brownian motion which is "attracted" to 0). Having to deal with transport theory, a supremum (over time) and the inductive definition of $\overline{\delta}$ makes it virtually impossible to compute any harder example. As we have seen in the second example (Section 8.2), it may still be possible to compare the behaviours of two states by stating for instance that "the behaviour of the process starting from x is closer to that starting from y than from z". The problem of transport theory and the Kantorovich metric also exists in discrete time and there are interesting ways to deal with it, for instance through the MICo distance [8].

One of the advantages of optimal transport is that the Kantorovich duality gives us a way of computing bounds. As one notes however, there is again some difficulty with having a supremum over time in the definition of our functional $\mathcal{F}_c$. In particular, we can only provide lower bounds for $\overline{\delta}$.

One avenue of work on this could be to study replacements for this supremum, such as integrals over time for instance. Note that the real-valued logic would also need to be adapted even though it seems to generalize discrete-time really well.

This work is, as far as we know, the first behavioural metric adapted to the continuous-time case. Clearly much remains to be done, particularly the exploration of examples and connexions to broader classes of processes, such as for example those defined by stochastic differential equations.

References

[1] R. B. Ash. *Real Analysis and Probability.* Academic Press, 1972.

[2] C. Baier and J.-P. Katoen. *Principles of Model Checking.* MIT Press, 2008.

[3] P. Billingsley. *Convergence of Probability Measures.* Wiley Interscience, 2nd edition, 1999.

[4] R. Blute, J. Desharnais, A. Edalat, and P. Panangaden. Bisimulation for labelled Markov processes. In *Proceedings of the Twelfth IEEE Symposium On Logic In Computer Science, Warsaw, Poland.*, 1997.

[5] A. Bobrowski. *Functional analysis for probability and stochastic processes: an introduction.* Cambridge University Press, 2005.

[6] A. N. Borodin and P. Salminen. Handbook of brownian motion-facts and formulae. *Journal of the Royal Statistical Society-Series A Statistics in Society*, 160(3):596, 1997.

[7] N. Bourbaki. *General Topology Chapters 5-10.* Springer-Verlag, 1989.

[8] P. S. Castro, T. Kastner, P. Panangaden, and M. Rowland. Mico: Improved representations via sampling-based state similarity for markov decision processes. *Advances in Neural Information Processing Systems*, 34:30113–30126, 2021.

[9] L. Chen, F. Clerc, and P. Panangaden. Bisimulation for feller-dynkin processes. *Electronic Notes in Theoretical Computer Science*, 347:45 – 63, 2019. Proceedings of the Thirty-Fifth Conference on the Mathematical Foundations of Programming Semantics.

[10] L. Chen, F. Clerc, and P. Panangaden. Towards a classification of behavioural equivalences in continuous-time Markov processes. *Electronic Notes in Theoretical Computer Science*, 2020. Proceedings of the Thirty-Sixth Conference on the Mathematical Foundations of Programming Semantics.

[11] L. Chen, F. Clerc, and P. Panangaden. Behavioural equivalences for continuous-time markov processes. *Mathematical Structures in Computer Science*, page 1–37, 2023.

[12] L. Chen, F. Clerc, and P. Panangaden. Behavioural pseudometrics for continuous-time diffusions. arXiv 2312.16729, 2024.

[13] J. Desharnais, A. Edalat, and P. Panangaden. A logical characterization of bisimulation for labelled Markov processes. In *proceedings of the 13th IEEE Symposium On Logic In Computer Science, Indianapolis*, pages 478–489. IEEE Press, June 1998.

[14] J. Desharnais, A. Edalat, and P. Panangaden. Bisimulation for labeled Markov processes. *Information and Computation*, 179(2):163–193, Dec 2002.

[15] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for labeled Markov systems. In *Proceedings of CONCUR99*, number 1664 in Lecture Notes in Computer Science. Springer-Verlag, 1999.

[16] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. The metric analogue of weak bisimulation for labelled Markov processes. In *Proceedings of the Seventeenth Annual IEEE Symposium On Logic In Computer Science*, pages 413–422, July 2002.

[17] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. A metric for labelled Markov processes. *Theoretical Computer Science*, 318(3):323–354, June 2004.

[18] N. Ferns, P. Panangaden, and D. Precup. Bisimulation metrics for continuous Markov decision processes. *SIAM Journal of Computing*, 40(6):1662–1714, 2011.

[19] V. Gupta, R. Jagadeesan, and P. Panangaden. Approximate reasoning for real-time probabilistic processes. In *The Quantitative Evaluation of Systems, First International Conference QEST04*, pages 304–313. IEEE Press, 2004.

[20] V. Gupta, R. Jagadeesan, and P. Panangaden. Approximate reasoning for real-time probabilistic processes. *Logical Methods in Computer Science*, 2(1):paper 4, 2006.

[21] I. Karatzas and S. Shreve. *Brownian motion and stochastic calculus*, volume 113. Springer Science and Business Media, 2012.

[22] K. G. Larsen and A. Skou. Bisimulation through probablistic testing. *Information and Computation*, 94:1–28, 1991.

[23] R. Milner. *A Calculus for Communicating Systems*, volume 92 of *Lecture Notes in Computer Science*. Springer-Verlag, 1980.

[24] B. Oksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.

[25] P. Panangaden. *Labelled Markov Processes.* Imperial College Press, 2009.

[26] D. Park. Concurrency and automata on infinite sequences. In *Proceedings of the 5th GI Conference on Theoretical Computer Science*, number 104 in Lecture Notes In Computer Science, pages 167–183. Springer-Verlag, 1981.

[27] L. C. G. Rogers and D. Williams. *Diffusions, Markov processes and martingales: Volume 1. Foundations.* Cambridge university press, 2nd edition, 2000.

[28] H. Royden and P. Fitzpatrick. Real analysis 4th edition. *Printice-Hall Inc, Boston*, 2010.

[29] D. Sangiorgi. On the origins of bisimulation and coinduction. *ACM Transactions on Programming Languages and Systems (TOPLAS)*, 31(4):15, 2009.

[30] F. van Breugel and J. Worrell. A behavioural pseudometric for probabilistic transition systems. *Theoretical Computer Science*, 331(1):115 – 142, 2005.

[31] C. Villani. *Optimal transport: old and new.* Springer-Verlag, 2008.

[32] W. Whitt. *An Introduction to Stochastic-Process Limits and their Applications to Queues.* Springer Series in Operations Research. Springer-Verlag, 2002.

A Proofs for Section 2

A.1 Lower semi-continuity

Lemma A.1. Assume there is an arbitrary family of continuous functions $f_i : X \longrightarrow \mathbb{R}$ $(i \in \mathcal{I})$ and define $f(x) = \sup_{i\in\mathcal{I}} f_i(x)$ for every $x \in X$. Then f is lower semi-continuous.

Proof. Pick $r \in \mathbb{R}$. Then

$$\begin{aligned} f^{-1}((r, +\infty)) &= \{x \mid \sup_{i\in\mathcal{I}} f_i(x) > r\} \\ &= \{x \mid \exists i \in \mathcal{I}\ f_i(x) > r\} = \bigcup_{i\in\mathcal{I}} f_i^{-1}((r, +\infty)). \end{aligned}$$

Since f_i is continuous, the set $f_i^{-1}((r+\infty))$ is open and therefore $f^{-1}((r, +\infty))$ is open which concludes the proof. □ ∎

The converse is also true, as Baire's theorem states:

Theorem A.2. If X is a metric space and if a function $f : X \longrightarrow \mathbb{R}$ is lower semi-continuous, then f is the limit of an increasing sequence of real-valued continuous functions on X.

A.2 Couplings

In order to prove Lemma 2.3, we will need to prove some more results first.

Proposition A.3. If P and Q are two measures on a Polish space X such that for every continuous and bounded function $f : X \longrightarrow \mathbb{R}$, we have $\int f \ \mathrm{d}P = \int f \ \mathrm{d}Q$, then $P = Q$.

Proof. For any open set U, its indicator function $\mathbb{1}_U$ is lower semicontinuous, which means that there exists an increasing sequence of continuous functions f_n converging pointwise to $\mathbb{1}_U$ (see Theorem A.2). Without loss of generality, we can assume that the functions f_n are non negative (consider the sequence $\max\{0, f_n\}$ instead if they are not non-negative). Using the monotone convergence theorem, we know that

$$\lim_{n\longrightarrow\infty} \int f_n \ \mathrm{d}P = \int \mathbb{1}_U \ \mathrm{d}P = P(U)$$

and similarly for Q. By our hypothesis on P and Q, we obtain that for every open set U, $P(U) = Q(U)$.

Since P and Q agree on the topology, they agree on the Borel algebra by Caratheodory's extension theorem. □ ∎

Lemma A.4. Given two probability measures P and Q on a Polish space X, the set of couplings $\Gamma(P, Q)$ is closed in the weak topology.

Proof. Consider a sequence $(\gamma_n)_{n\in\mathbb{N}}$ of couplings of P and Q weakly converging to a measure μ on $X \times X$, meaning that for every bounded continuous functions $f : X \times X \longrightarrow \mathbb{R}$, $\lim_{n\longrightarrow\infty} \int f \, \mathrm{d}\gamma_n = \int f \, \mathrm{d}\mu$. Note that μ is indeed a probability measure. We only have to prove that the marginals of μ are P and Q. We will only do it for the first one, i.e. showing that for every measurable set A, $\mu(A \times X) = P(A)$ since the case for Q works the same.

Let $\pi : X \times X \longrightarrow X$ be the first projection map. The first marginal of γ_n and μ are obtained as the pushforward measures $\pi_*\gamma_n$ and $\pi_*\mu$. Indeed, $\pi_*\mu$ is defined for all measurable set $A \subset X$ as $\pi_*\mu(A) = \mu(\pi^{-1}(A)) = \mu(A \times X)$ (and similarly for the γ_n's). This means that for every n, $\pi_*\gamma_n = P$.

For an arbitrary continuous bounded function $f : X \longrightarrow \mathbb{R}$, define the function $g : X \times X \longrightarrow \mathbb{R}$ as $g = f \circ \pi$, i.e. $g(x, y) = f(x)$. The function g is also continuous and bounded, so by weak convergence of the sequence $(\gamma_n)_{n\in\mathbb{N}}$ to μ, we get that

$$\lim_{n\longrightarrow\infty} \int f \circ \pi \, \mathrm{d}\gamma_n = \int f \circ \pi \, \mathrm{d}\mu.$$

Using the change of variables formula, we obtain that for any continuous bounded function f

$$\int f \, \mathrm{d}P = \lim_{n\longrightarrow\infty} \int f \, \mathrm{d}(\pi_*\gamma_n) = \int f \, \mathrm{d}(\pi_*\mu).$$

Using Proposition A.3, we get that $\pi_*\mu = P$. □ ■

Lemma A.5. Consider two probability measures P and Q on Polish spaces X and Y respectively. Then the set of couplings $\Gamma(P, Q)$ is tight: for all $\epsilon > 0$, there exists a compact set K in $X \times Y$ such that for all coupling $\gamma \in \Gamma(P, Q)$, $\gamma(K) > 1 - \epsilon$

Proof. First note that by Ulam's tightness theorem, the probability measures P and Q are tight.

Consider $\epsilon > 0$. Since P and Q are tight, there exist two compact sets K and K' such that $P(X \setminus K) \leq \epsilon/2$ and $Q(Y \setminus K') \leq \epsilon/2$. Define the set $C = K \times K'$. This set is compact as a product of two compact sets. Furthermore, for every $\gamma \in \Gamma(P, Q)$,

$$\begin{aligned}\gamma((X \times Y) \setminus C) &= \gamma\left([(X \setminus K) \times Y] \cup [X \times (Y \setminus K')]\right) \\ &\leq \gamma[(X \setminus K) \times Y] + \gamma[X \times (Y \setminus K')] \\ &= P(X \setminus K) + Q(Y \setminus K') \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon.\end{aligned}$$

This shows that the set $\Gamma(P, Q)$ is tight. □ ■

We can now prove Lemma 2.3. Let us start by restating it.

Lemma A.6. Given two probability measures P and Q on Polish spaces X and Y respectively, the set of couplings $\Gamma(P, Q)$ is compact under the topology of weak convergence.

Proof. Using Lemma A.5, we know that the set $\Gamma(P, Q)$ is tight. Applying Prokhorov's theorem (see Theorem A.7), we get that the set $\Gamma(P, Q)$ is precompact.

Since the set $\Gamma(P, Q)$ is also closed (see Lemma A.4), then it is compact. □ ■

We have used Prokhorov's theorem in the previous proof. Here is the version cited in [31]

Theorem A.7. Given a Polish space $\mathcal{X}$, a subset $\mathcal{P}$ of the set of probabilities on $\mathcal{X}$ is precompact for the weak topology if and only if it is tight.

A.3 Optimal transport theory

Remark A.8. The expression that we provided for the Kantorovich duality is not the exact expression in Theorem 5.10 of [31] but the one found in Particular Case 5.16. The former expression also applies to our case. Indeed, according to Theorem 5.10, the dual expression is

$$\min_{\gamma \in \Gamma(\mu,\nu)} \int c \ \mathrm{d}\gamma = \max_{\phi,\psi} \left(\int \phi \ \mathrm{d}\mu - \int \psi \ \mathrm{d}\nu \right)$$

where ϕ and ψ are such that $\forall x, y \ |\phi(x) - \psi(y)| \leq c(x, y)$. However, since $c(x, x) = 0$ for every $x \in \mathcal{X}$, then for any pair of functions ϕ and ψ considered, for all $x, \in X$, $|\phi(x) - \psi(x)| \leq c(x, x) = 0$, which implies that $\phi = \psi$.

Lemma A.9. If the cost function c is a 1-bounded pseudometric on $\mathcal{X}$, then $W(c)$ is a 1-bounded pseudometric on the space of probability distributions on $\mathcal{X}$.

Proof. The first expression $W(c)(\mu, \nu) = \min_{\gamma \in \Gamma(\mu,\nu)} \int c \ \mathrm{d}\gamma$ immediately gives us that $W(c)$ is 1-bounded (since we are working on the space of probability distributions on $\mathcal{X}$) and that $W(c)$ is symmetric (by the change of variable formula with the function on $\mathcal{X} \times \mathcal{X}$, $s(x, y) = (y, x)$).

The second expression $W(c)(\mu,\nu) = \max_{h\in\mathcal{H}(c)} \left|\int h \, \mathrm{d}\mu - \int h \, \mathrm{d}\nu\right|$ immediately gives us that $W(c)(\mu,\mu) = 0$ and for any three probability distributions μ, ν, θ and for any function $h \in \mathcal{H}(c)$,

$$\begin{aligned}\left|\int h \, \mathrm{d}\mu - \int h \, \mathrm{d}\theta\right| &\leq \left|\int h \, \mathrm{d}\mu - \int h \, \mathrm{d}\nu\right| + \left|\int h \, \mathrm{d}\nu - \int h \, \mathrm{d}\theta\right| \\ &\leq W(c)(\mu,\nu) + W(c)(\nu,\theta).\end{aligned}$$

Since this holds for every $h \in \mathcal{H}(c)$, we get the triangular inequality. □ ∎

Lemma A.10. Consider a Polish space $\mathcal{X}$ and a cost function $c : \mathcal{X} \times \mathcal{X} \to [0,1]$ such that there exists an increasing ($c_{k+1} \geq c_k$ for every k) sequence of continuous cost functions $c_k : \mathcal{X} \times \mathcal{X} \to [0,1]$ that converges to c pointwise. Then, given two probability distributions P and Q on $\mathcal{X}$,

$$\lim_{k\to\infty} W(c_k)(P,Q) = W(c)(P,Q).$$

Proof. For each c_k, the optimal transport cost $W(c_k)(P,Q)$ is attained for a coupling π_k. Using Lemma 2.3, we know that the space of couplings $\Gamma(P,Q)$ is compact. We can thus extract a subsequence that we will still denote by $(\pi_k)_{k\in\mathbb{N}}$ which converges weakly to some coupling $\pi \in \Gamma(P,Q)$. We will show that this coupling π is in fact an optimal transference plan.

By Monotone Convergence Theorem,

$$\int c \, \mathrm{d}\pi = \lim_{k\to\infty} \int c_k \, \mathrm{d}\pi.$$

Consider $\epsilon > 0$. There exists k such that

$$\int c \, \mathrm{d}\pi \leq \epsilon \int c_k \, \mathrm{d}\pi \tag{12}$$

Since π_k converges weakly to π and since c_k is continuous and bounded,

$$\int c_k \, \mathrm{d}\pi = \lim_{n\to\infty} \int c_k \, \mathrm{d}\pi_n,$$

which implies that there exists $n_k \geq k$ such that

$$\int c_k \, \mathrm{d}\pi \leq \epsilon + \int c_k \, \mathrm{d}\pi_{n_k}. \tag{13}$$

Putting Equations (12) and (13) together, we get

$$\begin{aligned}\int c \,\mathrm{d}\pi &\leq 2\epsilon + \int c_k \,\mathrm{d}\pi_{n_k} \\ &\leq 2\epsilon + \int c_{n_k} \,\mathrm{d}\pi_{n_k} \quad \text{since } (c_n)_{n\in\mathbb{N}} \text{ is increasing} \\ &= 2\epsilon + W(c_{n_k})(P, Q).\end{aligned}$$

This implies that

$$\int c \,\mathrm{d}\pi \leq \lim_{k\to\infty} W(c_k)(P, Q).$$

The other inequality is trivial since $c_k \leq c$. □ ∎

B Details for Section 3

Under the conditions of FD-semigroups, strong continuity is equivalent to the apparently weaker condition (see Lemma III. 6.7 in [27] for the proof):

$$\forall f \in C_0(E) \;\; \forall x \in E, \;\; \lim_{t\downarrow 0}(\hat{P}_t f)(x) = f(x)$$

The authors also offer the following useful extension (Theorem III.6.1):

Theorem B.1. A bounded linear functional ϕ on $C_0(E)$ may be written uniquely in the form

$$\phi(f) = \int_E f(x) \; \mu(\mathrm{d}x)$$

where μ is a signed measure on E of finite total variation.

As stated in the Riesz representation theorem, the above measure μ is inner regular. Theorem B.1 is also known as the Riesz-Markov-Kakutani representation theorem in some other references.

Theorem B.1 has the following corollary (Theorem III.6.2) where $b\mathcal{E}$ denotes the set of bounded, $\mathcal{E}$-measurable functions $E \longrightarrow \mathbb{R}$.

Corollary B.2. Suppose that $V : C_0(E) \longrightarrow b\mathcal{E}$ is a (bounded) linear operator that is sub-Markov in the sense that $0 \leq f \leq 1$ implies $0 \leq Vf \leq 1$. Then there exists a unique sub-Markov kernel (also denoted by) V on $(E, \mathcal{E})$ such that for all $f \in C_0(E)$ and $x \in E$

$$Vf(x) = \int f(y) \; V(x, \mathrm{d}y).$$

While the proof is left as an exercise in [27], we choose to explicitly write it down for clarity

Proof. For every x in E, write V_x for the functional $V_x(f) = Vf(x)$. This functional is bounded and linear which enables us to use Theorem B.1: there exists a signed measure μ_x on E of finite total variation such that

$$V_x(f) = \int_E f(y)\ \mu_x(\mathrm{d}y)$$

We claim that $V : (x, B) \mapsto \mu_x(B)$ is a sub-Markov kernel, i.e.

- For all x in E, $V(x, -)$ is a subprobability measure on $(E, \mathcal{E})$
- For all B in $\mathcal{E}$, $V(-, B)$ is $\mathcal{E}$-measurable.

The first condition directly follows from the definition of V: $V(x, -) = \mu_x$ which is a measure and furthermore $V(x, E) = \mu_x(E) = V_x(1)$ (where 1 is the constant function over the whole space E which value is 1). Using the hypothesis that V is Markov, we get that $V1 \leq 1$. We have to be more careful in order to prove that $V(x, B) \geq 0$ for every measurable set B: this is a consequence of the regularity of the measure μ_x (see Proposition 11 of section 21.4 of [28]). This shows that μ_x is a subprobability measure on $(E, \mathcal{E})$.

Now, we have to prove that for every $B \in \mathcal{E}$, $V(-, B)$ is measurable. Recall that the set E is σ-compact: there exists countably many compact sets K_k such that $E = \bigcup_{k\in\mathbb{N}} K_k$. For $n \in \mathbb{N}$, define $E_n = \bigcup_{k=0}^{n} K_k$ and $B_n = B \cap E_n$.

For every $n \in \mathbb{N}$, there exists a sequence of functions $(f_j^n)_{j\in\mathbb{N}} \subset C_0(E)$ that converges pointwise to $\mathbb{1}_{B_n}$, i.e. for every $x \in E$, $\mu_x(B_n) = \lim_{j\longrightarrow+\infty} Vf_j^n(x)$. Since the operator $V : C_0(E) \longrightarrow b\mathcal{E}$, the maps Vf_j^n are measurable which means that $V(-, B_n) : x \mapsto \mu_x(B_n)$ is measurable.

Since for every $x \in E$, $\mu_x(B) = \lim_{n\longrightarrow+\infty} \mu_x(B_n)$, this further means that $V(-, B) : x \mapsto \mu_x(B)$ is measurable. □ ∎

We can then use this corollary to derive Proposition 3.4 which relates these FD-semigroups with Markov kernels. This allows one to see the connection with familiar probabilistic transition systems.

C Some background on Brownian motion

C.1 Definition of Brownian motion

Brownian motion is a stochastic process describing the irregular motion of a particle being buffeted by invisible molecules. Now its range of applicability extends far beyond its initial application [21]. The following definition is from [21].

Definition C.1. A standard one-dimensional Brownian motion is a Markov process:

$$B = (B_t, \mathcal{F}_t), 0 \leq t < \infty$$

where $(\mathcal{F}_t)_{t\geq 0}$ is the natural filtration adapted to $(B_t)_{t\geq 0}$ and whe stochastic process $(B_t)_{t\geq 0}$ is defined on a probability space $(\Omega, \mathcal{F}, P)$ with the properties

1. $B_0 = 0$ almost surely,
2. for $0 \leq s < t$, $B_t - B_s$ is independent of $\mathcal{F}_s$ and is normally distributed with mean 0 and variance $t - s$.

In this very special process, one can start at any place, there is an overall translation symmetry which makes calculations more tractable. We denote $(B_t^x)_{t\geq 0}$ the standard Brownian motion on the real line starting from x in section 8.2.

In order to do any calculations we use the following fundamental formula: If the process is at x at time 0 then at time t the probability that it is in the (measurable) set D is given by

$$P_t(x, D) = \int_{y\in D} \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{(x-y)^2}{2t}\right) \mathrm{d}y.$$

The associated FD semigroup is the following: for $f \in C_0(\mathbb{R})$ and $x \in \mathbb{R}$,

$$\hat{P}_t(f)(x) = \int_y \frac{f(y)}{\sqrt{2\pi t}} \exp\left(-\frac{(x-y)^2}{2t}\right) \mathrm{d}y.$$

C.2 Hitting time for stochastic processes

We follow the definitions of Karatzas and Shreve in [21].

Assume that $(X_t)_{t\geq 0}$ is a stochastic process on $(\Omega, \mathcal{F})$ such that $(X_t)_{t\geq 0}$ takes values in a state space E equipped with its Borel algebra $\mathcal{B}(\mathcal{O})$), has right-continuous paths (for every time t, $\omega(t) = \lim_{s\longrightarrow t, s>t} \omega(s)$) and is adapted to a filtration $(\mathcal{F}_t)_{t\geq 0}$.

Definition C.2. Given a measurable set $C \in \mathcal{B}(\mathcal{O})$, the *hitting time* is the random time

$$T_C(\omega) = \inf\{t \geq 0 \mid X_t(\omega) \in C\}.$$

Intuitively $T_C(\omega)$ corresponds to the first time the trajectory ω "touches" the set C.

C.3 Hitting time for Brownian motion

Consider $(B_t^x)_{t \geq 0}$ the standard Brownian motion on the real line starting from $x > 0$. With probability 1, the trajectories of Brownian motion are continuous, which means that

$$\mathbb{P}^x(H_{(-\infty,0]} < t) = \mathbb{P}^x(H_{\{0\}} < t).$$

Here we denote $\mathbb{P}^x$ for the probability on trajectories for Brownian motion starting from x. We refer the reader to our previous papers [9, 10, 11] for a full description.

A standard result is that

$$\mathbb{P}^x(H_{\{0\}} < t) = \sqrt{\frac{2}{\pi}} \int_{x/\sqrt{t}}^{+\infty} e^{-s^2/2} ds.$$

We refer the reader to [6] for formulas regarding various hitting times of standard Brownian motion and many of its variants.

In section 8.2, we introduced the hitting time $\tau := \inf\{t \geq 0 : B_t^x = 0 \text{ or } B_t^x = 1\}$, with $x \in [0, 1]$. This corresponds to the hitting time of the set $(-\infty, 0] \cup [1, +\infty)$. We can then define the process $B_{t \wedge \tau}^x$ on the state space $[0, 1]$: the intuition is that this process behaves like Brownian motion until it reaches one of the "boundaries" (either 0 or 1) where it becomes "stuck".

Note that there are two main types of boundaries that one can consider for Brownian motion on the real line: when the process hits the boundary, it can either "get stuck" / vanish (boundary with absorption) or bounce back (boundary with reflection).

D Proofs for Section 6

Lemma D.1. For every f in Λ_c, there exists n such that for every x, y, $f(x) - f(y) \leq \delta_n^c(x, y)$.

Lemma D.2. Consider a continuous function $h : E \longrightarrow [0, 1]$ and two states z, z' such that there exists f in the logic Λ_c such that $|h(z) - h(z')| \leq |f(z) - f(z')|$. Then for every $\epsilon > 0$, there exists $g \in \Lambda_c$ such that $|h(z) - g(z)| \leq 2\epsilon$ and $|h(z') - g(z')| \leq 2\epsilon$.

E Computations for Section 8

- $\overline{\delta}(0, \partial)$: First of all, note that $\delta_0(0, \partial) = 1$ which means that

$$1 = \delta_0(0, \partial) \leq \overline{\delta}(0, \partial) \leq 1$$

and thus $\overline{\delta}(0, \partial) = 1$.

- $\overline{\delta}(0, x)$: First of all, $\delta_0(0, x) = 1 - r$. We can now compute $\delta_1(0, x)$:

$$\begin{aligned}\delta_1(0, x) &= \sup_{t \geq 0} e^{-\lambda t} \left((1 - e^{-\lambda t}) \times \delta_0(0, 0) + e^{-\lambda t} \times \delta_0(0, x) \right) \\ &= \sup_{t \geq 0} e^{-2\lambda t} \times (1 - r) = 1 - r\end{aligned}$$

  Since $\delta_1(0, x) = \delta_0(0, x) = 1 - r$, we also have that $\overline{\delta}(0, x) = 1 - r$.

- $\overline{\delta}(0, y)$: First of all, $\delta_0(0, y) = 1 - r$. We can now compute $\delta_1(0, y)$:

$$\begin{aligned}\delta_1(0, y) &= \sup_{t \geq 0} e^{-\lambda t} \delta_0(0, y) \\ &= \sup_{t \geq 0} e^{-\lambda t} \times (1 - r) = 1 - r\end{aligned}$$

  Since $\delta_1(0, y) = \delta_0(0, y) = 1 - r$, we also have that $\overline{\delta}(0, y) = 1 - r$.

- $\overline{\delta}(y, \partial)$: First of all, $\delta_0(y, \partial) = r$. We can now compute $\delta_1(y, \partial)$:

$$\begin{aligned}\delta_1(y, \partial) &= \sup_{t \geq 0} e^{-\lambda t} \delta_0(y, \partial) \\ &= \sup_{t \geq 0} e^{-\lambda t} r = r\end{aligned}$$

  Since $\delta_1(y, \partial) = \delta_0(y, \partial) = r$, we also have that $\overline{\delta}(y, \partial) = r$.

- $\overline{\delta}(x, y)$: First, by induction on n, we prove that $\delta_n(x, y) = \epsilon_n(1 - r)$ where the sequence $(\epsilon_n)_{n \in \mathbb{N}}$ is defined as

$$\epsilon_0 = 0 \qquad \epsilon_{n+1} = \frac{1}{4(1 - \epsilon_n)}$$

  and satisfies $0 < \epsilon_n \leq \frac{1}{2}$ for every $n \in \mathbb{N}$. Indeed, $\delta_0(x, y) = r - r = 0$ and assuming the result holds at rank n:

$$\begin{aligned}\delta_{n+1}(x, y) &= \sup_{t \geq 0} e^{-\lambda t} \left((1 - e^{-\lambda t}) \delta_n(0, y) + e^{-\lambda t} \delta_n(x, y) \right) \\ &= \sup_{0 < \theta \leq 1} \theta \left((1 - \theta)(1 - r) + \theta (1 - r) \epsilon_n \right) \\ &= (1 - r) \sup_{0 < \theta \leq 1} \theta \left(1 - \theta (1 - \epsilon_n) \right).\end{aligned}$$

  We are now left looking for the supremum of the function $\phi : \theta \mapsto \theta \left(1 - \theta(1 - \epsilon_n)\right)$ on $(0, 1]$. We have that

$$\phi'(\theta) = 1 - 2\theta(1 - \epsilon_n)$$

and we can thus conclude that its maximum is attained at $\frac{1}{2(1-\epsilon_n)} \in \left(\frac{1}{2}, 1\right]$ and thus

$$\delta_{n+1}(x,y) = (1-r)\phi\left(\frac{1}{2(1-\epsilon_n)}\right) = \frac{1}{2(1-\epsilon_n)} \times \frac{1}{2} \in \left(\frac{1}{4}, \frac{1}{2}\right].$$

We are now left to study the limit of the sequence $(\epsilon_n)_{n\in\mathbb{N}}$. We know that this limit exists so we only have to compute it. This limit l satisfies $l = \frac{1}{4(1-l)}$ and we can thus conclude that $\overline{\delta}(x,y) = \frac{1-r}{2}$.

- $\overline{\delta}(y,z)$: Similarly to what we did before, we obtain that $\delta_0 = 0$ and

$$\delta_{n+1}(y,z) = \sup_{0<\theta\leq 1} \theta\left(\frac{1-\theta}{2} + \theta\delta_n(y,z)\right) \leq \frac{1}{4}.$$

This supremum is attained for $\theta = \frac{1}{2(1-2\delta_n(y,z))} \leq 1$ and thus

$$\delta_{n+1}(y,z) = \frac{1}{8(1-2\delta_n(y,z))}$$

and finally $\overline{\delta}(y,z) = \frac{1}{4}$.

- $\overline{\delta}(x,\partial)$: First, we have that $\delta_0(x,\partial) = r - 0 = r$ and for every $n \geq 0$,

$$\begin{aligned}\delta_{n+1}(x,\partial) &= \sup_{t\geq 0} e^{-\lambda t}\left((1-e^{-\lambda t})\delta_n(0,\partial) + e^{-\lambda t}\delta_n(x,\partial)\right)\\ &= \sup_{0<\theta\leq 1} \theta\left((1-\theta) + \theta\delta_n(x,\partial)\right)\end{aligned}$$

We are now left looking for the supremum of the function

$$\phi : \theta \mapsto \theta\left((1-\theta) + \theta\delta_n(x,\partial)\right) \text{ on } (0,1].$$

Its derivative ϕ' has value 0 at $\theta_0 = \frac{1}{2(1-\delta_n(x,\partial))}$. From the theory, we know that $\delta_n(x,\partial) \leq 1$ and we therefore know that ϕ reaches its supremum in $\theta_0 \geq \frac{1}{2}$. There are now two cases to consider: either $\theta_0 > 1$ or $\theta_0 \leq 1$. We get that

$$\overline{\delta}(x,\partial) = \begin{cases} r & \text{if } r \geq \frac{1}{2} \\ \frac{1}{2} & \text{otherwise.}\end{cases}$$

- $\overline{\delta}(z,\partial)$: This case is similar to the previous ones: $\delta_0(z,\partial) = r$ and

$$\delta_{n+1}(x,\partial) = \sup_{0<\theta\leq 1} \theta\left(\frac{1-\theta}{2} + \theta\delta_n(z,\partial)\right)$$

  Similarly to the case of $\overline{\delta}(0,z)$. we have three cases to study, and we end up with

$$\overline{\delta}(z,\partial) = \begin{cases} r & \text{if } r \geq \frac{1}{4} \\ \frac{1}{4} & \text{otherwise.} \end{cases}$$

- $\overline{\delta}(x,z)$: due to a conflict of notation, we will write $\mathfrak{d}_b$ for the Dirac measure centered in b.

  First, we note that $\overline{\delta}(x,z) \leq \frac{1}{2}$. Similarly to what we have done so far, we have that $\delta_0(x,z) = 0$ and

$$\delta_{n+1}(x,z) = \sup_{0<\theta\leq 1} \theta W(\delta_n)(\theta\mathfrak{d}_x + (1-\theta)\mathfrak{d}_0, \theta\mathfrak{d}_z + \frac{1-\theta}{2}(\mathfrak{d}_0 + \mathfrak{d}_\partial)).$$

  We can now show by induction that for every n, $\delta_n(x,z) \leq \frac{1}{4}$: consider the coupling γ_θ defined by

$$\gamma(0,0)^5 = \gamma(0,\partial) = \frac{1-\theta}{2} \text{ and } \gamma(x,z) = \theta.$$

  We then have that

$$\begin{aligned} &W(\delta_n)\left(\theta\mathfrak{d}_x + (1-\theta)\mathfrak{d}_0, \theta\mathfrak{d}_z + \frac{1-\theta}{2}(\mathfrak{d}_0 + \mathfrak{d}_\partial)\right) \\ &\leq \int \delta_n \,\mathrm{d}\gamma_\theta \\ &= \frac{1-\theta}{2}\delta_n(0,\partial) + \theta\delta_n(x,z) \\ &= \frac{1}{2} - \theta\left(\frac{1}{2} - \delta_n(x,z)\right) \end{aligned}$$

  This means that

$$\delta_{n+1}(x,z) \leq \sup_{0<\theta\leq 1} \theta\left[\frac{1}{2} - \theta\left(\frac{1}{2} - \delta_n(x,z)\right)\right]$$

  And we get that $\delta_{n+1}(x,z) \leq \frac{1}{4}$ as usual.

[5] Due to how heavy the notations are already, we write $\gamma(0,0)$ instead of $\gamma(\{(0,0)\})$

We also have that

$$\begin{aligned}
\overline{\delta}(x,z) &\geq \sup_{0<\theta\leq 1} \theta \left| (\theta obs(x) + (1-\theta) obs(0)) - \left(\theta obs(z) + \frac{1-\theta}{2}(obs(\partial) + obs(0)) \right) \right| \\
&= \sup_{0<\theta\leq 1} \theta \left| (\theta r + (1-\theta)) - \left(\theta r + \frac{1-\theta}{2} \right) \right| \\
&= \frac{1}{2} \sup_{0<\theta\leq 1} \theta(1-\theta) = \frac{1}{8}
\end{aligned}$$

From this we conclude that $\frac{1}{8} \leq \overline{\delta}(x,z) \leq \frac{1}{4}$.