

Approximating Labelled Markov Processes Again!

Philippe Chaptut^{1,*}, Vincent Danos², Prakash Panangaden^{1,*},
and Gordon Plotkin^{2,**}

¹ School of Computer Science, McGill University

² School of Informatics, University of Edinburgh

Abstract. Labelled Markov processes are continuous-state fully probabilistic labelled transition systems. They can be seen as co-algebras of a suitable monad on the category of measurable space. The theory as developed so far included a treatment of bisimulation, logical characterization of bisimulation, weak bisimulation, metrics, universal domains for LMPs and approximations. Much of the theory involved delicate properties of analytic spaces.

Recently a new kind of averaging procedure was used to construct approximations. Remarkably, this version of the theory uses a dual view of LMPs and greatly simplifies the theory eliminating the need to consider analytic spaces. In this talk I will survey some of the ideas that led to this work.

1 Introduction

The study of continuous systems is becoming a more common part of computer science. Control theory, or systems theory has worked with an abstract notion of a system; see, for example, the text by Sontag [1]. These notions were general enough to include the continuous physical systems normally studied by control engineers as well as systems such as automata which were inherently discrete in nature. General concepts such as “state”, “transformation” and “observation” are used in both the discrete and the continuous settings. Labelled Markov processes (LMPs) [2,3] were invented as a formalism to study *interacting probabilistic systems* that may be either discrete or continuous.

The basic idea of LMPs is to consider probabilistic labelled transition systems with continuous state spaces. There are states and actions or labels. A state and an action determines a probability distribution over the next states. There is no other nondeterminism as seen, for example, in probabilistic automata[4]. Nor is there any attempt to assign probabilities to the actions; the nondeterminism is captured by the arbitrariness in the choice of actions. This model is very close to the notion of Markov decision process (MDP) [5] used in optimization theory and machine learning. The only difference is that we do not have a concept of

* Research supported by NSERC.

** Research supported by a Royal Society – Wolfson Research Merit Award.

“reward” or “payoff.” The discrete version of this model is due to Larsen and Skou [6] and is called by them the *reactive model*.

The systems that we consider - continuous state space and discrete time - are of interest for two reasons. First, this is a reasonable middle ground before we get to a completely continuous model. Several authors have looked at discrete space, continuous time models. It is clear that fully continuous models will force us to make a significant deepening of the mathematics. One will have to consider stochastic differential equations rather than transition systems to describe the evolution of systems: we are not there yet.

Second, these systems occur in natural examples. In the past one of us was involved with a case study from avionics where the discrete-time-continuous-space character of the system was clearly in evidence. Many control systems have this character. It is possible that this may even be fruitful for biology.

Very similar notions were introduced at around the same time by de Vink and Rutten[7]. The main difference is that they worked with ultrametric spaces; one can argue that these spaces are essentially discrete. For example, they are totally disconnected and are nothing like the spaces that arise in physical systems. Their important contribution was to emphasize the co-algebraic point of view – noted in passing in [3] – and show that bisimulation could be presented as a span of co-algebra homomorphisms. The same idea, span of “zigzags”, is used by Desharnais et al. [3]. In the present work we emphasize the importance of co-spans rather than spans.

The notion of bisimulation is central to the study of concurrent systems. While there is a variety of different equivalence relations between processes (two-way simulation, trace equivalence, failures equivalence and many more), bisimulation enjoys some fundamental mathematical properties, most notably its characterization as a fixed-point, which make it the most discussed process equivalence.

One might take the view that any automated analysis or logical reasoning must be inherently discrete in character. In particular, even if one is interested in reasoning about a physical system, one has to first discretize the system. In fact, this point of view actually provides a good argument for retaining the continuous view of the system. A given system may well be described in continuous terms. Without formalizing the continuous system - and having a notion of equivalence between discrete and continuous systems - how does one argue that the discretized system is a faithful model of the underlying continuous system? Even suppose one is willing to treat a discrete model as given, what if one needs to refine the model? For example, a given discretization may arise from some type of approximation based on a given *tolerance*; how does one refine the tolerance or discretize *adaptively*? Clearly the underlying continuous model has to be retained and used if we want to construct different discrete approximations.

In brief, a labelled Markov process can be described as follows. There is a set of states and a set of labels. The system is in a state at a point in time and moves between states. The state which it moves to is governed by which interaction with the environment is taking place and this is indicated by the labels. The

system evolves according to a probabilistic law. If the system interacts with the environment by synchronizing on a label it makes a transition to a new state governed by a transition probability distribution. So far, this is essentially the model developed by Larsen and Skou [6]. They specify the transitions by giving, for each label, a probability for going from one state to another. Bisimulation then amounts to matching the moves; this means that both the labels and the probabilities must be the same.

In the case of a continuous state space, however, one cannot simply specify transition probabilities from one state to another. In most interesting systems all such transition probabilities would be zero! Instead one must work with probability densities. In so doing, one has to confront the major issues that arose when probability theory was first formalized, such as the existence of subsets for which the notion of probability does not make sense. In the present case, we have to introduce a notion of set for which “probabilities make sense” (i.e. a σ -field) and instead of talking about probabilities of going from a state s to another state s' , we have to talk about going from a state s to a *set of* states A .

The notion of bisimulation for these systems is a mild modification of the definition of Larsen and Skou. One has to add a measure theoretic condition. In the logical characterization it was shown that two states are bisimilar if and only if they satisfy all the same formulas of a modal logic similar to Hennessy-Milner logic. The striking aspect of this result is that the logic is completely negation free. Even for purely discrete systems this result was new and quite unexpected. It shows that fully probabilistic systems are very close to determinate systems. The nature of the proof is quite different from proofs of other Hennessy-Milner type results.

In recent work 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 defined on it rather than as a transformer of the state. 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 $\int f(s')\tau(s, ds')$ over the state space. This is the probabilistic analogue of working with predicate transformers, a point of view advocated by Kozen [8] in a path-breaking early paper on probabilistic systems and logic.

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 [9].
3. It is possible to show that there is a minimal bisimulation equivalent to a process obtained as the limit of finite approximants.

2 Labelled Markov Processes

Labelled Markov processes are probabilistic versions of labelled transition systems. Corresponding to each label a Markov process is defined. The transition probability is given by a *stochastic kernel* (Feller’s terminology [10]) also commonly called a *Markov kernel*. Thus the indeterminacy has two sources: the “choice” made by the environment - no probabilities are attributed to this at all - and the probabilistic transitions made by the process. This is the “reactive” model of Larsen and Skou [6] who used it in a discrete state-space setting.

A key ingredient in the theory is the stochastic kernel or Markov kernel. We will call it a *transition probability function*.

Definition 1. A **transition (sub-)probability function** on a measurable space (S, Σ) is a function $\tau : S \times \Sigma \rightarrow [0, 1]$ such that for each fixed $s \in S$, the set function $\tau(s, \cdot)$ is a (sub-)probability measure, and for each fixed $X \in \Sigma$ the function $\tau(\cdot, X)$ is a measurable function.

One interprets $\tau(s, X)$ as the probability of the process starting in state s making a transition into one of the states in X . The transition probability is really a *conditional probability*; it gives the probability of the process being in one of the states of the set X after the transition, *given* that it was in the state s before the transition. In general the transition probabilities could depend on time, in the sense that the transition probability could be different at every step, but still independent of past history; we always consider the time-independent case.

We will work with *sub-probability* functions; i.e. with functions where $\tau(s, S) \leq 1$ rather than $\tau(s, S) = 1$. The mathematical results go through in this extended case. We view processes where the transition functions are only sub-probabilities as being *partially defined*, opening the way for a notion of approximation.

The stochastic systems studied in the literature are usually only the very special version where $\tau(s, S)$ is either 1 or 0. We call such processes *total* and the general processes are called *partial*. We capture the idea that an action is rejected by setting $\tau(s, S)$ to be 0.

Definition 2. A **labelled Markov process (LMP)** \mathcal{S} with label set \mathcal{A} is a structure $(S, i, \Sigma, \{\tau_a \mid a \in \mathcal{A}\})$, where S is the set of states, i is the initial state, and Σ is the Borel σ -field on S , and

$$\forall a \in \mathcal{A}, \tau_a : S \times \Sigma \longrightarrow [0, 1]$$

is a transition sub-probability function.

We will fix the label set to be \mathcal{A} once and for all. We will write (S, i, Σ, τ) for labelled Markov processes, instead of the more precise $(S, i, \Sigma, \{\tau_a \mid a \in \mathcal{A}\})$.

We give a simple example to illustrate the ideas.

Example 1. (From [11]) Consider a process with two labels $\{a, b\}$. The state space is upper right quadrant of the real plane, \mathbb{R}^2 together with a single extra

point. In order to describe this system conveniently we will pretend, at first, that the state space is the entire real plane. When the process makes an a -move from state (x_0, y_0) , it jumps to (x, y_0) , where the probability distribution for x is given by the density $K_\alpha \exp(-\alpha(x - x_0)^2)$, where $K_\alpha = \sqrt{\alpha/\pi}$ is the normalizing factor. When it makes a b -move it jumps from state (x_0, y_0) to (x_0, y) , where the distribution of y is given by the density function $K_\beta \exp(-\beta(y - y_0)^2)$. The meaning of these densities is as follows. The probability of jumping from (x_0, y_0) to a state with x -coordinate in the interval $[s, t]$ under an a -move is $\int_s^t K_\alpha \exp(-\alpha(x - x_0)^2) dx$. All points with $x < 0$ or $y < 0$ are identified as a single absorbing state. Once it is in this state no more transitions are possible. Note that the probability of jumping to any given point is, of course, 0. In this process the interaction with the environment controls whether the jump is along the x -axis or along the y -axis but the actual extent of the jump is governed by a probability distribution. If there were just a single label we would have an ordinary (time-independent) Markov process; in fact it would be a brownian motion with absorbing walls.

3 Bisimulation and Logic

The fundamental process equivalence that we consider is *strong probabilistic bisimulation*. Probabilistic bisimulation means matching the moves and probabilities *exactly* – thus each system must be able to make the same transitions with the same probabilities as the other. Larsen and Skou define a bisimulation relation R as an equivalence relation on the states satisfying the condition that, for each label a , equivalent states have equal probability of making an a -transition to any R -equivalence class of states. In the continuous case, we demand that equivalent states have equal probability of making an a -transition to any union of equivalence classes of states *provided that the union is measurable*.

Instead of talking about sets of equivalence classes we will instead use the notion of R -closed sets. Let R be a binary relation on a set S . We say a set $X \subseteq S$ is R -closed if $R(X) := \{t \mid \exists s \in X, sRt\}$ is a subset of X . If R is reflexive, this becomes $R(X) = X$. If R is an equivalence relation, a set is R -closed if and only if it is a union of equivalence classes.

Definition 3. Let $\mathcal{S} = (S, i, \Sigma, \tau)$ be a labelled Markov process. An equivalence relation R on S is a **bisimulation** if whenever sRs' , with $s, s' \in S$, we have that for all $a \in \mathcal{A}$ and every R -closed measurable set $X \in \Sigma$, $\tau_a(s, X) = \tau_a(s', X)$. Two states are bisimilar if they are related by a bisimulation relation.

Alternately, bisimulation on the states of a labelled Markov process can be viewed as the maximum fixed point of the following (monotone) functional F on the lattice of equivalence relations on $(S \times S, \subseteq)$:

$$s F(R) t \text{ if for all } a \in \mathcal{A}, \text{ and all } R\text{-closed } C \in \Sigma, \tau_a(s, C) \leq \tau_a(t, C).$$

It is easy to prove that bisimulation is an equivalence relation.

The intuition of this definition is that the relation R relates those states that can be “lumped” together. Bisimulation is the largest such relation. In fact the notion of bisimulation was known in the queuing theory community [12] under the name of “lumpability”.

One can define a simple modal logic and prove that two states are bisimilar if and only if they satisfy exactly the same formulas. As before we assume that there is a fixed set of “actions” \mathcal{A} . The logic is called \mathcal{L} and has the following syntax:

$$\top \mid \phi_1 \wedge \phi_2 \mid \langle a \rangle_q \phi$$

where a is an action and q is a rational number. This is the basic logic with which we establish the logical characterization.

Given a labelled Markov process $\mathcal{S} = (S, i, \Sigma, \tau)$ we write $s \models \phi$ to mean that the state s satisfies the formula ϕ . The definition of the relation \models is given by induction on formulas. The definition is obvious for the propositional constant \top and conjunction. We say $s \models \langle a \rangle_q \phi$ if and only if $\exists X \in \Sigma. (\forall s' \in X. s' \models \phi) \wedge (\tau_a(s, X) > q)$. In other words, the process in state s can make an a -move to a state, that satisfies ϕ , with probability strictly greater than q . We write $\llbracket \phi \rrbracket_{\mathcal{S}}$ for the set $\{s \in S \mid s \models \phi\}$, and $\mathcal{S} \models \phi$ if $i \models \phi$. We often omit the subscript when no confusion can arise.

The logic that Larsen and Skou used in [6] has more constructs including disjunction and some negative constructs. They show that for systems satisfying a “minimum deviation condition” – a uniform bound on the degree of branching everywhere – two *states* of the same process are bisimilar if and only if they satisfy the same formulas of their logic.

The main theorem relating \mathcal{L} and bisimulation is the following:

Theorem 4. *Let (S, i, Σ, τ) be a labelled Markov process. Two states $s, s' \in S$ are bisimilar if and only if they satisfy the same formulas of \mathcal{L} .*

It does not have any kind of minimal deviation condition, nor any negative constructs in the logic. In [13] a logical characterization of simulation was also proved.

4 Approximation by Unfolding

In this section I review the “old” view of approximation. The key idea is the construction of some approximants via an “unfolding” construction. As the approximation is refined there are more and more transitions possible. There are two parameters to the approximation, one is a natural number n , and the other is a positive rational ϵ . The number n gives the number of successive transitions possible from the start state. The number ϵ measures the accuracy with which the probabilities approximate the transition probabilities of the original process.

Given a labelled Markov process $\mathcal{S} = (S, i, \Sigma, \tau)$, an integer n and a rational number $\epsilon > 0$, we define $\mathcal{S}(n, \epsilon)$ to be an n -step unfolding approximation of \mathcal{S} .

Its state-space is divided into $n + 1$ levels which are numbered $0, 1, \dots, n$. Bisimulation is a fixed point and that one has - for each n - a level n approximation to bisimulation. At each level, say n , the states of the approximant is a partition of S ; these partitions correspond to the equivalence classes corresponding to the level n approximation to bisimulation. The initial state of $\mathcal{S}(n, \epsilon)$ is at level n and transitions only occur between a state of one level to a state of one lower level. Thus, in particular, states of level 0 have no outgoing transitions. In the following we omit the curly brackets around singletons.

Let (S, i, Σ, τ) be a labelled Markov process, $n \in \mathbf{N}$ and ϵ a positive rational. We denote the finite-state approximation by $\mathcal{S}(n, \epsilon) = (P, p_0, \rho)$ where P is a subset of $\Sigma \times \{0, \dots, n\}$. It is defined as follows, for $n \in \mathbf{N}$ and $\epsilon > 0$. $\mathcal{S}(n, \epsilon)$ has $n + 1$ levels. States are defined by induction on their level. Level 0 has one state $(S, 0)$. Now, given the sets from level l , we define states of level $l + 1$ as follows. Suppose that there are m states at level l , we partition the interval $[0, 1]$ into intervals of size ϵ/m . Let $(B_j)_{j \in I}$ stand for this partition; i.e. for $\{\{0\}, (0, \epsilon/m], (\epsilon/m, 2\epsilon/m], \dots\}$. States of level $l + 1$ are obtained by the partition of S that is generated by the sets $\tau_a(\cdot, C)^{-1}(B_j)$, for every set C corresponding to state at level l and every label $a \in \{a_1, \dots, a_n\}$, $i \in I$. Thus, if a set X is in this partition of S , $(X, l + 1)$ is a state of level $l + 1$. Transitions can happen from a state of level $l + 1$ to a state of level l , and the transition probability function is given by

$$\rho_a((X, k), (B, l)) = \begin{cases} \inf_{t \in X} \tau_a(t, B) & \text{if } k = l + 1, \\ 0 & \text{otherwise.} \end{cases}$$

The initial state p_0 of $\mathcal{S}(n, \epsilon)$ is the unique state (X, n) such that X contains i , the initial state of \mathcal{S} .

The next theorem is the main result.

Theorem 5. *If a state $s \in S$ satisfies a formula $\phi \in \mathcal{L}_\vee$, then there is some approximation $\mathcal{S}(n, \epsilon)$ such that $(X_s, n) \models \phi$.*

In the original approximation paper [13], a universal domain of LMPs was defined and a simulation order was given. It was shown that the approximants form a directed set and that any LMP could be given as a sup of the approximants. The approximants gave a countable base for the domain which was thus a continuous domain. It is a final co-algebra for a suitable category of LMPs. Independently, van Breugel and Worrell showed that one could define a final co-algebra for LMPs using other techniques.

In a later paper, Desharnais et al. [14] developed a metric and showed that the approximations converge in that metric. Thus this seems to strengthen the idea that this is a robust notion of convergence. The trouble is that the approximants are from below. In other words, if we cluster some of the states to form an approximate state say B , to estimate the transition probability from B to A we take the inf of $\tau(x, A)$ as x ranges over B . It is far more natural to average the states in B . This idea was developed in [9]. The construction that we develop next absorbs this idea but in a completely new way.

5 The Dual View

Definition 6. An abstract Markov process (AMP) on a probability space X is a ω -continuous linear map $\tau : L_\infty^+(X) \rightarrow L_\infty^+(X)$ 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 [15].

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

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

$$\begin{array}{ccc} L_\infty^+(X, p) & \xrightarrow{\quad} & \mathcal{M}^{\leq Kp}(X) \\ & \searrow \triangleright p & \\ & \downarrow \mathbb{E}_\alpha & \downarrow M_\alpha(-) \\ L_\infty^+(Y, q) & \xleftarrow{\frac{d}{dq}} & \mathcal{M}^{\leq Kq}(Y) \end{array}$$

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 $\lambda : (X, \Sigma, p) \rightarrow (X, \Lambda, p)$ which is the identity on the underlying set X . This map is in **Rad** $_\infty$ and it is easy to see that \mathbb{E}_λ 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.

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

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

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 .

6 Bisimulation

The notion of probabilistic bisimulation was introduced by Larsen and Skou [6] for discrete spaces and by Desharnais et al. [3] for continuous spaces.

Subsequently a dual notion called event bisimulation or probabilistic co-congruence was defined independently by Danos et al. [16] and by Bartels et al. [17]. 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 7. *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 [16].*

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 **Rad** $_\infty$ map α from (X, Σ, p) to (Y, Λ, q) , respectively equipped with AMPs τ and ρ , such that the following commutes:

$$\begin{array}{ccc} L_\infty^+(X, p) & \xrightarrow{\tau} & L_\infty^+(X, p) \\ \uparrow (-) \circ \alpha & & \uparrow (-) \circ \alpha \\ L_\infty^+(Y, q) & \xrightarrow{\rho} & L_\infty^+(Y, q) \end{array}$$

We will call such a map a *zigzag*.

Definition 8. *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*

$$\begin{aligned} \alpha &: (X, \Sigma, p, \tau) \rightarrow (Z, \Gamma, r, \pi) \\ \beta &: (Y, \Lambda, q, \rho) \rightarrow (Z, \Gamma, r, \pi) \end{aligned}$$

making a cospan diagram

$$\begin{array}{ccc} (X, \Sigma, p, \tau) & & (Y, \Lambda, q, \rho) \\ & \searrow \alpha & \swarrow \beta \\ & (Z, \Gamma, r, \pi) & \end{array}$$

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

Corollary 10. *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}, Ξ, r, ξ) 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}, Ξ, r, ξ) . It can be shown that such an object exists.

Proposition 11. *Let $\{\alpha_i : (X, \Sigma, p, \tau) \rightarrow (Y_i, \Lambda_i, q_i, \rho_i)\}$ be the set of all zigzags in AMP with domain (X, Σ, p, τ) . This yields a generalized pushout diagram, and as in Theorem 9, the pushout (\tilde{X}, Ξ, r, ξ) exists and the pushout maps are zigzags.*

A version of the logical characterization theorem follows immediately from this.

7 Approximation Revisited

In this section, we let the measurable map $i_\Lambda : (X, \Sigma) \rightarrow (X, \Lambda)$ be the identity on the set X , restricting the σ -field. The resulting AMP morphism is denoted as $i_\Lambda : (X, \Sigma, p, \tau) \rightarrow (X, \Lambda, p, \Lambda(\tau))$, 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 [18], subsets of Σ , as follows:

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

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

In fact one can use a theorem of Choksi’s from 1958 [19] to construct projective limits of finite approximants and establish the following fundamental theorem.

Theorem 13. *Given a labelled AMP (X, Σ, p, τ_a) , the projective limit of its finite approximants $(\text{proj lim } \tilde{X}, \Gamma, \gamma, \zeta_a)$ is isomorphic to its smallest bisimulation $(\tilde{X}, \Xi, r, \xi_a)$.*

8 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.

References

1. Sontag, E.: *Mathematical Control Theory*. Texts in Applied Mathematics, vol. 6. Springer, Heidelberg (1990)
2. Blute, R., Desharnais, J., Edalat, A., Panangaden, P.: Bisimulation for labelled Markov processes. In: *Proceedings of the Twelfth IEEE Symposium On Logic In Computer Science*, Warsaw, Poland (1997)
3. Desharnais, J., Edalat, A., Panangaden, P.: Bisimulation for labeled Markov processes. *Information and Computation* 179(2), 163–193 (2002)
4. Segala, R., Lynch, N.: Probabilistic simulations for probabilistic processes. In: Jonsson, B., Parrow, J. (eds.) *CONCUR 1994*. LNCS, vol. 836, pp. 481–496. Springer, Heidelberg (1994)
5. Puterman, M.L.: *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley, Chichester (1994)
6. Larsen, K.G., Skou, A.: Bisimulation through probabilistic testing. *Information and Computation* 94, 1–28 (1991)
7. de Vink, E., Rutten, J.J.M.M.: Bisimulation for probabilistic transition systems: A coalgebraic approach. In: *Proceedings of the 24th International Colloquium On Automata Languages And Programming* (1997)
8. Kozen, D.: A probabilistic PDL. *Journal of Computer and Systems Sciences* 30(2), 162–178 (1985)
9. Danos, V., Desharnais, J., Panangaden, P.: Conditional expectation and the approximation of labelled markov processes. In: Amadio, R.M., Lugiez, D. (eds.) *CONCUR 2003*. LNCS, vol. 2761, pp. 477–491. Springer, Heidelberg (2003)
10. Feller, W.: *An Introduction to Probability Theory and its Applications II*, 2nd edn. John Wiley and Sons, Chichester (1971)
11. Desharnais, J., Gupta, V., Jagadeesan, R., Panangaden, P.: The metric analogue of weak bisimulation for labelled Markov processes. In: *Proceedings of the Seventeenth Annual IEEE Symposium On Logic In Computer Science*, pp. 413–422 (2002)
12. Kemeny, J.G., Snell, J.L.: *Finite Markov Chains*. Van Nostrand (1960)

13. Desharnais, J., Gupta, V., Jagadeesan, R., Panangaden, P.: Approximating labeled Markov processes. *Information and Computation* 184(1), 160–200 (2003)
14. Desharnais, J., Gupta, V., Jagadeesan, R., Panangaden, P.: A metric for labelled Markov processes. *Theoretical Computer Science* 318(3), 323–354 (2004)
15. Hopf, E.: The general temporally discrete Markoff process. *J. Rational Math. Mech. Anal.* 3, 13–45 (1954)
16. Danos, V., Desharnais, J., Laviolette, F., Panangaden, P.: Bisimulation and cocongruence for probabilistic systems. *Information and Computation* 204(4), 503–523 (2006)
17. Bartels, F., Sokolova, A., de Vink, E.: A hierarchy of probabilistic system types. *Theoretical Computer Science* 327, 3–22 (2004)
18. Billingsley, P.: *Probability and Measure*. Wiley Interscience, Hoboken (1995)
19. Choksi, J.: Inverse limits on measure spaces. *Proc. London Math. Soc.* 8(3), 321–342 (1958)