Approximating Markov Processes by Averaging

Philippe Chaput[1,⋆], Vincent Danos[2], Prakash Panangaden[1,⋆], and Gordon Plotkin[2,⋆⋆]

[1] School of Computer Science; McGill University
[2] School of Informatics; University of Edinburgh

Abstract. We take a dual view of Markov processes – advocated by Kozen – as transformers of bounded measurable functions. We redevelop the theory of labelled Markov processes from this view point, in particular we explore approximation theory. We obtain three main results:
(i) It is possible to define bisimulation on general measure spaces and show that it is an equivalence relation. The logical characterization of bisimulation can be done straightforwardly and generally. (ii) A new and flexible approach to approximation based on averaging can be given. This vastly generalizes and streamlines the idea of using conditional expectations to compute approximation. (iii) It is possible to show that there is a minimal bisimulation equivalent to a process obtained as the limit of the finite approximants.

1 Introduction

Markov processes with continuous state spaces or continuous time evolution or both arise naturally in many areas of computer science: robotics, performance evaluation, modelling and simulation for example. For discrete systems there was a pioneering treatment of probabilisitic bisimulation and logical characterization [1]. The continuous case, however, was neglected for a time. For a little over a decade now, there has been significant activity among computer scientists [2–4] [5] [6–8] [9, 10] as it came to be realized that ideas from process algebra, like bisimulation and the existence of a modal characterization, would be useful for the study of continuous systems.

In [4] a theory of approximation for LMPs was initiated. Finding finite approximations is vital to give a computational handle on such systems. The previous work was characterized by rather intricate proofs that did not seem to follow from basic ideas in any straightforward way. For example, the logical characterization of (probabilistic) bisimulation requires subtle properties of analytic spaces.

In the present paper we take an entirely new approach, in some ways "dual" to the normal view of probabilistic transition systems. We think of a Markov process as a transformer of functions, rather than as a transformer of the state.

⋆ Research supported by NSERC
⋆⋆ Research supported by a Royal Society – Wolfson Research Merit Award.

Thus, instead of working directly with a Markov kernel $\tau(s, A)$ that takes a state s to a probability distribution over the state space, we think of a Markov process as transforming a function f into a new function $\int f(s')\tau(s, \mathrm{d}s')$ over the state space. This is the probabilistic analogue of working with predicate transformers, a point of view advocated by Kozen [11].

This new way of looking at things leads to three new results:

1. It is possible to define bisimulation on general spaces – not just on analytic spaces – and show that it is an equivalence relation with easy categorical constructions. The logical characterization of bisimulation can also be done generally, and with no complicated measure theoretic arguments.
2. A new and flexible approach to approximation based on averaging can be given. This vastly generalizes and streamlines the idea of using conditional expectations to compute approximation [5].
3. It is possible to show that there is a minimal bisimulation equivalent to a process obtained as the limit of the finite approximants.

2 Preliminary Definitions

Given a measurable space (X, Σ) with a measure μ we say two measurable functions are μ-equivalent if they differ on a set of μ-measure zero. Given two measurable real-valued functions f and g on X, we say $f \leq_\mu g$ if f is less than g except maybe on a set of measure zero. For $B \in \Sigma$, we let $\mathbf{1}_B$ be the indicator function of the set B. $L_1(X, \mu)$ stands for the space of *equivalence classes* of integrable functions. Similarly we write $L_1^+(X, \mu)$ for equivalence classes of functions that are positive μ-almost everywhere. $L_\infty(X, \mu)$ is the space of equivalence classes of μ-almost everywhere uniformly bounded functions on X, and $L_\infty^+(X, \mu)$ are the μ-almost everywhere positive functions of that space. Given two measures ν, μ on (X, Σ), if we have, for all $A \in \Sigma$, that $\mu(A) = 0 \Rightarrow \nu(A) = 0$, we say that ν is absolutely continuous with respect to μ, and write $\nu \ll \mu$.

Theorem 1. *[12] If $\nu \ll \mu$, where ν, μ are finite measures on (X, Σ) there is a positive measurable function h on X such that for every $B \in \Sigma$*

$$\nu(B) = \int_B h \, \mathrm{d}\mu \ .$$

The function h is defined uniquely, up to a set of μ-measure 0.

The function h is called the Radon-Nikodym derivative of ν with respect to μ; we write $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$ for the Radon-Nikodym derivative of the measure ν with respect to μ. Note that $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \in L_1(X, \mu)$.

Given a function $f \in L_1^+(X, \mu)$, we let $f \rhd \mu$ be the measure which has density f with respect to μ. According to the Radon-Nikodym theorem, given $\nu \ll \mu$, we have $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \rhd \mu = \nu$, and given $f \in L_1^+(X, \mu)$, $\frac{\mathrm{d}f \rhd \mu}{\mathrm{d}\mu} = f$ These two identities just say that the operations $- \rhd \mu$ and $\frac{\mathrm{d}}{\mathrm{d}\mu}$ are inverses of each other as operations from $L_1^+(X, \mu)$ to the space of finite measures on X.

Let **Prb** be the category of probability spaces and measurable maps; we will usually suppress the σ-algebra. There are no conditions relating to the measures but the categories of interest will be subcategories where the morphisms do have extra conditions related to the measures. Given a map $\alpha : (X, p) \longrightarrow (Y, q)$ in **Prb**, where p and q are probability measures, we denote by $M_\alpha(p)$ the image measure of p by α onto Y.

One normally works with vector spaces, but it is more convenient to work with cones. The following definition is due to Selinger [13].

Definition 1. *A* cone *is a set V on which a commutative and associative binary operation, written $+$, is defined and has a 0. Multiplication by positive real numbers is defined and it distributes over addition. The following cancellation law holds:*

$$\forall u, v, w \in V, v + u = w + u \Rightarrow v = w \ .$$

The following strictness property also holds: $v + w = 0 \Rightarrow v = w = 0$.

Cones come equipped with a natural partial order. If $u, v \in V$, a cone, one says $u \leq v$ if and only if there is an element $w \in V$ such that $u + w = v$. One can also put a norm on a cone, with the additional requirement that the norm be monotone with respect to the partial order.

Definition 2. *A* (ω-)complete normed cone *is a normed cone such that its unit ideal is a (ω-)dcpo.*

A (ω-)*continuous* linear map between two cones is one that preserves sups of (ω-)directed sets, i.e. is Scott-continuous. Note that in a (ω-)complete normed cone, the norm is (ω-)Scott-continuous. All the cones that we work with are complete normed cones. For instance, $L_\infty^+(X)$ is a complete normed cone, with the norm $\|-\|_\infty$ the usual essential supremum norm.

Let (X, Σ) be a measure space. We write $\mathcal{L}^+(X)$ for the cone of bounded measurable maps from X to $\mathbb{R}_+$; in this cone we have functions, not equivalence classes of functions. Let (X, Σ, μ) be a measure space. We define the cone $\mathcal{M}^{\leq K\mu}(X)$ to be the cone of all measures on (X, Σ) which are uniformly less than a multiple of the measure μ; this minimal multiple is the norm on this cone. The normed cones $\mathcal{M}^{\leq K\mu}(X)$ and $L_\infty^+(X, \Sigma, \mu)$ are isomorphic via the two maps $\frac{\mathrm{d}(-)}{\mathrm{d}\mu}$ and $(-) \rhd \mu$, which are also norm-preserving.

Markov processes can be viewed as linear maps on function spaces. Given τ a Markov process on X, we define $\hat{\tau} : \mathcal{L}^+(X) \longrightarrow \mathcal{L}^+(X)$, for $f \in \mathcal{L}^+(X)$, $x \in X$, as $\hat{\tau}(f)(x) = \int_X f(z)\tau(x, dz)$. This map is well-defined, as per our definition above, $\hat{\tau}(\mathbf{1}_B)$ is measurable for every $B \in \Sigma$. In fact, $\hat{\tau}(\mathbf{1}_B)(x) = \tau(x, B)$ is the probability of jumping from x to B. $\hat{\tau}$ is also linear and continuous and thus

$\hat{\tau}(f)$ is measurable for any measurable f. Conversely, any such functional L with $L(\mathbf{1}_X) \leq \mathbf{1}_X$ is a Markov process. *From now on, we shall only consider Markov processes from this functional point of view.*

3 Abstract Markov Processes and Conditional Expectation

In order to reduce the state space, we would like to project the space $\mathcal{L}^+(X)$ onto a smaller space. Let $\Lambda \subseteq \Sigma$ be a sub-σ-algebra, and let p be a finite measure on (X, Σ). We have a positive, linear and continuous map $\mathbb{E}_\Lambda : L_1^+(X, \Sigma, p) \longrightarrow L_1^+(X, \Lambda, p)$, the conditional expectation with respect to the sub-σ-algebra Λ. It can be restricted to L_∞^+, as it is a subcone of L_1^+. This map averages the function f over the sets of Λ. However, we cannot use the conditional expectation map in conjunction with Markov processes just yet, as Markov processes are defined as maps on $\mathcal{L}^+$, and not on any L_p^+ space. We therefore make the following definition:

Definition 3. *An* abstract Markov process *(AMP) on a probability space X is a ω-continuous linear map $\tau : L_\infty^+(X) \longrightarrow L_\infty^+(X)$ with $\tau(\mathbf{1}_X) \leq_p \mathbf{1}_X$.*

The condition that $\tau(\mathbf{1}_X) \leq_p \mathbf{1}_X$ is equivalent to requiring that the operator norm of τ be less than one, i.e. that $\|\tau(f)\|_\infty \leq \|f\|_\infty$ for all $f \in L_\infty^+(X)$. This is natural, as the function $\tau(\mathbf{1}_X)$, evaluated at a point x, is the probability of jumping from x to X, which is less than one.

AMPs are often called Markov operators in the literature, and have been first introduced in [14]. The novelty here is that our transition probabilities may be subprobabilities, as in the LMP literature, and we may then examine LMPs from this point of view.

It can be shown that a (usual) Markov process $\tau(x, B)$ on a probability space (X, Σ, p) can be expressed as an AMP if and only if for all $B \in \Sigma$ such that $p(B) = 0$, we have $\tau(x, B) = 0$, p-almost everywhere.

The simplest example of an AMP on a probability space (X, Σ, p) is the identity tranformation on $L_\infty^+(X)$, which sends any $f \in L_\infty^+(X)$ to itself. This AMP corresponds to the Markov process $\delta(x, B) = \mathbf{1}_B(x)$.

We now formalize the notion of conditional expectation. We work in a subcategory of **Prb**, called $\mathbf{Rad}_\infty$, where we require the image measure to be bounded by a multiple of the measure in the codomain; that is, measurable maps $\alpha : (X, \Sigma, p) \longrightarrow (Y, \Lambda, q)$ such that $M_\alpha(p) \leq Kq$ for some real number K.

Let us define an operator $\mathbb{E}_\alpha : L_\infty^+(X, p) \longrightarrow L_\infty^+(Y, q)$, as follows: $\mathbb{E}_\alpha(f) = \frac{\mathrm{d}M_\alpha(f \rhd p)}{\mathrm{d}q}$. As α is in $\mathbf{Rad}_\infty$, the Radon-Nikodym derivative is defined and is in $L_\infty^+(X, p)$. That is, the following diagram commutes by definition:

$$\begin{array}{ccc} L_\infty^+(X, p) & \xrightarrow[\rhd p]{} & \mathcal{M}^{\leq Kp}(X) \\ \Big\downarrow{\scriptstyle \mathbb{E}_\alpha} & & \Big\downarrow{\scriptstyle M_\alpha(-)} \\ L_\infty^+(Y, q) & \xleftarrow{\frac{\mathrm{d}}{\mathrm{d}q}} & \mathcal{M}^{\leq Kq}(Y) \end{array}$$

Note that if (X, Σ, p) is a probability space and $\Lambda \subseteq \Sigma$ is a sub-σ-algebra, then we have the obvious map $\lambda : (X, \Sigma, p) \longrightarrow (X, \Lambda, p)$ which is the identity on the underlying set X. This map is in $\mathbf{Rad}_\infty$ and it is easy to see that $\mathbb{E}_\lambda$ is precisely the conditional expectation onto Λ. Thus the operator $\mathbb{E}_-$ truly generalizes conditional expectation. It is easy to show that $\mathbb{E}_{\alpha\circ\beta} = \mathbb{E}_\alpha \circ \mathbb{E}_\beta$ and thus $\mathbb{E}_-$ is functorial.

Let us define, for any map $\alpha : (X, p) \longrightarrow (Y, q)$ in $\mathbf{Rad}_\infty$, a function $d(\alpha) = \mathbb{E}_\alpha(\mathbf{1}_X) = \frac{\mathrm{d}M_\alpha(p)}{\mathrm{d}q}$. Note that $d(\alpha)$ is in $L^+_\infty(Y, q)$. It can be shown that the operator norm of $\mathbb{E}_\alpha$ is $\|d(\alpha)\|_\infty$

Given an AMP on (X, p) and a map $\alpha : (X, p) \longrightarrow (Y, q)$ in $\mathbf{Rad}_\infty$, we thus have the following approximation scheme:

$$\begin{array}{ccc} L^+_\infty(Y,q) & \xrightarrow{\alpha(\tau)} & L^+_\infty(Y,q) \\ {\scriptstyle (-)\circ\alpha}\downarrow & & \uparrow{\scriptstyle \mathbb{E}_\alpha} \\ L^+_\infty(X,p) & \xrightarrow{\tau} & L^+_\infty(X,p) \end{array}$$

Note that $\|\alpha(\tau)\| \leq \|(-) \circ \alpha\| \cdot \|\tau\| \cdot \|\mathbb{E}_\alpha\| = \|\tau\| \cdot \|d(\alpha)\|_\infty$. Here the norm of $(\cdot) \circ \alpha$ is 1. As an AMP has a norm less than 1, we can only be sure that a map α yields an approximation for every AMP on X if $\|d(\alpha)\|_\infty \leq 1$. We call the AMP $\alpha(\tau)$ the projection of τ on Y.

4 Bisimulation

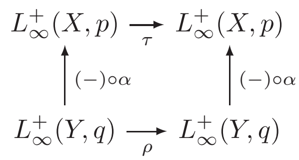
The notion of probabilistic bisimulation was introduced by Larsen and Skou [1] for discrete spaces and by Desharnais et al. [2] for continuous spaces. Subsequently a dual notion called event bisimulation or probabilistic co-congruence was defined independently by Danos et al. [9] and by Bartels et al. [15]. The idea of event bisimulation was that one should focus on the measurable sets rather than on the points. This meshes exactly with the view here.

Definition 4. *Given a (usual) Markov process (X, Σ, τ), an event-bisimulation is a sub-σ-algebra Λ of Σ such that (X, Λ, τ) is still a Markov process [9].*

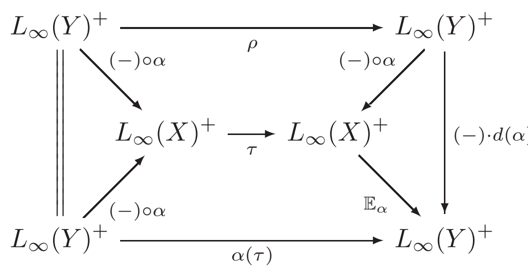
The only additional condition that needs to be respected for this to be true is that the Markov process $\tau(x, A)$ is Λ-measurable for a fixed $A \in \Lambda$. Translating this definition in terms of AMPs, this implies that the AMP τ sends the subspace $L^+_\infty(X, \Lambda, p)$ to itself, and so that the following commutes:

$$\begin{array}{ccc} L^+_\infty(X,\Sigma) & \xrightarrow[\tau]{} & L^+_\infty(X,\Sigma) \\ \uparrow & & \uparrow \\ L^+_\infty(X,\Lambda) & \xrightarrow[\tau]{} & L^+_\infty(X,\Lambda) \end{array}$$

A generalization to the above would be a $\mathbf{Rad}_\infty$ map α from (X, Σ, p) to (Y, Λ, q), respectively equipped with AMPs τ and ρ, such that the following commutes:



We will call such a map a *zigzag*. Note that if there is a zigzag from X to Y, then the AMP on Y is very closely related to the projection $\alpha(\tau)$ on Y. Indeed, we have the following diagram:



We have that $\mathbb{E}_\alpha(f \circ \alpha) = f \cdot d(\alpha)$ from a lemma in the full paper. This implies that $\alpha(\tau) = \rho \cdot d(\alpha)$. In particular, if $d(\alpha) = \mathbf{1}_Y$ - which happens if $M_\alpha(p) = q$ - then ρ is equal to $\alpha(\tau)$. Note that the condition $M_\alpha(p) = q$ means that the image measure is precisely the measure in the codomain of α.

We have developed the above theory in a very general setting where the maps between state spaces need only to respect some conditions; for instance, in $\mathbf{Rad}_\infty$, that the image measure be bounded by a multiple of the measure in the target space. From now on we shall considerably restrict the maps between the state spaces. Indeed, we have seen above that zigzags and projections coincided exactly given that the map of the state spaces was particularly well-behaved.

Definition 5. *A map* $\alpha : (X, p) \longrightarrow (Y, q)$ *in* **Prb** *is said to be* measure-preserving *if* $M_\alpha(p) = q$.

Clearly these maps form a subcategory of **Prb**. In effect, this ensures that the map α is essentially surjective. However, there is no reason why we would consider essentially surjective maps which are not surjective in the usual sense. We shall thus consider the subcategory of **Prb** consisting of the surjective measure-preserving maps. We will also augment this category with additional structure relevant to our situation.

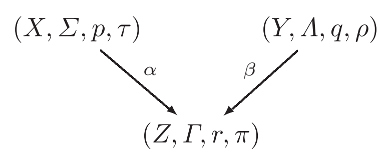
We define the category **AMP** of abstract Markov processes as follows. The objects consist of probability spaces (X, Σ, p), together with an abstract Markov process τ on X. The arrows $\alpha : (X, \Sigma, p, \tau) \longrightarrow (Y, \Lambda, q, \rho)$ are surjective measure-preserving maps from X to Y such that $\alpha(\tau) = \rho$. In words, this means that the Markov processes defined on the codomain are precisely the projection of the Markov processes τ on the domain through α. When working in this category, we will often denote objects by the state space, when the context is clear.

One can define a preorder on **AMP** as follows: given two AMPs (X, Σ, p, τ) and (Y, Λ, q, ρ), we say that $Y \preceq X$ if there is an arrow $\alpha : (X, \Sigma, p, \tau) \longrightarrow (Y, \Lambda, q, \rho)$ in **AMP**.

Definition 6. *We say that two objects of* **AMP***,* (X, Σ, p, τ) *and* (Y, Λ, q, ρ)*, are* bisimilar *if there is a third object* (Z, Γ, r, π) *with a pair of zigzags*

$$\alpha : (X, \Sigma, p, \tau) \longrightarrow (Z, \Gamma, r, \pi)$$
$$\beta : (Y, \Lambda, q, \rho) \longrightarrow (Z, \Gamma, r, \pi)$$

making a cospan diagram



Note that the identity function on an AMP is a zigzag, and thus that any zigzag between two AMPs X and Y implies that they are bisimilar.

The great advantage of cospans is that one needs pushouts to exist rather than pullbacks (or weak pullbacks); pushouts are much easier to construct. The following theorem shows that bisimulation is an equivalence.

Theorem 2. *Let* $\alpha : X \longrightarrow Y$ *and* $\beta : X \longrightarrow Z$ *be a span of zigzags. Then the pushout* W *exists and the pushout maps* $\delta : Y \longrightarrow W$ *and* $\gamma : Z \longrightarrow W$ *are zigzags.*

Corollary 1. *Bisimulation is an equivalence relation on the objects of* **AMP***.*

It turns out that there is a "smallest" bisimulation. Given an AMP (X, Σ, p, τ), one question one may ask is whether there is a "smallest" object $(\tilde{X}, \Xi, r, \xi)$ in **AMP** such that, for every zigzag from X to another AMP (Y, Λ, q, ρ), there is a zigzag from (Y, Λ, q, ρ) to $(\tilde{X}, \Xi, r, \xi)$. It can be shown that such an object exists.

Proposition 1. *Let* $\{\alpha_i : (X, \Sigma, p, \tau) \longrightarrow (Y_i, \Lambda_i, q_i, \rho_i)\}$ *be the set of all zigzags in* **AMP** *with domain* (X, Σ, p, τ)*. This yields a generalized pushout diagram, and as in Theorem 2, the pushout* $(\tilde{X}, \Xi, r, \xi)$ *exists and the pushout maps are zigzags.*

This object has important uniqueness properties.

Corollary 2. *Up to isomorphism, the object* $(\tilde{X}, \Xi, r, \xi)$ *the unique bottom element of* $\mathbf{ZZ}_X$*, the collection of all zigzags with* X *as domain. If* (W, Ω, q, ρ) *is another AMP such that there is a zigzag* μ *from* $\tilde{X}$ *to* W*, then* μ *is an isomorphism.*

Thus, we can say that $\tilde{X}$ is the meet (or infimum) of all objects Y_i which are bisimilar to the AMP X, with respect to the preorder $\preceq$. This "smallest" object is given in an abstract way; however, it can be constructed explicitly. Its construction is closely linked to a modal logic.

The logical characterization result from LMPs can easily be recast in the context of AMPs. We skip the proofs but give the basic definitions. Let us fix a finite set of labels $\mathcal{A}$ once and for all. We can then speak of objects in a category $\mathbf{AMP}_\mathcal{A}$ of *labelled* AMPs, consisting of a probability space (x, Σ, p) and a set of AMPs τ_a indexed by $\mathcal{A}$.

Definition 7. *We define a logic* $\mathcal{L}$ *as follows, with* $a \in \mathcal{A}$*:*

$$\mathcal{L} ::= \mathbf{T} | \phi \wedge \psi | \langle a \rangle_q \psi$$

Given a labelled AMP (X, Σ, p, τ_a)*, we associate to each formula* ϕ *a measurable set* $[\![\phi]\!]$*, defined recursively as follows:*

$$[\![\mathbf{T}]\!] = X \qquad [\![\phi \wedge \psi]\!] = [\![\phi]\!] \cap [\![\psi]\!]$$
$$\left[\!\left[\langle a \rangle_q \psi\right]\!\right] = \left\{ s : \tau_a(\mathbf{1}_{[\![\psi]\!]})(s) > q \right\}$$

We let $[\![\mathcal{L}]\!]$ *denote the measurable sets obtained by all formulas of* $\mathcal{L}$*.*

Theorem 3. (From [9]) *Given a labelled AMP* (X, Σ, p, τ_a)*, the* σ*-field* $\sigma([\![\mathcal{L}]\!])$ *generated by the logic* $\mathcal{L}$ *is the smallest event-bisimulation on* X*. That is, the map* $i : (X, \Sigma, p, \tau_a) \longrightarrow (X, \sigma([\![\mathcal{L}]\!]), p, \tau_a)$ *is a zigzag; furthermore, given any zigzag* $\alpha : (X, \Sigma, p, \tau_a) \longrightarrow (Y, \Lambda, q, \rho_a)$*, we have that* $\sigma([\![\mathcal{L}]\!]) \subseteq \alpha^{-1}(\Lambda)$*.*

Hence, the σ-field obtained on X by the "smallest object" $\tilde{X}$ is precisely the σ-field we obtain from the logic.

5 Approximations of AMPs

Given an arbitrary AMP, it may be very difficult to study its behavior if its state space is very large or uncountable. It is therefore crucial to devise a way to reduce the state space to a manageable size.

In this section, we let the measurable map $i_\Lambda : (X, \Sigma) \longrightarrow (X, \Lambda)$ be the identity on the set X, restricting the σ-field. The resulting AMP morphism is denoted as $i_\Lambda : (X, \Sigma, p, \tau) \longrightarrow (X, \Lambda, p, \Lambda(\tau))$, as p is just restricted on a smaller σ-field, with $\Lambda(\tau)$ being the projection of τ on the smaller σ-field Λ.

Let (X, Σ, p, τ_a) be a labelled AMP. Let $\mathcal{P}$ be a finite set of rationals in $[0,1]$; we will call it a *rational partition*. We define a family of finite π-systems [12], subsets of Σ, as follows:

$$\begin{aligned}
&\Phi_{\mathcal{P},0} = \{X, \emptyset\} \\
&\Phi_{\mathcal{P},n} = \pi\left(\left\{\tau_a(\mathbf{1}_A)^{-1}(q_i, 1] : q_i \in \mathcal{P}, A \in \Phi_{\mathcal{P},n-1}, a \in \mathcal{A}\right\} \cup \Phi_{\mathcal{P},n-1}\right)
\end{aligned}$$

where $\pi(\Omega)$ is the π-system generated by the class of sets Ω.

For each pair $(\mathcal{P}, M)$ consisting of a rational partition and a natural number, we define a σ-algebra $\Lambda_{\mathcal{P},M}$ on X as $\Lambda_{\mathcal{P},M} = \sigma(\Phi_{\mathcal{P},M})$, the σ-algebra generated by $\Phi_{\mathcal{P},M}$. We shall call each pair $(\mathcal{P}, M)$ consisting of a rational partition and a natural number an *approximation pair*. These σ-algebras have a very important property:

Proposition 2. *Given any labelled AMP (X, Σ, p, τ_a), the σ-field $\sigma\left(\bigcup \Lambda_{\mathcal{P},M}\right)$, where the union is taken over all approximation pairs, is precisely the σ-field $\sigma\,[\![\mathcal{L}]\!]$ obtained from the logic.*

Consider the σ-algebra $\Lambda_{\mathcal{P},M}$. We have the map

$$i_{\Lambda_{\mathcal{P},M}} : (X, \Sigma, p, \tau_a) \longrightarrow (X, \Lambda_{\mathcal{P},M}, p, \Lambda_{\mathcal{P},M}(\tau_a)) \ .$$

Now since $\Lambda_{\mathcal{P},M}$ is finite, it is atomic, and so it partitions our state space X, yielding an equivalence relation. Quotienting by this equivalence relation gives a map $\pi_{\mathcal{P},M} : (X, \Lambda_{\mathcal{P},M}, p, \Lambda_{\mathcal{P},M}(\tau_a)) \longrightarrow (\hat{X}_{\mathcal{P},M}, \Omega, q, \rho_a)$, where $\hat{X}_{\mathcal{P},M}$ is the (finite!) set of atoms of $\Lambda_{\mathcal{P},M}$ and Ω is just the powerset of $\hat{X}_{\mathcal{P},M}$. The measure q is just the image measure and AMPs ρ_a are the projections $\pi_{\mathcal{P},M}(\tau_a)$. Note that $\pi_{\mathcal{P},M}$ is a zigzag as ${\pi_{\mathcal{P},M}}^{-1}(\Omega) = \Lambda_{\mathcal{P},M}$.

As the σ-field on $\hat{X}_{\mathcal{P},M}$ is its powerset, we will refrain from writing Ω when involving a finite approximation. We thus have an approximation map $\phi_{\mathcal{P},M} = \pi_{\mathcal{P},M} \circ i_{\Lambda_{\mathcal{P},M}}$ from our original state space to a finite state space; furthermore it is clear that this map is an arrow in **AMP**.

Let us define an ordering on the approximation pairs by $(\mathcal{P}, M) \leq (\mathcal{Q}, N)$ if $\mathcal{Q}$ refines $\mathcal{P}$ and $M \leq N$. This order is natural as $(\mathcal{P}, M) \leq (\mathcal{Q}, N)$ implies $\Lambda_{\mathcal{P},M} \subseteq \Lambda_{\mathcal{Q},N}$, which is clear from the definition. Thus, this poset is a directed set: given $(\mathcal{P}, M)$ and $(\mathcal{Q}, N)$ two approximation pairs, then the approximation pair $(\mathcal{P} \cup \mathcal{Q}, \max(M, N))$ is an upper bound.

Given two approximation pairs such that $(\mathcal{P}, M) \leq (\mathcal{Q}, N)$, we have a map $i_{(\mathcal{Q},N),(\mathcal{P},M)} : (X, \Lambda_{\mathcal{Q},N}, \Lambda_{\mathcal{Q},N}(\tau_a)) \longrightarrow (X, \Lambda_{\mathcal{P},M}, \Lambda_{\mathcal{P},M}(\tau_a))$ which is well defined by the inclusion $\Lambda_{\mathcal{P},M} \subseteq \Lambda_{\mathcal{Q},N} \subseteq \Sigma$. We therefore have a projective system of such maps indexed by our poset of approximation pairs. It can be shown that these maps induce a map on the finite approximation spaces $\hat{X}_{\mathcal{P},M}$, say $j_{(\mathcal{Q},N),(\mathcal{P},M)} : (\hat{X}_{\mathcal{Q},N}, \phi_{\mathcal{Q},N}(\tau_a)) \longrightarrow (\hat{X}_{\mathcal{P},M}, \phi_{\mathcal{P},M}(\tau_a))$, such that the map $\phi_{(\mathcal{P},M)}$ factors through the map $\phi_{(\mathcal{Q},N)}$ as $\phi_{(\mathcal{P},M)} = j_{(\mathcal{Q},N),(\mathcal{P},M)} \circ \phi_{(\mathcal{Q},N)}$. Hence, the maps $j_{(\mathcal{Q},N),(\mathcal{P},M)}$ together with the approximants $\hat{X}_{(\mathcal{P},M)}$ also form a projective system with respect to our poset of approximation pairs.

A result of Choksi [16] allows us to construct projective limits of measure spaces. We consider the underlying probability spaces of the finite approximants of a labelled AMP (X, Σ, p, τ_a).

Proposition 3. *(From [16]) The probability spaces of finite approximants* $\hat{X}_{\mathcal{P},M}$ *of an AMP* (X, Σ, p, τ_a)*, indexed by the approximation pairs, form a projective system of surjective measure-preserving maps; furthermore, its projective limit* $(\operatorname{proj\,lim} \hat{X}, \Gamma, \gamma)$ *exists in* **Prb**

Concretely, $\operatorname{proj\,lim} \hat{X}$ is the projective limit in **Set**. Thus we have the usual projection maps, appropriately restricted, $\psi_{\mathcal{P},M} : \operatorname{proj\,lim} \hat{X} \longrightarrow \hat{X}_{\mathcal{P},M}$ for every approximation pair. We also have, in **Set**, a unique map $\kappa : X \longrightarrow \operatorname{proj\,lim} \hat{X}$ such that $\psi_{\mathcal{P},M} \circ \kappa = \phi_{\mathcal{P},M}$.

The σ-field Γ is the smallest σ-field making all of the maps $\psi_{\mathcal{P},M}$ measurable. We must show that $\kappa^{-1}(\Gamma) \subseteq \Sigma$. We shall show something stronger.

Proposition 4. *The* σ*-field* $\kappa^{-1}(\Gamma)$ *is precisely equal to* $\sigma [\![\mathcal{L}]\!]$*; in particular* κ *is measurable.*

Proof. The σ-field Γ is generated by the preimages of $\psi_{\mathcal{P},M}$. Taking the preimage of this through κ is equivalent to taking the preimage through the approximation maps $\phi_{\mathcal{P},M}$, which is exactly $\Lambda_{\mathcal{P},M}$. These σ-fields generate $\sigma [\![\mathcal{L}]\!]$.

We now need to show that κ is measure-preserving. γ was defined so that the maps $\psi_{\mathcal{P},M}$ were measure preserving [16]; thus γ and $M_\kappa(p)$ agree on all subsets of $\operatorname{proj\,lim} \hat{X}$ which are the preimage of a measurable set in a finite approximant $\hat{X}_{\mathcal{P},M}$. Since these sets generate Γ, and form a π-system, the uniqueness of measure theorem [12] implies that $\gamma = M_\kappa(p)$.

Finally, we define the AMP ζ_a on $\operatorname{proj\,lim} \hat{X}$ in the obvious way; that is, as the projection of τ_a through κ. Then the projection of ζ_a onto the finite approximants through $\psi_{\mathcal{P},M}$ is precisely equal to ρ_a as they were previously defined, since $\psi_{\mathcal{P},M} \circ \kappa = \phi_{\mathcal{P},M}$. Thus, the projective limit of measure spaces can be extended to a projective limit of AMPs.

Proposition 5. *The universal map* κ *obtained from the projective limit is a zigzag.*

Therefore, if we let $(\tilde{X}, \Xi, r, \xi_a)$ be the smallest bisimulation obtained as in proposition 1, we have a zigzag $\omega : (\operatorname{proj\,lim} \hat{X}, \Gamma, \gamma, \zeta_a) \longrightarrow (\tilde{X}, \Xi, r, \xi_a)$. This zigzag must be an isomorphism of σ-fields as Ξ is the smallest possible σ-field on $\tilde{X}$. We can show that there is a zigzag going in the other direction.

Proposition 6. *Let* $\alpha : (X, \Sigma, p, \tau_a) \longrightarrow (Y, \Theta, q, \rho_a)$ *be a zigzag. Then these two AMPs have the same finite approximants. In particular, two bisimilar AMPs have the same finite approximants.*

We conclude with the main result.

Theorem 4. *Given a labelled AMP* (X, Σ, p, τ_a)*, the projective limit of its finite approximants* $(\text{proj}\lim \hat{X}, \Gamma, \gamma, \zeta_a)$ *is isomorphic to its smallest bisimulation* $(\tilde{X}, \Xi, r, \xi_a)$*.*

6 Related Work and Conclusions

The main contribution of the present work is to show how one can obtain a powerful and general notion of approximation of Markov processes using the dualized view of Markov processes as transformers of random variables (measurable functions). We view Markov processes as "predicate transformers". Our main result is to show that this way of working with Markov processes greatly simplifies the theory: bisimulation, logical characterization and approximation. Working with the functions (properties) one is less troubled by having to deal with things that are defined only "almost everywhere" as happens when one works with states.

A very nice feature of the theory is the ability to show that a minimal bisimulation exists. Furthermore, this minimal object can be constructed as the projective limit of finite approximants.

Previous work on bisimulation-based approximation of Markov processes began with a paper by Desharnais et al. [4] where the approximation scheme was based on an unfolding of the transition system. The main technical result is that every formula satisfied by a process is satisfied by one of its finite approximants.

In [5] the idea of approximating by averaging was introduced and the main tool used to compute the approximation is the conditional expectation. The mathematical theory developed there is the bare beginnings of the theory developed here. There also the idea of averaging by conditional approximation was used; but none of the results relating to bisimulation and especially the result about constructing a minimal bisimulation by taking a projective limit of finite approximants was known. Moving to AMPs was crucial for all this to work.

One of the problems with any of the approximation schemes is that they are hard to implement. In a recent paper [7], an approach based on Monte Carlo approximation was used to "approximate the approximation." The point is that it is hard to compute τ^{-1} in practice. Our most pressing future work is to explore the possibility of implementing the approximation scheme and, perhaps using some technique like Monte Carlo, to compute the approximations concretely. It is curious that the abstract version of Markov processes makes it more likely that one can compute approximations in practice and is another argument in favour of a "pointless" view of processes.

References

1. Larsen, K.G., Skou, A.: Bisimulation through probablistic testing. Information and Computation **94** (1991) 1–28
2. Desharnais, J., Edalat, A., Panangaden, P.: Bisimulation for labeled Markov processes. Information and Computation **179**(2) (2002) 163–193
3. de Vink, E., Rutten, J.J.M.M.: Bisimulation for probabilistic transition systems: A coalgebraic approach. Theoretical Computer Science **221**(1/2) (1999) 271–293
4. Desharnais, J., Gupta, V., Jagadeesan, R., Panangaden, P.: Approximating labeled Markov processes. Information and Computation **184**(1) (2003) 160–200
5. Danos, V., Desharnais, J., Panangaden, P.: Conditional expectation and the approximation of labelled Markov processes. In Amadio, R., Lugiez, D., eds.: CONCUR 2003 - Concurrency Theory. Volume 2761 of Lecture Notes In Computer Science., Springer-Verlag (2003) 477–491
6. Ferns, N., Panangaden, P., Precup, D.: Metrics for Markov decision processes with infinite state spaces. In: Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence. (2005) 201–208
7. Bouchard-Côté, A., Ferns, N., Panangaden, P., Precup, D.: An approximation algorithm for labelled Markov processes: towards realistic approximation. In: Proceedings of the 2nd International Conference on the Quantitative Evaluation of Systems (QEST). (2005) 54–61
8. Cattani, S., Segala, R., Kwiatkowska, M., Norman, G.: Stochastic transition systems for continuous state spaces and non-determinism. In: Proceedings of the 8th International Conference on Foundations of Software Science and Computational Structures (FOSSACS). Number 3441 in Lecture Notes In Computer Science (2005) 125–139
9. Danos, V., Desharnais, J., Laviolette, F., Panangaden, P.: Bisimulation and cocongruence for probabilistic systems. Information and Computation **204**(4) (2006) 503 – 523 Seventh Workshop on Coalgebraic Methods in Computer Science 2004.
10. Goubault-Larrecq, J.: Continuous capacities on continuous state spaces. In: Proceedings of the 34th International Colloquium on Automata, Languages and Programming (ICALP'07). Volume 4596 of Lecture Notes In Computer Science., Springer-Verlag (2007) 764–776
11. Kozen, D.: A probabilistic PDL. Journal of Computer and Systems Sciences **30**(2) (1985) 162–178
12. Billingsley, P.: Probability and Measure. Wiley-Interscience (1995)
13. Selinger, P.: Towards a quantum programming language. Mathematical Structures in Computer Science **14**(4) (2004) 527–586
14. Hopf, E.: The general temporally discrete Markoff process. J. Rational Math. Mech. Anal. **3** (1954) 13–45
15. Bartels, F., Sokolova, A., de Vink, E.: A hierarchy of probabilistic system types. Theoretical Computer Science **327** (2004) 3–22
16. Choksi, J.: Inverse limits on measure spaces. Proc. London Math. Soc **8**(3) (1958) 321–342