Approximate reasoning for real-time probabilistic processes

Vineet Gupta
Google Inc.
vineet@google.com

Radha Jagadeesan*
School of CTI, DePaul University
rjagadeesan@cs.depaul.edu

Prakash Panangaden†
School of Computer Science, McGill University
prakash@cs.mcgill.ca

November 8, 2004

Abstract

We develop a pseudo-metric analogue of bisimulation for generalized semi-Markov processes. The kernel of this pseudo-metric corresponds to bisimulation; thus we have extended bisimulation for continuous-time probabilistic processes to a much broader class of distributions than exponential distributions. This pseudo-metric gives a useful handle on approximate reasoning in the presence of numerical information — such as probabilities and time — in the model.

We give a fixed point characterization of the pseudo-metric. This makes available coinductive reasoning principles for reasoning about distances. We demonstrate that our approach is insensitive to potentially ad hoc articulations of distance by showing that it is intrinsic to an underlying *uniformity*. We provide a logical characterization of this uniformity using a real-valued modal logic.

We show that several quantitative properties of interest are continuous with respect to the pseudo-metric. Thus, if two processes are metrically close, then observable quantitative properties of interest are indeed close.

1 Introduction

The starting point and conceptual basis for classical investigations in concurrency are the notions of equivalence and congruence of processes — when can two processes be considered the same and when can they be substituted for each other. Most investigations into timed [AH92, AD94] and probabilistic concurrent processes are based on equivalences of one kind or another, e.g. [CSZ92, Han94, Hil94, LS91, SL95, PLS00] to name but a few.

As has been argued before [JS90, DGJP99, DGJP04], this style of reasoning is fragile in the sense of being too dependent on the exact numerical values of times and probabilities. Previously this had pointed out for probability, but the same remarks apply, *mutis mutandis*, to real time as well. Consider the following two paradigmatic examples:

- Consider the probabilistic choice operator: $\mathcal{A}_1 +_p \mathcal{A}_2$, which starts $\mathcal{A}_1$ with probability p and $\mathcal{A}_2$ with probability $1-p$. Consider $\mathcal{A}_1 +_{p+\epsilon} \mathcal{A}_2$ and $\mathcal{A}_1 +_{p+2\epsilon} \mathcal{A}_2$. In traditional exact reasoning, the best that one can do is to say that all these three processes are inequivalent. Clearly, there is a gradation here: $\mathcal{A}_1 +_{p+\epsilon} \mathcal{A}_2$ is closer to $\mathcal{A}_1 +_p \mathcal{A}_2$ than $\mathcal{A}_1 +_{p+2\epsilon} \mathcal{A}_2$ is to $\mathcal{A}_1 +_p \mathcal{A}_2$.

*Research supported in part by NSF 0244901
†Research supported in part by NSERC and EPSRC.

- Consider the $\texttt{delay}_t.\mathcal{A}$ operator that starts $\mathcal{A}$ after a delay of t time units. Consider $\texttt{delay}_{t+\epsilon}.\mathcal{A}$ and $\texttt{delay}_{t+2\epsilon}.\mathcal{A}$. Again, in exact reasoning, the best that one can do is to say that all these three processes are inequivalent. Again, $\texttt{delay}_{t+\epsilon}.\mathcal{A}$ is intuitively closer to $\texttt{delay}_t.\mathcal{A}$ than $\texttt{delay}_{t+2\epsilon}.\mathcal{A}$ is to $\texttt{delay}_t.\mathcal{A}$.

In both examples, the intuitive reasoning behind relative distances is supported by calculated numerical values of quantitative observables — expectations in the probabilistic case and (cumulative) rewards in the timed case.

The fragility of exact equivalence is particularly unfortunate for two reasons: firstly, the timings and probabilities appearing in models should be viewed as numbers with some error estimate. Secondly, probability distributions over uncountably many states arise in even superficially discrete paradigms such as Generalized semi-markov processes (e.g. see [She87] for a textbook survey), and discrete approximations are used for algorithmic purposes [BHK99, HCH$^+$02]. These approximants do not match the continuous state model exactly and force us to think about approximate reasoning principles — e.g. when does it suffice to prove a property about an approximant?

Thus, we really want an "approximate" notion of equality of processes. In the probabilistic context, Jou and Smolka [JS90] propose that the correct formulation of the "nearness" notion is via a metric. Similar reasons motivate the study of Lincoln, Mitchell, Mitchell and Scedrov [LMMS98], our previous study of metrics for labelled Markov processes [DGJP99, DGJP04, DGJP02], the study of the fine structure of these metrics by van Breugel and Worrell [vBW01b, vBW01a] and the study of Alfaro, Henzinger and Majumdar of metrics for probabilistic games [dAHM03].

In contrast to these papers, in the present paper we focus on real-time probabilistic systems that combine *continuous time and probability.* We consider generalized semi-Markov processes (GSMPs). Semi-Markov processes strictly generalize continuous time Markov chains by permitting general (i.e. non exponential) probability distributions; GSMPs further generalize them by allowing competition between multiple events each driven by a different clock.

Following the format of the usual definition of bisimulation as a maximum fixed point, we define a metric on configurations of a GSMP as a maximum fixed point. This permits us to use analogues of traditional methods to reason about metric distances. For example, in exact reasoning, to deduce that two states are equivalent, it suffices to produce a bisimulation that relates the states. In our setting, to show that the distance between two states is less than ϵ, it suffices to produce a (metric) bisimulation that sets the distance between the states to be less than ϵ.

Viewing metric distance 0 as bisimilarity, we get a definition of bisimulation for GSMPs, a class that properly includes CTMCs. In contrast to existing work on bisimulation for general probability distributions (e.g. [BG02, Her02]) our definition accounts explicitly for the change of probability densities over time.

Secondly, we demonstrate that our study does not rely on "ad-hoc" construction of metric distances. Uniform spaces capture the essential aspects of metric distances by axiomatizing the structure needed to capture relative distances – e.g. statements of the form "x is closer to y than to z". A metric determines a uniform space but different metrics can yield the same uniform space. Uniform spaces represent more information than topological spaces but less than metric spaces, so we are identifying, as precisely as we can, the intrinsic source of the quantitative information. We present our maximal fixpoint construction as a construction on uniform spaces, showing that the numerical values of different metric representations are not used in an essential way. In particular, in our setting, it shows that the actual numerical values of the discount factors used in the definition of the metric do not play any essential role.

Thirdly, we provide a "logical" characterization of the uniformity using a real-valued modal logic. In analogy to traditional completeness results, we prove that the uniformity notion induced by the real-valued modal logic coincides with the uniformity induced by the metric defined earlier. Our logic is intentionally chosen to prove this completeness result. It is not intended to be used as an expressive specification formalism to describe properties of interest. Our framework provides an intrinsic characterization of the quantitative observables that can be accommodated – functions that are continuous with respect to the metric.

Finally, we illustrate the use of such studies in reasoning by showing that several quantitative properties of interest are continuous with respect to the metric. Thus, if two processes are close in the metric then observable quantitative properties of interest are indeed close. For expository purposes, the list considered in this paper includes expected hitting time, expected (cumulative and average) rewards. The tools used to establish these results are "continuous mapping theorems" from stochastic process theory, and provide a general recipe to tackle other observables of interest.

The rest of this paper is organized as follows. We begin with a review of the model of GSMPs in Section 2. We then give a review of the basic ideas from stochastic process theory – metrics on probability measures on metric spaces in Section 3.1 and the Skorohod J2 metrics on timed traces in Section 4. We define metric bisimulation in Section 6 and we show the completeness of the real-valued modal logic in Section 9. We discuss interesting quantitative observables are continuous functions in Section 8. Finally we present our construction in terms of uniform spaces in Section 7.

2 Generalized semi-Markov processes

GSMPs properly include finite state CTMCs while also permitting general probability distributions. We describe GSMPs informally here following the formal description of [Whi80]. The key point is that in each state there are possibly several events that can be executed. Each event has its own clock - running down at its own rate - and when the first one reaches zero that event is selected for execution. Then a probabilistic transition determines the final state and any new clocks are set according to given probability distributions: defined by conditional density functions. The probability distribution over the next states depends only on the current state the the event that has occurred: this is the "Markov" in semi-Markov. The clocks are reset according to an arbitrary distribution, not necessarily an exponential (memoryless) distribution: hence the "semi". We will consider finite-state systems throughout this paper.

A finite-state GSMP over a set of atomic propositions AP has the following ingredients:

1. A finite set S of states. Each state has an associated finite set of events $I(s)$, each with its own clock (we use the same letter for an event and its clock) and a non-zero rate for each clock in $I(s)$. A clock in $I(s)$ runs down at the constant rate associated with it.
2. A labelling function $\mathtt{Props} : S \to 2^{AP}$ that assigns truth values to atomic propositions in each state.
3. A continuous probability density function $f(t; s, i; s', i')$, over time, for each $i \in I(s)$, for each target state s' and $i' \in I(s')$. This is used to define how clocks are reset during a transition.
4. For each $i \in I(s)$, a probabilistic transition function $\mathtt{next}_i : S \times S \to [0,1]$. We require $\sum_{s' \in S} \mathtt{next}_i(s, s') = 1$.

We use $\tilde{\mathbf{c}}, \tilde{\mathbf{c'}}$ (resp. $\tilde{\mathbf{r}}$) for vectors of clock values (resp. rates). We use the vector operation $\tilde{\mathbf{c}} - \tilde{\mathbf{r_c}}t$ to indicate the clock vector resulting from evolution of each clock under its rate for time t.

Definition 2.1 *Let s be a state. A* ***generalized state*** *is of the form $\langle s, \tilde{\mathbf{c}} \rangle$ where c is a vector of clock values indexed by $i \in I(s)$ that satisfies a uniqueness condition: there is a unique clock in $I(s)$ that reaches 0 first.*

We write $\mathcal{T}(\langle s, \tilde{\mathbf{c}} \rangle)$ for the time required for the first clock (unique by the above definition) to reach 0. We use $\mathcal{G}$ for the set of generalized states, and $\mathbf{g_s}, \mathbf{g'_s}, \mathbf{g^1_s} \ldots$ for generalized states.

We describe the evolution starting in a generalized state $\langle s, \tilde{\mathbf{c}} \rangle$. Each clock in $\tilde{\mathbf{c}}$ decreases at its associated rate. By the uniqueness condition on generalized states, a unique clock reaches 0 first. Let this clock be $i \in I(s)$. The distribution on the next states is determined by the probabilistic transition function $\mathtt{next}_i : S \times S \to [0,1]$. For each target state s',

- The clocks $i' \in I(s) \setminus I(s')$ are discarded.
- The new clocks $i' \in I(s') \setminus [I(s) \setminus \{i\}]$, get new initial time values assigned according to the continuous probability density function $f(t; s, i; s', i')$.

- The remaining clocks carry forward their time values from s to s'

The continuity condition on probability distributions ensures that this informal description yields a legitimate Markov kernel [Whi80]. The semantics of a real-time probabilistic process can be described as a discrete-time Markov process on generalized states. For each generalized state, we associate a set of sequences of generalized states that arise following the prescription of the evolution given above.

3 Pseudometrics

Definition 3.1 *A pseudometric m on a state space $\mathcal{S}$ is a function $\mathcal{S} \times \mathcal{S} \to [0,1]$ such that:*

$$d(x,x) = 0,\ d(x,y) = d(y,x),\ d(x,z) \leq d(x,y) + d(x,z)$$

A function $f : (M,m) \to (M',m')$ is Lipschitz if $(\forall x,y)\ m'(f(x), f(y)) \leq m(x,y)$.

We consider a partial order on pseudometrics on a fixed set of states $\mathcal{S}$.

Definition 3.2 *$\mathcal{M}$ is the class of pseudometrics on $\mathcal{S}$ ordered as: $m_1 \preceq m_2$ if $(\forall s,t)\ m_1(s,t) \geq m_2(s,t)$.*

Lemma 3.3 *$(\mathcal{M}, \preceq)$ is a complete lattice.*

The top element $\top$ is the constant 0 function, and the bottom element is the discrete metric [DGJP02]. Thus, any monotone function F on $(\mathcal{M}, \preceq)$ has a complete lattice of fixed points.

3.1 Wasserstein metric

The Wasserstein metric is actually a prescription for lifting the metric from a given (pseudo)metric space to the space of probability distributions on the given space.

Let (M,m) be a pseudometric space, and let P, Q be probability measures on M. Then, $W(m)(P,Q)$ is defined by the solution to the following linear program ($h : M \to [0,1]$ is any function):

$$\begin{array}{rcl} W(m)(P,Q) & = & \sup_h \int h dP - \int h dQ \\ & & \text{subject to} : \forall s \in M.\ h(s) \leq 1 \\ & & \forall s, s'.\ |h(s) - h(s')| \leq m(s,s'). \end{array}$$

An easy calculation using the linear program shows that the distances on distributions satisfies symmetry and the triangle inequality, so we get a pseudometric - written $W(m)$ - on distributions.

By standard results (see Anderson and Nash [AN87]), this is equivalent to defining $W(m)(P,Q)$ as the solution to the following dual linear program (here ρ is any measure on $M \times M$, S and S' are any measurable subsets:

$$\begin{array}{rcll} W(m)(P,Q) & = & \inf_\rho \int m\ d\rho & \\ & & \text{subject to} : & \forall S.\rho(S \times M) = P(S) \\ & & & \forall S'.\rho(M \times S') = Q(S') \\ & & & \forall S, S'.\ \rho(S \times S') \geq 0. \end{array}$$

The Wasserstein construction is monotone on the lattice of pseudometrics.

Lemma 3.4 $m \preceq m' \Rightarrow W(m) \preceq W(m')$

Proof. Clearly every solution to the linear program for $m'(P,Q)$ is also a solution to the linear program for $m(P,Q)$. The result is now immediate. ∎

We discuss some concrete examples to illustrate the distances yielded by this construction. Let (M, m) be a 1-bounded metric space, ie. $m(x, y) \leq 1$, for all x, y. Let 1_x be the unit measure concentrated at x,

Example 3.5 *We calculate $W(m)(1_x, 1_{x'})$. The primal linear program - using the function $h : M \to [0,1]$ defined by $h(y) = m(x, y)$ - yields $W(m)(1_x, 1_{x'}) \geq m(x, x')$. The dual linear program - using the product measure - yields $W(m)(1_x, 1_{x'}) \leq m(x, x')$.*

Example 3.6 *Let P, Q be such that for all measurable U, $|P(U) - Q(U)| < \epsilon$. For any 1-bounded function h, $|\int hdP - \int hdQ| < \epsilon$ — since for any simple function g with finite range $\{v_1 \ldots v_n\}$ dominated by h:*

$$\begin{array}{l} \sum_i v_i P(g^{-1}(v_i)) - v_i Q(g^{-1}(v_i)) \\ \sum_i v_i [P(g^{-1}(v_i)) - Q(g^{-1}(v_i))] \\ \leq \sum_i P(g^{-1}(v_i)) - Q(g^{-1}(v_i)) \\ = P(g^{-1}(\{v_1 \ldots v_n\}) - Q(g^{-1}(\{v_1 \ldots v_n\})) \\ < \epsilon \end{array}$$

So, $W(m)(P, Q) < \epsilon$.

In 1-bounded metric spaces, the Wasserstein metric is closely related to the Prohorov metric, π, which metrizes the topology of weak convergence. We say that P_n weakly converges to P if for all bounded continuous real-valued functions f $\int f dP_n$ converges to $\int f dP$. For any Borel set A, we write A^ϵ for $\{u : \exists v \in A. d(u, v) < \epsilon\}$. The Prohorov metric $\pi(P, Q)$ between two measures is defined by

$$\inf_{\epsilon \geq 0} P(A) \leq Q(A^\epsilon) + \epsilon \text{ and } Q(A) \leq P(A^\epsilon) + \epsilon.$$

The connection between the Waserstein metric and the Prohorov metric is:

$$\sqrt{\pi(P, Q)} \leq W(m)(P, Q) \leq 2\pi(P, Q).$$

The following lemma is the key tool to approximate continuous probability distributions by discrete distributions in a separable metric space. We use the following lemma later with the $U_i, i > 1$'s being subsets of ϵ-neighborhoods of a point, and the $U_i, i > 1$ being a finite cover, wrt P, for all but ϵ of the space (which will be covered by U_0).

Lemma 3.7 *Let P be a probability measure on (M, m). Let $U_i, i = 0, 1, 2, \ldots n$ be a finite partition of the points of M into measurable sets such that:*

- $(\forall\ i \geq 1)$ $[\mathtt{diameter}(U_i) \leq \epsilon]$ [1]
- $P(U_0) \leq \epsilon$

Let x_i be such that $x_i \in U_i$. Define a discrete probability measure Q on (M, m) by: $Q(\{x_i\}) = P(U_i)$. Then:

$$W(m)(P, Q) \leq 2\epsilon$$

Proof. Let $h : M \to [0, 1]$ be any function that satisfies $\forall s, s'.\ |h(s) - h(s')| \leq m(s, s')$. Then

$$\begin{array}{rcl} \int hdP - \int hdQ &=& \sum_i \int_{U_i} hdP - \int_{U_i} hdQ \\ &=& \sum_i \int_{U_i} hdP - h(x_i)P(U_i) \\ &=& \sum_i \int_{U_i} (h(x) - h(x_i))dP \\ &\leq& \int_{U_0} 1dP + \sum_{i>0} \int_{U_i} \epsilon dP \\ &\leq& P(U_0) + \int \epsilon dP \\ &\leq& 2\epsilon \end{array}$$

The fourth inequality follows as $0 \leq h(x) \leq 1$ and from our assumption on $U_i, m(x, x_i) \leq \epsilon$, and from the constraint on h, $h(x) - h(x_i) \leq m(x, x_i) \leq \epsilon$. Thus $W(m)(P, Q) \leq 2\epsilon$. ∎

[1] The diameter of a set S is $\sup\{m(x, y) \mid x, y \in S\}$

4 Cadlag functions

Usually when one defines bisimulation one requires that "each step" of one process matches "each step" of the other process. What varies from situation to situation is the notion of step and of matching. In the present case there is no notion of atomic step: one has to match sequences instead. In the usual cases matching steps and matching sequences are equivalent so one works with steps as they are simpler. Here we have no choice: we have to work with timed sequences.

The timed sequences that one works with are functions from $[0, \infty)$ to the state space. Since we have discrete transitions these functions are not continuous. It turns out that the class of functions most often used are the *cadlag*[2] functions.

Definition 4.1 *Let (M, m) be a pseudometric space. $f : [0, \infty) \to M$ is* ***cadlag*** *if for any decreasing sequence $\{t\} \downarrow t_0$*

$$\lim_{t \to t_0} f(t) = f(t_0)$$

and for any increasing sequence $\{t\} \uparrow t_0$

$$\lim_{t \to t_0} f(t) \textit{ exists}$$

We write $\mathcal{D}^{(M,m)}[0, \infty)$ (or $\mathcal{D}^m[0, \infty)$, when M is clear from context) for cadlag functions with range (M, m).

These functions have very nice properties: for example, they have at most countably many discontinuities. More to the point perhaps, if one fixes an $\epsilon > 0$ there are at most finitely many jumps higher than ϵ: so all but finitely many jumps are small.

The study of metrics on spaces of these functions was initiated by Skorohod [Sko56]; see Whitt's book [Whi02] for an expository presentation. Skorohod defined several metrics: we use one called the J_2 metric. The most naive metric that one can define is the sup metric. This fails to capture the convergence properties that one wants: it insists on comparing two functions at the exact same points. Skorohod's first metric (the J_1 metric) allows one to perturb the time axis so that functions which have nearby values *at nearby points* are close. The J_1 metric also fails to satisfy certain convergence properties and we use the J_2 metric defined below, which like the J_1 metric, allows one to compare nearby time points.

Let (M, m) be a metric space. Let $|\cdot|$ be the metric on positive reals R^+ be defined by

$$|\cdot|(r, r') = |r - r'|$$

Definition 4.2 (Skorohod J_2 metric)
Let (M, m) be a metric space. Let f, g be cadlag functions: $[0, \infty) \to M$. $J(m)(f, g)$ is defined as:

$$\begin{aligned} \max(\quad & \sup_t \inf_{t'} [\max(m(f(t), g(t')), |t - t'|)], \\ & \sup_{t'} \inf_t [\max(m(f(t), g(t')), |t - t'|)] \quad) \end{aligned}$$

Thus the J_2 distance between two functions is the Hausdorff distance between their graphs (ie. the set of points $(x, f(x))$) in the space $[0, \infty) \times M$ equipped with the metric $d((x, s), (y, t)) = max(|x - y|, m(s, t))$.

The following lemma is immediate from definitions.

Lemma 4.3 $m_1 \preceq m_2 \Rightarrow J(m_1) \preceq J(m_2)$

The next lemma is standard, e.g. see Billingsley [Bil99].

[2] This is an acronym for the French phrase "continue á droite limites á gauche" meaning "continuous on the right with left limits."

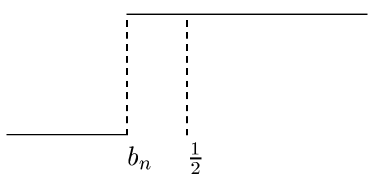
Lemma 4.4 (Skorohod)
If (M, m) is separable, $\mathcal{D}^M[0,\infty)$ is a separable space with a countable basis given by piecewise constant functions with finitely many discontinuities and finite range contained in a basis of M.

We consider a few examples, to illustrate the metric — see the book by Whitt [Whi02] for a detailed analysis of Skorohod's metrics. The first example shows that jumps/discontinuities can be matched by nearby jumps.

Example 4.5 *[Whi02] Let $\{b_n\}$ be an increasing sequence that converges to $\frac{1}{2}$. Consider:*

$$f_{b_n}(r) = \begin{cases} 0, r < b_n \\ 1, r \geq b_n \end{cases}$$

These are depicted in the picture below.



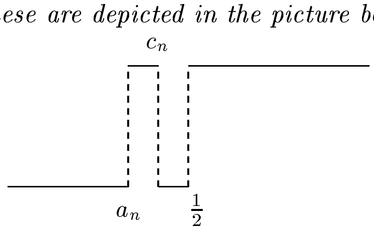
The sequence $\{f_{b_n}\}$ converges to $f_{\frac{1}{2}}$.

The next example shows that a single jump can be matched by multiple nearby jumps.

Example 4.6 *[Whi02] Let $\{a_n,\}, \{c_n\}$ be increasing sequences that converges to $\frac{1}{2}$ such that $a_n < c_n$. Let:*

$$g_n(r) = \begin{cases} 0, & r < a_n \\ 1, & a_n \leq r < c_n \\ 0, & c_n \leq r < \frac{1}{2} \\ 1, & r \geq \frac{1}{2} \end{cases}$$

These are depicted in the picture below.



The sequence $\{g_n\}$ converges to $f_{\frac{1}{2}}$.

The next two non-examples shows that "jumps are detected". In a later section, we develop a real-valued modal logic that captures the reasoning behind these two non-examples. Here, to provide preliminary intuitions, we provide a preview of this development in a specialized form. Given a cadlag function with range $[0,1]$ and the standard metric, and a Lipschitz function h on $[0,1]$ let $\mathcal{L}(h)(f)$ be defined as

$$\mathcal{L}(h)(f)(t) = \sup_{t'}\{h(f(t')) - |t' - t| \mid t' \in [0,\infty)\}$$

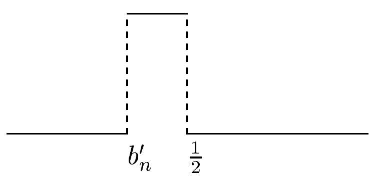
In this definition, view h as a test performed on the values taken by the function f — since h is a Lipschitz function on $[0,1]$, the results of such tests are smoothed out, and include the analogue of (logical) negation via the operation $1 - (\cdot)$ and smoothed conditionals via $h_q(x) = max(0, x - q)$ that correspond to a "greater than q" test. The $\mathcal{L}(h)(f)$ also performs an extra smoothing operation over time, so that the values of $\mathcal{L}(h)(f)$ at times t, t' differ by atmost $|t-t'|$. We can show that if $J(m)(f,g) < \epsilon$, then $\forall h.\forall t.|\mathcal{L}(h)(f)(t) - \mathcal{L}(h)(g)(t)| < 2\epsilon$. We will use this to establish non-convergence of function sequences in the J-metric in the next two examples.

The first non-example shows that jumps are detected — a sequence of functions with jumps cannot converge to a continuous function.

Example 4.7 *Let $\{b'_n\}$ be an increasing sequence that converges to $\frac{1}{2}$. Consider:*

$$f_{b'_n}(r) = \begin{cases} 0, & r < b_n \\ 1, & b'_n \leq r < \frac{1}{2} \\ 0, & \frac{1}{2} \leq r \end{cases}$$

These are depicted in the picture below.



The sequence $\{f_{b'_n}\}$ does not converge to the constant 0 *function* $\mathbf{0}$. *This is illustrated by considering the Lipschitz operator h'_ϵ on $[0,1]$ defined as:*

$$h'_\epsilon(r) = \max(0, \epsilon - |1-r|)$$

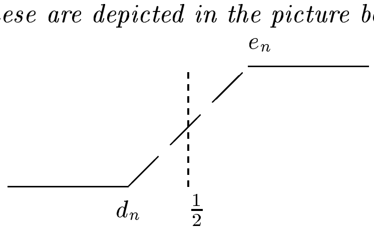
For all n, $\mathcal{L}(h)(g_n)(\frac{1}{2}) = \epsilon$, but $\mathcal{L}(h)(\mathbf{0})(\frac{1}{2}) = 0$.

The second non-example shows that continuous functions do not approximate a function with jumps.

Example 4.8 *[Whi02] Let $\{\epsilon_n\}$ be an decreasing sequences that converges to* 0. *Let $d_n = \frac{1}{2} - \frac{\epsilon_n}{2}$, $e_n \frac{1}{2} + \frac{\epsilon_n}{2}$. Consider:*

$$h_n(r) = \begin{cases} 0, & r < d_n \\ \frac{r-d_n}{\epsilon_n}, & d_n \leq r < e_n \\ 1, & r \geq e_n \end{cases}$$

These are depicted in the picture below.



The sequence $\{g_n\}$ does not converge to $f_{\frac{1}{2}}$. To analyze in terms of the operator $\mathcal{L}$, consider the Lipschitz operator h_ϵ on $[0,1]$ defined as:

$$h_\epsilon(r) = \max(0, \epsilon - |\frac{1}{2} - r|)$$

For all n, $\mathcal{L}(h)(g_n)(\frac{1}{2}) = \epsilon$, but $\mathcal{L}(h)(f_{\frac{1}{2}})(\frac{1}{2}) = 0$.

We conclude this section with a discussion of a delay operator on the space of cadlag functions.

Definition 4.9 *Let (M,m) be a metric space. Let $f \in \mathcal{D}^{(M,m)}[0,\infty)$. Let $s \in M, 0 \leq r$. Let $u : [0,r) \to M$ be a continuous function. Define $\mathtt{delay}_u(f) \in \mathcal{D}^{(M,m)}[0,\infty)$ as follows:*

$$\mathtt{delay}_u(f)(t) = \begin{cases} f(t-r), & t \geq r \\ u(t), & 0 \leq t < r \end{cases}$$

The distance between a cadlag function and its t-delayed version is no greater than t.

Lemma 4.10 *Let $0 \leq r$. Let $u : [0,r) \to M$ be continuous such that $(\forall 0 \leq r' < r)\ m(u(r'), f(0)) \leq r$. Then: $J(m)(\mathtt{delay}_u(f), f) \leq r$.*

Proof. Follows from $m(\mathtt{delay}_u(f)(t), f(t-r)) \leq r$, forall t. ∎

5 GSMPs and cadlag functions.

We deal with the temporal aspects of GSMPs next by constructing cadlag functions for paths: i.e. sequences of generalized states of a GSMP.

A sequence of generalized states is *finitely varying* if it is non-Zeno, i.e. for any i, $\sum_{j>i}^{\infty} \mathcal{T}(\langle s_j, \tilde{\mathbf{c_j}}\rangle$ diverges. Any finitely varying sequence of generalized states $\langle s_i, \tilde{\mathbf{c_i}}\rangle$ generates $f : [0,\infty) \to \mathcal{G}$ as follows:

- $f(t) = \langle s_i, \tilde{\mathbf{c}}\rangle$, where
$$\sum_{k=0}^{i} \mathcal{T}(\langle s_k, \tilde{\mathbf{c_k}}\rangle) \leq t < \sum_{k=0}^{i+1} \mathcal{T}(\langle s_k, \tilde{\mathbf{c_k}}\rangle)$$
and $\tilde{\mathbf{c}} = \tilde{\mathbf{c_i}} - \tilde{\mathbf{r_c}}|t - \sum_{k=0}^{i} \mathcal{T}(\langle s_k, \tilde{\mathbf{c_k}}\rangle)|$

Such finitely varying traces satisfy the following: for any interval $[t,t']$, that there is a finite partition $t = t_0 < t_1 < t_2 \ldots < t_n = t'$ such that:

- If $f(t_i) = \langle s, \tilde{\mathbf{c}}\rangle$, then: $(\forall t_i \leq t < t_{i+1}) f(t) = \langle s, \tilde{\mathbf{c}} - \tilde{\mathbf{r_c}}|t - t_i|\rangle$.

We write $\mathtt{Traces}\langle s, \tilde{\mathbf{c}}\rangle$ for the set of traces that start with $\langle s, \tilde{\mathbf{c}}\rangle$. The probability distributions associated with initial clock-values at states (f_s) and transitions ($\mathtt{next}$) induces a probability measure on $\mathtt{Traces}\langle s, \tilde{\mathbf{c}}\rangle$. The paths that are Zeno have measure zero, so the finitely-varying paths generate $\mathtt{Traces}\langle s, \tilde{\mathbf{c}}\rangle$ in measure.

Arbitrarily close approximations to the distance between finitely-varying functions $f, g \in \mathcal{D}^{(\mathcal{G},m)}[0,\infty)$ are forced by the distances between the values of f, g at finitely many points of time. This lemma is useful later on to show that our coinductive definition of metric has closure ordinal ω.

Lemma 5.1 *Let m be a pseudometric over generalized states. Let $f, g \in \mathcal{D}^{(\mathcal{G},m)}[0,\infty)$ be finitely-varying functions such that $J(m)(f,g) > \delta$.*

Then there is a finite subset $\mathcal{G}_{fin} \subseteq \mathcal{G}$ and $\epsilon > 0$ such that for any m', $J(m')(f,g) > \delta$ if $(\forall \mathtt{g}_1, \mathtt{g}_2 \in \mathcal{G}_{fin})$ $[m'(\mathtt{g}_1, \mathtt{g}_2) \geq m(\mathtt{g}_1, \mathtt{g}_2) - \epsilon]$.

Proof. If $J(m)(f,g) > \delta$, without loss of generality there is a t such that the Hausdorff distance of $(f(t), t)$ from $\mathtt{Graph}(g)$ is greater than $\delta + \gamma$ for some $\gamma > 0$. Consider the bounded interval $[t-\delta, t+\delta]$. By finite-variation of g, we have a partition $t_0 = t - \delta < t_1 < t_2 \ldots < t_n = t + \delta$ such that:

- $(\forall t_i \leq t < t_{i+1}) f(t) = \langle s, \tilde{\mathbf{c}} - \tilde{\mathbf{r_c}}(t - t_i)\rangle$, where $f(t_i) = \langle s, \tilde{\mathbf{c}}\rangle$
- $|t_i - t_{i+1}| < \frac{\gamma}{2}$
- If t', t'' are in the same partition, then $m(f(t'), f(t'')) < \frac{\gamma}{2}$

Construct $\mathcal{G}_{fin} = \{\langle s_1, \tilde{\mathbf{c_1}}\rangle, \ldots, \langle s_n, \tilde{\mathbf{c_n}}\rangle\}$ by choosing one $\langle s_i, \tilde{\mathbf{c_i}}\rangle$ each from $g[t_{i-1}, t_i]$ Choose $\epsilon = \frac{\gamma}{2}$. For any m' on generalized states such that:

$$(\forall i,j) m'(\langle s_i, \tilde{\mathbf{c_i}}\rangle, \langle s_j, \tilde{\mathbf{c_j}}\rangle) \geq m(\langle s_i, \tilde{\mathbf{c_i}}\rangle, \langle s_j, \tilde{\mathbf{c_j}}\rangle) - \epsilon$$

the Hausdorff distance of $(f(t), t)$ from $\mathtt{Graph}(g)$ is greater than δ by construction. ∎

6 Bisimulation style definition of metric

Let $\mathcal{M}$ be the class of pseudo-metrics on generalized states that satisfy:

$$\texttt{Props}(s) \neq \texttt{Props}(s') \Rightarrow m(\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c'}}\rangle) = 1$$

where $\texttt{Props}(s)$ is the set of atomic propositions true in state s. We order these pseudo-metrics as in section 3: $m_1 \preceq m_2$ if $(\forall s,t) m_1(s,t) \geq m_2(s,t)$. Fix $0 < k < 1$. Define a functional $\mathcal{F}_k$ on $\mathcal{M}$:

Definition 6.1 *$\mathcal{F}_k(m)(\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c'}}\rangle) < \epsilon$ if*

$$k \times W(J(m))(\texttt{Traces}\langle s, \tilde{\mathbf{c}}\rangle, \texttt{Traces}\langle s', \tilde{\mathbf{c'}}\rangle) < \epsilon$$

In this definition, view k as a discount factor. In the next section, we will show that the choice of k does not affect the essential character of the metric.

As an immediate consequence of lemmas 4.3 and 3.4:

Lemma 6.2 *$\mathcal{F}_k$ is monotone on $\mathcal{M}$.*

Since $(\mathcal{M}, \preceq)$ is a complete lattice, $\mathcal{F}_k$ has a maximum fixed point, $m_{\mathcal{F}_k}$.

Definition 6.3 *m is a metric-bisimulation if $m \preceq \mathcal{F}_k(m)$.*

It is well known that the greatest fixed point of $\mathcal{F}_k$ is given by:

$$m_{\mathcal{F}_k} = \bigsqcup\{m \mid m \text{ is a metric-bisimulation}\}$$

Thus, a metric-bisimulation m provides an upper bound on the distances assigned by $m_{\mathcal{F}_k}$.

Consider the equivalence relation $\simeq = \{(\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c'}}\rangle) \mid m_{\mathcal{F}_k}(\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c'}}\rangle) = 0\}$. $\simeq$ describes a notion of bisimulation and is explicitly defined as follows.

Let $\mathcal{M}_{\{0,1\}}$, the sublattice of $\mathcal{M}$ consisting of metrics whose range is $\{0,1\}$, i.e. all distances are either 0 or 1. $\mathcal{M}_{\{0,1\}}$ is essentially the class of equivalence relations. A simple proof shows that:

$$\simeq = \bigcup\{m^{-1}(0) \mid m \in \mathcal{M}_{\{0,1\}}, m \preceq \mathcal{F}_k(m)\}$$

As an example of metric-reasoning, we now show that generalized states with the same state, but clock values reflecting evolution for a time t are $m_{\mathcal{F}_k}$-close.

Lemma 6.4 *Define a pseudo-metric m on generalized states as follows:*

$$m(\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c'}}\rangle) = \begin{cases} \min(1,t), & \text{if } s = s' \text{ and} \\ \quad \tilde{\mathbf{c}} = \tilde{\mathbf{c'}} + \tilde{\mathbf{r_c}}t \text{ or} \\ \quad \tilde{\mathbf{c}} = \tilde{\mathbf{c'}} - \tilde{\mathbf{r_c}}t \\ 1, & \text{otherwise} \end{cases}$$

Then: $m \preceq \mathcal{F}_k(m)$.

Proof. It suffices to prove that for all generalized states $\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c'}}\rangle$

$$\mathcal{F}_k(m)(\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c'}}\rangle) \leq m(\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c'}}\rangle).$$

The only case to consider is when $s = s'$, wlog assume $\tilde{\mathbf{c}} \leq \tilde{\mathbf{c'}}$.

Define $u : [0,t) \to \mathcal{G}$ as follows.

$$u(t) = \langle s, \tilde{\mathbf{c'}} - \tilde{\mathbf{r_c}}t\rangle$$

With this definition, it is clear that:

$$\texttt{Traces}\langle s', \tilde{\mathbf{c'}}\rangle = \{\texttt{delay}_u(f) \mid f \in \texttt{Traces}\langle s, \tilde{\mathbf{c}}\rangle\}$$

with the distribution inherited from $\texttt{Traces}\langle s, \tilde{\mathbf{c}}\rangle$. Now this induces a matching between the traces in $\texttt{Traces}\langle s, \tilde{\mathbf{c}}\rangle$ and those of $\texttt{Traces}\langle s', \tilde{\mathbf{c'}}\rangle$. This in turn induces a distribution on the product space $\texttt{Traces}\langle s, \tilde{\mathbf{c}}\rangle \times \texttt{Traces}\langle s', \tilde{\mathbf{c'}}\rangle$. Using this distribution as the ρ in Lemma 3.5 - the dual form of the definition - we get the result. ∎

The role played by the discount constant k is captured in the following fact:

$$(\forall \langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c}'}\rangle)\ |m_{n+1}(\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c}'}\rangle) - m_n(\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c}'}\rangle)| \leq k^{n+1}.$$

This is the key step in the proof of the following lemma.

Lemma 6.5 *$m_{\mathcal{F}_k}$ is separable.*

Proof. We first note that if m, m' are such that $(\forall s)|m(s) - m'(s)| \leq \delta$, then:

- $(\forall P, Q)\ |W(m)(P,Q) - W(m')(P,Q)| \leq \delta$, this is immediate from the dual formulation
- $(\forall f, g)|J(m)(f,g) - J(m')(f,g)| \leq \delta$, this is immediate from the definition.

Let $m_0 = \top$ and $m_{i+1} = \mathcal{F}_k(m_i)$. We prove by induction on n that $(\forall \langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c}'}\rangle)\ |m_{n+1}(\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c}'}\rangle) - m_n(\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c}'}\rangle)| \leq k^{n+1}$.

- Base. $n = 0$. Follows from the fact that m_0 is the constant 0 function and m_1 is bounded above by k.
- Induction. Assume $(\forall \langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c}'}\rangle)\ |m_{n+1}(\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c}'}\rangle) - m_n(\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c}'}\rangle)| \leq k^{n+1}$. Thus, for any $\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c}'}\rangle)$, $W(J(m_{n+1})$ and $(\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c}'}\rangle)$ and $W(J(m_n))$ differ by atmost k^{n+1}. So:

$$|m_{n+2}(\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c}'}\rangle) - m_{n+1}((\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c}'}\rangle)| \leq k^{n+2}$$

Let $m = \sup m_i$. From above, $(\forall \langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c}'}\rangle)$

$$|m(\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c}'}\rangle) - m_n(\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c}'}\rangle)| \leq \frac{k^{n+1}}{1-k}$$

Thus, an ϵ ball around $\langle s, \tilde{\mathbf{c}}\rangle$ wrt the metric m can be realized as the countable union of open sets wrt the metrics m_n. The result now follows from the separability of the metrics m_n.

∎

The separability of $m_{\mathcal{F}_k}$ enables one to prove the analogue of lemma 5.1.

Lemma 6.6 (Finite detectability of distances)
Let m be a pseudometric on $\mathcal{G}$ with countable basis. Let $W(J(m))(\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c}'}\rangle) > \delta$.

Then there is a finite subset $\mathcal{G}_{fin} \subseteq \mathcal{G}$ and $\epsilon > 0$ such that for any m', $W(J(m'))(\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c}'}\rangle) > \delta$ if $(\forall \mathrm{g_s}, \mathrm{g'_s} \in \mathcal{G}_{fin})\ [m'(\mathrm{g_s}, \mathrm{g'_s}) \geq m(\mathrm{g_s}, \mathrm{g'_s}) - \epsilon]$.

Proof.
Let $W(J(m))(\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c}'}\rangle) > \delta + \gamma$, for $\gamma > 0$. Let h be an arbitrary, but fixed, finitely-varying function.

From separability of m, lemma 4.4 yields separability of $J(m)$. Using separability of $J(m)$ and lemma 3.7, we construct finite linear combinations $L_1[1 - \sum_i p_i]h + \sum_i p_i f_i$ and $L_2 = [1 - \sum_i p'_i]h + \sum_i p'_i f'_i$ such that:

- $\sum_i p_i > 1 - \frac{\gamma}{16}$, $\sum_i p'_i > 1 - \frac{\gamma}{16}$
- $W(J(m))(\mathtt{Traces}(\langle s, \tilde{\mathbf{c}}\rangle), L_1) < \frac{\gamma}{4}$
- $W(J(m))(\mathtt{Traces}(\langle s', \tilde{\mathbf{c}'}\rangle), L_2) < \frac{\gamma}{4}$

So, $W(J(m))(L_1, L_2) > \delta + \frac{\gamma}{2}$.

Using corollary 5.1 for every pair in $(\{f_i \mid i\} \cup \{h\}) \times (\{f'_j \mid j\} \cup \{h\})$, yields a finite set $\mathcal{G}_{fin} \subseteq \mathcal{G}$ and $\epsilon > 0$ such that for any m', if $(\forall \mathrm{g_s}, \mathrm{g'_s} \in \mathcal{G}_{fin})\ [m'(\mathrm{g_s}, \mathrm{g'_s}) \geq m(\mathrm{g_s}, \mathrm{g'_s}) - \epsilon]$, then:

- $J(m')(f_i, f'_j) > J(m)(f_i, f'_j) - \frac{\gamma}{4}$ for all i, j
- $J(m')(f_i, h) > J(m)(f_i, h) - \frac{\gamma}{4}$
- $J(m')(f'_i, h) > J(m)(f'_i, h) - \frac{\gamma}{4}$ for all i

A simple use of the dual form of the Wasserstein metric yields $W(J(m'))(L_1, L_2) > W(J(m))(L_1, L_2) - \frac{\gamma}{4}$ under these conditions. ∎

Lemma 6.7 *$\mathcal{F}_k$ has closure ordinal ω.*

Proof. The proof proceeds by showing that the maximum fixed point m is given by $m = \sqcup_i m_i$, where $m_0 = \top$ and $m_{i+1} = \mathcal{F}_k(m_i)$.

Let $m(\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c}'}\rangle) > \epsilon$. From lemma 6.6, we deduce the finitely many conditions of the form

$$m'(\langle s_i, \tilde{\mathbf{c_i}}\rangle, \langle s_i', \tilde{\mathbf{c_i'}}\rangle) > \epsilon_i$$

that suffice to ensure that $F(m') > \epsilon$. Each of these finitely many conditions are met at a finite index, therefore by ω they are all met and the result follows. ∎

7 Uniform spaces

A metric captures a quantitative notion of distance or "nearness", a topology captures a qualitiative notion of nearness: a topology is enough to talk about convergence and continuity. A topology is, however, not enough to capture a notion of relative distance. One cannot say "x is closer to y than it is to z" on the basis of a topology alone. A uniform space – see, for example, [Ger85] for a quick survey – captures the essence of the relative distance notion in metric spaces: if there are points x, y, z such that x is closer to y than to z, uniform spaces have enough data to capture this without committing to the actual numerical values of the distances. The aim of this section is to show that our treatment is "upto uniformity" – this is a formal way of showing that there is no ad-hoc treatment of the quantitative metric distances. In particular, we show that different discount factors k yield the same uniformity.

Let S be a set.

Definition 7.1 *A* pseudo-uniformity*, $\mathcal{U}$ is a collection of subsets of $S \times S$, called entourages, that satisfies:*

- $(\forall E \in \mathcal{U})\ (\forall x \in S)\ (x, x) \in E$
- $E \in \mathcal{U} \Rightarrow\ E^{-1} \in \mathcal{U}$
- $E \in \mathcal{U} \Rightarrow\ (\exists E' \in \mathcal{U})\ E'E' \subseteq E$
- $E, E' \in \mathcal{U} \Rightarrow\ E \cap E' \in \mathcal{U}$
- $E \in \mathcal{U}, E \subseteq E' \Rightarrow\ E' \in \mathcal{U}$

One can think of the entourages as defining approximations to the identity relation - just as the neighbourhood of a point can be thought of as an approximation of the point. The first axiom says this, the second axiom is symmetry and the third is a truncated version of transitivity and the final axiom says that a superset of an approximation to the identity is also an approximation to the identity. The usual presentation of uniformities also includes $\bigcap_{E \in \mathcal{U}} E = \{(x, x) \mid x \in S\}$, but this condition is not appropriate to our pseudo-metric setting.

Definition 7.2 *A pseudo-uniform space is a pair $(\mathcal{S}, \mathcal{U})$ where $\mathcal{U}$ is a pseudouniformity on $\mathcal{S}$.*

There is a natural notion of map between uniform spaces. A morphism between uniform spaces generalizes uniformly-continuous functions.

Definition 7.3 *A morphism f between (pseudo)uniform spaces $f : (S_1, \mathcal{U}_1) \to (S_2, \mathcal{U}_2)$ is a function $f : S_1 \to S_2$ such that*

$$(\forall E_2 \in \mathcal{U}_2) \hat{f}^{-1}(E_2) \in \mathcal{U}_1$$

where $\hat{f} : S_1 \times S_1 \to S_2 \times S_2$ is given by $\hat{f}(x, y) = (f(x), f(y))$.

To gain intuition into this definition, we describe how a pseudometric generates a pseudo-uniformity. Given a pseudo-metric m on $\mathcal{S}$, let $K_m^\epsilon = \{(x,y) \mid m(x,y) < \epsilon\}$ for $\epsilon > 0$. We get a pseudo-uniformity by considering [Ger85, p.218]:

$$\mathcal{U} = \{E \subseteq \mathcal{S} \times \mathcal{S} \mid (\exists \epsilon) K_m^\epsilon \subseteq E\}$$

Thus, if $m(x,y) < \epsilon$ and $m(x,z) > \epsilon$, there is an entourage that contains (x,y) but not (x,z). Clearly if one just scales the metric this construction yields the same uniformity.

Thus, all pseudometrics induce pseudo-uniformities, but the converse is not true: there are pseudo-uniformities that are not induced by metrics. Two metrics m, m' on the same set induce the same uniformity if and only if the identity map is uniformly continuous in both directions.

7.1 The lattice of uniformities

Consider uniformities induced by pseudo-metrics on a fixed set of states $\mathcal{S}$.

Definition 7.4 *$\mathcal{PMU}$ is the class of pseudo-metrizable uniformities $\{\mathcal{U}_i\}$ on $\mathcal{S}$ ordered as follows.*

$$\mathcal{U}_1 \leq \mathcal{U}_2 \quad \textit{if} \quad \mathcal{U}_2 \subseteq \mathcal{U}_1$$

This order is closely related to the order on the lattice of pseudometrics.

Lemma 7.5 *Let pseudometrics m_1, m_2 induce $\mathcal{U}_1$,$\mathcal{U}_2$ respectively, Let $\mathcal{U}_2 \subseteq \mathcal{U}_1$. Then pseudometric m defined as $m(s,t) = \max(m_2(s,t), m_1(s,t))$ also induces $\mathcal{U}_1$.*

Proof. Let $\mathcal{U}$ be the uniformity induced by m. We need to show that $\mathcal{U} = \mathcal{U}_1$. Since $K_m^\epsilon \subseteq K_{m_1}^\epsilon$, $\mathcal{U} \supseteq \mathcal{U}_1$.

We now show that $\mathcal{U} \subseteq \mathcal{U}_1$. Consider an entourage $E \in \mathcal{U}$. Thus, there is an ϵ such that $E \supseteq K_m^\epsilon$, i.e. $E \supseteq K_{m_1}^\epsilon \cap K_{m_2}^\epsilon$. But $K_{m_2}^\epsilon \in \mathcal{U}_2$ and by assumption $\mathcal{U}_2 \subseteq \mathcal{U}_1$, $K_{m_2}^\epsilon \in \mathcal{U}_1$. So, $K_{m_1}^\epsilon \cap K_{m_2}^\epsilon \in \mathcal{U}_1$. So, $E \in \mathcal{U}_1$. ∎

7.2 Wasserstein uniformity

Lemma 7.6 *Let $(M,m), (M,m')$ be such that the uniformities induced by m, m' are the same. Then, $W(m), W(m')$ induce the same uniformity.*

Proof. Since the uniformities induced by m, m' are the same, the identity map $(M,m) \to (M,m')$ is uniformly continuous. We need to show that the identity map on distributions with metrics $W(m)$ and $W(m')$ is uniformly continuous.

Since the uniformity induced by $W(m)$ (resp. $W(m')$) is the same as the uniformity induced by the Prohorov metric $\pi(m)$ (resp. $\pi(m')$), it suffices to prove that the identity map on distributions with metrics $\pi(m)$ and $\pi(m')$ is uniformly continuous.

Let $\pi(m')(P,Q) \leq \epsilon$. Let δ_ϵ be such that: $m(x,y) < \delta_\epsilon \Rightarrow m'(x,y) < \epsilon$. Let $\gamma = \min(\epsilon, \delta_\epsilon$. Then:

$$\begin{aligned}
\pi(m)(P,Q) \leq \gamma \quad &\Rightarrow \quad P(U) \leq Q(U_m^\gamma) + \gamma \\
&\Rightarrow \quad P(U) \leq Q(U_{m'}^\epsilon) + \gamma \quad [Q(U_m^\gamma) \subseteq Q(U_{m'}^\epsilon)] \\
&\Rightarrow \quad P(U) \leq Q(U_{m'}^\epsilon) + \epsilon \quad (\gamma \leq \epsilon) \\
&\Rightarrow \quad \pi(m')(P,Q) \leq \epsilon
\end{aligned}$$

∎

In the light of this theorem, we write $W(\mathcal{U})$ for the uniformity generated by a pseudo-metrizable uniformity $\mathcal{U}$. As a direct consequence of lemma 3.4, we have:

Corollary 7.7 $\mathcal{U} \leq \mathcal{U}' \Rightarrow W(\mathcal{U}) \leq W(\mathcal{U}')$.

7.3 Skorohod J2 uniformity

A similar result holds for the Skorohod metric.

Lemma 7.8 *The uniformity induced by the Skorohod J2 metric depends only the uniformity induced by m on M.*

Proof. For each $f \in \mathcal{D}^M[0,\infty)$ consider its graph, $\texttt{Graph}(f) = \{(f(t), t)\}$. For a relation R and a set X, write $R(Y) = \{x \mid (x,y) \in R, y \in Y\}$. Similarly, $(X)R = \{y \mid (x,y) \in R, x \in X\}$. These operations are monotone, under the subset ordering, in both X and R.

Let the space $\mathcal{S} = [0,\infty) \times M$ be equipped with the metric $d((x,s),(y,t)) = max(|x-y|, m(s,t))$. This metric induces a uniformity $U(\mathcal{S})$ on $\mathcal{S}$.

Let $E \subseteq \mathcal{S} \times \mathcal{S}$ be an entourage in $U(\mathcal{S})$. Consider the subset of $J(E)$ of $\mathcal{D}^M[0,\infty) \times \mathcal{D}^M[0,\infty)$ induced by E as follows. $J(E)$ is the set of all (f,g) such that:

- $(\texttt{Graph}(f))E \supseteq \texttt{Graph}(g)$
- $E(\texttt{Graph}(g)) \supseteq \texttt{Graph}(f)$

Consider $\mathcal{U} = \{S \mid S \supseteq J(E), E \subseteq \mathcal{S} \times \mathcal{S}, E \in U(\mathcal{S})\}$. We will show that $\mathcal{U}$ is the uniformity generated by $J(m)$.

Rewriting $J(m)$ in this style: $J(m)(f,g) < \epsilon \Leftrightarrow$

- $(\texttt{Graph}(f))K^{\epsilon}_{\langle m, |\cdot|\rangle} \supseteq \texttt{Graph}(g)$
- $K^{\epsilon}_{\langle m, |\cdot|\rangle}(\texttt{Graph}(g)) \supseteq \texttt{Graph}(f)$

Clearly, the uniformity generated by $J(m)$ is a subset of $\mathcal{U}$, since $K^{\epsilon}_{\langle m, |\cdot|\rangle}$ is an entourage of $\mathcal{S} \times \mathcal{S}$, and thus considered in the definition of $\mathcal{U}$. Furthermore, any arbitrary entourage E in $U(\mathcal{S})$ is a superset of these basic entourages, and yields a superset by monotonicity of the relational operations.

The result now follows by the upward-closure axiom in the definition of $\mathcal{U}$ and the uniformity generated by $J(m)$. ∎

The proof of this theorem also shows that the construction $J(m)$ yields the same uniformity for other definitions of metrics on M that yield the same uniformity, e.g. the metric $d((x,s),(y,t)) = |x-y| + m(s,t))$.

In light of this theorem, we write $J(\mathcal{U})$ for the uniformity generated by a pseudo-metrizable uniformity $\mathcal{U}$. As a direct consequence of lemma 4.3, we have:

Corollary 7.9 $\mathcal{S}_1 \leq \mathcal{S}_2 \Rightarrow J(\mathcal{S}_1) \leq J(\mathcal{S}_2)$

7.4 A functional on the lattice of uniformities

Combining the above results, we deduce the existence of a monotone function $\mathcal{F}_k$ on the lattice of uniformities. This function is insensitive to the actual numerical value of the discount factor k.

Lemma 7.10 $(\forall 0 < k, k' < 1)$ $\mathcal{F}_k = \mathcal{F}_{k'}$.

Proof. By Lemmas 7.6 and 7.8 we see that the functional $\mathcal{F}_k$ is defined upto uniformity. If we change k to k' we are simply rescaling the metric, this clearly gives the same uniformity. ∎

Furthermore, for any discount factor $0 < k < 1$, we get the same maximum fixed point in the lattice of uniformities.

Theorem 7.11 *The maximum fixpoint in $(\mathcal{PMU}, \leq)$ is the uniformity induced by $m_{\mathcal{F}_k}$, the maximum fixpoint in $(\mathcal{M}, \preceq)$.*

Proof. It suffices to show that the greatest lower bound of the $\mathcal{U}_i$, $\cup\mathcal{U}_i$ is the uniformity induced by $\sup m_i$. Let $\mathcal{V}$ be the uniformity induced by $\sup m_i$. Since $m_i \preceq \sum m_i$ for all i we have that $\mathcal{U}_i \subseteq \mathcal{V}$ for all i. Thus $\cup\mathcal{U}_i \subseteq \mathcal{V}$ and hence the identity function from $\mathcal{V}$ to $\mathcal{U}_i$ is uniformly continuous.

For the converse, we will show that the identity function from $\cup\mathcal{U}_i \rightarrow \mathcal{V}$ is uniformly continuous. We know - from the proof of Lemma 6.5 - that for any generalized states $\mathbf{g_s}$ and $\mathbf{g}'_\mathbf{s}$

$$\sup m_i(\mathbf{g_s}, \mathbf{g}'_\mathbf{s}) - m_i(\mathbf{g_s}, \mathbf{g}'_\mathbf{s}) \leq \frac{k^{n+1}}{1-k}.$$

We choose n and δ such that $\delta + \frac{k^{n+1}}{1-k} < \epsilon$. Now for any such δ and n we have

$$m_n(\mathbf{g_s}, \mathbf{g}'_\mathbf{s}) < delta \Rightarrow \sup m_i(\mathbf{g_s}, \mathbf{g}'_\mathbf{s}) < epsilon.$$

This shows that for any $K^{\epsilon}_{\sup m_i}$ is contained in a $K^{\delta}_{m_n}$ and hence that the identity function is uniformly continuous. ∎

This proof relies on the fact that $k < 1$ otherwise the δ would not be defined.

8 Examples

In this section, we discuss several examples of the use of approximate reasoning techniques. The general approach in this section is to identify natural quantitative observables, already explored in the literature, that are amenable to approximation — i.e. to calculate the observable at a state $\langle s, \tilde{\mathbf{c}}\rangle$ upto ϵ, it suffices to calculate it a close-enough state $\langle s', \tilde{\mathbf{c}'}\rangle$. This is clearly implied by continuity of the observable w.r.t. the metric $m_{\mathcal{F}_k}$.

The main technical tool that we use to establish continuity of observables is a continuous mapping theorem, e.g. see [Whi02, She87] for an introductory exposition.

Theorem 8.1 (Continuous mapping theorem)
Let P_n be a sequence of probability distributions on X that weakly converge to P. Let U be a continuous function $X \rightarrow R$. Then $\int U dP_n$ converges to $\int U dP$.

8.1 Expected time to hit a proposition

Let $\mathbf{p}$ be a proposition. We consider the expected time required to hit a p-state, i.e. a state where the proposition $\mathbf{p}$ is true. Define $\mathtt{Hit}_p : \mathcal{D}^{m_F}[0, \infty) \rightarrow [0, \infty]$:

$$\mathtt{Hit}_p(f) = \inf\{t \mid f(t) = \langle s, \tilde{\mathbf{c}}\rangle, \mathbf{p} \ \text{ true at } s\}$$

$\mathtt{Hit}_p$ is a continuous function — if $J(m_F)(f, g) < \epsilon$, then $|\mathtt{Hit}_p(f) - \mathtt{Hit}_p(g)| < \epsilon$.

So, using the continuous mapping theorem, we deduce that if $\{\langle s_i, \tilde{\mathbf{c_i}}\rangle\}$ converges to $\langle s, \tilde{\mathbf{c}}\rangle$ then the sequence of expected times to hit a p-state from $\{\langle s_i, \tilde{\mathbf{c_i}}\rangle\}$ converges to the expected time to hit a p-state from $\{\langle s, \tilde{\mathbf{c}}\rangle\}$. In fact, since in this case, $\mathtt{Hit}_p$ is a 1-Lipschitz function, we can also deduce the rate of convergence using [Whi02] — if $m_F(\langle s_i, \tilde{\mathbf{c_i}}\rangle, \langle s, \tilde{\mathbf{c}}\rangle) < \epsilon$, then the expected times to hit a p-state differ by atmost 2ϵ.

8.2 Expected rewards

Let r_i be an assignment of rewards to states s_i such that if $r_i \neq r_j$, then states s_i, s_j differ in the truth-assignment of at least one proposition. This restriction can be viewed purely as a modelling constraint.

Define a function $R : \mathcal{G} \to [0, \infty)$ by:

$$R(\langle s_i, \tilde{\mathbf{c}}\rangle = r_i$$

Under the hypothesis that distinct rewards are distinguished propositionally, R defines a continuous function.

For any finitely-varying f, consider $\texttt{CumR}(f)$, a continuous function of t defined as follows:

$$\texttt{CumR}(f)(T) = \int_0^T R(f(t))dt$$

By standard results — e.g. see [Whi02] — $\texttt{CumR}$ is a continuous function from $\mathcal{D}^{m_F}[0, \infty)$ to $(C, \texttt{unif})$ where C is the space of continuous functions from $[0, \infty) \to [0, \infty)$ with the uniform metric:

$$\texttt{unif}(f, g) = \sup_t |f(t) - g(t)|$$

Consider the following continuous functions from $(C, \texttt{unif})$ to $[0, \infty)$:

- For a fixed T, cumulative reward at time T.
- For a fixed T, average reward per unit time at T.
- The supremum of the times T at which cumulative reward is less than a fixed v, for some value v.
- The supremum of the times T at which the average reward is less than a fixed v, for some value v.

In each of these cases, by composing with $\texttt{CumR}$, we get a continuous function from $\mathcal{D}^{m_F}[0, \infty)$ to $[0, \infty)$. So, the continuous mapping theorem applies, and we deduce that if $\{\langle s_i, \tilde{\mathbf{c}}_\mathbf{i}\rangle\}$ converges to $\langle s, \tilde{\mathbf{c}}\rangle$ then the sequence of expected values from $\{\langle s_i, \tilde{\mathbf{c}}_\mathbf{i}\rangle\}$ converges to the expected value at $\{\langle s, \tilde{\mathbf{c}}\rangle\}$.

9 Functional characterization of uniformity

In an early treatment of metrics for LMPs [DGJP99, DGJP04] the metric was defined through a class of functions closely related to a modal logic. The idea was that, in a probabilistic setting, random variables play a role analogous to modal formulas. A class of random variables (measurable functions) was defined on the state space and the metric was obtained by taking the sup over this class of functions. The coinductive definition of the metric came later [vBW01b, DGJP02] and was shown to be the same as the metric defined logically. In the present work we develop the subject along similar lines. We already have the fixed-point version of the metric: we now give the "logical" view. In this section, we provide an explicit construction of the maximum fixed point by considering a class of $[0, 1]$ valued functions.

9.1 Function expressions

Definition 9.1 Fix $0 < k < 1$. The syntax of function expressions is given by:

$$\begin{array}{rcl} F_k & ::= & \mathbf{1} \mid \mathbf{p} \mid \min(F_k, F_k) \mid h \circ F_k \mid \int G_k \\ G_k & ::= & \mathcal{L}(F_k)(t) \mid \min(G_k(t), G_k(t')) \mid h \circ G_k(t) \end{array}$$

where $\mathbf{p}$ ranges over atomic propositions, h is any Lipschitz function on $[0, 1]$, $t \in [0, \infty)$.

The subscript k gives the discount factor. We will not usually write this factor explicitly. Intuitively the F-function expressions are evaluated at generalized states, and the G-function expressions are evaluated on finitely-varying paths at the times shown. In a temporal logic with state and path formulas, like CTL*, the path formulas are *implicitly* evaluated at the first time. This may not seem to be the case with a formula like Gp ($\Box p$ in LTL notation) but is clear with a formula like Xp ($\bigcirc p$). In our G-formulas we cannot have a first time: we provide the time as an explicit parameter. One can imagine a much richer language of path formulas: for example, one might have time averages along a path. However, the present language suffices for the definition of the metric.

As preliminary intuition, $\mathbf{1}$ corresponds to the formula `true`, $\min(\cdot,\cdot)$ corresponds to conjunction[3], and $h \circ f$ encompasses both testing (via $h(x) = \max(x - q, 0)$) and negation (via $h(x) = 1 - x$). At a generalized state $\langle s, \tilde{\mathbf{c}}\rangle$, $\int G(t)$ yields the (discounted) expectation of $G(t)$ wrt the distribution of $\texttt{Traces}\langle s, \tilde{\mathbf{c}}\rangle$. The intuition underlying $\mathcal{L}(F)(t)$ has been discussed in section 4 — at a finitely-varying function f, $\mathcal{L}(F)(t)$ yields the evaluation at time t of a time-smoothed variant of f.

We formalize these intuitions below. The interpretations of F-function expressions and G-function expressions yield maps whose range is the interval $[0, 1]$.

- The domain of F-function expressions is $\mathcal{G}$, the set of generalized states.
- The domain of G-function expressions is the set of finitely-varying functions with range $\mathcal{G}$.

Fix a GSMP. F-function expressions are evaluated as follows at a generalized state $\langle s, \tilde{\mathbf{c}}\rangle$:

$$
\begin{array}{rcl}
\mathbf{p}(\langle s, \tilde{\mathbf{c}}\rangle) & = & 1, \text{ iff } \texttt{p} \text{ true at } s \\
\mathbf{1}(\langle s, \tilde{\mathbf{c}}\rangle) & = & 1 \\
\min(F1, F2)(\langle s, \tilde{\mathbf{c}}\rangle) & = & \min(F1(\langle s, \tilde{\mathbf{c}}\rangle), F2(\langle s, \tilde{\mathbf{c}}\rangle)) \\
h \circ F(\langle s, \tilde{\mathbf{c}}\rangle) & = & h(F(\langle s, \tilde{\mathbf{c}}\rangle)) \\
(\int G(t))(\langle s, \tilde{\mathbf{c}}\rangle) & = & k \times \int G(t)(f)\mathrm{d}\mu
\end{array}
$$

where μ is the distribution of $\texttt{Traces}(\langle s, \tilde{\mathbf{c}}\rangle)$. Note that in this definition f varies among the paths of $\texttt{Traces}(\langle s, \tilde{\mathbf{c}}\rangle)$ so $G(t)$ is a measurable function on the space of these paths and μ is a measure on these paths.

G-function expressions are evaluated as follows at a finitely-varying function f:

$$
\begin{array}{rcl}
\mathcal{L}(F)(t)(f) & = & \sup_{t'}\{F(f(t')) - |t' - t|\} \\
\min(G1(t_1), G2(t_2))(f) & = & \min(G1(t_1)(f), G2(t_2)(f)) \\
h \circ G(t)(f) & = & h(G(t)(f))
\end{array}
$$

Thus, for $f \in \texttt{Traces}(\langle s, \tilde{\mathbf{c}}\rangle)$, $\mathcal{L}(F)(t)(f)$ is the upper Lipschitz approximation to $F_k^s \circ f$ evaluated at t.

9.2 A pseudometric from function expressions

We define a pseudometric d_k as follows.

Definition 9.2
$d_k(\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c}'}\rangle) = \sup_{F_k} |F_k(\langle s, \tilde{\mathbf{c}}\rangle) - F_k(\langle s', \tilde{\mathbf{c}'}\rangle)|$.

We proceed to show that the uniformities defined by these two metrics agree. Unlike the case with discrete time systems the metrics themselves do not agree: it is the uniformity that is common to the two of them.

[3] $\max(\cdot,\cdot)$ is definable as $1 - \min(1 - \cdot, 1 - \cdot)$ in both classes of function expressions.

Theorem 9.3 *The uniformity induced by d_k coincides with the uniformity induced by $m_{\mathcal{F}_k}$, the maximum fixed point of $\mathcal{F}_k$.*

Proof.

We demonstrate that the identity function is a uniformly continuous isomorphism between $\mathcal{G}$ equipped with the metrics d_k and $m_{\mathcal{F}_k}$.

Consider the identity function from domain with metric $m_{\mathcal{F}_k}$ and range with metric d_k. From remarks following lemma 7.8 and theorem 7.11, we can consider the metric $m_{\mathcal{F}_k}$ constructed using the following definition of $J(m)$:

$$\begin{aligned}\max(\quad & \sup_t \inf_{t'} [m(f(t), g(t')) + |t - t'|], \\ & \sup_{t'} \inf_t [m(f(t), g(t')) + |t - t'|)] \quad)\end{aligned}$$

A simple mutual inductive proof shows that:

- Every F-function-expression F satisfies:

$$|F(\langle c, \tilde{\mathbf{s}}\rangle) - F(\langle c', \tilde{\mathbf{s}'}\rangle)| \leq m_{\mathcal{F}_k}(\langle c, \tilde{\mathbf{s}}\rangle, \langle c', \tilde{\mathbf{s}'}\rangle)$$

- Every G-function-expression G satisfies

$$|G(t)(f) - G(t)(g)| \leq 2 \times J(m_{\mathcal{F}_k})(f, g)$$

This shows that the identity function from domain with metric $m_{\mathcal{F}_k}$ and range with metric d_k is uniformly continuous.

We prove the converse below. We show that $m_{\mathcal{F}_k}$ (using the standard definition) is dominated by d_k. We use the fact that the closure ordinal of $\mathcal{F}_k$ is ω. Let $m_0 = \top, m_{i+1} = \mathcal{F}_k(m)$. We show by induction on i that each m_i is dominated by d_k. We proceed in the following two steps:

- Let f, g be such that $J(m_i)(f, g) > \epsilon$, where $\epsilon > 0$. We show that there is a G-function expression such that for some t, $G(t)(g) = 0$ and $G(t)(g) > \epsilon$.

  If $J(m_i)(f, g) = \epsilon + \gamma$ where $\epsilon, \gamma > 0$, without loss of generality we can assume that there is a t such that:

  $$(\forall t')\ |t - t'| <= \epsilon + \gamma \Rightarrow\ m_i(f(t), g(t')) >= \epsilon + \gamma$$

  From finite-variance, we get a partition

  $$t_0 = t - (\epsilon + \gamma), t_1, t_2, \ldots, t_n + \epsilon + \gamma$$

  such that: such that:

  - Let $g(t_i) = \langle s, \tilde{\mathbf{c}}\rangle$. Then: $(\forall t_i \leq t < t_{i+1}) g(t) = \langle s, \tilde{\mathbf{c}} - \tilde{\mathbf{r_c}}|t - t_i|\rangle$.
  - $|t_i - t_{i+1}| < \frac{\gamma}{2}$

  From the assumption that m_i is dominated by d_k, there are F-function expression $F1, \ldots, Fn$ such that $|Fj(f(t)) - Fj(g(t_i))| > \epsilon + \frac{\gamma}{2}$. Using lemma 6.4, we have $(\forall t_i \leq t' < t_{i+1}) m_{\mathcal{F}_k}(g(t_i), g(t')) < \frac{\gamma}{2}$, and thus we deduce that $|Fj(f(t)) - Fj(g(t'))| > \epsilon$, for all $t_i \leq t' < t_{i+1}$. Using min and $h\circ$, without loss of generality, we can assume that $Fj(f(t)) > \epsilon$ and $Fj(g(t_i)) = 0$. Consider $\min(F1, \ldots, Fn)$ and $\mathcal{L}(F)(t)$. It evaluates to 0 on g and to a value $> \epsilon$ on f.

- Let $\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c}'}\rangle$ be such that $W(J(m_i))(\langle s, \tilde{\mathbf{c}}\rangle, \langle s', \tilde{\mathbf{c}'}\rangle) = \epsilon + \gamma$, where $\gamma > 0$. We show that there is a F-function expression such that , $F(\langle s', \tilde{\mathbf{c}'}\rangle) = 0$ and $F(\langle s, \tilde{\mathbf{c}}\rangle) > \epsilon$.

  Let h be an arbitrary, but fixed, finitely-varying function. From separability of m_i, we deduce separability of $J(m_i)$. Using lemma 3.7, there is a finite linear combination $L_1[1-\sum_i p_i]h+\sum_i p_i f_i$ such that $\sum_i p_i > 1 - \frac{\gamma}{16}$ and $W(J(m_i))(\texttt{Traces}(\langle s, \tilde{\mathbf{c}}\rangle), L_1) < \frac{\gamma}{4}$. Similarly, there is a finite linear combination $L_2 = [1-\sum_i p'_i]h + \sum_i p'_i f'_i$ such that $\sum_i p'_i > 1 - \frac{\gamma}{16}$ and $W(J(m_i))(\texttt{Traces}(\langle s', \tilde{\mathbf{c}'}\rangle), L_2) < \frac{\gamma}{4}$. So, we have $W(J(m_i))(L_1, L_2) > \epsilon + \frac{\gamma}{2}$ and it suffices to prove the result for finite linear combinations.

  From the above item, for each pair of summands (one from L_1 and the other from L_2), there are G-function expressions, G_{ij} that are non-zero only on $f_1, \ldots, f_n$ and zero on $f'_1, \ldots, f'_m$ and which yield arbitrarily close approximations to the distance between the f_i, f'_j pairs. The result now follows by considering $\max(G^p_{ij})$.

∎

10 Conclusions

We have given a pseudo-metric analogue of bisimulation for GSMPs. We have shown that this really depends on the underlying uniformity and that quantities of interest are continuous in this metric. We have given a coinduction principle and a logical characterization reminiscent of previous work for weak bisimulation of a discrete time concurrent Markov chain.

The previous approaches to bisimulation work well for CTMCs, precisely because of the fact that the distribution is memoryless; at any given instant the expected duration in a state and the transition probabilities only depend on the current state of the system, and thus one *can* define a bisimulation on the state space. In contrast, the problem of describing bisimulation for real-time processes that have general distributions, rather than memoryless distributions, has been vexing. In the present work, we have shifted emphasis to the generalized states that incorporate time and not tried to define a bisimulation on the ordinary states. Because the generalized states embody the quantitative temporal information we *have to work metrically*; an attempt to define bisimulation directly would have fallen afoul of the approximate nature of the timing information.

If we want to move to continuous state spaces and stochastic hybrid systems, the whole dynamical formalism has to be different: one can no longer think of paths as cadlag functions. We will have to use stochastic differential equations to describe the systems and the space of sample paths for the trajectories. That is a subject for future work and one that we have been heading towards from the inception of our work on LMPs [BDEP97, DEP02].

References

[AD94] R. Alur and D. Dill. A theory of timed automata. *Theoretical Computer Science*, 126:183–235, 1994.

[AH92] R. Alur and T. Henzinger. Logics and models of real-time: a survey. In G. Huizing, W. P. de Roever, and G. Rozenberg, editors, *REX workshop Real-time, theory and practice*, number 600 in Lecture Notes In Computer Science. Springer-Verlag, 1992.

[AN87] E. J. Anderson and P. Nash. *Linear Programming in Infinite-dimensional Spaces.* Discrete Mathematics and Computation. Wiley-Interscience, 1987.

[BDEP97] 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.

[BG02] M. Bravetti and R. Gorreiri. The theory of interactive semi-Markov processes. *Theoretical Computer Science*, 281(2):5–32, 2002.

[BHK99] C. Baier, H. Hermans, and J-P. Katonen. Approximative symbolic model checking of continuous-time Markov chains. In J. C. M. Baeten and S. Mauw, editors, *Proceedings of the Tenth International Conference on Concurrency Theory, CONCUR99*, number 1664 in Lecture Notes In Computer Science, pages 146–161. Springer-Verlag, 1999.

[Bil99] P. Billingsley. *Convergence of Probability Measures*. Wiley Interscience, 2nd edition edition, 1999.

[CSZ92] R. Cleaveland, S. Smolka, and A. Zwarico. Testing preorders for probabilistic processes. In W. Kuich, editor, *Automata, Languages and Programming (ICALP 92)*, number 623 in Lecture Notes in Computer Science, pages 708–719. Springer-Verlag, 1992.

[dAHM03] L. de Alfaro, T. Henzinger, and R. Majumdar. Discounting the future in systems theory. In J. Baeten, J. K. Lenstra, J. Parrow, and G. J. Woeginger, editors, *Thirtieth International Colloquium On Automata Languages And Programming*, number 2719 in Lecture Notes In Computer Science, pages 1022–1037. Springer-Verlag, 2003.

[DEP02] J. Desharnais, A. Edalat, and P. Panangaden. Bisimulation for labeled Markov processes. *Information and Computation*, 179(2):163–193, Dec 2002.

[DGJP99] 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.

[DGJP02] 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.

[DGJP04] Josée Desharnais, Vineet Gupta, Radhakrishnan Jagadeesan, and Prakash Panangaden. A metric for labelled Markov processes. *Theoretical Computer Science*, 318(3):323–354, June 2004.

[Ger85] Robert Geroch. *Mathematical Physics*. Chicago Lectures in Physics. University of Chicago Press, 1985.

[Han94] Hans A. Hansson. *Time and Probability in Formal Design of Distributed Systems*, volume 1 of *Real-time Safety-critical Systems*. Elseiver, 1994.

[HCH+02] B. R. Haverkort, L. Cloth, H. Hermans, J.-P. Katoen, and C. Baier. Model checking performability properties. In *Proceedings of the International COnference on Dependable Systems and Networks 2002*, pages 102–113. IEEE Computer Society, June 2002.

[Her02] H. Hermanns. *The Quest for Quantified Quality*. Number 2428 in Lecture Notes In Computer Science. Springer-Verlag, 2002.

[Hil94] J. Hillston. *A Compositional Approach to Performance Modelling*. PhD thesis, University of Edinburgh, 1994. Published as a Distinguished Dissertation by Cambridge University Press in 1996.

[JS90] C.-C. Jou and S. A. Smolka. Equivalences, congruences, and complete axiomatizations for probabilistic processes. In J.C.M. Baeten and J.W. Klop, editors, *CONCUR 90 First International Conference on Concurrency Theory*, number 458 in Lecture Notes In Computer Science. Springer-Verlag, 1990.

[LMMS98] P. D. Lincoln, J.C. Mitchell, M. Mitchell, and A. Scedrov. A probabilistic poly-time framework for protocol analysis. In *ACM Computer and Communication Security (CCS-5)*, 1998.

[LS91] K. G. Larsen and A. Skou. Bisimulation through probablistic testing. *Information and Computation*, 94:1–28, 1991.

[PLS00] A. Philippou, I. Lee, and O. Sokolsky. Weal bisimulation for probabilistic processes. In C. Palamidessi, editor, *Proceedings of CONCUR 2000*, number 1877 in Lecture Notes In Computer Science, pages 334–349. Springer-Verlag, 2000.

[She87] G. S. Shedler. *Regeneration and Networks of Queues*. Springer-Verlag, 1987.

[Sko56] A. V. Skorohod. Limit theorems for stochastic processes. *Theory of Probability and Applications*, 1:261–290, 1956.

[SL95] R. Segala and N. Lynch. Probabilistic simulations for probabilistic processes. *Nordic Journal of Computing*, 2(2):250–273, 1995.

[vBW01a] Franck van Breugel and James Worrell. An algorithm for quantitative verification of probabilistic systems. In K. G. Larsen and M. Nielsen, editors, *Proceedings of the Twelfth International Conference on Concurrency Theory - CONCUR'01*, number 2154 in Lecture Notes In Computer Science, pages 336–350. Springer-Verlag, 2001.

[vBW01b] Franck van Breugel and James Worrell. Towards quantitative verification of probabilistic systems. In *Proceedings of the Twenty-eighth International Colloquium on Automata, Languages and Programming*. Springer-Verlag, July 2001.

[Whi80] W. Whitt. Continuity of generalized semi-Markov processes. *Mathematics of Operations Research*, 5(4):494–501, 1980.

[Whi02] W. Whitt. *An Introduction to Stochastic-Process Limits and their Applications to Queues*. Springer Series in Operations Research. Springer-Verlag, 2002.

A Proofs of theorems on uniform spaces

Definition A.1 *A pseudouniform space is a pair* $(\mathcal{S},\mathcal{U})$ *where* $\mathcal{U}$ *is a pseudouniformity on* $\mathcal{S}$.

A morphism between uniform spaces generalizes uniformly-continuous functions.

Definition A.2 *A morphism* f *between (pseudo)uniform spaces* $f : (\mathcal{S}_1,\mathcal{U}_1) \to (\mathcal{S}_2,\mathcal{U}_2)$ *is a function* $f : \mathcal{S}_1 \to \mathcal{S}_2$ *such that*

$$(\forall E_2 \in \mathcal{S}_2) f^{-1}(E_2) \in \mathcal{U}_1$$

Products. Let $\mathcal{S}_1,\mathcal{S}_2$ be sets of states. Let $\mathcal{U}_1,\mathcal{U}_2$ be uniformities induced by pseudometrics m_1, m_2. The metric $\langle m_1, m_2\rangle$ on $\mathcal{S}_1 \times \mathcal{S}_2$ is defined by:

$$\langle m_1, m_2\rangle((x,y),(x',y')) = \max(m_1(x,x'), m_2(y,y'))$$

Lemma A.3 *The uniformity induced by* $\langle m_1, m_2\rangle$ *is the product of* $(\mathcal{S}_1,\mathcal{U}_1)$ *and* $(\mathcal{S}_2,\mathcal{U}_2)$.

Proof. Let $\mathcal{U}$ be the uniformity on $\mathcal{S}_1 \times \mathcal{S}_2$ induced by $\langle m_1, m_2\rangle$.

We first show that the projections are morphisms of uniform spaces. Let $E \in \mathcal{U}_1 \supseteq K^{\epsilon}_{m_1}$. Then, $\pi_i^{-1}(E) = E \times \mathcal{S}_2 \supseteq K^{\epsilon}_{\langle m_1, m_2\rangle}$, so is in $\mathcal{U}$.

Let $f_i : (\mathcal{S}_i,\mathcal{U}_i) \to (\mathcal{S}',\mathcal{U}')$, be morphisms of uniform spaces. Consider $\langle f_1, f_2\rangle$ defined as

$$\langle f_1, f_2\rangle(x_1,x_2) = (f_1(x_1), f_2(x_2))$$

We will show that $\langle f_1, f_2\rangle : (\mathcal{S}_1 \times \mathcal{S}_2,\mathcal{U}) \to (\mathcal{S}',\mathcal{U}')$ is a morphism of uniform spaces. Let $E' \in \mathcal{U}'$. Since

$$(\langle f_1, f_2\rangle)^{-1} = \{((x_1, x_2), (x_1', x_2')) \mid (x_i, x_i') \in (f_i)^{-1}(E), i = 1, 2\}$$

It suffices to show that $E_1 \in \mathcal{U}_1, E_2 \in \mathcal{U}_2$ implies that:

$$\{((x_1, x_2), (x_1', x_2')) \mid (x_i, x_i') \in E_i, i = 1, 2\} \in \mathcal{U}$$

Since $\mathcal{U}$ is closed under intersection, it suffices to show that if $E_1 \in \mathcal{U}_1$, then

$$E_1 \times \mathcal{S}_2\{((x_1, x_2), (x_1', x_2')) \mid (x_i, x_i') \in E_1\} \in \mathcal{U}$$

Similarly for E_2.

This is immediate since

$$\begin{aligned} E_1 \in \mathcal{U}_1 &\Rightarrow E_1 \supseteq K^{\epsilon}_{m_1} \\ &\Rightarrow E_1 \times \mathcal{S}_2 \supseteq K^{\epsilon}_{\langle m_1, m_2\rangle} \end{aligned}$$

∎

Lemma A.4 *$(\mathcal{PMU}, \leq)$ is a complete lattice.*

Proof. The least element is given by the discrete metric: $\perp(s,t) = 0$ if $s = t, 1$ otherwise. The top element has only one entourage $\mathcal{S}$ and is induced by the pseudometric given by $(\forall s, t)\top(s,t) = 0$. The greatest lower bounds of $\{\mathcal{U}_i\}$ is given by the $\cup_i \mathcal{U}_i$. ∎