Approximating Markov Processes By Averaging

PHILIPPE CHAPUT
McGill University
VINCENT DANOS
University of Edinburgh
PRAKASH PANANGADEN
McGill University
GORDON PLOTKIN
University of Edinburgh

October 14, 2013

Abstract

Normally, one thinks of probabilistic transition systems as taking an initial probability distribution over the state space into a new probability distribution representing the system after a transition. We, however, take a dual view of Markov processes as transformers of bounded measurable functions. This is very much in the same spirit as a "predicate-transformer" view, which is dual to the state-transformer view of transition systems.

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 approximations. (iii) We show that there is a minimal process bisimulation-equivalent to a given process, and this minimal process is 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 probabilistic bisimulation and logical characterization by Larsen and Skou [LS91]. The continuous case, however, was neglected for a time. For a little over a decade there has been significant activity among computer scientists [DEP02, dVR99, DGJP00] [MOPW04, vBW01a, DDP03] [FPP05, BCFPP05, CSKN05] [DDLP06, GL07a, Dob03] 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 such systems. In [BDEP97] continuous-state Markov processes with labels to capture interactions were christened labelled Markov processes (LMPs). Some of this material has appeared in book form [Pan09, Dob10]. There is a vast literature on timed systems, hybrid systems, robotics and control theory that also refer to systems with continuous state spaces.

A labelled Markov process is a discrete time dynamical system combining nondeterministic and probabilistic behavior. The intuitive picture is the following. The system evolves within a state space X. A user can control this system via a set of actions $\mathcal{A}$, assumed to be finite. To each action is associated a probabilistic transition within the system. The system undergoes these transitions when the user chooses the corresponding action. For each action, the transitions are Markov and time homogeneous, and thus only depend on the current state of the system. The user has full control over which action to choose; the nondeterminism of the system stems from the user interaction.

However, there is a crucial difference in the way such systems are interpreted in comparison to usual stochastic processes or dynamical systems. Typically, the current position in the state space is what one keeps track of; in our case, we are concerned with the interaction between the user and the actions. Indeed, at each point in the state space, the actions may have a nonzero probability of being disabled, and the user knows when the action he chose was disabled. Furthermore, this information about actions is the *only* information the user can obtain from the system, as the system's state is internal and not visible to the user.

In [DGJP00] and [DGJP03] a theory of approximation for LMPs was initiated and was refined and extended in [DD03] and [DDP03]. Finding finite approximations is vital to give a computational handle on such systems. These techniques were adapted to Markov decision processes (MDPs) and applied to find good estimates of value functions [FPP05]. 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 proved first in [DEP98] requires subtle properties of analytic spaces and rather awkward and ad-hoc seeming constructions [Eda99][1]. Proofs of basic results in approximation theory also seemed to be more difficult than they should be.

In the present paper we take an entirely new approach: we consider Markov processes as transformers of measurable functions on the state space rather than as transformers of probability distributions on the state space. This is in some ways "dual" to the normal view of probabilistic transition systems. It is akin to the relationship between predicate-transformer semantics and state-transformer semantics. However, both styles of semantics can be accommodated in our viewpoint; it is purely because the theory is slightly smoother in the predicate-transformer view that we develop that viewpoint in the paper. 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 [Koz85] 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 [DDP03].

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

[1] Later these results were put together in a much more systematic way by [Dob10] using the machinery of descriptive set theory.

There is a key mathematical fact that allows these results to be established and it hinges on duality. In the usual theory of L_p spaces in functional analysis one defines the space $L_p(X, \mu)$ as the space of functions[2] whose absolute values raised to the pth power are integrable with respect to μ. Now if $1 < p, q < \infty$ the space of continuous linear functionals on L_p is isomorphic to L_q if $\frac{1}{p} + \frac{1}{q} = 1$; the spaces L_p and L_q are duals; for example L_2 is self-dual. However, for L_1 and L_∞ one does not have a duality. In the present paper we consider cones rather than vector spaces. One can think of cones as subsets of vector spaces consisting of the "positive" vectors; of course, this needs to be axiomatized properly. When one has such a cone, say C, the vector space V can be viewed as having a partial order defined on it by the simple device of saying $u \leq v$ if $v - u \in C$. One can now use order-theoretic continuity to strengthen the requirements on the spaces and obtain a perfect duality between the L_1 and the L_∞ spaces. In fact, we will axiomatize cones *ab initio* rather than viewing them as subsets of vector spaces; this will allow us to work with the space of all positive measures as a cone rather than artificially embedding it into some vector space. The ability to switch between these dual views is very useful and allows easy proofs of many facts.

A second main innovation in the present paper is a functorial view of the conditional expectation. Some of the key properties of conditional expectation turn out to be nothing more than functoriality. This facilitates the view of conditional expectation as a coarsening of the description of the system and hence makes it a key step in the approximation process. It also provides a unified view of bisimulation and approximation.

The rest of the paper is organized as follows. In Section 2 we review some of the background needed to read the paper. In Section 3 we describe categories of cones and develop duality theory for these categories. In Section 4 we define conditional expectation functorially. In Section 5 we define labelelled abstract Markov processes (LAMPs) and we define the notion of approximation of LAMPs in Section 6. In Section 7 we define bisimulation and we show that it is an equivalence relation in Section 8. In Section 9 we obtain the minimal realization of a LAMP from which the logical characterization follows. In Section 11 we develop the theory of approximation and show that the limit of the finite approximants gives the minimal realization of a process. In Section 12 we review the history of LMPs and review other

[2] We are only considering real-valued functions, in functional analysis one usually considers complex-valued functions.

related work.

2 Background

In this section we review some of the mathematical background needed for this paper. We need some basic measure theory and functional analysis.

2.1 Measure theory

We assume that the reader is familiar with the definitions of σ-algebras, measurable spaces (set equipped with a σ-algebra), measures, measurable functions, integration and basic concepts from topology [Bil95, Dud89, KT66, Rud66, Wil91]. By a *finite* measure we mean a measure that assigns a finite value as the measure of the whole space on which it is defined. We recall the definition of measurable function to avoid a common confusion.

Definition 2.1 *A function f from a measurable space (X, Σ) to a measurable space (Y, Λ) is said to be* ***measurable*** *if $f^{-1}(B) \in \Sigma$ whenever $B \in \Lambda$.*

Note this is *not* the definition in [Hal74], but is the one used by most modern authors. Halmos's definition has the annoying property that the composite of two measurable functions need not be measurable; a price he is willing to pay in order to integrate a few more functions.

We define the category **Mes** where the objects are measurable spaces and the morphisms are measurable functions. There is an obvious forgetful functor into **Set** which preserves limits.

Definition 2.2 *A* ***probability triple*** *(X, Σ, p) is a measurable space with a measure p with $p(X) = 1$; such a measure is called a probability measure.*

We also use the term *subprobability* measure on (X, Σ) to mean a finite measure q with $q(X) \leq 1$. Given a measurable space (X, Σ) we write $\mathcal{M}(X)$ for the space of finite measures on X. We will always work with finite measures, usually – but not always – probability or subprobability measures.

We say a real-valued measurable function f on a space (X, Σ) equipped with a measure μ is *integrable* if the integral $\int f \mathrm{d}\mu$ is finite. Since we are working with finite measures, positive *bounded* measurable functions are always integrable.

Given (X, Σ, p) and (Y, Λ) and a measurable function $f : X \longrightarrow Y$ we obtain a measure q on Y by $q(B) = p(f^{-1}(B))$. This is written $M_f(p)$ and is called the *image measure* of p under f. We say that a map $f : (X, \Sigma, p) \longrightarrow (Y, \Lambda, q)$ is *measure preserving* if $M_f(p) = q$.

In measure theory it is more convenient to work with equivalence classes of functions that are equal "almost everywhere." 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. $L_1(X, \mu)$ stands for the space of *equivalence classes* of integrable functions. Similarly we write $L_1^+(X, \mu)$ for equivalence classes of integrable functions that are positive μ-almost everywhere. We will often write just $L_1(X)$ if the μ is clear from context and similarly for the variations that crop up. The space L_1 is a real vector space but the space $L_1^+(X)$ is not; it is a *cone*, a concept to be defined below.

We need a bit more standard measure theory for the approximation results. A π-system is a family of sets closed under finite intersection. The following proposition appears as Theorem 10.3 in [Bil95].

Proposition 2.3 *If two measures agree on a π-system they agree on the σ-algebra generated by the π-system.*

2.2 The Radon-Nikodym theorem

Given a measurable function $\alpha : (X, \Sigma, p) \longrightarrow (Y, \Lambda, q)$ recall that we denote by $M_\alpha(p)$ the image measure of p by α onto Y.

The Radon-Nikodym theorem [Rud66] is a central result in measure theory allowing one to define a "derivative" of a measure with respect to another measure.

Definition 2.4 *We say that a measure ν is **absolutely continuous** with respect to another measure μ if for any measurable set A, $\mu(A) = 0$ implies that $\nu(A) = 0$. We write $\nu \ll \mu$.*

Theorem 2.5 *If $\nu \ll \mu$, where ν, μ are finite measures on a measurable space (X, Σ) there is a positive measurable function h on X such that for every measurable set B*

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

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 denote it by $\frac{\mathrm{d}\nu}{\mathrm{d}\mu}$. Since ν is finite, $\frac{\mathrm{d}\nu}{\mathrm{d}\mu} \in L_1^+(X, \mu)$.

The Radon-Nikodym theorem applies to a more general class of measures called σ-finite measures: these are measures where the total space can be written as the countable union of sets of finite measure. In this more general case it will not be true in general that the Radon-Nikodym derivative is in L_1^+.

Given an (almost-everywhere) positive function[3] $f \in L_1(X, p)$, we let $f \cdot p$ be the measure which has density f with respect to p. Two identities that we get from the Radon-Nikodym theorem are:

- given $q \ll p$, we have $\frac{\mathrm{d}q}{\mathrm{d}p} \cdot p = q$.
- given $f \in L_1^+(X, p)$, $\frac{\mathrm{d}f \cdot p}{\mathrm{d}p} = f$

These two identities just say that the operations $(-) \cdot p$ and $\frac{\mathrm{d}(-)}{\mathrm{d}p}$ are inverses of each other as maps between $L_1^+(X, p)$ and $\mathcal{M}^{\ll p}(X)$ the space of finite measures on X that are absolutely continuous with respect to p.

2.3 Conditional expectation

A *random variable* on a measurable space is just a measurable function. We will use the language of measurable functions rather than random variables, because our emphasis is more measure theoretic than probabilistic. In the probability literature everything is usually stated in terms of random variables.

The expectation $\mathbb{E}_p(f)$ of a measurable function f is the average computed by $\int f \mathrm{d}p$ and therefore it is just a number. The *conditional* expectation is not a mere number but a random variable. It is meant to measure the expected value in the presence of additional information.

The additional information takes the form of a sub-σ algebra, say Λ, of Σ. In what way does this represent "additional information"? The idea is that an experimenter is trying to compute probabilities of various outcomes of a random process. The process is described by (X, Σ, p). However she may

[3] Of course, one should really say "equivalence class of functions" but it is common to abuse the terminology in this fashion.

only have partial information in advance, by knowing that the outcome is in a measurable set Q. Now she may try to recompute her expectation values based on this information. To know that the outcome is in Q also means that it is *not* in Q^c. Note that $\{\emptyset, Q, Q^c, X\}$ is in fact a (tiny) sub-σ-algebra of Σ. Thus one can generalize this idea and say that for some given sub-σ-algebra Λ of Σ she knows for every $Q \in \Lambda$ whether the outcome is in Q or not. Now she can recompute the expectation values given this information. The point of requiring Λ-measurability is that it "smooths out" variations that are too rapid to show up in Λ.

It is an immediate consequence of the Radon-Nikodym theorem that such conditional expectations exist.

Theorem 2.6 (Kolmogorov) *Let (X, Σ, p) be a measure space with p a finite measure, f be in $L_1(X, \Sigma, p)$ and Λ be a sub-σ-algebra of Σ, then there exists a $g \in L_1(X, \Lambda, p)$ such that for all $B \in \Lambda$*

$$\int_B f \mathrm{d}p = \int_B g \mathrm{d}p.$$

This function g is usually denoted by $\mathbb{E}(f|\Lambda)$.

We clearly have $f \cdot p \ll p$ so the required g is simply $\frac{\mathrm{d}f \cdot p}{\mathrm{d}p|_\Lambda}$, where $p\ |_\Lambda$ is the restriction of p to the sub-σ-algebra Λ. The conditional expectation is *linear*, *increasing* with respect to the point wise order and possesses other pleasing properties to be described below. It is defined uniquely p-almost everywhere.

2.4 Markov kernels

We begin with some preliminary definitions. Let (X, Σ) and (Y, Λ) be measurable spaces. We define a stochastic transition from X to Y:

Definition 2.7 *A **Markov kernel** from X to Y is a map*

$$\tau : X \times \Lambda \longrightarrow [0, 1]$$

such that:

- *for all $x \in X$, $\tau(x, \cdot)$ is a subprobability measure on Y*
- *for all $B \in \Lambda$, $\tau(\cdot, B)$ is a measurable function*

The interpretation of such functions is that $\tau(x, B)$ is the probability of jumping from the point x to the set B. Thus, if $(X, \Sigma) = (Y, \Lambda)$, the Markov kernel may be iterated to determine the evolution of a discrete-time and time-homogeneous Markov process where the state is a point in X; we will call such a Markov kernel a *Markov kernel on X*. Note that this definition is slightly different from the usual definition of a Markov process on a measurable space, as we allow our transition probabilities to be *subprobabilities*. One may interpret this difference as follows: given a point x with $\tau(x, Y) = k \leq 1$, the process τ has a probability $1 - k$ to be disabled at the point x.

We now give the definition of a labelled Markov process, first given in this form in [BDEP97].

Definition 2.8 *A **labelled Markov process** (LMP) on a measurable space (X, Σ) is a collection of Markov kernels τ_a on X, indexed by a finite or countable set $\mathcal{A}$, called the set of **actions**.*

Note that the set of labels $\mathcal{A}$ will be fixed once and for all.

2.5 Cones

Cones are a way of combining order structure with linear structure. The idea is that a subset of a vector space is designated as the set of "positive" vectors. A cone, viewed as a subset of a vector space, will need to satisfy some natural closure properties. Then we can define $u \leq v$ for two vectors u and v by saying that $v - u$ is positive. We can, however, define cones intrinsically without reference to an ambient vector space. This is sometimes important particularly in speaking of probability distributions where subtraction is not always defined. Cones are well known in the functional analysis literature; however, we base the definition and discussion of cones below, on a paper by [Sel04] which we found particularly apt for our purposes, partly because it introduces cones abstractly rather than as subsets of vector spaces. We discuss related concepts of cones in the related work section.

Definition 2.9 *A **cone** is a set V on which a commutative and associative binary operation, written $+$, is defined and on which multiplication by positive real numbers is defined. There is a distinguished element $0 \in V$, which is an identity for the $+$ operation; in short, $(V, +, 0)$ forms a commutative monoid. Multiplication by reals distributes over addition and 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 2.10 *A **normed cone** C is a cone with a function $||\cdot|| : C \longrightarrow \mathbb{R}_+$ satisfying the usual conditions:*

1. $||v|| = 0$ *if and only if* $v = 0$
2. $\forall r \in \mathbb{R}_+, v \in C, ||r \cdot v|| = r||v||$
3. $||u + v|| \leq ||u|| + ||v||$
4. $u \leq v \Rightarrow ||u|| \leq ||v||$.

The only slight difference from the usual definition of norm is the requirement that r be positive. Owing to the lack of a subtraction operation, it is not possible to speak of a sequence being Cauchy in the usual sense; however, order-theoretic concepts can be used instead.

Definition 2.11 *An ω-**complete normed cone** is a normed cone such that*

1. *if $\{a_i \mid i \in I\}$ is an increasing sequence with $\{||a_i||\}$ bounded then the lub $\bigvee_{i \in I} a_i$ exists and*
2. $\bigvee_{i \in I} ||a_i|| = ||\bigvee_{i \in I} a_i||$.

The norm gives a notion of convergence as does the notion of lub of a chain. The following lemma from [Sel04] relates the two.

Lemma 2.12 *Suppose that u_i is a countable chain with a least upper bound in an ω-complete normed cone and u is an upper bound of the u_i. Suppose furthermore that $\lim_{i \longrightarrow \infty} \|u - u_i\| = 0$. Then $u = \bigvee_i u_i$.*

A linear map of cones is precisely what one would expect: i.e. a map that preserves the linear operations. Note than any such map is monotone.

Definition 2.13 *An ω-**continuous** linear map between two cones is one that preserves least upper bounds of countable chains. More precisely if C and D are cones and $f : C \longrightarrow D$ is linear we say that it is ω-continuous if for every countable chain $\{a_i\}$ in C such that $\bigvee_i a_i$ exists then so does $\bigvee_i f(a_i)$ and $f(\bigvee_i a_i) = \bigvee_i f(a_i)$.*

We will also want to restrict our attention to *bounded* linear maps of normed cones. A *bounded* linear map of normed cones $f : C \longrightarrow D$ is one such that for all u in C, $||f(u)|| \leq K||u||$ for some real number K. A lemma in [Sel04] shows that any linear map of ω-complete normed cones is bounded; it is thus superfluous to mention boundedness when discussing a map of ω-complete normed cones. The norm of a bounded linear map $f : C \longrightarrow D$ is defined as $||f|| = \sup\{||f(u)|| : u \in C, ||u|| \leq 1\}$; this is analogous to the operator norm for bounded linear maps between vector spaces.

We need the concept of dual cone; indeed it is one of the central concepts of the present work. Given an ω-complete normed cone C, its dual C^* is the set of all ω-continuous linear maps from C to $\mathbb{R}_+$. We define the norm on C^* to be the operator norm. It is not hard to show that this cone is a ω-complete normed cone as well, and that the cone order corresponds to the point wise order. For the latter one needs to show that if g is less than f point wise then $f - g$ is also an ω-continuous map. If $\{x_i\}$ is an increasing sequence in C with sup x we need to show that $\sup\{(f-g)(x_i)\} = (f-g)(\sup\{x_i\})$. This follows from the fact that $\sup\{(f-g)(x_i)\} = \sup\{f(x_i) - g(x_i)\} = \sup\{f(x_i)\} - \sup\{g(x_i)\} = f(\sup\{x_i\}) - g(\sup\{x_i\})$, where the last equality follows from the continuity of f and g and the one before that is an elementary "ϵ argument."

The ω-complete normed cones, along with ω-continuous linear maps, form a category which we shall denote $\omega\mathbf{CC}$. If we define the subcategory $\omega\mathbf{CC}_1$ of $\omega\mathbf{CC}$ as the one where the norms of the maps are all bounded by 1 then isomorphisms in this category are always isometries. It is easy to see that given any linear map F between normed spaces, if F^{-1} exists and has bounded norm then $||F| \cdot ||F^{-1}|| \geq 1$. Thus if we are working in $\omega\mathbf{CC}_1$ this condition implies that both F and F^{-1} have norm 1. Many of the cones of interest and the maps between them live in $\omega\mathbf{CC}_1$.

In $\omega\mathbf{CC}$, the dual operation becomes a contravariant functor; if $f : C \longrightarrow D$ is a map of cones, we define $f^* : D^* \longrightarrow C^*$ as follows. Given a map L in D^*, we define a map f^*L in C^* as $f^*L(u) = L(f(u))$. Now $||L(f(u))|| \leq ||L|| \cdot ||f|| \cdot ||u||$ and thus $||f^*|| \leq ||f||$.

Note that this dual is stronger than the dual in usual Banach spaces, where we only require the maps to be bounded. This has nice consequences with respect to the cones we are considering. For instance, we shall see that the dual to $L_\infty^+(X,\Sigma,\mu)$ (to be defined below) is isomorphic to $L_1^+(X,\Sigma,\mu)$, which is not the case with the Banach space $L_\infty(X,\Sigma,p)$.

Next, we introduce the cones that we use in the present work. They are all ω-complete normed cones.

3 Cones of measures and of measurable functions

Let (X,Σ) be a measure space. We write $\mathcal{L}^+(X,\Sigma)$ for the cone of bounded measurable maps from X to $\mathbb{R}_+$. This is an ω-complete normed cone as the supremum of countably many measurable functions is measurable. Closely related to this is the cone $\mathcal{M}(X,\Sigma)$ of finite measures on (X,Σ). The ordering on this cone is the cone order as defined in the previous section. Explicitly, $\mu \leq \nu$ if there is a finite measure λ such that $\nu = \mu + \lambda$; note this is not the same as the pointwise order. The cone order implies the pointwise order but the reverse may not be the case. The norm of a measure μ is just $\mu(X)$.

Proposition 3.1 $\mathcal{M}(X,\Sigma)$ *is an ω-complete normed cone.*

Proof . Checking the norm axioms is routine. Suppose that $\mu_{i+1} = \mu_i + \theta_i$ for all i. We can define $\theta^{(k)} := \sum_{i=k+1}^{\infty} \theta_i$. It is straightforward to verify that all the $\theta^{(k)}$ are finite measures and that for all k, $\mu = \mu_K + \theta^{(k)}$ so μ is an upper bound in the cone order and since the cone order implies the pointwise order, it is the least upper bound in the cone order. ∎

We will usually just write $\mathcal{L}(X)$ and $\mathcal{M}(X)$. The real action occurs in subcones of these cones.

If μ is a measure on X, then one has the well-known Banach spaces L_1 and L_∞ mentioned above. These can be restricted to cones by considering the μ-almost everywhere positive functions. We will denote these cones by $L_1^+(X,\Sigma,\mu)$ and $L_\infty^+(X,\Sigma)$; if the context is clear we will drop the Σ and often the measure as well. These also are complete normed cones.

We also work with cones of measures on a space. Let (X,Σ,p) be a measure space with finite measure p. We denote by $\mathcal{M}^{\ll p}(X)$, the cone of all measures on (X, Σ, p) that are absolutely continuous with respect to p[4]. If q is such a measure, we define its norm to be $q(X)$. It is easy to see that this norm coincides precisely with the norm on $L_1^+(X, \Sigma, p)$ if q is viewed as a density function through the Radon-Nikodym theorem. Hence $\mathcal{M}^{\ll p}(X)$ is also an ω-complete normed cone. In fact, one can say more; it is easy to show that the maps $\frac{\mathrm{d}(-)}{\mathrm{d}p} : \mathcal{M}^{\ll p}(X) \to L_1^+(X, \Sigma, p)$ and $(-) \cdot p : L_1^+(X, \Sigma, p) \to \mathcal{M}^{\ll p}(X)$ are both ω-continuous maps of cones which are furthermore norm-preserving. Thus the cones $\mathcal{M}^{\ll p}(X)$ and $L_1^+(X, \Sigma, p)$ are isometrically isomorphic in $\omega\mathbf{CC}$.

Similarly, one can consider $\mathcal{M}^p_{\mathsf{UB}}(X)$, the cone of all measures on (X, Σ) that are uniformly less than a multiple of the measure p; in other words, $q \in \mathcal{M}^p_{\mathsf{UB}}$ means that for some real constant $K > 0$ we have $q \leq Kp$. For such a measure q, we can define the norm of q to be the infimum of all constants K such that $q \leq Kp$, which coincides with the norm on $L_\infty^+(X, \Sigma, p)$ when q is considered as a density function; thus $\mathcal{M}^p_{\mathsf{UB}}(X)$ is an ω-complete normed cone. As with $\mathcal{M}^{\ll p}(X)$, the cones $\mathcal{M}^p_{\mathsf{UB}}(X)$ and $L_\infty^+(X, \Sigma, p)$ are isomorphic. The two maps $\frac{\mathrm{d}(-)}{\mathrm{d}p}$ and $(-) \cdot p$ also are norm-preserving.

Proposition 3.2 *The dual of the cone $L_\infty^+(X, \Sigma, p)$ is isometrically isomorphic to $\mathcal{M}^{\ll p}(X)$.*

Proof . Let L be an element of $L_\infty^{+,*}(X)$. We define a measure q on X as follows:

$$q(B) = L\left(\mathbf{1}_B\right)$$

The countable additivity of q is a direct consequence of the ω-continuity of L: given a countable collection of disjoint measurable sets B_i, we have that

$$\mathbf{1}_{\cup_{i=1}^n B_i} = \sum_{i=1}^n \mathbf{1}_{B_i}$$

Clearly the functions $\mathbf{1}_{\cup_{i=1}^n B_i}$ form an increasing sequence, and are bounded by $\mathbf{1}_X$ because the B_is are disjoint. We can write $q\left(\bigcup_{i=1}^\infty B_i\right)$ as $L\left(\sup_n \sum_{i=1}^n \mathbf{1}_{B_i}\right)$. Since $\mathbf{1}_X$ has finite norm in $L_\infty^+(X)$, we have

$$L\left(\sup_n \sum_{i=1}^n \mathbf{1}_{B_i}\right) = \sup_n L\left(\sum_{i=1}^n \mathbf{1}_{B_i}\right) = \sup_n \sum_{i=1}^n L\left(\mathbf{1}_{B_i}\right) = \sum_{i=1}^\infty L\left(\mathbf{1}_{B_i}\right).$$

[4]Since a cone has to be closed under multiplication by positive reals this cone cannot consist of just probability measures; we have to consider general finite measures

This shows countable additivity of q. Furthermore, $q(\emptyset) = L(0) = 0$, and thus q is a measure.

We want to show that the operator norm of L is $q(X)$. We have that

$$\|L\| = \sup_{\|f\|_\infty \leq 1} L(f) = L(\mathbf{1}_X) = q(X)$$

since L is monotone and $\mathbf{1}_X$ is the least upper bound of the unit ball of $L_\infty^+(X)$.

Finally, if $p(B) = 0$, we have that $\mathbf{1}_B = 0$ in $L_\infty^+(X)$, and thus q is absolutely continuous with respect to p.

Thus, each element of $L_\infty^{+,*}(X)$ can be associated with a measure in $\mathcal{M}^{\ll p}(X)$ via a map, which we call ϕ, such that, in the above discussion, we have $\phi(L) = q$.

It is easy to check that ϕ is linear and ω-continuous. Furthermore, we just showed that it was norm-preserving. On the other hand, it is clear that every element q of $\mathcal{M}^{\ll p}(X)$ corresponds to an unique element of $L_\infty^{+,*}(X)$. If u is the Radon-Nikodym derivative of q, we have the functional $f \mapsto \int_X f u \,\mathrm{d}p$ on $L_\infty^+(X)$ which is bounded by Hölder's inequality. Thus ϕ is an isometric isomorphism. ∎

Since $\mathcal{M}^{\ll p}(X)$ is isometrically isomorphic to $L_1^+(X)$, an immediate corollary is that $L_\infty^{+,*}(X)$ is isometrically isomorphic to $L_1^+(X)$, which is of course false in general in the context of Banach spaces.

The following proposition is proved analogously:

Proposition 3.3 *The dual of the cone $L_1^+(X, \Sigma, p)$ is isometrically isomorphic to $\mathcal{M}^p_{\mathsf{UB}}(X)$.*

We will not give the proof but we will note a minor lemma that is used in the proof.

Lemma 3.4 *If $\alpha : (X, p) \longrightarrow (Y, q)$ satisfies $M_\alpha(p) \leq Kq$ for some real positive constant K (i.e. $M_\alpha(p) \in \mathcal{M}^q_{\mathsf{UB}}$) then $\frac{\mathrm{d}M_\alpha(p)}{\mathrm{d}q}$ is in $L_\infty^+(Y, q)$.*

Proof . We write h for $\frac{\mathrm{d}M_\alpha(p)}{\mathrm{d}q}$. The Radon-Nikodym theorem tells us that h is in $L_1(Y, q)$. For any g in $L_1^+(Y, q)$ we have $\int g \mathrm{d}M_\alpha(p) \leq K \int g \mathrm{d}q$. Fix a positive real η and define $Z_\eta = \{y | h(y) > \eta\}$, then

$$\eta q(Z_\eta) \leq \int_{Z_\eta} h \mathrm{d}q = M_\alpha(p)(Z_\eta) \leq K q(Z_\eta).$$

So if $q(Z_\eta) \neq 0$ we have $\eta \leq K$; thus, except for a set of q-measure 0, h is bounded by K; i.e. h in $L^+_\infty(Y, q)$. ∎

As above, as $\mathcal{M}^p_{\mathsf{UB}}(X)$ is isometrically isomorphic to $L^+_\infty(X)$, an immediate corollary is that $L^{+,*}_1(X)$ is isometrically isomorphic to $L^+_\infty(X)$.

Definition 3.5 *There is a map from the product of the cones $L^+_\infty(X, p)$ and $L^+_1(X, p)$ to $\mathbb{R}^+$ defined as follows:*

$$\forall f \in L^+_\infty(X, p), g \in L^+_1(X, p) \quad \langle f,\ g\rangle = \int fg \mathrm{d}p.$$

This map is bilinear and is continuous and ω-continuous in both arguments; we refer to it as the pairing.

This pairing allows one to express the dualities in a very convenient way. For example, the isomorphism between $L^+_\infty(X, p)$ and $L^{+,*}_1(X, p)$ sends $f \in L^+_\infty(X, p)$ to $\lambda g.\langle f,\ g\rangle = \lambda g. \int fg\mathrm{d}p$. A trivial but useful lemma about the pairing function is that it is multiplicative.

Lemma 3.6 *For all $g, h \in L^+_\infty$ and $f \in L^+_1$, $\langle g, hf\rangle = \langle gh, f\rangle$.*

The proof is immediate from the definition; the only point to note is that the product of a function in L^+_1 and a function in L^+_∞ is again in L^+_1.

Using the pairing the following is a consequence of the duality of $L^+_1(X)$ and $L^+_\infty(X)$ in $\omega\mathbf{CC}$.

Proposition 3.7 *Given $A : L^+_1(X, p) \longrightarrow L^+_1(Y, q)$ in $\omega\boldsymbol{CC}$, there is a unique adjoint arrow $L^+_\infty(X, p) \leftarrow L^+_\infty(Y, q) : A^\dagger$ in $\omega\boldsymbol{CC}$, such that:*

$$\langle g,\ Af\rangle_Y = \langle A^\dagger g,\ f\rangle_X$$

for all $f \in L^+_1(X, p)$, $g \in L^+_\infty(Y, q)$. Similarly, given $L^+_\infty(X, p) \leftarrow L^+_\infty(Y, q) : A^\dagger$ in $\omega\boldsymbol{CC}$, there is a unique adjoint $A : L^+_1(X, p) \longrightarrow L^+_1(Y, q)$ such that the above holds.

Proof . Suppose A is given, we define:

$$g \in L^+_\infty(Y, q) \mapsto \lambda f \in L^+_1(X, p).\langle g,\ Af\rangle_Y.$$

The right hand side is linear, continuous and ω-continuous in f so is in $L^{+,*}_1(X, p) \sim L^+_\infty(X, p)$. This defines $A^\dagger g$ in dual form; this definition is unique because it is forced by the adjointness relation. This map is clearly linear and continuous as:

$$\left\|\widetilde{A^\dagger g}\right\| = \left\|\lambda f \in L^+_1(X, p).\langle g,\ Af\rangle_Y\right\|$$

where the tilde indicates that it is defined in the dual space. Now the right hand side of the above is equal to

$$\sup_{f \in L_1^+(X,p)} \langle g,\ Af \rangle_Y / \|f\|_1 \leq \|g\|_\infty \|A\| .$$

which tells us in passing that $\|A^\dagger\| \leq \|A\|$, and is ω-continuous as for all $f \in L_1^+(X,p)$ and for all sequences g_n converging from below to g in $L_\infty^+(Y,q)$

$$\langle g_n,\ Af \rangle_Y \longrightarrow \langle g,\ Af \rangle_Y$$

by the monotone convergence theorem.

The dual version is essentially the same. ∎

We define two categories $\mathbf{Rad}_\infty$ and $\mathbf{Rad}_1$ that will be needed for the functorial definition of conditional expectation.

Definition 3.8 *The category $\mathbf{Rad}_\infty$ has as objects probability spaces, and as arrows $\alpha : (X,p) \longrightarrow (Y,q)$, measurable maps such that $M_\alpha(p) \leq Kq$ for some real number K. The category $\mathbf{Rad}_1$ has as objects probability spaces and as arrows $\alpha : (X,p) \longrightarrow (Y,q)$, measurable maps such that $M_\alpha(p) \ll q$.*

The reason for choosing the names $\mathbf{Rad}_1$ and $\mathbf{Rad}_\infty$ is that $\alpha \in \mathbf{Rad}_x$ maps to $d/dqM_\alpha(p) \in L_x^+(Y,q)$ (here x is 1 or ∞). For $x = 1$ this is true by the Radon-Nikodym theorem while for $x = \infty$ it follows from Lemma 3.4. The fact that the category $\mathbf{Rad}_\infty$ embeds in $\mathbf{Rad}_1$ reflects the fact that L_∞^+ embeds in L_1^+.

When we define bisimulation we will need the subcategory of $\mathbf{Rad}_\infty$ consisting of measure-preserving maps. We call this category $\mathbf{Rad}_=$.

3.1 Summary of spaces and their relationships

We summarize the various categories that we have defined and the relationships between them which we have proved in this previous section. All the spaces are ω-complete normed cones, thus, isomorphism always means isomorphism in the category of ω-complete normed cones.

We fix a probability triple (X, Σ, p) and focus on six spaces of cones that are based on them. They break into two natural groups of three isomorphic spaces. The first three spaces are:

A1 $\mathcal{M}^{\ll p}(X)$ - the cone of all measures on (X, Σ, p) that are absolutely continuous with respect to p,

A2 $L_1^+(X,p)$ - the cone of integrable almost-everywhere positive functions,

A3 $L_\infty^{+,*}(X,p)$ - the dual cone of the the cone of almost-everywhere positive bounded measurable functions.

The first space above, $\mathcal{M}^{\ll p}(X)$ is clearly a subspace of $\mathcal{M}(X)$, the space of all finite measures on X.

The next group of three isomorphic spaces are:

B1 $\mathcal{M}^p_{\mathsf{UB}}(X)$ - the cone of all measures that are uniformly less than a multiple of the measure p,

B2 $L_\infty^+(X,p)$ - the cone of almost-everywhere positive functions in the normed vector space $L_\infty(X,p)$,

B3 $L_1^{+,*}(X,p)$ - the dual of the cone of almost-everywhere positive functions in the normed vector space $L_1(X,p)$.

The functions that arise in the equivalence classes of functions constituting $L_\infty^+(X,p)$ and $L_1^+(X,p)$ are contained in $\mathcal{L}^+(X)$ the space of non-negative real-valued functions on X.

The spaces defined in A1, A2 and A3 are dual to the spaces defined in B1, B2 and B3 respectively. The situation may be depicted in the diagram

$$\begin{tikzcd}
\mathcal{M}^{\ll p}(X) \arrow[r, leftrightarrow, "\sim"] \arrow[d, leftrightarrow, dotted] & L_1^+(X,p) \arrow[r, leftrightarrow, "\sim"] \arrow[d, leftrightarrow, dotted] & L_\infty^{+,*}(X,p) \arrow[d, leftrightarrow, dotted] \\
\mathcal{M}^p_{\mathsf{UB}} \arrow[r, leftrightarrow, "\sim"] & L_\infty^+(X,p) \arrow[r, leftrightarrow, "\sim"] & L_1^{+,*}(X,p)
\end{tikzcd} \tag{1}$$

where the vertical arrows represent dualities and the horizontal arrows represent isomorphisms. The proofs of the isomorphism go through the first column, but once they are established, we can mainly work with the second column.

The traditional theory of labelled Markov processes (LMP) was formulated in terms of the spaces $\mathcal{M}(X)$ and $\mathcal{L}^+(X)$. The Markov kernels used in the definition of an LMP are of the form $\tau(x, A)$: they are subprobability measures for each x and positive bounded measurable functions for each A. The essential shift of viewpoint that we propose in this paper is to work with the spaces in A2 and B2 instead: this will be the key definition in the next section.

4 Conditional expectation functorially

There is a very pleasant view of conditional expectation as a functor; this view sets the stage for the approximation theory. The key ingredient is the duality between the cones L_1^+ and L_∞^+ as captured by the pairing map $\langle\cdot,\ \cdot\rangle : L_\infty^+ \times L_1^+ \longrightarrow \mathbb{R}^+$.

First, recall the categories $\mathbf{Rad}_1$ and $\mathbf{Rad}_\infty$ defined in Def. 3.8 and the pairing function defined in Def. 3.5. We have the isomorphism between $L_\infty^+(X,p)$ and $L_1^{+,*}(X,p)$ mediated by the pairing function:

$$f \in L_\infty^+(X,p) \mapsto \lambda g : L_1^+(X,p).\langle f,\ g\rangle = \int fg\mathrm{d}p.$$

Now, precomposition with α in $\mathbf{Rad}_\infty$ gives a map $P_1(\alpha)$ from $L_1^+(Y,q)$ to $L_1^+(X,p)$. To see this consider $\alpha \in \mathbf{Rad}_\infty$ and $g \in L_1^+(Y,q)$. Now

$$\int P_1(\alpha)(g)\,\mathrm{d}p = \langle \tfrac{\mathrm{d}}{\mathrm{d}q}\cdot M_\alpha(p),\ g\rangle_Y$$

which shows that $P_1(\alpha)(g)$ is in $L_1^+(X,p)$. Dually, given $\alpha \in \mathbf{Rad}_1 : (X,p) \longrightarrow (Y,q)$ and $g \in L_\infty^+(Y,q)$ we have

$$\int P_\infty(\alpha)(g)\,\mathrm{d}p = \langle g,\ \tfrac{\mathrm{d}}{\mathrm{d}q}\cdot M_\alpha(p)\rangle_Y$$

which implies that $P_\infty(\alpha)(g) \in L_\infty^+(X,p)$. Thus the subscripts on the two precomposition functors describe the *target* categories. Using the $*$-functor we get a map $(P_1(\alpha))^*$ from $L_1^{+,*}(X,p)$ to $L_1^{+,*}(Y,q)$ in the first case and dually we get $(P_\infty(\alpha))^*$ from $L_\infty^{+,*}(X,p)$ to $L_\infty^{+,*}(Y,q)$.

We are now ready to define the expectation value map.

Definition 4.1 *The **functor** $\mathbb{E}_\infty(\cdot)$ is a functor from $\mathbf{Rad}_\infty$ to $\omega\boldsymbol{CC}$ which, on objects, maps (X,p) to $L_\infty^+(X,p)$ and on maps is given as follows. Given $\alpha : (X,p) \longrightarrow (Y,q)$ in $\mathbf{Rad}_\infty$ the action of the functor is to produce the map $\mathbb{E}_\infty(\alpha) : L_\infty^+(X,p) \longrightarrow L_\infty^+(Y,q)$ obtained by composing $(P_1(\alpha))^*$ with* *the isomorphisms between $L_1^{+,*}$ and L_∞^+ as shown in the diagram below*

$$\begin{tikzcd} L_1^{+,*}(X,p) \arrow[d, "(P_1(\alpha))^*"'] & L_\infty^+(X,p) \arrow[l, dotted] \arrow[d, "\mathbb{E}_\infty(\alpha)"] \\ L_1^{+,*}(Y,q) \arrow[r, dotted] & L_\infty^+(Y,q) \end{tikzcd} \tag{2}$$

It is an immediate consequence of the definitions that

Proposition 4.2 *for any $f \in L_\infty^+(X,p)$ and $g \in L_1^+(Y,q)$*

$$\langle \mathbb{E}_\infty(\alpha)(f), g\rangle_Y = \langle f, P_1(\alpha)(g)\rangle_X.$$

One can informally view this functor as a "left adjoint" in view of this proposition. Note that since we started with α in $\mathbf{Rad}_\infty$ we get the expectation value as a map between the L_∞^+ cones.

We calculate $\mathbb{E}_\infty(\alpha)(\mathbf{1}_X)$ to illustrate the definition. We start with $\mathbf{1}_X \in L_\infty^+(X,p)$. Under the $*$ isomorphism it maps to $\lambda g : L_1^+(X,p). \int g \mathrm{d}p$, which is an element of $L_1^{+,*}(X,p)$. Then under the action of $P_1(\alpha)^*$it maps to $\lambda h : L_1^+(Y,q). \int (h \circ \alpha)\ \mathrm{d}p$ which is in $L_1^{+,*}(Y,q)$. Note that because α satisfies $M_\alpha(p) \leq Kq$ for some K, it follows that $h \circ \alpha$ is in $L_1^+(X,p)$. Finally taking the iso back we get $\frac{\mathrm{d}M_\alpha(p)}{\mathrm{d}q}$ as the value of $\mathbb{E}_\infty(\alpha)(\mathbf{1}_X)$, which is in $L_\infty^+(Y,q)$.

It is a well-known elementary fact that $\int_X g \circ \alpha\ \mathrm{d}p = \int_Y g\ \mathrm{d}q$ if and only if α is measure preserving. It follows then that $\mathbb{E}_\infty(\alpha)(\mathbf{1}_X) = \mathbf{1}_Y$ if and only if α is measure preserving. The general statement is

$$\forall f \in L_\infty^+(X,p). \mathbb{E}_\infty(\alpha)(f) = \frac{\mathrm{d}}{\mathrm{d}q} \cdot M_\alpha(f \cdot p).$$

In exactly the same way we can define a functor from $\mathbf{Rad}_1$ to $\omega\mathbf{CC}$.

Definition 4.3 *The* ***functor*** $\mathbb{E}_{\mathbf{1}}(\cdot)$ *is a functor from* $\mathbf{Rad}_1$ *to* $\omega\boldsymbol{CC}$ *which maps the object (X,p) to $L_1^+(X,p)$ and on maps is given as follows: Given $\alpha : (X,p) \longrightarrow (Y,q)$ in* $\mathbf{Rad}_1$ *the action of the functor is to produce the map $\mathbb{E}_1(\alpha) : L_1^+(X,p) \longrightarrow L_1^+(Y,q)$ obtained by composing $(P_\infty(\alpha))^*$ with* *the isomorphisms between $L_\infty^{+,*}$ and L_1^+ as shown in the diagram below*

$$\begin{tikzcd} L_\infty^{+,*}(X,p) \arrow[d, "(P_\infty(\alpha))^*"'] & L_1^+(X,p) \arrow[l, dotted] \arrow[d, "\mathbb{E}_1(\alpha)"] \\ L_\infty^{+,*}(Y,q) \arrow[r, dotted] & L_1^+(Y,q) \end{tikzcd} \tag{3}$$

Once again we have an "adjointness" statement; this time it is a right adjoint.

Proposition 4.4 *Given $f \in L_\infty^+(Y,q)$ and $g \in L_1^+(X,p)$ we have*

$$\langle f, \mathbb{E}_1(\alpha)(g)\rangle_Y = \langle P_\infty(\alpha)(f), g\rangle_X.$$

The relationship between these two expectation value functors and the corresponding precomposition functors is given by the following proposition.

Proposition 4.5 *Given $\alpha \in \mathbf{Rad}_\infty[(X,p),(Y,q)]$ we have*

(a) $\mathbb{E}_1(\alpha)(f \circ \alpha) = \mathbb{E}_\infty(\alpha)(\mathbf{1}_X)f,$ *for $f \in L_1^+(Y,q)$ and*
(b) $\mathbb{E}_\infty(\alpha)(f \circ \alpha) = \mathbb{E}_1(\alpha)(\mathbf{1}_X)f,$ *for $f \in L_\infty^+(Y,q)$.*

Proof . We prove the first, the second is virtually identical, one just has to dualize every step; in fact they are the same up to adjunction.

In view of the duality, it suffices to show that for any $g \in L_\infty^+(Y,q)$ we have

$$\langle g, \mathbb{E}_1(\alpha)(f \circ \alpha)\rangle = \langle g, \mathbb{E}_\infty(\alpha)(\mathbf{1}_X)f\rangle.$$

We calculate as follows:

$$\begin{array}{rcll} \langle g, \mathbb{E}_1(\alpha)(f\circ\alpha)\rangle & = & \langle g\circ\alpha, f\circ\alpha\rangle & \text{right-adjointness of } \mathbb{E}_1() \\ & = & \langle \mathbf{1}_X(g\circ\alpha), f\circ\alpha\rangle & \text{obvious} \\ & = & \langle \mathbf{1}_X, (g\circ\alpha)(f\circ\alpha)\rangle & \text{Lemma 3.6} \\ & = & \langle \mathbf{1}_X, (gf)\circ\alpha\rangle & \text{pointwise multiplication} \\ & = & \langle \mathbb{E}_\infty(\alpha)(\mathbf{1}_X), gf\rangle & \text{left-adjointness of } \mathbb{E}_\infty() \\ & = & \langle g, \mathbb{E}_\infty(\alpha)(\mathbf{1}_X)f\rangle & \text{Lemma 3.6 twice.} \end{array}$$

∎

One last detail that needs to be tied up is the calculation of the norm of some operators. We start with an almost immediate observation; we write $\|\cdot\|$ for the operator norm.

Lemma 4.6 *Given any linear* $F : L_\infty^+(X) \longrightarrow L_\infty^+(Y)$, $\|F\| = \|F(\mathbf{1}_X)\|_\infty$.

Proof . If $f \in L_\infty^+(X)$ we have $f \leq \|f\|_\infty \mathbf{1}_X$, where $\leq$ is the cone order. In particular, if $\|f\|_\infty = 1$ we have $f \leq \mathbf{1}_X$. For such an f and for monotone F, we have $F(f) \leq F(\mathbf{1}_X)$, so by monotonicity of the norm we have $\|F(f)\|_\infty \leq \|F(\mathbf{1}_X)\|_\infty$. Hence by definition of the operator norm $\|F\| = \|F(\mathbf{1}_X)\|_\infty$. ∎

We have two immediate consequences.

Lemma 4.7 *Suppose we have a map* $\alpha : (X, p) \longrightarrow (Y, q)$ *in* $\mathbf{Rad}_\infty$. *Then:*

1. $\mathbb{E}_\infty(\alpha) : L_\infty^+(X) \longrightarrow L_\infty^+(Y)$ *has norm* $\|\mathbb{E}_\infty(\alpha)(\mathbf{1}_X)\|_\infty$.
2. *The map* $P_\infty(\alpha) : L_\infty^+(Y, q) \longrightarrow L_\infty^+(X, p)$ *has norm* 1.

We have already seen that there is a dagger functor introduced in Proposition 3.7. This adjoint is a contravariant functor which is defined on the subcategories that arise as L_1^+ and L_∞^+.

5 Labelled abstract Markov processes

5.1 Markov processes as function transformers

It is a pleasing fact that Markov kernels can be viewed as linear maps on function spaces. This idea was first elaborated by [YK41] and underlies much of the present work.

Given τ a Markov kernel from (X, Σ) to (Y, Λ), we define $T_\tau : \mathcal{L}^+(Y) \longrightarrow \mathcal{L}^+(X)$, for $f \in \mathcal{L}^+(Y)$, $x \in X$, as $T_\tau(f)(x) = \int_Y f(z)\tau(x, dz)$. This map is well-defined, linear and ω-continuous. If we write $\mathbf{1}_B$ for the indicator function of the measurable set B we have that $T_\tau(\mathbf{1}_B)(x) = \tau(x, B)$ and hence is measurable for every $B \in \Lambda$. Thus $T_\tau(f)$ is measurable for any measurable f by the usual argument starting from simple functions and using first linearity and then the monotone convergence theorem.

Conversely, any ω-continuous morphism L with $L(\mathbf{1}_Y) \leq \mathbf{1}_X$ can be cast as a Markov kernel by reversing the process above. The interpretation of L is that $L(\mathbf{1}_B)$ is a measurable function on X such that $L(\mathbf{1}_B)(x)$ is the probability of jumping from x to B. Thus L does encode a transition probability.

We can also define an operator on $\mathcal{M}(X)$ by using τ the other way. We define $\bar{T}_\tau : \mathcal{M}(X) \longrightarrow \mathcal{M}(Y)$, for $\mu \in \mathcal{M}(X)$ and $B \in \Lambda$, as $\bar{T}_\tau(\mu)(B) = \int_X \tau(x, B) \, \mathrm{d}\mu(x)$. It is easy to show that this map is linear and ω-continuous.

The two operators T_τ and $\bar{T}_\tau$ have interesting interpretations. The operator $\bar{T}_\tau$ transforms measures "forwards in time"; if μ is a measure on X representing the current state of the system, $\bar{T}_\tau(\mu)$ is the resulting measure on Y after a transition through τ.

On the other hand, the operator T_τ may be interpreted as a transformer of random variables that propagates information "backwards", just as we expect from predicate transformers. This inversion can be seen from the reversal of X and Y in the definition of the operator. Note that $T_\tau(f)(x)$ is just the expected value of f after one τ-step given that one is at x. Thus, we have an expectation-value transformer.

5.2 Abstract Markov processes

If our measurable spaces X and Y are endowed with measures p and q, respectively, which we shall assume finite, it is tempting to consider positive operators on L_1^+ and L_∞^+ instead of on $\mathcal{L}^+$: we call these abstract Markov processes because they operate on equivalence classes of functions rather than on the concrete functions, but, in view of the isomorphisms discussed in Section 2, they can also be regarded as operating on spaces of measures.

This view was first explored by [Hop54]. We will slightly modify the classical definitions in order to work with cones; the interested reader may consult standard sources [Sch74, AGG$^+$86, Haw06] for the usual framework in Banach spaces or Banach lattices.

Definition 5.1 *A **Markov operator** from a state space (X, Σ, μ) to a state space (Y, Λ, ν) is a linear map $T : L_1^+(X) \longrightarrow L_1^+(Y)$ such that $\|T\| \leq 1$.*

Note that the operator norms of both T_τ and $\bar{T}_\tau$ are less than one. Here $\bar{T}_\tau : \mathcal{M}(X) \longrightarrow \mathcal{M}(Y)$ and $T_\tau : \mathcal{L}^+(Y) \longrightarrow \mathcal{L}^+(X)$ and the operator norms are computed using the norms on the cones $\mathcal{M}(X), \mathcal{M}(Y), \mathcal{L}^+(X)$ and $\mathcal{L}^+(Y)$.

This is the analog of the measure transforming operator $\bar{T}_\tau$ above, as the elements of $L_1^+(X)$ correspond to measures which are absolutely continuous with respect to our given measure μ (and similarly for $L_1^+(Y)$). In this case the map is automatically order-continuous.

Proposition 5.2 *If $F : L_1^+(X,\mu) \longrightarrow L_1^+(Y,\nu)$ is linear and has finite operator norm, i.e. it is a continuous linear map, then F is ω-continuous.*

Proof . Suppose that we have an increasing sequence $\{f_i\}$ with a pointwise lub f, then by the monotone convergence theorem we have that $\int f_i \,\mathrm{d}\mu$ converges to $\int f \,\mathrm{d}\mu$. Since F is monotone, $F(f_i)$ is increasing and is bounded by $F(f)$. Since F has finite operator norm we have $\|F(f) - F(f_i)\|_1 \leq \|F\| \, \|f - f_i\|_1$ and by the monotone convergence theorem we have $\lim_{i\longrightarrow\infty} \|f - f_i\|_1 = 0$. Now note that $F(f_i) \leq F(f)$ since F is linear, hence monotone. Also, from the definition of $\|F\|$, we have $\|F(f_i)\|_1 \leq \|F\| \cdot \|f_i\|_1$ so the sequence $\|F(f_i)\|_1$ is bounded and, since the cone is complete, has a least uper bound. Thus from Lemma 2.12 we have $F(f) = \bigvee_i F(f_i)$. ∎

From the "backwards transformation" point of view the operator we work with is the equivalent of T_τ. We have the following definition:

Definition 5.3 *An* ***abstract Markov kernel*** *from (X, Σ, p) to (Y, Λ, q) is an ω-continuous linear map $\tau : L_\infty^+(Y) \longrightarrow L_\infty^+(X)$ with $\|\tau\| \leq 1$.*

Definition 5.4 *A* ***labelled abstract Markov process*** *on a probability space (X, Σ, p) with a set of labels (or actions) $\mathcal{A}$ is a family of abstract Markov kernels $\tau_a : L_\infty^+(X,p) \longrightarrow L_\infty^+(X,p)$ indexed by elements a of $\mathcal{A}$.*

Requiring that $\|\tau\|$ be less than 1 is equivalent to requiring that $\tau\mathbf{1}_X \leq \mathbf{1}_X$. Hence, an abstract Markov kernel is an arrow in the category $\omega\mathbf{CC}$. Note the inversion of Y and X in the definition.

In this definition, we require that τ be ω-continuous in addition to being linear. Unlike the L_1^+ case, linearity does not guarantee ω-continuity; [Sel04] gives a counter example. It is worth understanding the counter-example because it sheds light on why we have a perfect duality in our setting. We work with the space $L_\infty^+(\mathbb{N}, \#)$, where $\mathbb{N}$ is the natural numbers and $\#$ represents the counting measure. We write l_∞^+ for this space: it consists of bounded sequences of real numbers. We write s for such a sequence and $s[i]$ for the ith element of the sequence. Let $\mathcal{U}$ be a non-principal ultrafilter on $\mathbb{N}$. We define a function $\lim_{\mathcal{U}} : l_\infty^+ \longrightarrow \mathbb{R}^+$ as follows:

$$\lim_{\mathcal{U}}(s) = \sup\{x \mid \{i \mid s[i] \geq x\} \in \mathcal{U}\}.$$

It is not obvious but one can show that $\lim_{\mathcal{U}}$ is linear. Consider the increasing chain of sequences $s_n = [1, 1, \ldots, 1, 0, 0, \ldots]$ where the first n entries of s_n are 1s. Since $\mathcal{U}$ is a non-principal ultrafiler we have $\lim_{\mathcal{U}} s_n = 0$ for all n. However, the limit is the constant sequence of 1s and the $\lim_{\mathcal{U}}$ of this is 1. Thus this functional is not continuous. It was important to have a non-principal ultrafilter for this example to work. Note that this example shows that just taking bounded linear maps to construct dual spaces will not give us the perfect duality that we have. The ω-continuity controls the dual more stringently than the usual norm continuity and gives us duality.

The following corollary, though not needed for any of the results, gives the relation between Markov operators and abstract Markov kernels.

Corollary 5.5 *Given finite measure spaces (X, Σ, μ) and (X, Λ, ν), there is a bijection between Markov operators from X to Y and abstract Markov kernels from X to Y. The bijection is given by the adjoint operation.*

Remark 5.6 *One can find a similar bilinear form which demonstrates that the operators $\bar{T}_\tau$ and T_τ are adjoints.*

We can relate Markov operators to a special type of Markov kernel. If X and Y are endowed with measures μ and ν, respectively, a Markov kernel from X to Y is *nonsingular* if, for all measurable sets $B \subseteq Y$ such that $\nu(B) = 0$, we have $\tau(x, B) = 0$, μ-almost everywhere. The following result is essentially due to [Hop54], one has to make very minor modifications to adapt it to the cone situation:

Proposition 5.7 *Every Markov operator from (X, Σ, μ) to (Y, Λ, ν) corresponds uniquely to a nonsingular Markov kernel from X to Y.*

As an immediate corollary, one obtains a one-to-one correspondence between nonsingular Markov kernels and abstract Markov kernels from X to Y. Informally, one obtains a Markov kernel $\hat{\tau}$ from an abstract Markov kernel τ from X to Y as follows: given a measurable set B in Λ, we let $\tau(\mathbf{1}_B)(x) = \hat{\tau}(x, B)$; this is precisely the interpretation we had for the operator T_τ.

The above proposition is not completely trivial because the functions $\tau(\mathbf{1}_B)(x)$ are only defined μ-almost everywhere. The proof of this proposition will be omitted; however, we give an intuitive justification of why it holds. If $\hat{\tau}$ is a nonsingular Markov kernel from X to Y, we require that $\nu(B) = 0 \Rightarrow \hat{\tau}(x, B) =_\mu 0$. Interpreting $\hat{\tau}$ as an abstract Markov kernel, we thus require that $\tau(\mathbf{1}_B) =_\mu 0$ if $\nu(B) = 0$, or if $\mathbf{1}_B =_\nu 0$. This is a necessary condition for τ to be linear; the proposition above shows that it is sufficient.

6 The approximation map on LAMPs

The expectation value functors essentially project a probability space onto another one with a possibly coarser σ-algebra. This is what we use to define the notion of approximation. Given an AMP on (X,p) and a map $\alpha : (X,p) \to (Y,q)$ in $\mathbf{Rad}_\infty$, we have the following approximation scheme:

$$
\begin{tikzcd}
L_\infty^+(X,p) \arrow[r, "\tau_a"] & L_\infty^+(X,p) \arrow[d, "\mathbb{E}_\infty(\alpha)"] \\
L_\infty^+(Y,q) \arrow[u, "P_\infty(\alpha)"] \arrow[r, dotted, "\alpha(\tau_a)"] & L_\infty^+(Y,q)
\end{tikzcd}
\tag{4}
$$

Here we write τ_a for all the Markov kernels associated with the AMP. Thus any a that appears is intended to be universally quantified. It follows from Prop. 4.5 that if α is measure preserving then $\alpha(Id) = Id$ where Id is the identity on L_∞^+. There is no reason why α should be a functor though. Note that $\|\alpha(\tau_a)\| \leq \|P_\infty(\alpha)\| \cdot \|\tau_a\| \cdot \|\mathbb{E}_\alpha\| = \|\tau_a\| \cdot \|\mathbb{E}_\infty(\alpha)(\mathbf{1}_X)\|_\infty$. Thus, if α is measure preserving we get $\|\alpha(\tau_a)\| \leq \|\tau_a\|$.

A special case of this is when we have (X,Σ) and (X,Λ), i.e. the two spaces have the same underlying point set but are equipped with different σ-algebras and $\Lambda \subset \Sigma$, now the identity function id from (X,Σ) to (X,Λ) is measurable and we can define an approximation by moving to a coarser σ-algebra. In our set up we are approximating along any measurable function rather than just identity maps between the same spaces but with different σ-algebras.

In the same situation as in the previous paragraph, the map $\mathbb{E}_1(id) : L_1^+(X,\Sigma,p) \to L_1^+(X,\Lambda,p)$ is the exactly function that is traditionally written $\mathbb{E}(\cdot|\Lambda)$ [Bil95]. The functoriality of the expectation value is what is called the "tower law of conditional expectation" in probability theory [Wil91].

The notion of approximation immediately applies to LAMPs. Given probability spaces (X,p) and (Y,q) and a $\mathbf{Rad}_\infty$ map α from (X,p) to (Y,q) we can project each τ_a of a LAMP on (X,p) to one on (Y,q) as described just above. Since an AMP has a norm less than 1, we can only be sure that α yields an approximation for every AMP on X if $\|\mathbb{E}_\infty(\alpha)(\mathbf{1}_X)\|_\infty \leq 1$. We call the AMP $\alpha(\tau_a)$ the projection of τ_a on Y.

7 Bisimulation

The notion of probabilistic bisimulation was introduced by [LS91] for discrete spaces and by [BDEP97] (see also [DEP02]) for continuous spaces. Subsequently a dual notion called event bisimulation or probabilistic cocongruence was defined independently by [DDLP06] and by [BSdV04]. For a more detailed discussion of the history see Section 12. 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.

7.1 The category AMP

We have developed the functorial theory of conditional expectation in a fairly general setting with mild conditions on the maps: for example, in $\mathbf{Rad}_\infty$, the image measure is bounded by a multiple of the measure in the target space. From now on, we consider a category where the objects are LAMPs that will be relevant to the approximation theory. We will work with probability spaces equippaed with abstract Markov processes. The maps will be measure-preserving maps. These maps are essentially surjective but there is no real reason not to restrict to maps that are not surjective in the usual sense.

Definition 7.1 *We define the category* **AMP** *as follows. The objects consist of probability spaces (X, Σ, p), along with an abstract Markov process τ_a on X. The arrows $\alpha : (X, \Sigma, p, \tau_a) \longrightarrow (Y, \Lambda, q, \rho_a)$ are surjective measurable measure-preserving maps from X to Y such that $\alpha(\tau_a) = \rho_a$.*

In words, this means that the Markov processes defined on the codomain are precisely the projection of the Markov processes τ_a on the domain through α. When working in this category, we will often denote objects by the state space, when the context is clear.

7.2 Event bisimulation and Zigzags

We begin with the definition of *event bisimulation* which comes from [DDLP06] where it was developed for LMPs.

Definition 7.2 *Given a LMP (X, Σ, τ_a), an event-bisimulation is a sub-σ-algebra Λ of Σ such that (X, Λ, τ_a) is still an LMP [DDLP06].*

More explicitly, the condition that needs to hold for Λ to be an event bisimulation is that $\tau(x, A)$ is Λ-measurable for a fixed $A \in \Lambda$. This is the case if and only if $\tau_a : L_\infty^+(X, \Sigma, p) \longrightarrow L_\infty^+(X, \Sigma, p)$ sends the subspace $L_\infty^+(X, \Lambda, p)$ to itself, where we are now viewing τ_a as a map on the function space. In other words, the following diagram commutes:

$$\begin{tikzcd}
L_\infty^+(X, \Sigma, p) \arrow[r, "\tau_a"] & L_\infty^+(X, \Sigma, p) \\
L_\infty^+(X, \Lambda, p) \arrow[r, "\tau_a"] \arrow[u, hook] & L_\infty^+(X, \Lambda, p) \arrow[u, hook]
\end{tikzcd} \tag{5}$$

This is the notion we need for LAMPS.

We can generalize the notion of event bisimulation by using maps other than the identity map on the underlying sets. This would be a map α from (X, Σ, p) to (Y, Λ, q), equipped with LMPs τ_a and ρ_a respectively, such that the following commutes:

$$\begin{tikzcd}
L_\infty^+(X, \Sigma, p) \arrow[r, "\tau_a"] & L_\infty^+(X, \Sigma, p) \\
L_\infty^+(Y, \Lambda, q) \arrow[r, "\rho_a"] \arrow[u, "P_\infty(\alpha)"] & L_\infty^+(Y, \Lambda, q) \arrow[u, "P_\infty(\alpha)"']
\end{tikzcd} \tag{6}$$

This corresponds to a morphism of coalgebras in the concrete case. Note that if, in Diagram 6, we consider the special case where α is the identity map $(X, \Sigma) \longrightarrow (X, \Lambda)$, we get Diagram 5.

We will refrain from calling these maps bisimulation maps yet; we will call such maps *zigzags*; they are essentially the same as zigzags for labelled Markov processes [DEP02].

Definition 7.3 *A* ***zigzag*** *from an abstract Markov process* (X, Σ, p, τ_a) *to another abstract Markov process* (Y, Λ, q, ρ_a) *is a measurable, measure-preserving surjective function from* X *to* Y *such that Diagram 6 commutes.*

Note that if there is a zigzag α from X to Y, then the LAMP on Y is very closely related to the projection of τ_a onto Y via α, i.e. to $\alpha(\tau_a) =$

$\mathbb{E}_\infty(\alpha) \circ \tau_a \circ P_\infty(\alpha)$. We have the following commuting diagram:

$$\begin{tikzcd}
L^+_\infty(Y) \arrow[rrr, "\rho_a"] \arrow[dd, equal] \arrow[dr, "P_\infty(\alpha)"] & & & L^+_\infty(Y) \arrow[dl, "P_\infty(\alpha)"'] \arrow[dd, "\mathbb{E}_1(\alpha)(\mathbf{1}_X)\cdot(-)"] \\
& L^+_\infty(X) \arrow[r, "\tau_a"] & L^+_\infty(X) \arrow[dr, "\mathbb{E}_\infty(\alpha)"'] & \\
L^+_\infty(Y) \arrow[ur, "P_\infty(\alpha)"'] \arrow[rrr, "\alpha(\tau_a)"] & & & L^+_\infty(Y)
\end{tikzcd} \tag{7}$$

We have that $\mathbb{E}_\infty(\alpha)(f \circ \alpha) = \mathbb{E}_1(\alpha)(\mathbf{1}_X)f$ from the second equation of Prop. 4.5. This implies that $\alpha(\tau_a) = \rho_a \cdot \mathbb{E}_1(\alpha)(\mathbf{1}_X)$. In particular, if $\mathbb{E}_1(\alpha)(\mathbf{1}_X) = \mathbf{1}_Y$ – which happens if *and only if* $M_\alpha(p) = q$ – then ρ_a is equal to $\alpha(\tau_a)$, the projection of τ onto Y. Note that the condition $M_\alpha(p) = q$ means by definition that the image measure is precisely the measure in the codomain of α. In short if we "approximate" along a measure-preserving zigzag then the approximation is the same as the exact result. This means that approximations and bisimulations live in the same universe and bisimulations appear as special approximations, or, put another way, the approximations are really approximate bisimulations. This explains why we restricted to the measure-preserving case in this section.

We record the fact that zigzags are arrows in **AMP** as a Lemma.

Lemma 7.4 *If $\alpha : (X, \Sigma, p, \tau_a) \longrightarrow (Y, \Lambda, q, \rho_a)$ is a zig-zag then $\alpha(\tau_a) = \rho_a$, which is to say that α is a morphism of* **AMP**.

Proof . From the Diagram 7 we have that

$$\alpha(\tau_a) = \rho_a \cdot \mathbb{E}_\infty(\alpha)(\mathbf{1}_X).$$

Since α is measure preserving we have $\mathbb{E}_\infty(\alpha)(\mathbf{1}_X) = \mathbf{1}_Y$ so we get $\alpha(\tau_a) = \rho_a$. ∎

7.3 Bisimulation Defined on AMP

It should be noticed that surjective measure-preserving maps between probability spaces typically involve information loss. This information loss is encoded in the requirement that the maps be measurable: one only asks for the preimages of the measurable sets to be measurable. To recall the situation that we discussed earlier; consider the identity map on a set X equipped with two σ-algebras $\Lambda \subset \Sigma$. This map id induces the conditional expectation operator $\mathbb{E}(\cdot|\Lambda) : L_1^+(X, \Sigma, p) \longrightarrow L_1^+(X, \Lambda, p)$ or $\mathbb{E}_1(id)$ which effectively "pixelizes" the functions in the sense that Σ-measurable functions become only Λ-measurable.

The existence of a zigzag is a very strong condition, too strong for a reasonable theory; bisimulation as originally defined is a relation. The relational aspect is captured by using cospans[5].

Definition 7.5 *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*

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

giving a cospan diagram

$$\begin{tikzcd} (X, \Sigma, p, \tau) \arrow[dr, "\alpha"] & & (Y, \Lambda, q, \rho) \arrow[dl, "\beta"] \\ & (Z, \Gamma, r, \pi) & \end{tikzcd} \tag{8}$$

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.

8 Bisimulation is an equivalence

This section is devoted to establishing that bisimulation is an equivalence relation. The crucial step is Theorem 8.1 which shows that one can paste together cospans of zigzags in order to show transitivity.

Theorem 8.1 *The category* **AMP** *has pushouts. Furthermore, if the morphisms in the span are zigzags then the morphisms in the pushout diagram are also zigzags. More explicitly, let* $\alpha : (X, \Sigma, p, \tau_a) \longrightarrow (Y, \Lambda, q, \rho_a)$ *and* $\beta : (X, \Sigma, p, \tau_a) \longrightarrow (Z, \Gamma, r, \kappa_a)$ *be a span in* **AMP***. Then there is an object*

[5] When bisimulation was developed for LMPs [DEP02], the authors used *spans* rather than co-spans. Later [DDLP06] it was realized that the theory is smoother with co-spans. The two notions turn out to be equivalent on analytic spaces but are not the same if the underlying σ-algebra does not arise as the Borel algebra of an analytic space. See the historical review for more discussion of this.

*(W, Ω, μ, π_a) of **AMP** and **AMP** maps $\delta : Y \longrightarrow W$ and $\gamma : Z \longrightarrow W$ such that the diagram*

$$\begin{tikzcd}
 & (X, \Sigma, p, \tau_a) \arrow[dl, "\alpha"] \arrow[dr, "\beta"] & \\
(Y, \Lambda, q, \rho_a) \arrow[dr, "\delta"] & & (Z, \Gamma, r, \kappa_a) \arrow[dl, "\gamma"] \\
 & (W, \Omega, \mu, \pi_a) &
\end{tikzcd} \tag{9}$$

*commutes. If (U, Ξ, ν, λ_a) is another **AMP** object and $\phi : Y \longrightarrow U$ and $\psi : Z \longrightarrow U$ are **AMP** maps such that α, β, ϕ and ψ form a commuting square, then there is a unique **AMP** map $\theta : W \longrightarrow U$ such that the diagram*

$$\begin{tikzcd}
 & (X, \Sigma, p, \tau_a) \arrow[dl, "\alpha"] \arrow[dr, "\beta"] & \\
(Y, \Lambda, q, \rho_a) \arrow[dr, "\delta"] \arrow[ddr, "\phi"'] & & (Z, \Gamma, r, \kappa_a) \arrow[dl, "\gamma"'] \arrow[ddl, "\psi"] \\
 & (W, \Omega, \mu, \pi_a) \arrow[d, "\theta"] & \\
 & (U, \Xi, \nu, \lambda_a) &
\end{tikzcd} \tag{10}$$

commutes. Furthermore, if α and β are zigzags, then so are γ and δ.

We will present the proof in stages. First we observe that pushouts can be constructed in the category **Set**. This can then be lifted to the category **Mes**, then we show that this construction can be lifted to $\mathbf{Rad}_=$ and finally to **AMP**. In fact, the pushout object in each case will be built on the previous one and the maps will be the same. Thus the couniversality property that we need for **AMP** follows from that of **Set**, once we show that the mediating morphism constructed in **Set** has the right properties to qualify as an **AMP** morphism.

Proof . It is straightforward to show [DEP02, DDLP06, Pan09] that pushouts exist in the category of measurable spaces: it is the usual pushout in **Set**, equipped with the largest σ-algebra making the pushout maps measurable. We thus have the following pushout diagram in **Mes**, the category of measurable spaces:

$$
\begin{tikzcd}
 & (X,\Sigma) \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] \arrow[dd, "\eta"] & \\
(Y,\Lambda) \arrow[dr, "\delta"'] & & (Z,\Gamma) \arrow[dl, "\gamma"] \\
 & (W,\Omega) &
\end{tikzcd}
\tag{11}
$$

Note here that, of course, $\eta = \delta \circ \alpha = \gamma \circ \beta$. Couniversality is captured by the following diagram:

$$
\begin{tikzcd}
 & (X,\Sigma) \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\
(Y,\Lambda) \arrow[dr, "\delta"] \arrow[ddr, bend right, "\phi"'] & & (Z,\Gamma) \arrow[dl, "\gamma"'] \arrow[ddl, bend left, "\psi"] \\
 & (W,\Omega) \arrow[d, dashed, "\theta"] & \\
 & (U,\Xi) &
\end{tikzcd}
\tag{12}
$$

where θ, the mediating morphism, is unique. It is also constructed exactly as in **Set**; it can be readily verified that when the other maps in the diagram are measurable it is also measurable.

We have to construct a measure on W such that the maps δ and γ are measure preserving, we already know that they are surjective by the construction of the pushout in **Set**. Let us define on (W, Ω) the measure μ in the obvious way: for $B \in \Omega$, $\mu(B) = p(\eta^{-1}(B))$. Note that by the definition of η and the fact that α and β are measure-preserving, we have[6] $\mu(B) = p(\eta^{-1}(B)) = p(\alpha^{-1}(\delta^{-1}(B))) = q(\delta^{-1}(B)) = p(\beta^{-1}(\gamma^{-1}(B)) = r(\gamma^{-1}(B))$ and so we automatically have that γ and δ are measure-preserving. In short we have shown that we have a commuting square in the category $\mathbf{Rad}_=$. To

[6]We have used the explicit definition of the image measure here, i.e. we write, for example, $p(\eta^{-1}(B)$ instead of $M_\eta(p)(B)$ in order to make the calculations clearer.

show the couniversality property we consider the diagram

$$
\begin{tikzcd}
& (X,\Sigma,p) \arrow[dl, "\alpha"'] \arrow[dr, "\beta"] & \\
(Y,\Lambda,q) \arrow[dr, "\delta"] \arrow[ddr, bend right, "\phi"'] & & (Z,\Gamma,r) \arrow[dl, "\gamma"'] \arrow[ddl, bend left, "\psi"] \\
& (W,\Omega,\mu) \arrow[d, dashed, "\theta"] & \\
& (U,\Xi,\nu) &
\end{tikzcd}
\tag{13}
$$

where now all the maps, except θ, are assumed to be measure preserving. We need to show that θ is also measure preserving. Let $A \in \Xi$ be a measurable subset of U, we need to show $\nu(A) = \mu(\theta^{-1}(A))$. We calculate as follows

$$\nu(A) = q(\phi^{-1}(A)) = q(\delta^{-1}(\theta^{-1}(A))) = \mu(\theta^{-1}(A))$$

where the first equality holds because ϕ is measure preserving, the second from $\phi = \theta \circ \delta$ and the last because δ is measure preserving.

Finally, we have to construct kernels π_a on (W, Ω, μ) in such a way that δ and γ are **AMP** morphisms. We take $\pi_a = \eta(\tau_a)$. Thus, for all f in $L^+_\infty(W)$, we have $\pi_a(f) = \mathbb{E}_\infty(\eta)(\tau_a(f \circ \eta))$. Note that as $\mathbb{E}_\infty(-)$ is a functor and α is an arrow in **AMP**, we have $\pi_a(f) = \mathbb{E}_\infty(\delta)(\mathbb{E}_\infty(\alpha)(\tau_a((f \circ \delta) \circ \alpha))) = \mathbb{E}_\infty(\delta)(\rho_a(f \circ \delta)) = \delta(\rho_a)(f)$, and thus δ is an arrow in **AMP** as well. The same argument works for γ. Thus we have a commuting square in **AMP**.

To show that θ is an **AMP** morphism we calculate similarly. Let $h \in L^+_\infty(U, \Omega, \nu)$, then

$$
\begin{aligned}
\lambda_a(h) &= \phi(\rho_a)(h) \\
&= \mathbb{E}_\infty(\phi)(\rho_a(h \circ \phi)) \\
&= \mathbb{E}_\infty(\theta)(\mathbb{E}_\infty(\delta)(\rho_a((h \circ \theta) \circ \delta))) \\
&= \mathbb{E}_\infty(\theta)(\pi_a(h \circ \theta)) = \theta(\pi_a)(h).
\end{aligned}
$$

This completes the proof that we have pushouts in **AMP**.

We now need to show that if the morphisms α and β are zigzags then so are δ and γ. This requires some preliminary lemmas.

Lemma 8.2 *Let X be a set and (Y, Λ) be a measurable space. let $\alpha : X \longrightarrow Y$ be a surjective function and let $\Lambda' = \alpha^{-1}(\Lambda)$ be the induced σ-algebra on X. Then for all $h : X \longrightarrow \mathbb{R}$, h is Λ'-measurable if and only if h factors as $h' \circ \alpha$ for some measurable $h' : Y \longrightarrow \mathbb{R}$.*

Proof . The right to left direction is immediate since the definition of Λ' clearly makes α measurable and h' is assumed measurable. For the reverse direction we start with the claim that if $\alpha(x) = \alpha(x')$ for any x and x' in X then $h(x) = h(x')$. Consider the set $B = \{h(x)\}$, which is Borel-measurable. Since h is assumed measurable we have that $A = h^{-1}(B)$ is Λ'-measurable. By the definition of Λ', there is some C in Λ with $A = \alpha^{-1}(C)$. Now $x \in A$ so $\alpha(x) \in C$, but since $\alpha(x) = \alpha(x')$ we have that $x' \in A$ so $h(x') \in B$, i.e. $h(x) = h(x')$. This means that h is constant on subsets of X of the form $\alpha^{-1}(\{y\})$. Thus we can define $h' : Y \longrightarrow \mathbb{R}$ by $h'(y) = h(x)$ for any x in $\alpha^{-1}(y)$. This map clearly satisfies $h = h' \circ \alpha$. We need to show that h' is measurable. Let B be some Borel subset of $\mathbb{R}$ and let $A = h'^{-1}(B)$. Then $\alpha^{-1}(A) = h^{-1}(B)$ is in Λ' since h is Λ'-measurable, so $\alpha^{-1}(A) = \alpha^{-1}(C)$ for some $C \in \Lambda$, but since α is surjective we have that $A = C$ so $h'^{-1}(B) = C$ is in Λ, hence h' is measurable. ∎

Note that h and h' have the same image, and α is measure-preserving so if $h \in L^+_\infty(X)$ then $h' \in L^+_\infty(Y)$, in fact the essential sups coincide so we even have $\|h\|_\infty = \|h'\|_\infty$.

Lemma 8.3 *Let $\alpha : (X, \Sigma, p) \longrightarrow (Y, \Lambda, q)$ be a measure-preserving map of probability spaces. Then for all $h \in L^+_\infty(X)$, $\mathbb{E}_\infty(\alpha)(h) \circ \alpha = h \Leftrightarrow h$ is $\alpha^{-1}(\Lambda)$-measurable.*

Proof . We know that precomposition and conditional expectation functors compose to the identity if we have a measure preserving map, i.e. $\mathbb{E}_\infty(\alpha) \circ P_\infty(\alpha) = id$ if α is measure preserving. This follows from the remark just after Diagram 7 which in turn follows from Proposition 4.5. So the statement of the lemma is equivalent to saying that h is in the image of $P_\infty(\alpha)$ iff it is $\alpha^{-1}(\Lambda)$ measurable, but this is just what Lemma 8.2 says. ∎

Lemma 8.4 *Let $\alpha : (X, \Sigma, p, \tau_a) \longrightarrow (Y, \Lambda, q, \rho_a)$ be an arrow in* **AMP**. *Then α is a zigzag if and only if $P_\infty(\alpha) \circ \mathbb{E}_\infty(\alpha) = id$, i.e. if and only if for all $f \in L^+_\infty(Y)$, $\mathbb{E}_\alpha(\tau_a(f \circ \alpha)) \circ \alpha = \tau_a(f \circ \alpha)$.*

Proof . If α is a zigzag, the following diagram commutes:

$$\begin{tikzcd}
L^+_\infty(X) \arrow[r, "\tau_a"] & L^+_\infty(X) \arrow[d, "\mathbb{E}_\infty(\alpha)"] \\
L^+_\infty(Y) \arrow[u, "P_\infty(\alpha)"] \arrow[r, "\rho_a"] \arrow[d, "P_\infty(\alpha)"'] & L^+_\infty(Y) \arrow[d, "P_\infty(\alpha)"'] \\
L^+_\infty(X) \arrow[r, "\tau_a"] & L^+_\infty(X)
\end{tikzcd} \tag{14}$$

and the diagram shows the "only if part". The reverse direction is trivial, as $\mathbb{E}_\infty(\alpha)(\tau_a(f \circ \alpha)) = \rho_a(f)$ since α is an arrow in **AMP**. Thus $\rho_a(f) \circ \alpha = \tau_a(f \circ \alpha)$ and α is a zigzag. ∎

Corollary 8.5 *$\alpha : (X, \Sigma, p, \tau_a) \longrightarrow (Y, \Lambda, q, \rho_a)$ in* **AMP** *is a zigzag if and only if for all $f \in L^+_\infty(Y)$, $\tau(f \circ \alpha)$ is $\alpha^{-1}(\Lambda)$-measurable.*

Lemma 8.6 *If $\alpha : (X, \Sigma, p, \tau_a) \longrightarrow (Y, \Lambda, q, \rho_a)$ in* **AMP** *is a zigzag, $\beta : (Y, \Lambda, q, \rho_a) \longrightarrow (Z, \Gamma, r, \kappa_a)$ is a map in* **AMP***, and $\gamma = \beta \circ \alpha$ is a zigzag, then β is a zigzag.*

Proof .

$$\begin{aligned}
\kappa_a(f) \circ \beta \circ \alpha &= \kappa_a(f) \circ \gamma \\
&= \tau_a(f \circ \gamma) && \gamma \text{ is a zigzag} \\
&= \tau_a(f \circ \beta \circ \alpha) \\
&= \rho_a(f \circ \beta) \circ \alpha && \alpha \text{ is a zigzag}
\end{aligned}$$

Now α is surjective, hence epi, which means right-cancellable, and thus $\kappa_a(f) \circ \beta = \rho_a(f \circ \beta)$ and β is a zigzag. ∎

We are now ready to complete the proof of Theorem 8.1 by showing that δ and γ are zigzags. Let f be in $L^+_\infty(W)$, then we have

$$\begin{aligned}
\tau_a(f \circ \eta) &= \tau_a(f \circ \delta \circ \alpha) = \rho_a(f \circ \delta) \circ \alpha && \text{as } \alpha \text{ is a zigzag} \\
&= \tau_a(f \circ \gamma \circ \beta) = \kappa_a(f \circ \gamma) \circ \beta && \text{as } \beta \text{ is a zigzag.}
\end{aligned}$$

Let $\rho_a(f \circ \delta) = g$ and $\kappa_a(f \circ \gamma) = h$. We have the following diagram in **Mes**:

$$\begin{tikzcd}
 & (X,\Sigma) \arrow[dl, "\alpha"'] \arrow[dd, "\eta"] \arrow[dr, "\beta"] & \\
(Y,\Lambda) \arrow[dr, "\delta"] \arrow[ddr, "g"'] & & (Z,\Gamma) \arrow[dl, "\gamma"'] \arrow[ddl, "h"] \\
 & (W,\Omega) \arrow[d, dotted, "j"] & \\
 & \mathbb{R} &
\end{tikzcd} \tag{15}$$

As this is a pushout diagram, there is a unique measurable map $j : W \longrightarrow \mathbb{R}$ such that $g = j \circ \delta$ and $h = j \circ \gamma$. Thus $\tau_a(f \circ \eta) = g \circ \alpha = j \circ \delta \circ \alpha = j \circ \eta$. Thus $\tau_a(f \circ \eta)$ is $\eta^{-1}(\Omega)$ measurable and, from Corollary 8.5 we have that η is a zigzag. Now from Lemma 8.6 we conclude that δ and γ are zigzags. ∎

The main point of Theorem 8.1 is to show the following corollary.

Corollary 8.7 *Bisimulation is an equivalence relation on the objects of* **AMP**.

Proof . Clearly bisimulation is reflexive and symmetric, so we only need to check transitivity. We will label objects in **AMP** by just their state spaces to avoid cluttering up the diagram. Suppose X and Y are bisimilar, and that Y and Z are bisimilar. Then we have two cospans of zigzags, as in the following diagram:

$$\begin{tikzcd}
X \arrow[dr, "\alpha"'] & & Y \arrow[dl, "\beta"] \arrow[dr, "\delta"'] & & Z \arrow[dl, "\gamma"] \\
 & W \arrow[dr, dotted, "\zeta"'] & & U \arrow[dl, dotted, "\eta"] & \\
 & & V & &
\end{tikzcd} \tag{16}$$

The pushouts of the zigzags β and δ yield two more zigzags ζ and η (and the pushout object V). As the composition of two zigzags is a zigzag, X and Z are bisimilar. Thus bisimulation is transitive. ∎

It is worth noting that this proof did not require any assumptions about the nature of the measure spaces. In [DEP02], the proof of transitivity requires the σ-algebras of the measure spaces to be the Borel algebra of an analytic space. There are counter-examples showing that transitivity fails for the span definition of bisimulation on non-analytic spaces. We discuss this in the related work section.

Another point worth noting is that these pushouts exist in the category **AMP**, thus we can compose not just bisimulations, which are cospans of zigzags, but any cospans. In particular, this means that one can compose *approximate* bisimulations.

9 Minimal Realization

There is a very pleasing bisimulation-minimal realization theory for AMPs. Of course the notion of "minimal" cannot be based on counting the number of states, instead it is based on a suitable universal property. Given an AMP (X, Σ, p, τ_a), one may ask whether there is a "smallest" object in **AMP** up to bisimulation.

The precise definition is as follows.

Definition 9.1 *Given an AMP (X, Σ, p, τ_a), a* ***bisimulation-minimal realization*** *of this abstract Markov process is an AMP $(\tilde{X}, \Gamma, r, \pi_a)$ and a zigzag in* **AMP** *$\eta : X \longrightarrow \tilde{X}$ such that for every zigzag β from X to another AMP (Y, Λ, q, ρ_a), there is a unique zigzag γ from (Y, Λ, q, ρ_a) to $(\tilde{X}, \Gamma, r, \pi_a)$ with $\eta = \gamma \circ \beta$.*

If we think of a zigzag as defining a quotient of the original space then $\tilde{X}$ is the "most collapsed" version of X.

We now proceed to the proof that such an object exists for every AMP (X, Σ, p, τ_a).

Theorem 9.2 *Given any AMP (X, Σ, p, τ_a) there exists another AMP $(\tilde{X}, \Gamma, r, \pi_a)$ and a zigzag η in* **AMP***, $\eta : X \longrightarrow \tilde{X}$ such that $(\tilde{X}, \Gamma, r, \pi_a)$ and η define a bisimulation-minimal realization of (X, Σ, p, τ_a).*

Proof . We first note that the intersection of event bisimulations on (X, Σ, p, τ_a) (or any AMP) is again an event bisimulation so there is a well-defined *least* event bisimulation Ω. We define an equivalence relation R on X by xRx' if for every $A \in \Omega$, $x \in A \iff x' \in A$. We define the set $\tilde{X}$ as the quotient X/R. Let Q be the canonical surjection $Q : X \longrightarrow \tilde{X}$. We equip $\tilde{X}$ with a σ-algebra Γ, defined to be the finest (largest) σ-algebra making Q measurable; i.e. a subset C of $\tilde{X}$ is in Γ if and only if $Q^{-1}(C) \in \Omega$. We define the measure r by $\forall B \in \Gamma, r(B) = p(Q^{-1}(B))$; this makes Q a surjective, measurable, measure-preserving map.

We need to define π_a in such a way as to make Q a zigzag. This requires that $\forall h \in L_\infty^+(\tilde{X}), \tau_a(h \circ Q) = \pi_a(h) \circ Q$. Now $h \circ Q$ is constant on R-equivalence classes, by definition of Q; we claim that $\tau_a(h \circ Q)$ is also constant on R-equivalence classes. Since Ω is an event-bisimulation we know that $\tau_a(h \circ Q)$ is Ω-measurable. Let $x \in X$ and let $\tau_a(h \circ Q)(x) = u \in \mathbb{R}$. Then $(\tau_a(h \circ Q))^{-1}(u)$ is in Ω, call this set A; clearly $x \in A$. Suppose that xRx', then by the definition of R, $x' \in A$ so $(\tau_a(h \circ Q))(x') = u$; i.e. the claim is true. We can define $\forall w \in \tilde{X}, \pi_a(h)(w) = \tau_a(h \circ Q)(x)$ where x is such that $Q(x) = w$, this is well defined since Q is surjective and by virtue of the claim just proved. By construction, this establishes Q as a zigzag. The identity map of the underlying sets $\mathrm{id} : (X, \Sigma, p, \tau_a) \longrightarrow (X, \Omega, p \mid_\Omega, \tau_a)$ is a zigzag because Ω is an event bisimulation.

Now we claim that $\eta \stackrel{\text{def}}{=} Q \circ \mathrm{id}$ and $(\tilde{X}, \Gamma, r, \pi_a)$ is a minimal realization of (X, Σ, p, τ_a). Let $\beta : (X, \Sigma, p, \tau_a) \longrightarrow (Y, \Lambda, q, \rho_a)$ be a zigzag. We claim that if $\beta(x_1) = \beta(x_2)$ then $Q(x_1) = Q(x_2)$ for any x_1, x_2 in X. Since β is a zigzag, we have that $\beta^{-1}(\Lambda)$ is an event bisimulation and hence that $\Omega \subseteq \beta^{-1}(\Lambda)$. Now suppose that $\beta(x_1) = \beta(x_2)$, then there cannot be a set in $\beta^{-1}(\Lambda)$ that separates x_1 and x_2. Since $\Omega \subseteq \beta^{-1}(\Lambda)$ there cannot be a set in Ω that separates them either, hence $x_1 R x_2$ or $Q(x_1) = Q(x_2)$. Now we can define $\gamma(y)$ to be $Q(x)$, where x is an member of $\beta^{-1}(\{y\})$, this is well defined and surjective. Let A be a measurable set in Γ, $\gamma^{-1}(A) = \beta(Q^{-1}(A))$. Since Q is measurable, $Q^{-1}(A) \in \Omega$, hence $Q^{-1}(A) \in \beta^{-1}(\Lambda)$ from which it follows that $\beta(Q^{-1}(A))$ is in Λ, thus γ is measurable. Also for $A \in \Gamma$ we have

$$q(\gamma^{-1}(A)) = q(\beta(Q^{-1}(A))) = p(Q^{-1}(A)) = r(A)$$

hence γ is measure preserving. The first equality is by definition of γ, the second because β is a zigzag and the third because Q is a zigzag. Now from Lemma 8.6 it follows that γ is a zigzag. Clearly it is the only map that one could have defined to make the equation $\gamma \circ \beta = \eta$ hold. ∎

The minimal realization is unique up to isomorphism; this is an immediate consequence of the universal property.

Corollary 9.3 *Up to isomorphism,* $(\tilde{X}, \Gamma, r, \pi)$ *and* η *is the unique minimal realization of* (X, Σ, p, τ_a).

Another immediate corollary is that the minimal realization is terminal in an appropriate category.

Corollary 9.4 *The map η is the terminal object in the category where the objects are zigzags $\beta : (X, \Sigma, p, \tau_a) \longrightarrow (Y, \Lambda, q, \rho_a)$ from (X, Σ, p, τ_a) and a morphism from β to $\beta' : (X, \Sigma, p, \tau_a) \longrightarrow (Y', \Lambda', q', \rho'_a)$ is a zigzag $\gamma : (Y, \Lambda, q, \rho_a) \longrightarrow (Y', \Lambda', q', \rho'_a)$ such that $\beta' = \gamma \circ \beta$.*

A slight restatement of these is the following corollary.

Corollary 9.5 *If $\zeta : (\tilde{X}, \Gamma, r, \pi_a) \longrightarrow (W, \Xi, r, \lambda_a)$ is a zigzag then it is an isomorphism in* **AMP**.

Proof . The composed map $\zeta \circ \eta$ is a zigzag from X to W. Hence by the universal property of $(\tilde{X}, \eta)$ there is a unique map $\gamma : W \longrightarrow \tilde{X}$ such that $\gamma \circ (\zeta \circ \eta) = \eta$, hence, since η is an epi, $\gamma \circ \zeta = \mathrm{id}_{\tilde{X}}$. Now we also have $\zeta \circ (\gamma \circ \zeta) = (\zeta \circ \gamma) \circ \zeta) = \zeta$ and since ζ is an epi, we have $\zeta \circ \gamma = \mathrm{id}_W$. Thus ζ is an isomorphism in **AMP**. ∎

The most important consequence of the minimal realization theory is the following proposition that will be crucial in the approximation theory of Section 11.

Proposition 9.6 *Two AMPs (X, Σ, p, τ_a) and (Y, Λ, q, ρ_a) are bisimilar if and only if their minimal realizations $(\tilde{X}, \Gamma, r, \pi_a)$ and $(\tilde{Y}, \Delta, s, \theta_a)$ respectively are isomorphic.*

Proof . If $(\tilde{X}, \Gamma, r, \pi_a)$ and $(\tilde{Y}, \Delta, s, \theta_a)$ are isomorphic we immediately have the cospan

$$\begin{tikzcd} X \arrow[dr, "\eta"'] & & Y \arrow[dl, "\beta"] \\ & \tilde{X} & \end{tikzcd} \tag{17}$$

showing that X and Y are bisimilar. If X and Y are bisimilar we have the following diagram

$$\begin{tikzcd} X \arrow[dr, "\alpha"'] & & Y \arrow[dl, "\beta"] \\ & Z & \end{tikzcd} \tag{18}$$

where all the arrows are zigzags in **AMP**. Now consider the minimal realizations of X and Y, namely $\eta : X \longrightarrow (\tilde{X}, \Gamma, r, \pi_a)$ and $\xi : Y \longrightarrow (\tilde{Y}, \Delta, s, \theta_a)$ respectively. By the universality property for each one, we get zigzags $\gamma : Z$

$\longrightarrow \tilde{X}$ and $\delta : Z \longrightarrow \tilde{Y}$ such that $\gamma \circ \alpha = \eta$ and $\delta \circ \beta = \xi$ as shown in the diagram below.

$$\begin{tikzcd}
X \arrow[dr, "\alpha"] \arrow[dd, "\eta"'] & & Y \arrow[dl, "\beta"'] \arrow[dd, "\xi"] \\
& Z \arrow[dl, "\gamma"] \arrow[dr, "\delta"'] & \\
\tilde{X} \arrow[dr] & & \tilde{Y} \arrow[dl] \\
& W &
\end{tikzcd} \tag{19}$$

The span formed by $Z, \tilde{X}$ and $\tilde{Y}$ has a pushout with, say W, at the vertex. By Corollary 9.5 the maps from $\tilde{X}$ and $\tilde{Y}$ to W (to which we have not given explicit names) are both isos and hence $\tilde{X}$ and $\tilde{Y}$ are isomorphic.

∎

Here are two lemmas that are useful for the approximation theory of Section 11. The relation between event bisimulations and zigzags can be made precise now using a lemma proved in Section 8.

Lemma 9.7 *Suppose $\alpha : (X, \Sigma, p, \tau_a) \longrightarrow (Y, \Lambda, q, \rho_a)$ is a map in* **AMP** *such that $\alpha^{-1}(\Lambda) = \Sigma$. Then α is a zigzag.*

Proof . This is a direct consequence of corollary 8.5. Given f in $L^+_\infty(Y)$, $\tau(f \circ \alpha)$ is in $L^+_\infty(X)$ and thus is Σ-measurable. Hence it is $\alpha^{-1}(\Lambda)$-measurable, and so α is a zizag. ∎

Lemma 9.8 *Let $\alpha : (X, \Sigma, p, \tau_a) \longrightarrow (Y, \Lambda, q, \rho_a)$ be a zigzag. Then α factors into two maps as follows: $i_\alpha : (X, \Sigma, p, \tau_a) \longrightarrow (X, \alpha^{-1}(\Lambda), p, \tau_a)$, which is the identity on X, reducing the σ-algebra; and $\hat{\alpha} : (X, \alpha^{-1}(\Lambda), p, \tau_a) \longrightarrow (Y, \Lambda, q, \rho_a)$ which is the same as α above on the sets, but in which the σ-algebras are isomorphic.*

Proof . $\hat{\alpha}$ is a zigzag by virtue of the previous lemma; i_α is a zigzag by corollary 8.5. ∎

10 Logical characterization of bisimulation

One important consequence of the minimal realization theory is that one gets a logical characterization theorem for bisimulation. [DDLP06] showed that a simple modal logic gives a characterization of event bisimulation. This result can be presented in the framework of the present paper. We omit the proofs as they are all in [DDLP06]. As always we have some fixed set of actions $\mathcal{A}$.

Definition 10.1 *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:

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

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

Theorem 10.2 (From [DDLP06]) *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 event bisimulation is precisely the σ-field we obtain from the logic.

11 Approximations of AMPs

In this section we develop a theory of approximating AMPs using "finite" systems. In previous work [DGJP00, DGJP03] the idea was to collapse the state space to a finite set of equivalence classes. One could view the approximation construction as using an approximate version of bisimulation. Here we think of finite approximations in terms of finite σ-algebras. We have defined a category, **AMP** in which the maps defining bisimulation and the maps defining approximations are on the same footing: the viewpoint of the earlier papers pushed to its logical conclusion.

11.1 Preliminary lemmas

Before we begin, we need some elementary preliminary lemmas. The first one is a simple observation.

Lemma 11.1 *Suppose $\alpha : (X, \Sigma) \longrightarrow (Y, \Lambda)$ is a surjective measurable map such that $\alpha^{-1}(\Lambda) = \Sigma$. Then the forward image of every measurable set is measurable; that is, if $A \in \Sigma$, $\alpha(A) := B$ is measurable, and $\alpha^{-1}(B) = A$.*

Thus a surjective map which preserves the σ-algebras is an isomorphism of σ-algebras.

The next lemma gives a pushout diagram which we will need later in relating approximations and minimal realizations.

Lemma 11.2 *Suppose $\alpha : (X, \Sigma) \longrightarrow (Y, \Lambda)$ is surjective and $\alpha^{-1}(\Lambda) = \Sigma$. Suppose that $\Omega \subseteq \Sigma$ is a sub-σ-algebra of Σ. Then the following is a pushout square in the category* **Mes***:*

$$\begin{tikzcd}
& (X, \Sigma) \arrow[dl, "\alpha"'] \arrow[dr, "i_\Omega"] & \\
(Y, \Lambda) \arrow[dr, "j_{\alpha(\Lambda)}"'] & & (X, \Omega) \arrow[dl, "\hat{\alpha}"] \\
& (Y, \alpha(\Omega)) &
\end{tikzcd} \tag{20}$$

where $\alpha(\Omega) = \{B \subseteq Y \mid \alpha^{-1}(B) \in \Omega\}$ is a σ-algebra, i_Ω is the identity on X, $j_{\alpha(\Lambda)}$ is the identity on Y, and $\hat{\alpha}$ is the same as α on X.

Proof . For any α, $\alpha^{-1}(B^c) = (\alpha^{-1}(B))^c$, so $\alpha(\Omega)$ is closed under complements since Ω is. It is also easy to see that $\alpha(\Omega)$ is closed under countable intersections so $\alpha(\Omega)$ is a σ-algebra.

We know pushouts exist in **Mes**, so we need to show that this object satisfies the pushout conditions. Clearly, Y is the pushout in **Set**, with the maps described. In **Mes**, a pushout has the same underlying set as the corresponding pushout in **Set** equipped with the largest σ-algebra making the maps measurable. By the definition of $\alpha(\Omega)$ and the fact that α is measurable it follows that $\alpha(\Omega) \subseteq \Lambda$ hence the map $j_{\alpha(\Lambda)}$ is measurable and also that $\hat{\alpha}$ is measurable. Clearly if we added any measurable sets to the σ-algebra $\alpha(\Omega)$ the map $\hat{\alpha}$ would cease to be measurable since we have already included every set whose inverse image is in Ω.

To show the (co)universality property of pushouts in **Mes** we consider the following diagram

$$
\begin{tikzcd}
 & (X,\Sigma) \arrow[dl, "\alpha"'] \arrow[dr, "i_\Omega"] & \\
(Y,\Lambda) \arrow[dr, "i_{\alpha(\Lambda)}"] \arrow[ddr, bend right, "f"'] & & (X,\Omega) \arrow[dl, "\hat{\alpha}"'] \arrow[ddl, bend left, "g"] \\
 & (Y,\alpha(\Omega)) \arrow[d, dashed, "h"] & \\
 & (W,\Gamma) &
\end{tikzcd}
\tag{21}
$$

where the outer square commutes and the maps are all measurable. Clearly the only choice for h that can make the diagram commute is for $h = f$ as a set theoretic map. Now let $C \in \Gamma$, we need to show that $h^{-1}(C) \in \alpha(\Omega)$ in order to show that h is measurable. This follows from $\hat{\alpha}^{-1}(h^{-1}(C)) = g^{-1}(C) \in \Omega$. Thus in **Mes** there is a unique measurable mediating morphism h. ∎

11.2 Finite approximations

In this section we construct finite approximations of a LAMP by constructing first finite σ-algebras and then finite spaces from them.

Let (X, Σ, p, τ_a) be a LAMP. Let $\mathcal{P} = 0 < q_1 < q_2 < \ldots < q_k < 1$ be a finite partition of the unit interval with each q_i a rational number. We call these *rational partitions*. We define a family of finite π-systems, subsets of Σ, as follows:

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

where $\pi(\Omega)$ means the π-system generated by the family 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 $\Lambda_{\mathcal{P},M} = \sigma\left(\Phi_{\mathcal{P},M}\right)$, the σ-algebra generated by $\Phi_{\mathcal{P},M}$. We call each pair $(\mathcal{P}, M)$ consisting of a rational partition and a natural number an *approximation pair*.

The following result links the finite approximation with the formulas of the logic used in the characterization of bisimulation.

Proposition 11.3 *Given any labelled AMP (X, Σ, p, τ_a), the σ-algebra $\sigma\left(\bigcup \Phi_{\mathcal{P},M}\right)$, where the union is taken over all approximation pairs, is precisely the σ-algebra $\sigma\,[\![\mathcal{L}]\!]$ obtained from the logic.*

Proof . $\Phi_{\mathcal{P},M}$ contains precisely the measurable sets associated with formulas of length at most M, using rational numbers contained in $\mathcal{P}$, and so $\bigcup \Phi_{\mathcal{P},M} = [\![\mathcal{L}]\!]$. The conclusion is then clear. ∎

In order to describe the maps that arise it will be convenient to use the following notation. When $\Lambda \subseteq \Sigma$ are σ-algebras on a space X we have the measurable identity map $i : (X, \Sigma) \longrightarrow (X, \Lambda)$. If we have a LAMP τ_a on the space (X, Σ, p) we can define a LAMP on the space (X, Λ, p) as described in Diagram 4. We will write $\Lambda(\tau_a)$ rather than $i(\tau_a)$ since there will be many identity maps inducing LAMPs and it will not be helpful to label all the induced LAMPs with an i.

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

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

which is obtained from Diagram 4. Now since $\Lambda_{\mathcal{P},M}$ is finite, it is atomic, and so it partitions the 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 and AMPs ρ_a are defined in the obvious way, that is, q is the image measure through $\pi_{\mathcal{P},M}$ and $\rho_a = \pi_{\mathcal{P},M}(\Lambda_{\mathcal{P},M}(\tau_a))$. Note that $\pi_{\mathcal{P},M}$ is a zigzag as ${\pi_{\mathcal{P},M}}^{-1}(\Omega) = \Lambda_{\mathcal{P},M}$.

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**. When we collapse the space X to one of the quotient spaces, say $\hat{X}_{\mathcal{P},M}$ the map $\phi_{\mathcal{P},M}$ induces a projected version of the LAMP τ_a which we denote as usual as $\phi_{\mathcal{P},M}(\tau_a)$.

11.3 A Projective System of Finite Approximations

We 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. This poset is a directed set:

given $(\mathcal{P}, M)$ and $(\mathcal{Q}, N)$ two approximation pairs, then the approximation pair $(\mathcal{P} \cup \mathcal{Q}, L)$ is an upper bound, where L is $\max(M, N)$.

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 $\Lambda_{\mathcal{P},M} \subseteq \Lambda_{\mathcal{Q},N} \subseteq \Sigma$. The fact that it is an arrow in the category is clear from the functoriality of conditional expectation. Furthermore if $(\mathcal{P}, M) \leq (\mathcal{Q}, N) \leq (\mathcal{R}, K)$ the maps compose to give

$$i_{(\mathcal{R},K),(\mathcal{P},M)} = i_{(\mathcal{R},K),(\mathcal{Q},N)} \circ i_{(\mathcal{Q},N),(\mathcal{P},M)}.$$

This also follows from funcotiality. In short we have a projective system of such maps indexed by our poset of approximation pairs.

We can induce maps between the approximation spaces as follows. Recall that an element of $\hat{X}_{\mathcal{P},M}$ is an equivalence class of X where two points are equivalent if no sets in the σ-algebra separate them. If $(\mathcal{P}, M)$ is refined by $(\mathcal{Q}, N)$ then the σ-algebra $\Lambda_{\mathcal{P},M}$ is refined by the σ-algebra $\Lambda_{\mathcal{Q},N}$ hence an equivalence class represented by an element of $\hat{X}_{\mathcal{Q},N}$ is contained in a unique equivalence class represented by an element of $\hat{X}_{\mathcal{P},M}$; this correspondence defines a map $j_{(\mathcal{Q},N),(\mathcal{P},M)} : (\hat{X}_{\mathcal{Q},N}, p, \phi_{\mathcal{Q},N}(\tau_a)) \longrightarrow (\hat{X}_{\mathcal{P},M}, p, \phi_{\mathcal{P},M}(\tau_a))$ such that the following commutes:

$$\begin{tikzcd}
(X, \Lambda_{\mathcal{Q},N}, \Lambda_{\mathcal{Q},N}(\tau_a)) \arrow[r, "i_{(\mathcal{Q},N),(\mathcal{P},M)}"] \arrow[d, "\pi_{\mathcal{Q},N}"'] & (X, \Lambda_{\mathcal{P},M}, \Lambda_{\mathcal{P},M}(\tau_a)) \arrow[d, "\pi_{\mathcal{P},M}"] \\
(\hat{X}_{\mathcal{Q},N}, \phi_{\mathcal{Q},N}(\tau_a)) \arrow[r, "j_{(\mathcal{Q},N),(\mathcal{P},M)}"'] & (\hat{X}_{\mathcal{P},M}, \phi_{\mathcal{P},M}(\tau_a))
\end{tikzcd} \tag{22}$$

Hence, the maps $j_{(\mathcal{Q},N),(\mathcal{P},M)}$ along with the approximants $\hat{X}_{(\mathcal{P},M)}$ also form a projective system with respect to our poset of approximation pairs. In addition, the approximation map $\phi_{(\mathcal{P},M)}$ factors through the approximation map $\phi_{(\mathcal{Q},N)}$ as $\phi_{(\mathcal{P},M)} = j_{(\mathcal{Q},N),(\mathcal{P},M)} \circ \phi_{(\mathcal{Q},N)}$ so that maps $\phi_{\mathcal{P},M}$ form a cone above the projective system.

One can understand this functorially as follows. Given a measurable space (X, Σ) one can define an induced equivalence relation R by xRx' if for every measurable set B $x \in B \iff x' \in B$; this is the same equivalence relation that was introduced in the proof of the minimal realization theorem. It might be the case that R is the identity relation, for example this happens with the Borel algebra on the real line. In this case one says that Σ separates points. In any case, the quotient $X \longrightarrow X/R$ is actually an endofunctor on **Mes**. To see this consider a measurable function $f : (X, \Sigma) \longrightarrow (Y, \Lambda)$ and let the equivalence relations induced by Σ and Λ be R and T respectively. Then we can define the map $\hat{f} : X/R \longrightarrow Y/T$ by $\hat{f}([x]_R) = [f(x)]_T$; this is easily seen to be well-defined and measurable using arguments similar to the ones in the proof of Theorem 9.2. The preservation of composition is clear so we are entitled to call this functor $\mathbf{F} : \mathbf{Mes} \longrightarrow \mathbf{Mes}$. The statements in the paragraph above assert that $\mathbf{F}$ preserves projective diagrams. Later we will show that $\mathbf{F}$ preserves projective limits.

11.4 Existence of the Projective Limit

The existence of projective limits of our family of approximants rests on a result of Choksi [Cho58]; we need to be careful about exactly which category we are talking about however. The following proposition is from his paper. In stating his result we skip any mention of the LAMPs for the moment. A *topological measure space* is a topological space where the σ-algebra is induced by the open sets of the topology. A compact Hausdorff topological measure space is simply one where the topology is compact Hausdorff.

Proposition 11.4 *Suppose that we have a projective system of compact Hausdorff topological measure spaces* (X_i, Λ_i, p_i) *with measurable measure preserving maps* $\phi_{ji} : X_j \longrightarrow X_i$. *There is a topological measure space* $(X_\infty, \Gamma, \gamma)$ *also compact Hausdorff, and maps* $\psi_i : X_\infty \longrightarrow X_i$ *that are also measurable and measure preserving such that the entire diagram formed by the* ϕ*s and* ψ*s commutes.*

In his work, as was typical for analysis at the time, there is no proof that this "limit" object satisfies any kind of universal property. The finite approximants to the measure space underlying a LAMP have a projective limit in the category $\mathbf{Rad}_=$; recall this is the category where the objects are measure spaces and the morphisms are measurable and measure preserving maps. We will consider the LAMPs later.

Theorem 11.5 *The probability spaces of finite approximants* $\hat{X}_{\mathcal{P},M}$ *of a measure space* (X, Σ, p, τ_a) *each equipped with the discrete* σ*-algebra (i.e. the* σ*-algebra of all subsets) indexed by the approximation pairs, form a projective system in the category* $\mathbf{Rad}_=$. *This system of finite approximants* *to the LAMP* (X, Σ, p, τ_a) *has a projective limit in the category* $\mathbf{Rad}_=$.

Proof . The situation is shown in the diagram below:

$$
\begin{tikzcd}
 & (Y, \Xi, r) \arrow[ddl, "f_{\mathcal{P},M}"'] \arrow[ddr, "f_{\mathcal{Q},N}"] \arrow[d, dashed, "\lambda"] & \\
 & (\operatorname{proj\,lim} \hat{X}, \Gamma, \gamma) \arrow[dl, "\psi_{\mathcal{P},M}"'] \arrow[dr, "\psi_{\mathcal{Q},N}"] & \\
\hat{X}_{\mathcal{P},M} & & \hat{X}_{\mathcal{Q},N} \arrow[ll, "j_{(\mathcal{Q},N),(\mathcal{P},M)}"] \\
(X, \Lambda_{\mathcal{P},M}, p, \Lambda_{\mathcal{P},M}(\tau_a)) \arrow[u, "\pi_{\mathcal{P},M}"] & & (X, \Lambda_{\mathcal{Q},N}, p, \Lambda_{\mathcal{Q},N}(\tau_a)) \arrow[ll, "i_{(\mathcal{Q},N),(\mathcal{P},M)}"] \arrow[u, "\pi_{\mathcal{Q},M}"']
\end{tikzcd}
\tag{23}
$$

In order to make the diagram fit on the page we have written $\hat{X}_{\mathcal{P},M}$ instead of $(\hat{X}_{\mathcal{P},M}, \Omega_{\mathcal{P},M}, \hat{p}_{\mathcal{P},M}, \pi_{\mathcal{P},M}(\Lambda_{\mathcal{P},M}(\tau_a)))$ and $\hat{X}_{\mathcal{Q},N}$ instead of $(\hat{X}_{\mathcal{Q},N}, \Omega_{\mathcal{Q},N}, \hat{p}_{\mathcal{Q},N}, \pi_{\mathcal{Q},N}(\Lambda_{\mathcal{Q},N}(\tau_a)))$.

The spaces $(X, \Lambda_{\mathcal{P},M}, p, i_{\Lambda_{\mathcal{P},M}}(\tau_a))$ are only shown to remind the reader where the finite approximants come from; they are not part of the projective diagram whose limit we are taking. The measure space (Y, Ξ, r) is any[7] measure space and the family of maps $f_{\mathcal{P},M}$ are assumed to be measurable and measure preserving. Note that we are not claiming the existence of a projective limit in **AMP**. For this reason we consider only a measure space and show that we have a unique mediating morphism λ which is measurable and measure preserving.

The projective limit in **Mes** is constructed from the projective limit in **Set** in much the same way as pushouts in **Set** can be made into pushouts in **Mes**. Concretely, $\operatorname{proj\,lim} \hat{X}$ is the projective limit in **Set** – that is, the subset of the product $\prod \hat{X}_{\mathcal{P},M}$ which is compatible with the maps $j_{(\mathcal{Q},N),(\mathcal{P},M)}$ of the projective system. We have the usual projection maps in **Set** $\psi_{\mathcal{P},M} : \operatorname{proj\,lim} \hat{X} \longrightarrow \hat{X}_{\mathcal{P},M}$ for every approximation pair. The spaces $\hat{X}_{\mathcal{P},M}$ are finite sets equipped with the discrete σ-algebra. They can be viewed as topological measure spaces with the discrete topology, which, of course, generates the discrete σ-algebra. Viewed as such these finite approximants are compact Hausdorff spaces and Choksi's Theorem 11.4 applies, so we get a σ-algebra Γ and measure γ which makes the ψs $\mathbf{Rad}_=$ morphisms.

[7]Recall that all the measures are finite in this paper.

The σ-algebra Γ is the smallest σ-algebra that makes the ψ's measurable. If (Y, Ξ) is a measurable space and $f_{\mathcal{P},M}$ is a family of measurable maps from Y to $(\hat{X}_{\mathcal{P},M}, \Omega_{\mathcal{P},M})$ there is a measurable function $\lambda : Y \longrightarrow \operatorname{proj\,lim} \hat{X}$ making the diagram commute. To see this we use the same λ that one obtains in **Set** from the universality of the projective limit in **Set**. Γ is generated by sets of the form $\psi^{-1}_{\mathcal{P},M}(\{x\})$ where x is an element of $\hat{X}_{\mathcal{P},M}$. In order to check that a map is measurable it suffices to check that the inverse image of a set in the generating family of the σ-algebra is measurable. Thus we need to check that $\lambda^{-1}(A)$ is in Ξ for any set of the form $A = \psi^{-1}_{\mathcal{P},M}(\{x\})$. Now we can write $\lambda^{-1}(A)$ as

$$\lambda^{-1}(\psi^{-1}_{\mathcal{P},M}(\{x\})) = f^{-1}_{\mathcal{P},M}(\{x\})$$

which is in Ξ because the fs are measurable.

Now we know that λ is measurable, we need to show that it is measure preserving. The collection of sets of the form $\psi^{-1}_{\mathcal{P},M}(A_{\mathcal{P},M})$, where each $A_{\mathcal{P},M}$ is a measurable subset of $\hat{X}_{\mathcal{P},M}$[8] generates the σ-algebra Γ; we use Δ to refer to this collection of subsets of $\operatorname{proj\,lim} \hat{X}$. We claim that Δ forms a π-system of sets. Accordingly we only need to check that λ preserves the measures of these sets to conclude that it is measure preserving. To establish the claim it suffices to show that the intersection of two sets of the form $\psi^{-1}_{\mathcal{P},M}(x)$ is in Δ. Consider $\psi^{-1}_{\mathcal{P},M}(x)$ and $\psi^{-1}_{\mathcal{Q},N}(y)$. Because we have a projective system we have some $(\mathcal{K}, K)$ such that $(\mathcal{Q}, N), (\mathcal{P}, M) \leq (\mathcal{K}, K)$; of course $(\mathcal{K}, K)$ could be one of $(\mathcal{P}, M))$ or $(\mathcal{Q}, N)$ but that is a special case. For brevity we temporarily write m, n, k for the subscripts $(\mathcal{P}, M), (\mathcal{Q}, N)$ and $(\mathcal{K}, K)$ respectively. Now the maps j_{km} and j_{kn} are surjective. Define $B = j^{-1}_{km}(\{x\}) \cap j^{-1}_{kn}(\{y\})$. Now since the entire diagram commutes we have

$$\begin{aligned}
\psi_k^{-1}(B) &= \psi_k^{-1}(j^{-1}_{km}(\{x\}) \cap j^{-1}_{kn}(\{y\}) \\
&= (\psi_k^{-1}((j^{-1}_{km}(\{x\})) \cap \psi_k^{-1}(j^{-1}_{kn}(\{y\})) \\
&= \psi_m^{-1}(\{x\}) \cap \psi_n^{-1}(\{y\}).
\end{aligned}$$

We have shown that Δ a π-system.

Now a set in Δ looks like $\psi^{-1}_{\mathcal{P},M}(A_{\mathcal{P},M})$. Let the elements of $A_{\mathcal{P},M}$ be $\{x_1, \ldots, x_k\}$. The sets $\psi^{-1}_{\mathcal{P},M}(\{x_i\})$ for $i = 1, \ldots, k$ are all disjoint. Consider any one of these, say x_i. We have

$$r(\lambda^{-1}(\psi^{-1}_{\mathcal{P},M}(\{x_i\}))) = r(f^{-1}_{\mathcal{P},M}(\{x_i\})) = \hat{p}_{\Lambda_{\mathcal{P},M}}(\{x_i\}) = \gamma(\psi^{-1}_{\mathcal{P},M}(\{x_i\}))$$

[8] Of course this just means any subset.

where the second equality holds because the fs are assumed to be measure preserving and the last because the ψs are measure preserving. Thus λ is measure preserving on sets of this form. But a generic set in Δ is the disjoint union of sets like this so we have

$$
\begin{aligned}
\gamma(\psi^{-1}_{\mathcal{P},M}(A_{\mathcal{P},M})) &= \sum_{i=1}^{k} \gamma(\psi^{-1}_{\mathcal{P},M}(\{x_i\}) \\
&= \sum r(\lambda^{-1}(\psi^{-1}_{\mathcal{P},M}(\{x_i\}))) \\
&= r(\lambda^{-1}(\psi^{-1}_{\mathcal{P},M}(A_{\mathcal{P},M}))).
\end{aligned}
$$

Thus the measures $r \circ \lambda^{-1}$ and γ agree on the sets of Δ. Since Δ is a π-system generating Γ, the two measures agree on all of Γ by Prop. 2.3. We have completed the proof of the universal property in $\mathbf{Rad}_=$. ∎

We can now consider the LAMP structure. We do not get a universal property in the category **AMP**, however, the universality of the construction in $\mathbf{Rad}_=$ almost forces the structure of a LAMP on the projective limit constructed in $\mathbf{Rad}_=$.

Proposition 11.6 *A LAMP can be defined on the projective limit constructed in* $\mathbf{Rad}_=$ *so that the cone formed by this limit object and the maps to the finite approximants yields a commuting diagram in the category* **AMP**.

Proof . We can define the LAMP ζ_a on $\mathrm{proj}\lim \hat{X}$ as follows. We recall that we get a cone over the projective system of finite approximants from the LAMP (X, Σ, p, τ_a) with which we started as shown in the diagram below:

$$
\begin{tikzcd}
 & (X, \Sigma, p, \tau_a) \arrow[ddl, "\phi_{\mathcal{P},M}"'] \arrow[ddr, "\phi_{\mathcal{Q},N}"] \arrow[d, dashed, "\kappa"] & \\
 & (\mathrm{proj}\lim \hat{X}, \Gamma, \gamma, \zeta_a) \arrow[dl, "\psi_{\mathcal{P},M}"'] \arrow[dr, "\psi_{\mathcal{Q},N}"] & \\
\hat{X}_{\mathcal{P},M} & & \hat{X}_{\mathcal{Q},N} \arrow[ll, "j_{(\mathcal{Q},N),(\mathcal{P},M)}"] \\
(X, \Lambda_{\mathcal{P},M}, p, i_{\Lambda_{\mathcal{P},M}}(\tau_a)) \arrow[u, "\pi_{\mathcal{P},M}"] & & (X, \Lambda_{\mathcal{Q},N}, p, i_{\Lambda_{\mathcal{Q},N}}(\tau_a)) \arrow[u, "\pi_{\mathcal{Q},N}"'] \arrow[ll]
\end{tikzcd}
\tag{24}
$$

From universality in $\mathbf{Rad}_=$ we have a unique map $\kappa : X \longrightarrow \mathrm{proj}\lim \hat{X}$ such that $\psi_{\mathcal{P},M} \circ \kappa = \phi_{\mathcal{P},M}$, i.e., the approximation maps from X factor through κ. This κ is measurable and measure preserving being a $\mathbf{Rad}_=$ morphism.

We define the AMP ζ_a on $\mathrm{proj}\lim \hat{X}$ 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 equal to $\pi_{\mathcal{P},M}(i_{\Lambda_{\mathcal{P},M}}(\tau_a))$ since $\psi_{\mathcal{P},M} \circ \kappa = \phi_{\mathcal{P},M}$. This shows that the diagram formed by the projective limit, the finite approximants and the maps $\psi_{\mathcal{P},M}$ and the $j_{(\mathcal{Q},N),(\mathcal{P},M)}$ form a commuting diagram in **AMP**. ∎

Note that the finite approximants coming from the logic do not play a special role here. If we had used any other family of finite approximants we would still construct some kind of limit which would itself be an approximant. The special properties of the approximants that we are using comes across in the next subsection.

11.5 Minimal realization and Finite Approximation

The main result in this section is that the LAMP obtained by forming the projective limit in the category $\mathbf{Rad}_=$ and then defining a LAMP on it is isomorphic to the minimal realization of the original LAMP. This gives a very pleasing connection between the approximation process and the minimal realization.

Theorem 11.7 *Given an AMP (X, Σ, p, τ_a), the projective limit of its finite approximants* $(\mathrm{proj}\lim \hat{X}, \Gamma, \gamma, \zeta_a)$ *is isomorphic to its minimal realization* $(\tilde{X}, \Xi, r, \xi_a)$.

In order to prove this we need some preliminary results. It already follows from universality that κ is measurable, but we can show something slightly stronger.

Proposition 11.8 *The σ-algebra $\kappa^{-1}(\Gamma)$ is precisely equal to $\sigma\, [\![\mathcal{L}]\!]$; in particular κ is measurable.*

Proof . The σ-algebra Γ is the generated by the inverse images of $\psi_{\mathcal{P},M}$; letting $\Omega_{\mathcal{P},M}$ be the σ-algebra on $\hat{X}_{\mathcal{P},M}$, we have $\Gamma = \sigma(\bigcup {\psi_{\mathcal{P},M}}^{-1}(\Omega_{\mathcal{P},M}))$, where the union is over all approximation pairs. Now we know that

$$\psi_{\mathcal{P},M} \circ \kappa = \phi_{\mathcal{P},M} = \pi_{\mathcal{P},M} \circ i_{\Lambda_{\mathcal{P},M}}.$$

Since preimages preserve intersection, union and complement we have,

$$\begin{aligned}
\kappa^{-1}(\Gamma) &= \kappa^{-1}\left(\sigma\left(\bigcup {\psi_{\mathcal{P},M}}^{-1}\left(\Omega_{\mathcal{P},M}\right)\right)\right)\\
&= \sigma\left(\bigcup\left(\kappa^{-1}\left({\psi_{\mathcal{P},M}}^{-1}\left(\Omega_{\mathcal{P},M}\right)\right)\right)\right)\\
&= \sigma\left(\bigcup\left({i_{\Lambda_{\mathcal{P},M}}}^{-1}\left({\pi_{\mathcal{P},M}}^{-1}\left(\Omega_{\mathcal{P},M}\right)\right)\right)\right)\\
&= \sigma\left(\bigcup\left({i_{\Lambda_{\mathcal{P},M}}}^{-1}\left(\Lambda_{\mathcal{P},M}\right)\right)\right)\\
&= \sigma\left(\bigcup \Lambda_{\mathcal{P},M}\right)\\
&= \sigma\left([\![\mathcal{L}]\!]\right)
\end{aligned}$$

where the last step is justified by Proposition 11.3. Note that $\sigma([\![\mathcal{L}]\!])$ is indeed a sub-σ-algebra of Σ as can easily be shown by induction on the structure of formulas. ∎

Proposition 11.9 *The map $\kappa : (X, \Sigma, p, \tau_a) \longrightarrow (\text{proj}\lim \hat{X}.\Gamma, \gamma, \zeta_a)$ obtained from the projective limit diagram is a zigzag in* **AMP**.

Proof . As $\kappa^{-1}(\Gamma) = \sigma([\![\mathcal{L}]\!])$, we can factor κ as $\hat{\kappa} \circ i_\kappa$, where

$$\begin{aligned}
&i_\kappa : (X, \Sigma, p, \tau_a) \longrightarrow (X, \sigma([\![\mathcal{L}]\!]), p, \tau_a)\\
&\hat{\kappa} : (X, \sigma([\![\mathcal{L}]\!]), p, \tau_a) \longrightarrow (\text{proj}\lim \hat{X}, \Gamma, \gamma, \zeta_a)
\end{aligned}$$

i_κ is a zigzag as $\sigma([\![\mathcal{L}]\!])$ is an event bisimulation; $\hat{\kappa}$ is a zigzag by Lemma 9.7. Thus κ is a zigzag. ∎

If we let $(\tilde{X}, \Xi, r, \xi_a)$ be the minimal realization obtained as in proposition 9.2, we have a zigzag $\omega : (\text{proj}\lim \hat{X}, \Gamma, \gamma, \zeta_a) \longrightarrow (\tilde{X}, \Xi, r, \xi_a)$ from Corollary 9.4. The proof of Theorem 11.7 will establish that there is a zigzag in the other direction.

Proof (of Theorem 11.7). As X and $\tilde{X}$ are bisimilar, they have the same approximants, and thus the projective limits of these approximants $(\text{proj}\lim \hat{X}, \Gamma, \gamma, \zeta_a)$ is the same. Therefore, by Proposition 11.9 there is a zigzag $\epsilon : (\tilde{X}, \Xi, r, \xi_a)$ $\longrightarrow (\text{proj}\lim \hat{X}, \Gamma, \gamma, \zeta_a)$. Hence, by Corollary 9.3, ϵ is an isomorphism of AMPs. ∎

There are a number of other facts that show that the approximations capture something that is intrinsic to bisimulation equivalent LAMPS.

Theorem 11.10 *Let $\alpha : (X, \Sigma, p, \tau_a) \longrightarrow (Y, \Theta, q, \rho_a)$ be a zigzag. Then these two LAMPs have the same finite approximants.*

Corollary 11.11 *Two bisimilar AMPs have the same finite approximants.*

In order to prove Theorem 11.10 we need some preliminary lemmas.

Lemma 11.12 *Let $\alpha : (X, \Sigma, p, \tau_a) \longrightarrow (Y, \Theta, q, \rho_a)$ be a zigzag. Let $A \in \Theta$ and q be a rational number. Then*

$$\alpha^{-1}\left(\{y : \rho_a\left(\mathbf{1}_A\right)(y) > q\}\right) = \left\{x : \tau_a\left(\mathbf{1}_{\alpha^{-1}(A)}\right) > q\right\}$$

Proof .

$$\begin{aligned}
\alpha^{-1}\left(\{y : \rho_a\left(\mathbf{1}_A\right)(y) > q\}\right) &= \alpha^{-1}\left(\rho_a\left(\mathbf{1}_A\right)^{-1}(q, 1]\right) \\
&= \left(\rho_a\left(\mathbf{1}_A\right) \circ \alpha\right)^{-1}(q, 1] \\
&= \left(\tau_a\left(\mathbf{1}_A \circ \alpha\right)\right)^{-1}(q, 1] \\
&= \left(\tau_a\left(\mathbf{1}_{\alpha^{-1}(A)}\right)\right)^{-1}(q, 1] \\
&= \left\{x : \tau_a\left(\mathbf{1}_{\alpha^{-1}(A)}\right) > q\right\}
\end{aligned}$$

∎

Lemma 11.13 *Let (X, Σ, p, τ_a) be a labelled AMP and $\Omega \subseteq \Sigma$ be an event-bisimulation. Then (X, Ω, p, τ_a) and (X, Σ, p, τ_a) have the same finite approximants.*

Proof . The finite σ-algebras $\Lambda_{\mathcal{P},M}$ yielding the approximants are sub-σ-algebras of $\sigma\left([\![\mathcal{L}]\!]\right)$. As $\sigma\left([\![\mathcal{L}]\!]\right)$ is the smallest event-bisimulation, we have the inclusion

$$\Lambda_{\mathcal{P},M} \subseteq \sigma\left([\![\mathcal{L}]\!]\right) \subseteq \Omega \subseteq \Sigma$$

and so the approximation maps from (X, Σ, p, τ_a) factor through the approximation maps from (X, Ω, p, τ_a) ∎

Proof of proposition 11.10. Consider the following diagram of LAMPs:

$$\begin{tikzcd}
(X, \Sigma, p, \tau_a) \arrow[r, "\alpha"] \arrow[d, "i_{\Lambda_{\mathcal{P},M}}"'] & (Y, \Theta, q, \rho_a) \arrow[d, "j_{\alpha(\Lambda)}"] \\
(X, \Lambda_{\mathcal{P},M}) \arrow[r, "\hat{\alpha}"] \arrow[d, "\pi^X_{\mathcal{P},M}"'] & (Y, \alpha(\Lambda_{\mathcal{P},M})) \arrow[dl, "\pi^Y_{\mathcal{P},M}"] \\
\hat{X}_{\mathcal{P},M} &
\end{tikzcd} \tag{25}$$

The measures and LAMPS on the approximants are defined in the manner described in the approximation construction described above. We suppress explicit mention of them to make the diagram less cluttered. By

Lemma 11.13 and the factoring property of zigzags (by Lemma 9.8), we need only verify our claim on a zigzag $\alpha : (X, \Sigma, p, \tau_a) \longrightarrow (Y, \Theta, q, \rho_a)$ such that $\alpha^{-1}(\Theta) = \Sigma$. By Lemma 11.1, α is an isomorphism of σ-algebras. Let $\Lambda_{\mathcal{P},M} \subseteq \Sigma$ be an approximating σ-algebra on X.

By Lemma 11.2, the upper square in Diagram 25 commutes and is a pushout.

Note that $\alpha\left(\Lambda_{\mathcal{P},M}\right)$ is precisely the approximating σ-algebra obtained on Y by the approximation pair $(\mathcal{P}, M)$. This follows from Lemma 11.12 as expressions of the form $(\{y : \rho_a\left(\mathbf{1}_A\right)(y) > q\})$ generate the approximating σ-algebras. This shows that the right hand side of the diagram in indeed part of the approximation of (Y, Θ, q, ρ_a).

Finally, the quotienting map $\pi^X_{\mathcal{P},M}$ reducing the measure space $(X, \Lambda_{\mathcal{P},M})$ to a finite state space factors through the similar map from Y, $\pi^Y_{\mathcal{P},M}$, as α is surjective. This factorization extends to LAMPs, and so the bottom triangle of the above diagram commutes; thus the two original LAMPs (X, Σ, p, τ_a) and (Y, Θ, q, ρ_a) have the same finite approximations. ∎

12 Related Work

12.1 History of labelled Markov processes

We review the history of the theory of labelled Markov processes as described in the recent expository book [Pan09]. It is not necessary to read this section to follow the technical development of the present paper. Some of the points made here are repeated in the main text in order that a reader can read the rest of the paper without having to read this section.

The earliest work on incorporating probability in the theory of verification of transition systems is Vardi's work on concurrent Markov chains [Var85]. This is aimed at adapting techniques like model checking developed for finite transition systems to the probabilistic situation. The theory of bisimulation for probabilistic systems was initiated by [LS91] who described a modal logic for characterizing probabilistic bisimulation and explored the relation with testing. This prescient paper began the modern era of exploration of the field. There is a significant literature exploring variants like weak bisimulation and real-time systems: all this was done for discrete transition systems. A good review of model checking for discrete probabilistic systems appears in Chapter 12 of the recent text book *Principles of Model Checking* [BK08].

The theory was extended to continuous state spaces by [BDEP97] and by [dVR97]. The latter worked on ultrametric spaces and used the machinery of ultrametric spaces to show that bisimulation – defined in terms of spans – is transitive. In our opinion, ultrametric spaces are not at all like the continuous spaces that arise in physical systems: they are totally disconnected, for example. However, that work did emphasize the coalgebraic nature of the theory and that was a very important step.

The work begun in [BDEP97] was elaborated in [DEP02] and later papers [DGJP03, DD03, DDP03, DGJP04] where theories of approximation and of metrics were developed. Much of the work of [DEP02] was reworked by Doberkat in a series of papers that use powerful tools from descriptive set theory to put the theory in a more elegant, general and pleasing form. This work appeared in several papers and are summarized in two recent books [Dob07, Dob10].

12.1.1 Labelled Markov processes and bisimulation

There are two main approaches to bisimulation, and they are closely linked. The first is to equate *states*, that is, to determine which states behave the same with respect to the user. Loosely speaking, two states are bisimilar if they indistinguishable from the user's perspective. The other approach is to equate LMPs among themselves. In this higher level point of view, two LMPs are bisimilar if each state in one is bisimilar to a state in the other; or, in other words, if the two LMPs contain states which have the same behaviour. Note that we shall always assume that when speaking of bisimulation between different LMPs, the action set $\mathcal{A}$ will be fixed.

For each of these points of view, different definitions of bisimulation have been postulated. We review these briefly, following [DDLP06].

LMPs are the coalgebras of a monad, essentially discovered by Lawvere and discussed in detail by [Gir81]. The notion of zigzag that we have used comes from there, it is exactly the homomorphism notion for the coalgebras of Giry's monad [RdV97, dVR99, DEP02].

Generally speaking, a morphism f from a LMP (X, Σ, τ_a) to another (Y, Λ, ρ_a) is a measurable map of the underlying measurable spaces, which is assumed to respect some compatibility condition relative to the Markov kernels. The idea of a zigzag morphism is that we should be able to specify a condition on f which would imply that the two LMPs are bisimilar. Specifically, we have the following definition:

Definition 12.1 *A* ***zigzag morphism*** *from a LMP* (X, Σ, τ_a) *to another* (Y, Λ, ρ_a) *is a surjective measurable map* $f : (X, \Sigma) \longrightarrow (Y, \Lambda)$ *such that, for all* $a \in \mathcal{A}$, $x \in X$, $B \in \Lambda$,

$$\tau_a\left(x, f^{-1}(B)\right) = \rho_a\left(f(x), B\right)$$

Hence, the transition probabilities are essentially the same in both systems. However, information is still lost across a zigzag morphism. This loss is twofold; first, as the map is surjective (but not necessarily injective), different points in the domain space are sent to the same point in the target space and thus equated. Secondly, as f is measurable, we have that $f^{-1}(\Lambda) \subseteq \Sigma$, and thus the complexity of the σ-algebra may decrease. Nevertheless, note that since $\rho_a(y, B)$ must be a Λ-measurable function for a fixed set B, Λ cannot be trivial. Following the notion of bisimulation via open maps [JNW93], [DEP02] defined two LMPs to be bisimilar if there exists a *span* of zigzags between them.

Definition 12.2 *Two LMPs* (X, Σ, τ_a) *and* (Y, Λ, ρ_a) *are* ***bisimilar*** *if there exists a LMP* (U, Ω, σ_a) *such that there is a zigzag morphism* f *from* U *to* X *and another zigzag morphism* g *from* U *to* Y.

As the identity map from a LMP to itself is trivially a zigzag, any two LMPs with a zigzag between them are bisimilar. The reasoning behind the use of spans stems from the idea that bisimulation is often interpreted as an equivalence relation between states. Given two sets X and Y, any relation $R \subseteq X \times Y$ can be viewed as a span of functions from a set R to X and Y.

Example 12.3 *Let* (X, Σ) *be any measurable space. Define on* X *a Markov kernel* τ *such that* $\tau(x, X) = 1$ *for all* $x \in X$. *We thus have a labelled Markov process with a single action. Our condition on* τ *means that the single action of this process is never disabled. Let* $(\{\star\}, \Omega)$ *be a one point space with the obvious* σ*-algebra, and define a Markov kernel on* π *on* $\{\star\}$ *as* $\pi(\{\star\}, \{\star\}) = 1$. *Then the obvious map* $f : (X, \Sigma) \longrightarrow (\{\star\}, \Omega)$ *is a zigzag; indeed, we need only check the zigzag condition on the set* $\{\star\}$. *Thus, the two LMPs* (X, Σ, τ) *and* $(\{\star\}, \Omega, \pi)$ *are bisimilar.*

The main difficulty with the above definition of bisimulation is proving that it is a transitive relation among LMPs; it is clearly reflexive and symmetric. Transitivity could only be shown when the measurable spaces were analytic spaces with their Borel algebra.

In [DGJP03], bisimulation was defined as a relation on states of an LMP, in the spirit of [LS91]. One has to tie in measurability with the relation, but showing transitivity of the bisimulation is quite straightforward. In the paper of [DDLP06], a new definition of bisimulation, called event bisimulation, appeared. Its intent also is to relate similar states, but instead of thinking in terms of points one works with measurable sets.

Definition 12.4 *Given an LMP (X, Σ, τ_a), an **event bisimulation** is a sub-σ-algebra $\Lambda \subseteq \Sigma$ such that (X, Λ, τ_a) is still a LMP.*

In order to be an event bisimulation, the only condition that Λ needs to respect is that, for fixed action a and measurable set $B \in \Lambda$, $\tau_a(x, B)$ is a Λ-measurable function.

Event bisimulation and zigzag morphisms are closely related, as the following propositions show ([DDLP06]).

Proposition 12.5 *Given an LMP (X, Σ, τ_a), the σ-algebra Λ is an event bisimulation if and only if the map $i_\Lambda : (X, \Sigma) \longrightarrow (X, \Lambda)$, which is the identity as a set function, is a zigzag.*

The proof is straightforward. The above proposition can be generalized:

Proposition 12.6 *Given a zigzag morphism $f : (X, \Sigma, \tau_a) \longrightarrow (Y, \Lambda, \rho_a)$, the σ-algebra $f^{-1}(\Lambda) \subseteq \Sigma$ is an event-bisimulation.*

Thus, every event bisimulation comes from a zigzag morphism, and every zigzag morphism yields an event bisimulation; thus one can view an event-bisimulation as the "signature" of a zigzag morphism. If the idea of a zigzag morphism is to be central to the theory of LMPs, then event-bisimulation truly is the notion of state equivalence that we want to use, and is, in this context, the right notion of "measurable relation". It appears naïve to us to generalize the usual concept of an equivalence relation on a finite state space to a continuous state space; indeed, on a finite state space, every topology and every σ-algebra can be construed as an equivalence relation, and thus it is not clear how a concept of equivalence relation should generalize to a larger space while respecting the relevant structure. More details about the relationship between event bisimulation and state simulation (as a relation) are available in [DDLP06].

12.2 Logical characterization of bisimulation

The results of [vB76] and [HM85] established a characterization of ordinary (non-probabilistic) bisimulation in terms of a modal logic. Later [LS91] established such a characterization for probabilistic bisimulation using a probabilistic modal logic; of course, this was only for the case of discrete transition systems.

It turns out that a modal logic $\mathcal{L}$ characterizes bisimulation for labelled Markov processes as well [DEP98]. The logic has the following grammar, with $a \in \mathcal{A}$ and $q \in \mathbb{Q}$:

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

The logic is interpreted on states as follows. Every state satisfies $\mathbf{T}$. Conjunction is clear, so the last construct is the only one requiring explanation. A state s in a particular labelled Markov process (X, Σ, τ_a) is said to satisfy $\left\langle a \right\rangle_q \psi$ if, following an a transition from s, the probability of being in a state satisfying ψ is strictly larger than q, a rational number. More precisely, one can associate to each formula $\psi \in \mathcal{L}$ a measurable set $[\![\psi]\!]$ consisting of all points satisfying this formula. These sets are defined recursively as follows:

$$\begin{aligned} [\![\mathbf{T}]\!] &= X \\ [\![\phi \wedge \psi]\!] &= [\![\phi]\!] \cap [\![\psi]\!] \\ \left[\!\!\left[\left\langle a \right\rangle_q \psi \right]\!\!\right] &= \{ s : \tau_a \left(s, [\![\psi]\!] \right) > q \} \end{aligned}$$

and thus a state s satisfies ψ if and only if $s \in [\![\psi]\!]$.

As an example, consider the formula $\psi = \left\langle a \right\rangle_{\frac{1}{2}} \left\langle b \right\rangle_{\frac{3}{4}} \mathbf{T}$. A state satisfies ψ if it has a probability higher than $\frac{1}{2}$ to accept an a action *and*, by doing so, to transition to a state which has a probability higher than $\frac{3}{4}$ to accept a b action and to transition to another state where $\mathbf{T}$ is trivially satisfied.

The logic $\mathcal{L}$ characterizes bisimulation in the following sense. Given the restrictions on the underlying state spaces (specifically, the space must be an analytic space), two LMPs X and Y are bisimilar in the sense of definition 12.2 if and only if for each state in one LMP, there is a state in the other satisfying precisely the same formulas [DGJP03]. Keeping the same restriction on the state space, the logic also characterizes the relational definition of [DGJP03]: two states are bisimilar if and only if they satisfy the same formulas of $\mathcal{L}$.

If the underlying state space is not analytic it is possible to construct a variety of counter-examples. One can show that the basic constructions that allow one to prove that the traditional notion of bisimulation is transitive fail. One can show that the state and event bisimulation notions do not coincide. One can show that the modal logic does not in fact characterize bisimulation. These counterexamples are not very difficult to describe and should be appearing in print soon.

However the most interesting property of the logic $\mathcal{L}$ is that it *unconditionally* characterizes event-bisimulation. We let $[\![\mathcal{L}]\!]$ denote the measurable sets obtained by all formulas of $\mathcal{L}$. We state the results of [DDLP06].

Theorem 12.7 *Given any LMP (X, Σ, τ_a), the σ-algebra $\sigma([\![\mathcal{L}]\!])$ generated by the logic $\mathcal{L}$ is the smallest event-bisimulation on X. That is, the map $i : (X, \Sigma, \tau_a) \longrightarrow (X, \sigma([\![\mathcal{L}]\!]), \tau_a)$ is a zigzag; furthermore, given any zigzag $\alpha : (X, \Sigma, \tau_a) \longrightarrow (Y, \Lambda, \rho_a)$, we have that $\sigma([\![\mathcal{L}]\!]) \subseteq \alpha^{-1}(\Lambda)$.*

This generality survives in the present paper. In fact the earlier paper was a strong hint to work with a dualized point of view; a hint that we have finally taken in the present paper.

12.3 Approximation of labelled Markov processes

Approximation is a key aspect of the theory of Markov processes, especially if one is interested in applying all the tools developed for discrete systems to systems with continuous state spaces.

The first such theory was developed by [DGJP03]. The main idea was that one can focus on the behaviour of the LMP until a fixed upper bound of transitions; that is, we only care about the behaviour for the first N action choices. One can then discretize the space with respect to the Markov kernels and obtain an approximation of the starting LMP as a finite directed tree. Given an action depth N, this directed tree is split into $N + 1$ levels, from 0 to N, in such a way that a transition in this tree must increase the current level by one; hence, level N consists of a single point where no further transition is possible. The idea is that one typically chooses an initial state at level 0; thus, if the original LMP allows it, one can perform at most N transitions until being forced into a state where all actions are disabled. The transition probabilities are chosen to be an underestimate of the actual transition probabilities in the full system, which allows the approximants to be placed in a poset of LMPs.

The main drawback of this technique is that every level of the tree consists of a finite partition of the original state space; we are thus stuck with $N+1$ "finite copies" of X. This is particularly problematic for simple systems. Consider the LMP consisting of one point and one action; if the transition probability is nonzero, any finite approximation using the above scheme will consist of a chain of length $N+1$, which is counterintuitive.

Thus, it appeared that the best strategy to approximate LMPs would be to aggregate the states into a finite number of chunks; thus, a one-point space would remain a one-point space under any approximation. The problem with such a scheme is twofold; first of all, one needs an appropriate notion of state aggregation, and, ideally, a scheme to create this partition. Secondly, given a method to aggregate states, one needs to define transition probabilities on these aggregates.

One approximation scheme [DD03] is to define an equivalence relation on X which respects some compatibility property with respect to the σ-algebra of the LMP; the space of the approximate LMP is obviously the quotient space. Once this partition is defined, the transition probabilities are given by an infimum construction, again so that the approximate probabilities are an underestimate of the actual probabilities. However, one quickly runs into problems, as this technique does not yield probability measures on the approximate spaces, but what the authors call a pre-probability, yielding a new class of processes called pre-LMPs.

Another paper [DDP03] described a third method of approximation, which contains some of the ideas of the present paper in a primitive form. Given a way to aggregate the states, we would like to compute an "average" transition probability in between the lumped states and of course, this means that one needs to use conditional expectations.

Given an LMP (X, Σ, τ_a), suppose that we have a probability distribution p on the underlying measurable space. As argued in the discussion of event-bisimulation, the appropriate notion of an equivalence relation that we want to use is a σ-algebra. Thus, in order to reduce the state space X, one needs only consider a sub-σ-algebra $\Lambda \subseteq \Sigma$. Then, in order to approximate our given LMP, one needs only project the Σ-measurable functions $\tau_a(x, B)$, for each $a \in \mathcal{A}$ and $B \in \Lambda$, to a Λ-measurable function, by conditioning on Λ through the measure p. Of course, some difficulties arise; in particular, conditional expectation only yields a function which is defined p-almost-everywhere. To circumvent this difficulty, one can impose on the sub-σ-algebra that every set in Λ have nonzero measure, thereby forcing the conditional expectation operation to yield a unique function. In order to generate a sub-σ-algebra for the given LMP, the authors use the measurable sets given by a fragment of the logic $\mathcal{L}$.

12.4 Other related work

In the area of continuous state spaces there has been some substantial contributions from other authors as well. [vBW01b] developed the coalgebraic theory of transition systems using metric spaces (not ultrametric spaces). In later work [vBMOW03] they gave an intrinsic characterization of approximate bisimilarity. [DPW06] studied testing equivalences and made the connection with process logics. [MOPW04] developed a beautiful theory of duality for labelled Markov processes which relates LMPs to C^*-algebras.

A monumental program to combine probability and nondeterminism has been undertaken by Jean Goubault-Larrecq. He has written several papers [GL07a, GL07c, GL07b, GL08b, GL08a] which represent a small part of a massive unpublished book available (in French) on his web page.

There is an extensive literature on probabilistic model checking, on weak bisimulation on discrete spaces, on applications to machine learning all of which are part of the general area but it would take us too far afield to review them all here.

In the stochastic process literature entities like LAMPs have been studied under the name of Markov operators [Fog80] and approximation techniques for them have been studied by [Kim72]. The approximations introduced by Kim are of a different kind – they are not finite in any sense – and are aimed at finding a dense subset, in the weak* topology of the space of Markov operators. He also explores uniform approximation and convergence in the strong operator topology for related operators. There is no connection to logic or bisimulation.

13 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). Following [Koz85], one has the following analogy between ordinary logic and probability theory: truth values correspond to $[0, 1]$, states correspond to distributions, predicates correspond to measurable functions and satisfaction corresponds to integration. Carrying the analogy further, we have that Markov processes viewed as function transformers as we have done, is the "predicate transformer" view of probabilistic processes. Our main result is to show that this way of working with Markov processes greatly simplifies the theory: bisimulation, logical characterization and approximation. The key point is that 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 realization exists. Furthermore, this minimal object can be constructed as the projective limit of finite approximants.

In our development the duality between L_∞^+ and L_1^+ plays a key role and allows one to move back and forth. The theory could have been developed with an L_1^+ version of "predicate transformers" and worked out in a strikingly analogous fashion. We have, in fact sketched this out to the extent that it is clear that one could have gone either way. It may be that the other approach gives a better handle on constructing limits in **AMP** but in either case that seems to require substantially deeper results in measure theory to settle one way or another. It is possible that a forward version of the theory could have been developed as well; we have not investigated this thoroughly as yet.

One of the problems with any of the approximation schemes is that they are hard to implement. In a paper [BCFPP05] a few years ago, an approach based on Monte Carlo approximation was used to "approximate the approximation." The point is that it hard to compute the approximations based on applying τ^{-1} in practice. What happens is that there are lots of sets of very small measure. A sampling based technique will not see these sets and the method becomes more practical.

One line of 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.

Acknowledgements

Prakash Panangaden and Vincent Danos have benefited greatly from numerous interactions with François Laviolette and Jean Goubault-Larrecq and most especially Josée Desharnais. In addition Panangaden would like to acknowledge the long collaboration with Desharnais together with Radha Jagadeesan and Vineet Gupta where many of these ideas had their genesis. We also owe a special debt to Dexter Kozen for his insights into probabilistic systems. We have benefited from discussions with Frank van Breugel, Ernst-Erich Doberkat, Mike Mislove and James (Ben) Worrell. We thank Peter Selinger for explaining some of the basic facts about cones. We would like to thank all three referees for their comments. One of the referees produced an extraordinarily detailed and insightful referee report which led to a major reorganization of the material. We are deeply grateful to this referee for her or his contribution to this work. This research was supported by grants from Natural Sciences and Engineering Research Council (Canada) and from the Office of Naval Research (USA).

References

[AGG+86] W. Arendt, A. Grabosch, G. Greiner, U. Groh, H.P. Lotz, U. Moustakas, R. Nagel, F. Neubrander, and U. Schlotterbeck. *One-parameter Semigroups of Positive Operators.* Springer-Verlag, 1986.

[BCFPP05] Alexandre Bouchard-Côté, Norm Ferns, Prakash Panangaden, and Doina Precup. An approximation algorithm for labelled Markov processes: towards realistic approximation. In *Proceedings of the 2nd International Conference on the Quantitative Evaluation of Systems (QEST)*, pages 54–61, September 2005.

[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.

[Bil95] P. Billingsley. *Probability and Measure.* Wiley-Interscience, 1995.

[BK08] Christel Baier and Joost-Pieter Katoen. *Principles of Model Checking*. MIT Press, 2008.

[BSdV04] F. Bartels, A. Sokolova, and E.P. de Vink. A hierarchy of probabilistic system types. *Theoretical Computer Science*, 327:3–22, 2004.

[Cho58] J.R. Choksi. Inverse limits on measure spaces. *Proc. London Math. Soc*, 8(3):321–342, July 1958.

[CSKN05] S. Cattani, R. Segala, M. Kwiatkowska, and G. Norman. 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, pages 125–139, 2005.

[DD03] Vincent Danos and Josée Desharnais. Labeled Markov Processes: Stronger and faster approximations. In *Proceedings of the 18^{th} Symposium on Logic in Computer Science*, Ottawa, 2003. IEEE.

[DDLP06] Vincent Danos, Josée Desharnais, François Laviolette, and Prakash Panangaden. Bisimulation and cocongruence for probabilistic systems. *Information and Computation*, 204(4):503–523, 2006.

[DDP03] Vincent Danos, Josée Desharnais, and Prakash Panangaden. Conditional expectation and the approximation of labelled Markov processes. In Roberto Amadio and Denis Lugiez, editors, *CONCUR 2003 - Concurrency Theory*, volume 2761 of *Lecture Notes In Computer Science*, pages 477–491. Springer-Verlag, 2003.

[DEP98] J. Desharnais, A. Edalat, and P. Panangaden. A logical characterization of bisimulation for labelled Markov processes. In *proceedings of the 13th IEEE Symposium On Logic In Computer Science, Indianapolis*, pages 478–489. IEEE Press, June 1998.

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

[DGJP00] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Approximation of labeled Markov processes. In *Proceedings of the Fifteenth Annual IEEE Symposium On Logic In Computer Science*, pages 95–106. IEEE Computer Society Press, June 2000.

[DGJP03] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Approximating labeled Markov processes. *Information and Computation*, 184(1):160–200, July 2003.

[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.

[Dob03] E.-E. Doberkat. Semi-pullbacks and bisimulations in categories of stochastic relations. In J. C. M. Baeten, J. K. Lenstra, J. Parrow, and G. J. Woeinger, editors, *Proceedings of the 27th International Colloquium On Automata Languages And Programming, ICALP'03*, number 2719 in Lecture Notes In Computer Science, pages 996–1007. Springer-Verlag, July 2003.

[Dob07] E.-E. Doberkat. *Stochastic Relations. Foundations for Markov Transition Systems*. Chapman and Hall, New York, 2007.

[Dob10] Ernst-Erich Doberkat. *Stochastic Coalgebraic Logic*. Springer-Verlag, 2010.

[DPW06] M. Mislove D. Pavlovic and J. B. Worrell. Testing semantics: Connecting processes and process logics. In *Proceedings of the 11th International Conference on Algebraic Methodology and Software Technology (AMAST)*, number 4019 in Lecture Notes In Computer Science, pages 308–322. Springer-Verlag, July 2006.

[Dud89] R. M. Dudley. *Real Analysis and Probability*. Wadsworth and Brookes/Cole, 1989.

[dVR97] E. de Vink and J. J. M. M. Rutten. Bisimulation for probabilistic transition systems: A coalgebraic approach. In *Proceedings of the 24th International Colloquium On Automata Languages And Programming*, 1997.

[dVR99] E. de Vink and J. J. M. M. Rutten. Bisimulation for probabilistic transition systems: A coalgebraic approach. *Theoretical Computer Science*, 221(1/2):271–293, June 1999.

[Eda99] Abbas Edalat. Semi-pullbacks and bisimulation in categories of Markov processes. *Mathematical Structures in Computer Science*, 9(5):523–543, 1999.

[Fog80] Shaul R. Foguel. *Selected topics in the study of Markov operators.* Carolina Lecture Series, 9. University of North Carolina, Chapel Hill, North Carolina, 1980.

[FPP05] Norm Ferns, Prakash Panangaden, and Doina Precup. Metrics for Markov decision processes with infinite state spaces. In *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence*, pages 201–208, July 2005.

[Gir81] M. Giry. A categorical approach to probability theory. In B. Banaschewski, editor, *Categorical Aspects of Topology and Analysis*, number 915 in Lecture Notes In Mathematics, pages 68–85. Springer-Verlag, 1981.

[GL07a] Jean Goubault-Larrecq. 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*, pages 764–776. Springer-Verlag, 2007.

[GL07b] Jean Goubault-Larrecq. Continuous previsions. In *Proceedings of the 16th Annual EACSL Conference on Computer Science Logic (CSL'07)*, volume 4646 of *Lecture Notes In Computer Science*, pages 542–557. Springer-Verlag, 2007.

[GL07c] Jean Goubault-Larrecq. On Noetherian spaces. In *Proceedings of the 22nd Annual IEEE Symposium on Logic in Computer Science (LICS'07)*, pages 453–462. IEEE Computer Society Press, 2007.

[GL08a] Jean Goubault-Larrecq. Prevision domains and convex powercones. In *Proceedings of the 11th International Conference on Foundations of Software Science and Computation Structures (FoSSaCS'08)*, volume 4962 of *Lecture Notes In Computer Science*, pages 318–333. Springer-Verlag, 2008.

[GL08b] Jean Goubault-Larrecq. Simulation hemi-metrics between infinite-state stochastic games. In *Proceedings of the 11th International Conference on Foundations of Software Science and Computation Structures (FoSSaCS'08)*, volume 4962 of *Lecture Notes In Computer Science*, pages 50–65. Springer-Verlag, 2008.

[Hal74] P. Halmos. *Measure Theory.* Number 18 in Graduate Texts in Mathematics. Springer-Verlag, 1974. Originally published in 1950.

[Haw06] Peter Hawke. Markov operators on Banach lattices. Master's thesis, University of the Witwatersrand, Johannesburg, 2006.

[HM85] M. Hennessy and R. Milner. Algebraic laws for nondeterminism and concurrency. *Journal of the ACM*, 32(1):137–162, 1985.

[Hop54] Eberhard Hopf. The general temporally discrete Markoff process. *J. Rational Math. Mech. Anal.*, 3:13–45, 1954.

[JNW93] A. Joyal, M. Nielsen, and G. Winskel. Bisimulation and open maps. In *Proceedings of 8th Annual IEEE Symposium On Logic In Computer Science*, pages 418–427, 1993.

[Kim72] Choo-Whan Kim. Approximation theorems for Markov operators. *Probability Theory and Related Fields*, 21(3):207–214, 1972.

[Koz85] D. Kozen. A probabilistic PDL. *Journal of Computer and Systems Sciences*, 30(2):162–178, 1985.

[KT66] J. F. C. Kingman and S. J. Taylor. *Introduction to Measure and Probability.* Cambridge University Press, 1966.

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

[MOPW04] M. Mislove, J. Ouaknine, D. Pavlovic, and J. Worrell. Duality for labelled Markov processes. In I. Walukiewicz, editor, *Foundations of Software Science and Computation Structures, FOSSACS*, volume 2987 of *Lecture Notes In Computer Science*, pages 393–407, 2004.

[Pan09] Prakash Panangaden. *Labelled Markov Processes*. Imperial College Press, 2009.

[RdV97] J. J. M. M. Rutten and E. de Vink. Bisimulation for probabilistic transition systems: a coalgebraic approach. In P. Degano, editor, *Proceedings of ICALP 97*, number 1256 in Lecture Notes In Computer Science, pages 460–470. Springer-Verlag, 1997.

[Rud66] W. Rudin. *Real and Complex Analysis*. McGraw-Hill, 1966.

[Sch74] H. H. Schaefer. *Banach Lattices and Positive Operators*. Springer-Verlag, 1974.

[Sel04] Peter Selinger. Towards a semantics for higher-order quantum computation. In *Proceedings of the 2nd International Workshop on Quantum Programming Languages, Turku, Finland*, pages 127–143, June 2004. Available from `www.mathstat.dal.ca/~selinger/papers.htm`.

[Var85] Moshe Vardi. Automatic verification of probabilistic concurrent finite-state programs. In *26th IEEE Symposium On Foundations Of Computer Science*, pages 327–338, 1985.

[vB76] Johan van Benthem. *Modal correspondence theory*. PhD thesis, University of Amsterdam, 1976.

[vBMOW03] Franck van Breugel, Michael Mislove, Joël Ouaknine, and James Worrell. An intrinsic characterization of approximate probabilistic bisimilarity. In *Proceedings of FOSSACS 03*, volume 2620 of *Lecture Notes In Computer Science*. Springer-Verlag, 2003.

[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.

[Wil91] David Williams. *Probability with Martingales*. CUP, Cambridge, 1991.

[YK41] Kosaku Yosida and Shizuo Kakutani. Operator-theoretical treatment of Markoff's process and mean ergodic theorem. *The Annals of Mathematics*, 42(1):188–228, 1941.