Approximating Markov Processes by Averaging

Philippe Chaput

Master of Science

Mathematics and Statistics

McGill University

Montreal,Quebec

June 2009

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements of the degree of Master of Science

ACKNOWLEDGEMENTS

I would like to thank: Prakash Panangaden, for being an awesome supervisor and scotch buff; Vincent Danos and Gordon Plotkin, who inspired this work; Kaitlyn Urquhart, for her keen 'I'; Thomas Getgood, Tayeb Aïssiou for reading and commenting my work; caffeine, for obvious reasons; my friends and family, for being there; and finally, my parents, for supporting me and encouraging me. This research has been supported by NSERC.

ABSTRACT

We recast the theory of labelled Markov processes in a new setting, in a way "dual" to the usual point of view. Instead of considering state transitions as a collection of subprobability distributions on the state space, we view them as transformers of real-valued functions. By generalizing the operation of conditional expectation, we build a category consisting of labelled Markov processes viewed as a collection of operators; the arrows of this category behave as projections on a smaller state space. We define a notion of equivalence for such processes, called bisimulation, which is closely linked to the usual definition for probabilistic processes. We show that we can categorically construct the smallest bisimilar process, and that this smallest object is linked to a well-known modal logic. We also expose an approximation scheme based on this logic, where the state space of the approximants is finite; furthermore, we show that these finite approximants categorically converge to the smallest bisimilar process.

ABRÉGÉ

Nous reconsidérons les processus de Markov étiquetés sous une nouvelle approche, dans un certain sens “dual” au point de vue usuel. Au lieu de considérer les transitions d’état en état en tant qu’une collection de distributions de sous-probabilités sur l’espace d’états, nous les regardons en tant que transformations de fonctions réelles. En généralisant l’opération d’espérance conditionelle, nous construisons une catégorie où les objets sont des processus de Markov étiquetés regardés en tant qu’un rassemblement d’opérateurs; les flèches de cette catégorie se comportent comme des projections sur un espace d’états plus petit. Nous définissons une notion d’équivalence pour de tels processus, que l’on appelle bisimulation, qui est intimement liée avec la définition usuelle pour les processus probabilistes. Nous démontrons que nous pouvons construire, d’une manière catégorique, le plus petit processus bisimilaire à un processus donné, et que ce plus petit object est lié à une logique modale bien connue. Nous développons une méthode d’approximation basée sur cette logique, où l’espace d’états des processus approximatifs est fini; de plus, nous démontrons que ces processus approximatifs convergent, d’une manière catégorique, au plus petit processus bisimilaire.

TABLE OF CONTENTS

CHAPTER 1
Introduction

The main objects studied in this thesis are probabilistic systems. In particular, we are interested in knowing when two such systems can be deemed equivalent, and what would be an appropriate notion of approximation for these systems. These questions are particularly nontrivial when the state space of the system is not discrete, which is the situation which will be investigated in this work.

The systems that we shall be studying are interactive. In other words, the interpretation of the execution of such a system requires the existence of a "user" for this system (this "user" can be considered to be anything external to the system, e.g. the environment). The system is assumed to evolve only when the system and the user interact (there are no so-called hidden actions). Although the interaction between system and user can only be made through well-defined channels, the behavior of the user or environment is not itself modeled in any way, and thus the user is purely nondeterministic.

Throughout this thesis, time will be discrete, and so, upon user interaction, the system instantaneously changes its internal state via a probabilistic transition. However, the user is not directly aware of the state change effected by the system; the only feedback the user obtains is through the interaction channels.

The initial motivation for this work was to develop a mathematical framework that would streamline the approximation of probabilistic systems via an appropriate notion of average behavior. Surprisingly, this framework allowed us to weave behavioral approximation, behavioral equivalence and logic-based reasoning under a categorical description.

This work was inspired by private communication with Vincent Danos and Gordon Plotkin [DP06], and will be in part published in [CPDP09].

1.1 Labelled transition systems and an introduction to bisimulation

In order to motivate our work, we first introduce some relevant concepts in the context of a simpler kind of system.

Definition 1.1.1 *A labelled transition system (LTS) is a finite or countable state space X, together with a finite set $\mathcal{A}$ of actions or labels, to which we associate a set of transition relations $\rightarrow_a \subseteq X \times X$, for $a \in \mathcal{A}$. For every pair (x, y) of states, we shall write $x \rightarrow_a y$ if $(x, y) \in \rightarrow_a$. For a given state x, if there is no state y such that $x \rightarrow_a y$, we write $x \nrightarrow_a$ and we say that action a is* disabled *at x.*

We interpret such systems as follows, following Milner [Mil80]. Suppose the system is currently in a state, x. The user has access to buttons which are labelled by $\mathcal{A}$; these buttons are the channels through which user and system interact. If the user presses the button labelled by action a, one of two things can happen. If $x \nrightarrow_a$, the button physically jams and the user then knows that action a is disabled. On the other hand, if action a is not disabled at x, the system nondeterministically changes its state to any state y such that $x \rightarrow_a y$.

Although it is not necessary, labelled transition systems are often pointed; that is, a particular state in the state space is singled out, and is called the initial state.

Note that the above definition of an LTS does not encode the above interpretation; indeed, an LTS is nothing but a finite collection of relations on a countable set. We thus need a notion of *equivalence*, with respect to the above interpretation, in order to do any further mathematical analysis. One such equivalence concept is called *bisimulation*. The strategy is to determine which states are indistinguishable from the user's perspective. We begin with the following definition, due to Milner [Mil80].

Definition 1.1.2 *Given a labelled transition system on a state space X, a relation $R \subseteq X \times X$ is said to be a* bisimulation *if, for all pairs of states $(x, y) \in R$*

1. *If $x \to_a x'$, there exists a state y' such that $y \to_a y'$ and $(x', y') \in R$*
2. *If $y \to_a y'$, there exists a state x' such that $x \to_a x'$ and $(x', y') \in R$*

Two states x, y, are said to be bisimilar *if there exists a bisimulation R with $(x, y) \in R$.*

Intuitively speaking, one can think of a bisimulation R as an equivalence relation; indeed, it is easy to see from the definition that for any bisimulation R, the equivalence relation generated by R is also a bisimulation. Thus, the rôle of a bisimulation is to lump together, in equivalence classes, states which are indistinguishable with respect to the user. Indeed, not only can two bisimilar states perform the same actions, but for any enabled action on these states, the sets of reachable states associated to this action are contained in the same equivalence classes.

Park [Par81] defined bisimulation as the greatest fixed point of an operator on the space of relations on X, and it is a pleasing observation that Park's fixed point definition coincides with the above definition.

Another way to describe bisimulation is to translate the user experience into a modal logic; any finite description of the process from the user's point of view would be encodable in a suitable formula. Hennessy and Milner [HM85] defined such a logic, and they proved that two states are bisimilar with respect to definition 1.1.2 if and only if they satisfy the same formulas of this logic. We omit the details as we shall encounter a similar logic in the probabilistic case.

1.2 The probabilistic case

Extending the idea of labelled transition systems to probabilistic systems is fairly straightforward. Instead of associating, to each state-action pair, a subset of potential next states, we associate a probability distribution on the state space. Thus, when the user selects an action, the system effects a transition following this probability distribution.

Larsen and Skou [LS91] first analysed interactive systems in a probabilistic setting. The state space considered was discrete and thus there were no measure theoretic considerations. As in LTS's, actions are deterministically disabled; in a given state, the probability of effecting a transition to anywhere in the state space is either zero or one. The authors defined bisimulation for these probabilistic systems, and wrote out a logic characterizing this bisimulation relation.

Nondiscrete state spaces were first examined by Blute et al. [BDEP97] and de Vink and Rutten [RdV97]. However, de Vink and Rutten's work concentrated on ultrametric spaces and thus was of limited applicability compared to the work of Blute et al. Indeed, the latter defined a notion of probabilistic process on a much larger class of state spaces, along with a notion of bisimulation and the introduction of the now widespread use of the term "labelled Markov processes". Further publications [DEP98, DGJP99, DGJP00] developed different notions of bisimulation for these processes and defined a modal logic characterizing bisimulation. However, the proofs relating the logic to bisimulation required restrictions on the structure of the state space and the results were quite technical.

These papers also devised approximation schemes for labelled Markov processes. Indeed, in order to do computations on these systems, one needs discrete but reliable approximations. One promising approximation scheme of Danos et al. [DDP03] involves partitioning the state space into a finite number of chunks, and to average transition probabilities over these chunks. As averaging requires the use of a conditional probability operator, a note by Danos and Plotkin [DP06] considered using linear operators to represent transition probabilities as well. It turns out that this point of view is in a precise sense "dual" to the usual interpretation of Markov processes, as these linear operators transform functional expressions on the state space "backwards in time". This point of view also meshes very well with standard concepts in the labelled Markov processes literature.

1.3 Outline of thesis

In Chapter 2, we review the mathematical and computer theoretical background necessary for our work. In particular, we cover labelled Markov processes in detail and expose different definitions of bisimulation. We also give an overview of Markov operators, which will play an important rôle.

In Chapter 3, we define the probabilistic processes we shall study; we call them abstract Markov processes, because the transition probabilities are bundled into positive linear operators. We then show that the operation of conditional expectation can be generalized to a functor, and we use this to describe a very general approximation scheme induced by a measurable map of the underlying state space.

In Chapter 4, we create a category that contains all abstract Markov processes, which allows us to define bisimulation for these probabilistic processes. Bisimulation between processes is shown to be a transitive relation. Furthermore, we show that there is a minimal bisimilar process to any given process. The structure of this minimal bisimilar process is partly described by a modal logic.

In Chapter 5, we use the modal logic to explicitly construct finite approximants to abstract Markov processes. Furthermore, we show that these finite approximants converge to the original abstract Markov process in a categorical sense.

Chapter 6 is a discussion of related work.

Finally, in Chapter 7, we give a final overview of our contributions and discuss future work.

CHAPTER 2
Background

We shall assume that the reader is knowledgable in measure theory and category theory. We review the concepts that we will need.

2.1 Domain theory

As most of the structures we use come with a partial ordering, we will want to use some domain theoretical concepts.

Definition 2.1.1 *A* directed set X *in a poset* $(P, \leq)$ *is a non-empty subset of* P *such that for all* $x, y \in X$ *there exists* $z \in X$ *with* $x \leq z$ *and* $y \leq z$*.*

We can also simply speak of a directed set X, in which case we mean a poset which is itself a directed set.

Many of the directed sets that arise are increasing sequences (ω-chains).

Definition 2.1.2 *A* dcpo *is a poset in which every directed set has a least upper bound.*

Definition 2.1.3 *A* Scott-continuous function *between two dcpos is a function that preserves least upper bounds of directed sets.*

Scott-continuous functions are automatically monotonic. Note that Scott-continuity is also known as order-continuity. We may also speak of ω-dcpo and ω-Scott-continuous functions, in which case it is understood that the directed sets in question are countable.

2.2 Categories and measure theory

This thesis is couched in the language of category theory; for reference, the reader may consult[Mac98].

Let us denote the category of sets by **Set**.

If f is a function from a measurable space (X, Σ) to a measurable space (Y, Λ), we denote by $f^{-1}(\Lambda)$ the set of of all subsets A of X such that $A = f^{-1}(B)$ for some $B \in \Lambda$.

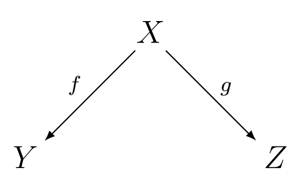
We recall the definition of a measurable function to avoid a common confusion.

Definition 2.2.1 *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 $B \in \Lambda$, i.e. if $f^{-1}(\Lambda) \subseteq \Sigma$.*

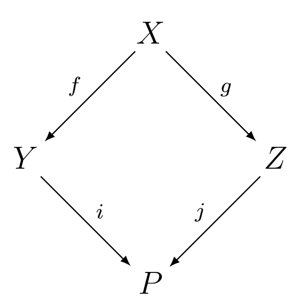
Note this is *not* the definition in Halmos [Hal74], but is the one used by most modern authors.

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**, the category of sets and functions, which preserves limits and colimits. Indeed, the limits (resp. colimits) in **Mes** are precisely the limits (resp. colimits) in **Set**, equipped with the smallest (resp. largest) σ-algebra making the maps measurable. We give an example.

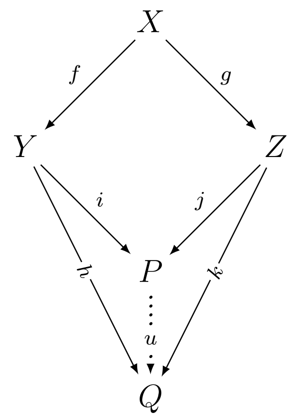
Definition 2.2.2 *In a category* **C**, *the* pushout *of a span of two morphisms $f : X \longrightarrow Y$ and $g : X \longrightarrow Z$*



is an object P together with two morphisms $i : Y \longrightarrow P$ and $j : Z \longrightarrow P$ such that the following diagram commutes:



Furthermore, the pushout (P, i, j) has the following universal property: given any other object Q with maps $h : Y \longrightarrow Q$ and $k : Z \longrightarrow Q$ with the above properties, there is a unique morphism $u : P \longrightarrow Q$ through which h and k factor



It is well-known that pushouts always exists in **Set**. As the pushout is a colimit, the pushout also exists in **Mes**; as we mentioned above, it is the same pushout set P equipped with the largest σ-algebra making the maps i, j measurable. Thus, in this case, the σ-algebra on P is generated by all sets $A \subseteq P$ such that both $i^{-1}(A)$ and $j^{-1}(A)$ are measurable in their respective measurable spaces.

We also review a particular case of a categorical limit.

Definition 2.2.3 *A* projective system *in a category* **C** *is a set of objects $\{X_i\}$, indexed by a directed set I, along with arrows $f_{ij} : X_j \longrightarrow X_i$ for every pair of points in I such that $i \leq j$. Furthermore, if $i \leq j \leq k$, we have that $f_{ij} \circ f_{jk} = f_{ik}$. We also have that $f_{ii} = id_{X_i}$, the identity arrow.*

Definition 2.2.4 *Given a projective system X_i over a directed set I, the projective limit of $\{X_i\}$ is an object* $\operatorname{proj\,lim} X_i = X_\infty$*, along with, for every $i \in I$, an arrow $f_{i\infty} : X_\infty \longrightarrow X_i$ such that if $i \leq j$, $f_{i\infty} = f_{ij} \circ f_{j\infty}$.*

This projective limit is universal in the following sense: if Y is an object in C, together with, for every $i \in I$, an arrow $g_i : Y \longrightarrow X_i$ such that if

$i \leq j$, $g_i = f_{ij} \circ g_j$, there is a unique map $u : Y \longrightarrow X_\infty$ such that $g_i = f_{i\infty} \circ u$ for all $i \in I$.

Of course, projective limits exist in **Set**; for a projective system $\{X_i\}_{i \in I}$, the projective limit X_∞ is the subset of the product set $\prod_{i \in I} X_i$ consisting of all I-indexed tuples $(x_i)_{i \in I}$ such that if $i \leq j$, $f_{ij}(x_j) = x_i$. As mentioned above, the projective limit of a projective system $\{X_i, \Sigma_i\}$ also exists in **Mes**. It is the projective limit X_∞ of **Set**, together with the smallest σ-algebra Σ_∞ making all of the $f_{i\infty}$ maps measurable. It is easy to see that Σ_∞ is the σ-algebra generated by the collection $\left\{ f_{i\infty}^{-1}(\Sigma_i) \right\}$ of σ-algebras induced on X_∞ by the limit maps $f_{i\infty}$.

2.2.1 The Radon-Nikodym Theorem and notation

Definition 2.2.5 *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 let **Prb** be the category of probability spaces and measurable maps.

Given (X, Σ, p) and (Y, Σ') and a measurable function $f : X \longrightarrow Y$ we obtain a measure $M_f(p)$ on Y defined as $M_f(p)(B) = p(f^{-1}(B))$. The axioms of a measure are easy to verify on $M_f(p)$ thanks to the well-behavedness of preimages. This measure is called the *image measure* of p under f.

Given a measurable space (X, Σ) with a measure μ, we say two measurable functions are μ-equivalent if they differ on a set of μ-measure zero. Given two measurable real-valued functions f and g on X, we say $f \leq_\mu g$ if f is less than g except maybe on a set of measure zero. For $B \in \Sigma$, we let $\mathbf{1}_B$ be the indicator function of the set B. $L_1(X, \mu)$ stands for the space of *equivalence classes* of integrable functions. Similarly we write $L_1^+(X, \mu)$ for equivalence classes of functions that are positive μ-almost everywhere.

$L_\infty(X,\mu)$ is the space of equivalence classes of μ-almost everywhere uniformly bounded functions on X.Given two measures ν,μ on (X,Σ), if we have, for all $A\in\Sigma$, that $\mu(A)=0\Rightarrow\nu(A)=0$, we say that ν is absolutely continuous with respect to μ, and write $\nu\ll\mu$. g The Radon-Nikodym theorem is a central result in measure theory allowing one to define a "derivative" of a measure with respect to another measure.

Theorem 2.2.6 *[Bil95] If $\nu\ll\mu$, where ν,μ are finite measures on (X,Σ), there is a positive measurable function h on X such that for every $B\in\Sigma$*

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

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

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

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

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

These two identities just say that the operations $-\rhd\mu$ and $\frac{\mathrm{d}}{\mathrm{d}\mu}$ are inverses of each other as operations from $L_1^+(X,\mu)$ to the space of finite measures on X.

Furthermore, the Radon-Nikodym derivative has a "chain-rule"-like property, in that if $\mu\ll\nu\ll\lambda$, then $\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}=\frac{\mathrm{d}\mu}{\mathrm{d}\nu}\frac{\mathrm{d}\nu}{\mathrm{d}\lambda}$, λ-almost everywhere.

2.2.2 Conditional expectation

We quickly recall the definition of conditional expectation of a measurable function with respect to a sub-σ-algebra.

Theorem 2.2.7 (Kolmogorov) *Let* (X, Σ, p) *be a probability triple,* f *be in* $L_1(X, \Sigma, p)$ *and* Λ *be a sub-*σ*-algebra of* Σ*. There exists* $g \in L_1(X, \Lambda, p)$ *such that for all* $B \in \Lambda$*,*

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

This theorem is proved using the Radon-Nikodym theorem. We shall write $g = \mathbb{E}_\Lambda(f)$. Conditional expectation is thus an operator from $L_1(X, \Sigma, p)$ to $L_1(X, \Lambda, p)$ which is linear and positive. It is defined uniquely almost everywhere.

2.3 Markov kernels and labelled Markov processes

We carefully review the required background on probabilistic processes, in particular, labelled Markov processes, which is the class in which we are interested.

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.

2.3.1 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 2.3.1 *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.

A Markov kernel is said to be *deterministic* if for all $x \in X$, $B \in \Lambda$, $\tau(x, B) = 0$ or 1. 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.

Example 2.3.2 *Let $f : (X, \Sigma) \longrightarrow (Y, \Lambda)$ be a measurable function. We can define a Markov kernel τ_f from X to Y as*

$$\tau_f(x, B) = \mathbf{1}_{f^{-1}(B)}(x)$$

This is a deterministic Markov kernel where the transition from X to Y follows the function f. τ_f is nonsingular if $\mathbf{1}_{f^{-1}(B)}(x) = 0$, μ-almost everywhere, for all $B \subseteq Y$ such that $\nu(B) = 0$. This is equivalent to requiring that $\mu(f^{-1}(B)) = 0$ for such B.

2.3.2 Labelled Markov processes and bisimulation

We now give the definition of a labelled Markov processes, first given in this form in [DEP98].

Definition 2.3.3 *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 in this thesis.

As with labelled transition systems, the interpretation of the dynamics of such systems is not encoded in the definition above; in fact, the above definition appears very innocuous. We thus need to define a notion of equivalence or 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 LMP's among themselves. In this higher level point of view, two LMP's are bisimilar if each state in one is bisimilar to a state in the other; or, in other words, if the two LMP's contain states which have the same behaviour. Note that we shall always assume that when speaking of bisimulation between different LMP's, 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].

First and foremost is the concept of a *zigzag* morphism [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 LMP's are bisimilar. Specifically, we have the following definition:

Definition 2.3.4 *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. The intuition behind zigzags is that neither of these information losses are visible to the user, as he may only witness whether an action is enabled, which only comes from the transition probabilities. We thus obtain a smaller system with the same behavior.

Following Joyal et al. [JNW93], Desharnais et al. [DEP02] defined two LMP's to be bisimilar if there exists a *span* of zigzags between them. However, for this definition to work, it was necessary to slightly loosen the measurability condition to introduce the concept of a *generalized* Markov process. We omit the details as we are mostly interested in the ideas at this point.

Definition 2.3.5 *Two LMP's* (X, Σ, τ_a) *and* (Y, Λ, ρ_a) *are* bisimilar *if there exists a generalized 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 LMP's 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 2.3.6 *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 LMP's* (X, Σ, τ) *and* $(\{\star\}, \Omega, \pi)$ *are bisimilar.*

Although this may look surprising, consider the user's point of view; the single button is always enabled, and thus, with respect to the user's experience, the state space could very well be a single point.

The main difficulty coming from the above definition of bisimulation is proving that it is a transitive relation among LMP's, as it is trivially reflexive and symmetric. The transitivity could only be shown when the measurable spaces were generated by a particular class of topological spaces; furthermore, as we have seen, the definition of a labelled Markov process had to be slightly generalized. These conceptual difficulties indicated that there may be a more elegant definition of bisimulation for LMP's.

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 [DDLP06], a new definition of bisimulation, called event bisimulation, appeared. Its intent also is to relate similar states, but its main idea is measure-theoretic, thus effectively avoiding the problem the [DGJP03] definition had.

Definition 2.3.7 *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 intimately linked, as the following propositions show ([DDLP06]).

Proposition 2.3.8 *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 2.3.9 *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 LMP's, 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". Indeed, it appears naïve to us to generalize what is an equivalence relation on a finite state space into an equivalence relation on 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. More details about the relationship between event bisimulation and state simulation (as a relation) are available in [DDLP06].

2.3.3 Logical characterization of bisimulation

As the results of [HM85] and [LS91] suggested, there may be a modal logic which would also characterize bisimulation in the case of a general labelled Markov process. In the two above cases, two states were bisimilar (in their respective context) if and only if they satisfied the same formulas of the logic.

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 $q \in \mathbb{Q}$:

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

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 $\langle a \rangle_q \psi$ 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:

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

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

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

The logic $\mathcal{L}$ characterizes bisimulation in the following sense. Given some 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 2.3.5 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 $\mathcal{L}$.

However, for our purposes, the most interesting property of the logic $\mathcal{L}$ is that it unconditionally characterizes event-bisimulation; indeed, we do not need any restriction on the state space. We let $[\![\mathcal{L}]\!]$ denote the measurable sets obtained by all formulas of $\mathcal{L}$. We state the results of [DDLP06], which we shall use later in this thesis:

Theorem 2.3.10 *[DDLP06] 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* $\sigma([\![\mathcal{L}]\!]) \subseteq \alpha^{-1}(\Lambda)$*.*

2.3.4 Approximation of labelled Markov processes

Given a labelled Markov process, it may be time-consuming or impossible to perform computational experiments if the state space is too large. Thus, there has been a lot of interest in developing techniques to construct finite approximations to labelled Markov processes.

The first such attempt was done in [DGJP03]. The main idea was that one can focus on the behavior of the LMP until a fixed upper bound of transitions; that is, we only care about the behavior 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 developed in [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] exposed a third method of approximation, which we shall expand in this thesis. Given a way to aggregate the states, we would like to compute an "average" transition probability in between the lumped states. As we are working in a measure-theoretic setting, the obvious method that we want to use is conditional expectation. Given an LMP (X, Σ, τ_a), suppose that we have a probability distribution p on the underlying measurable space. As argued in the event-bisimulation section, 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 $\mathcal{L}$; this is also the direction we shall take.

2.4 Cones

Cones are a way of marrying order structure with linear structure. The idea is that a subset of a vector space is designated as the set of "positive" vectors. This set will need to satisfy some natural closure properties. We can then define $u \leq v$ for two vectors u and v by saying that $v - u$ is positive. We base this discussion of cones on the paper by Selinger [Sel04].

Definition 2.4.1 *A* cone *is an abelian group* $(V, +)$ *on which multiplication by positive real numbers is defined. 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 2.4.2 *A* normed cone C *is a cone with a function*

$||\cdot|| : C \longrightarrow \mathbb{R}_+$ *satisfying:*

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. $u \leq v \Rightarrow ||u|| \leq ||v||$.

Owing to the lack of a subtraction operation, it is not possible to speak of a sequence being Cauchy in the usual sense. Let the *unit ideal* of a cone be the set of all elements with norm less than or equal to 1.

Definition 2.4.3 *A* (ω-)complete normed cone *is a normed cone such that its unit ideal is a (ω-)dcpo.*

It is then immediate that any (countable) norm-bounded directed set in an (ω-)complete normed cone has a least upper bound.

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. An *order-continuous* linear map between two cones is one that preserves least upper bounds of directed sets, i.e. is Scott-continuous. Similarly, we may also speak of ω-order-continuous linear maps of cones. Note that in a (ω-)complete normed cone, the norm is (ω-)order-continuous. As our work is measure-theoretic in nature, a morphism or map will be said to be *order-continuous* if it is ω-order-continuous.

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 complete (or 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 the same as the operator norm for bounded linear maps between vector spaces.

The ω-complete normed cones, along with ω-order-continuous bounded linear maps, form a category which we shall denote $\omega\mathbf{CC}$. We shall now introduce the cones which we shall use in this text. They are all ω-complete normed cones.

Let (X, Σ) be a measure space. Then one can speak of the cone $\mathcal{L}^+(X)$ of bounded measurable maps from X to $\mathbb{R}_+$; this cone has actual functions and not equivalence classes as in, for example, $L_\infty(X)$. This is a ω-complete normed cone as the supremum of countably many measurable functions is measurable; the norm is the supremum of the function over X.

If μ is a measure on X, then one has the usual L_p spaces, which can be restricted to cones by considering the μ-almost everywhere positive functions. We shall denote these cones by $L_p^+(X, \Sigma, \mu)$ or $L_p^+(X)$ if the context is clear. These also are complete normed cones, either by the monotone convergence theorem for $1 \leq p < \infty$, or as we pointed out above for the sup norm which corresponds to $p = \infty$. The norm on these spaces is denoted by $\|-\|_p$

We will also talk about cones of measures on a space. Let (X, Σ) be a measurable space. We denote by $\mathcal{M}(X)$ the cone of finite measures on X; the norm of a measure μ is just the measure of the space $\mu(X)$. Let us equip X with a finite measure p. We shall denote by $\mathcal{M}^{\ll p}(X)$, the cone of all measures on (X, Σ, p) which are absolutely continuous with respect to p. As with $\mathcal{M}(X)$, if μ is such a measure, we define its norm to be $\mu(X)$. It is easy to see that this norm coincides precisely with the norm on $L_1^+(X, \Sigma, p)$ if one considers the density function of μ through the Radon-Nikodym theorem. Hence $\mathcal{M}^{\ll p}(X)$ is also a ω-complete normed cone. In fact, we can say more; it is easy to show that the maps $\frac{\mathrm{d}(-)}{\mathrm{d}p} : \mathcal{M}^{\ll p}(X) \longrightarrow L_1^+(X, \Sigma, p)$ and $(-) \rhd p : L_1^+(X, \Sigma, p) \longrightarrow \mathcal{M}^{\ll p}(X)$ are both order-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 $\mathcal{M}^{\leq Kp}(X)$, the cone of all measures on (X, Σ) which are uniformly less than a multiple of the measure p; that is, $\mu \in \mathcal{M}^{\leq Kp}(X)$ if there is some constant K such that $\mu(B) \leq Kp(B)$ for all measurable sets B. For such a measure μ, we can define the norm of μ to be the infimum of all such constants K, which, analogously to the above instance, coincides with the norm on $L_\infty^+(X, \Sigma, p)$ when one considers the density function of μ ; thus $\mathcal{M}^{\leq Kp}(X)$ is a ω-complete normed cone. As with $\mathcal{M}^{\ll p}(X)$, the cones $\mathcal{M}^{\leq Kp}(X)$ and $L_\infty^+(X, \Sigma, p)$ are isometrically isomorphic. The two maps $\frac{\mathrm{d}(-)}{\mathrm{d}p}$ and $(-) \rhd p$ also are norm-preserving.

Finally, we shall need the concept of dual cone. Given a ω-complete normed cone C, its dual C^* is the set of all order-continuous linear maps from C to $\mathbb{R}_+$. This dual cone is sometimes denoted $C \multimap \mathbb{R}_+$. 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 pointwise order.

In $\omega\mathbf{CC}$, the dual operation becomes a contravariant functor; indeed, if $f : C \longrightarrow D$ is a map of cones, we can define $f^* : D^* \longrightarrow C^*$ in the usual way. That is, 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. Let us denote the dual of $L_p^+(X)$ by $L_p^{+,*}(X)$.

Proposition 2.4.4 *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 μ on X as follows:

$$\mu(B) = L(\mathbf{1}_B)$$

The countable additivity of μ is a direct consequence of the ω-continuity of L; indeed, given a countable collection of 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_i's are disjoint. As $\mathbf{1}_X$ has finite norm in $L_\infty^+(X)$, we have

$$\begin{aligned}
\mu\left(\bigcup_{i=1}^\infty B_i\right) &= L\left(\sup_n \sum_{i=1}^n \mathbf{1}_{B_i}\right) \\
&= \sup_n L\left(\sum_{i=1}^n \mathbf{1}_{B_i}\right) \\
&= \sup_n \sum_{i=1}^n L(\mathbf{1}_{B_i}) \\
&= \sum_{i=1}^\infty L(\mathbf{1}_{B_i})
\end{aligned}$$

Furthermore, $\mu(\emptyset) = L(0) = 0$, and thus μ is a measure.

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

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

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

Finally, if $p(B) = 0$, we have that $\mathbf{1}_B = 0$ in $L_\infty^+(X)$, and thus μ 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 may call ϕ, such that, in the above exposition, we have $\phi(L) = \mu$.

It is easy to check that ϕ is linear and order-continuous. Furthermore, we just checked that it was norm-preserving.

On the other hand, it is clear that every element μ of $\mathcal{M}^{\ll p}(X)$ corresponds to an element of $L_\infty^{+;*}(X)$. Indeed, if u is the Radon-Nikodym derivative of μ, we have the functional $f \mapsto \int_X fu \, \mathrm{d}x$ on $L_\infty^+(X)$ which is bounded by Hölder's inequality. Thus ϕ is an isometric isomorphism. ∎

As $\mathcal{M}^{\ll p}(X)$ 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 less surprising and is proved in the same way:

Proposition 2.4.5 *The dual of the cone $L_1^+(X, \Sigma, p)$ is isometrically isomorphic to $\mathcal{M}^{\leq Kp}(X)$.*

We shall denote the isomorphism from $L_1^+(X, \Sigma, p)$ to $\mathcal{M}^{\leq Kp}(X)$ by ϕ as well, as the construction is precisely the same.

As above, as $\mathcal{M}^{\leq Kp}(X)$ is isometrically isomorphic to $L_\infty^+(X)$, an immediate corollary is that $L_1^{+,*}(X)$ is isometrically isomorphic to $L_\infty^+(X)$.

2.5 Markov operators

It is a pleasing fact that Markov kernels can be viewed as linear maps on function spaces. This idea was first elaborated by Yosida and Kakutani [YK41].

Given τ a Markov kernel from X to Y, 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, as per our definition above, $T_\tau(\mathbf{1}_B)$ is measurable for every $B \in \Lambda$.

It is also linear and order-continuous, and thus $T_\tau(f)$ is measurable for any measurable f by going through simple functions. Note that we have $T_\tau(\mathbf{1}_B)(x) = \tau(x, B)$. Conversely, any order-continuous morphism L with $L(\mathbf{1}_Y) \leq \mathbf{1}_X$ can be cast as a Markov kernel by reversing the process above. Indeed, 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.

We can also define an operator on $\mathcal{M}(X)$ by using τ the other way. Indeed, 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 order-continuous.

This operator allows us to characterize nonsingular Markov kernels:

Lemma 2.5.1 *A Markov kernel from (X, Σ, μ) to (Y, Λ, ν) is nonsingular if and only if $\bar{T}_\tau(\mu) \ll \nu$*

Proof . Given a set $B \in \Sigma$ such that $\mu(B) = 0$, we only need to show that $\tau(x, B) = 0$, μ-almost everywhere if and only if $\bar{T}_\tau(\mu)(B) = 0$; but this is clear from the definition, as $\tau(x, B)$ is a positive function. ∎

Note that if the state spaces X and Y are finite, linear operators are just matrices; we thus obtain the standard stochastic matrices used for Markov chains on graphs, for instance.

Remark 2.5.2 *The two operators T_τ and $\bar{T}_\tau$ have a notable, if informal, interpretation. 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 likelihood transformer which works "backwards in time". This time inversion can be seen from the reversal of X and Y in the definition of the operator. Consider the function $\mathbf{1}_B$ on Y. It is a likelihood function for the event of "being *in the set B", where the parameter is the point in Y. Indeed, if we let $\mathcal{Y}$ be a random variable taking values in Y, we have that $P(\mathcal{Y} \in B \,|\mathcal{Y} = y) = \mathbf{1}_B$. Given a Markov kernel τ, we have that $T_\tau(\mathbf{1}_B)(x) = \tau(x, B)$. That is, $T_\tau(\mathbf{1}_B)$ is the probability of being in the set B, but after one stochastic transition through τ, as a function of the state $x \in X$. This is again a likelihood function, but where the parameter is one step back in time.*

Note that the operator norms of both T_τ and $\bar{T}_\tau$ are less than one.

If our measurable spaces X and Y are endowed with measures μ and ν, respectively, which we shall assume finite, it is tempting to consider positive operators on L_p-spaces instead than on $\mathcal{L}^+$. This was first explored by Hopf [Hop54]. We will slightly modify classical definitions in order to work within cones; the interested reader may consult [Sch74, AGG+86, Haw06] for the usual framework in Banach spaces or Banach lattices.

Definition 2.5.3 *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 $\|T\| \leq 1$*

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 $L_1^+(Y)$). Note that this map is automatically order-continuous, as if a sequence f_n increases to $f \in L_1^+(X)$, we have $\|f - f_n\|_1 \longrightarrow 0$. However, for our purposes, the operator we will want to work with is the equivalent of T_τ; that is, we will want an operator on L_∞^+ cones. We make the following definition:

Definition 2.5.4 *An* abstract Markov kernel *from (X, Σ, μ) to (Y, Λ, ν) is an order-continuous linear map $\tau : L_\infty^+(Y) \longrightarrow L_\infty^+(X)$ with $\|\tau\| \leq 1$.*

Asking 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 order-continuous. Indeed, one can find bounded positive linear operators on $L_\infty(\mathbb{R})$, when considered as a Banach space, which are not order-continuous. Consider the space $\mathcal{C}_{\lim}(\mathbb{R})$ of bounded continuous functions whose limit at $+\infty$ exists. This can be considered as a closed subspace of $L_\infty(\mathbb{R})$. Define a bounded linear functional L on $\mathcal{C}_{\lim}(\mathbb{R})$ defined as $L(f) = lim_{x \longrightarrow \infty} f(x)$. This functional is not order-continuous. Extending this functional to $L_\infty(\mathbb{R})$ by the Hahn-Banach theorem yields a bounded linear functional which is not order-continuous either.

We define a bilinear form $\langle \cdot, \cdot \rangle : L_\infty^+(X) \times L_1^+(X) \longrightarrow \mathbb{R}_+$ by

$$\langle f, u \rangle = \int_X f \, u \, \mathrm{d}\mu$$

The following is direct consequence of the duality of $L_1^+(X)$ and $L_\infty^+(X)$ in $\omega\mathbf{CC}$:

Lemma 2.5.5 *To every linear operator A from $L_1^+(X)$ to $L_1^+(Y)$ there corresponds a unique adjoint operator $A^\dagger$ from $L_\infty^+(Y)$ to $L_\infty^+(X)$ such that, for all $f \in L_\infty^+(Y)$ and $u \in L_1^+(X)$,*

$$\langle f, Au \rangle = \left\langle A^\dagger f, u \right\rangle$$

Furthermore, $A^\dagger$ is linear and order-continuous.

Similarly, to every order-continuous linear operator B from $L_\infty^+(X)$ to $L_\infty^+(Y)$ there corresponds an adjoint operator $B^\dagger$ from $L_1^+(Y)$ to $L_1^+(X)$, linear as well, such that, for all $g \in L_\infty^+(X)$ and $v \in L_1^+(Y)$,

$$\langle Bg, v \rangle = \left\langle g, B^\dagger v \right\rangle$$

And hence the following corollary is of particular importance:

Corollary 2.5.6 *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 2.5.7 *One can find a similar bilinear form which demonstrates that the operators* $\bar{T}_\tau$ *and* T_τ *are adjoints.*

The following result is due to Hopf [Hop54]:

Proposition 2.5.8 *Every Markov operator from* (X, Σ, μ) *to* (Y, Λ, ν) *corresponds uniquely to a nonsingular Markov kernel from* X *to* Y*.*

As a 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 $\tau(\mathbf{1}_B)(x) = \tau(x, B)$; this is precisely the interpretation we had for the operator T_τ.

Nevertheless, the above proposition is not 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 shall 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.

CHAPTER 3
Abstract Markov Processes and A Generalization of Conditional Expectation

In the preceding chapter, we have seen that a natural way to approximate labelled Markov processes is to reduce the complexity of the state space by averaging over a smaller σ-algebra. To do this, we needed to use the conditional expectation operator. If the state space is a probability space (X, Σ, p), we have, for any sub-σ-algebra $\Lambda \subseteq \Sigma$, a linear operator $\mathbb{E}_\Lambda : L_1(X, \Sigma) \longrightarrow L_1(X, \Lambda)$. Note that we used this operator to approximate the functions $\tau_a(x, B)$, for fixed a and B; these functions are bounded and positive. Hence, we truly only need the operator to act on the cone $L_\infty^+(X, \Sigma)$. Indeed, this linear operator is positive and thus can be restricted to a linear, order-continuous map of cones $\mathbb{E}_\Lambda : L_\infty^+(X, \Sigma) \longrightarrow L_\infty^+(X, \Lambda)$. Furthermore, as we have seen, a Markov kernel $\hat{\tau}$ on (X, Σ, p) can be cast as an abstract Markov kernel τ on X if it is nonsingular. This abstract Markov kernel is a map $\tau : L_\infty^+(X, \Sigma) \longrightarrow L_\infty^+(X, \Sigma)$, where, for any $B \in \Sigma$, $\tau(\mathbf{1}_B)$ is interpreted as the function $\hat{\tau}(x, B)$. Hence, postcomposing an abstract Markov kernel with the conditional expectation operator is precisely the approximation scheme of [DDLP06]. We shall formalize this point of view in this chapter.

3.1 Labelled abstract Markov processes

We begin by redefining our main object of study, labelled Markov processes, by using Markov operators.

Definition 3.1.1 *A* labelled abstract Markov process *is a probability space* (X, Σ, p) *on which we define a family of abstract Markov kernels* τ_a*, indexed by a finite or countable set* $\mathcal{A}$*, called the set of actions.*

In this chapter, we shall consider, for simplicity, labelled abstract Markov processes with only one action; we shall call these abstract Markov processes, or AMP's. We shall also, for readability, abuse the definitions slightly and call an abstract Markov kernel on X an AMP as well.

As hinted in the above introduction, given an AMP τ on (X, Σ, p), one can obtain an AMP $\Lambda(\tau)$ on (X, Λ, p), with $\Lambda \subseteq \Sigma$, by precomposing by the inclusion map $i : L_\infty^+(X, \Lambda, p) \longrightarrow L_\infty^+(X, \Sigma, p)$ and postcomposing by the conditional expectation map $\mathbb{E}_\Lambda$:

$$\begin{array}{ccc} L_\infty^+(X,\Lambda,p) & \xrightarrow{\Lambda(\tau)} & L_\infty^+(X,\Lambda,p) \\ \downarrow i & & \uparrow \mathbb{E}_\Lambda \\ L_\infty^+(X,\Sigma,p) & \xrightarrow{\tau} & L_\infty^+(X,\Sigma,p) \end{array}$$

So one may ask, how are the dynamics of this "projected" AMP related to the dynamics of the original AMP? To answer this question, we will first study the conditional expectation map in greater generality. The following work is based on [DP06].

3.2 Three functors

We will restrict our attention to two subcategories of **Prb**. Note that the restriction to probability spaces is not necessary, only a finite measure on the spaces considered is required. The results extend trivially to this more general case; however, since the subsequent work will be done in probability spaces, this restriction will improve readability.

Definition 3.2.1 *The category* $\mathbf{Rad}_1$ *has as objects probability spaces, and as arrows* $\alpha : (X, p) \longrightarrow (Y, q)$ *measurable maps such that* $M_\alpha(p) \ll q$

Definition 3.2.2 *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.

Note that $\mathbf{Rad}_\infty$ is a subcategory of $\mathbf{Rad}_1$.

Given an arrow $\alpha : (X, p) \longrightarrow (Y, q)$ in $\mathbf{Rad}_1$, we define a measurable function $d(\alpha)$ on Y as the density of the image measure of p onto Y:

$$d(\alpha) = \frac{\mathrm{d}M_\alpha(p)}{\mathrm{d}q}$$

The Radon-Nikodym derivative exists because $M_\alpha(p) \ll q$. Note that we have that $d(\alpha) \in L_1^+(Y)$. Furthermore, if α is an arrow in $\mathbf{Rad}_\infty$, we have that $d(\alpha) \in L_\infty^+(Y)$. This explains our choice of names for the categories.

Remark 3.2.3 *If* $\alpha : (X, p) \longrightarrow (Y, q)$ *is an arrow in* $\mathbf{Rad}_1$*, we may define a Markov kernel* τ_α *as in example 2.3.2. This Markov kernel is nonsingular if and only if* α *is in* $\mathbf{Rad}_1$.

We begin with some easy lemmas:

Lemma 3.2.4 *For all* $\alpha : (X, p) \longrightarrow (Y, q)$ *in* $\mathbf{Prb}$, $f \in L_1^+(Y, q)$ *and* $B \in \Sigma_Y$,

$$(f \circ \alpha) \cdot \mathbf{1}_{\alpha^{-1}(B)} = (f \cdot \mathbf{1}_B) \circ \alpha$$

The proof is a simple verification.

Lemma 3.2.5 (Change of variables) *Suppose* $\alpha : (X, p) \longrightarrow (Y, q)$ *is in* $\mathbf{Rad}_\infty$ *and* $\mu \in \mathcal{M}^{\leq Kp}(X)$*. Then for all* $g \in L_1^+(Y, q)$,

$$\int_Y g \,\mathrm{d}(M_\alpha(\mu)) = \int_X g \circ \alpha \,\mathrm{d}\mu$$

and both integrals are finite.

Proof . Let B be a measurable set in Y. Then

$$\begin{aligned}
\int_Y \mathbf{1}_B \,\mathrm{d}(M_\alpha(\mu)) &= \int_B \mathrm{d}(M_\alpha(\mu)) \\
&= M_\alpha(\mu)(B) \\
&= \mu\left(\alpha^{-1}(B)\right) \\
&= \int_X \mathbf{1}_{\alpha^{-1}(B)} \,\mathrm{d}\mu \\
&= \int_X \mathbf{1}_B \circ \alpha \,\mathrm{d}\mu
\end{aligned}$$

Thus both integrals of the statement are equal by passing through simple functions.

To show the integrals are finite, note that $\frac{\mathrm{d}\mu}{\mathrm{d}p} \in L_\infty^+(X)$. Thus, we have

$$\begin{aligned}
\int_X g \circ \alpha \,\mathrm{d}\mu &= \int_X (g \circ \alpha) \cdot \frac{\mathrm{d}\mu}{\mathrm{d}p} \,\mathrm{d}p \\
&\leq \left\| \frac{\mathrm{d}\mu}{\mathrm{d}p} \right\|_\infty \int_X g \circ \alpha \,\mathrm{d}p \\
&= \left\| \frac{\mathrm{d}\mu}{\mathrm{d}p} \right\|_\infty \int_X g \,\mathrm{d}M_\alpha(p) \\
&= \left\| \frac{\mathrm{d}\mu}{\mathrm{d}p} \right\|_\infty \int_X g \cdot d(\alpha) \,\mathrm{d}q \\
&\leq \left\| \frac{\mathrm{d}\mu}{\mathrm{d}p} \right\|_\infty \|d(\alpha)\|_\infty \|g\|_1
\end{aligned}$$

and we are done ∎

The above lemma has a "dual" lemma for α in $\mathbf{Rad}_1$:

Lemma 3.2.6 *Suppose $\alpha : (X,p) \longrightarrow (Y,q)$ is in $\mathbf{Rad}_1$ and $\mu \in \mathcal{M}^{\ll p}(X)$. Then for all $g \in L_\infty^+(Y,q)$,*

$$\int_Y g \,\mathrm{d}(M_\alpha(\mu)) = \int_X g \circ \alpha \,\mathrm{d}\mu$$

and both integrals are finite.

We will now define three functors from $\mathbf{Rad}_\infty$ to $\omega\mathbf{CC}$. Let us fix a map $\alpha : (X,p) \longrightarrow (Y,q)$ in $\mathbf{Rad}_\infty$. Let $f \in L_1^+(Y)$. Then the map $f \circ \alpha$ is in $L_1^+(X)$, as provided by lemma 3.2.5. In fact, one can show that the map $(-)\circ\alpha : L_1^+(Y) \longrightarrow L_1^+(X)$ is a linear ω-continuous map of cones. Hence we have a contravariant functor from $\mathbf{Rad}_\infty$ to $\omega\mathbf{CC}$ which sends a probability space (X,p) to $L_1^+(X)$ and sends α to $(-)\circ\alpha$. We shall call this functor the precomposition functor.

Composing the precomposition functor with the dual functor * on $\omega\mathbf{CC}$ yields another functor from $\mathbf{Rad}_\infty$ to $\omega\mathbf{CC}$, which is covariant and gives us an arrow $((-)\circ\alpha)^* : L_1^{+,*}(X) \longrightarrow L_1^{+,*}(Y)$. Concretely, given a functional $M \in L_1^{+,*}(X)$ and a function $g \in L_1^+(Y)$, we have $(((-)\circ\alpha)^*(M))(g) = M(((-)\circ\alpha)(g)) = M(g\circ\alpha)$.

Given a measure $\mu \in \mathcal{M}^{\leq Kp}(X)$, we can consider the image measure $M_\alpha(\mu)$ onto Y. Then given B a measurable set in Y, there are constants K and $\tilde{K}$ such that we have

$$M_\alpha(\mu)(B) = \mu(\alpha^{-1}(B)) \leq Kp(\alpha^{-1}(B)) = KM_\alpha(p)(B) \leq K\cdot\tilde{K}q(B)$$

since we know $M_\alpha(p) \leq \tilde{K}q$. We thus have a map from $\mathcal{M}^{\leq Kp}(X)$ to $\mathcal{M}^{\leq Kq}(Y)$. This map is linear and order-continuous. Thus, the image measure operation M_- is a functor from $\mathbf{Rad}_\infty$ to $\omega\mathbf{CC}$.

Lemma 3.2.5 and proposition 2.4.4 can be combined in the following commutative diagram, showing that the image measure functor and the dual of the precomposition functor are naturally isomorphic. Recall that we denoted by $\phi(L)$ the measure obtained from a functional L; likewise, the functional obtained from a measure μ is denoted $\phi^{-1}(\mu)$

$$\begin{array}{ccc} \mathcal{M}^{\leq Kp}(X) & \underset{\phi^{-1}}{\overset{\phi}{\leftrightarrows}} & L_1^{+,*}(X,p) \\ \downarrow{\scriptstyle M_\alpha(-)} & & \downarrow{\scriptstyle ((-)\circ\alpha)^*} \\ \mathcal{M}^{\leq Kq}(Y) & \underset{\phi}{\overset{\phi^{-1}}{\rightleftarrows}} & L_1^{+,*}(Y,q) \end{array}$$

The middle square commutes by 3.2.5, and the two horizontal pairs of arrows are mutual inverses which also act as the arrows of a natural isomorphism, by proposition 2.4.4.

Note that all of the above functors can equally be defined one $\mathbf{Rad}_1$; we use lemmas 3.2.6 and 2.4.5 instead. We let the precomposition functor map the probability spaces of $\mathbf{Rad}_1$ to the cones L_∞^+, thus making the arrow $((-)\circ\alpha)^*$ go from $L_\infty^{+,*}(X,p)$ to $L_\infty^{+,*}(Y,q)$. We also consider the greater cones of measures $\mathcal{M}^{\ll p}(X)$ and $\mathcal{M}^{\ll q}(Y)$ instead. We get that the following diagram commutes:

$$\begin{array}{ccc} \mathcal{M}^{\ll p}(X) & \underset{\phi^{-1}}{\overset{\phi}{\leftrightarrows}} & L_\infty^{+,*}(X,p) \\ \big\downarrow {\scriptstyle M_\alpha(-)} & & \big\downarrow {\scriptstyle ((-)\circ\alpha)^*} \\ \mathcal{M}^{\ll q}(Y) & \underset{\phi}{\overset{\phi^{-1}}{\rightleftarrows}} & L_\infty^{+,*}(Y,q) \end{array}$$

3.3 The functor $\mathbb{E}_{(-)}$

We define a functor which will generalize conditional expectation. Let us define, for a map $\alpha : (X,p) \longrightarrow (Y,q)$ in $\mathbf{Rad}_\infty$, an operator $\mathbb{E}_\alpha : L_\infty^+(X,p) \longrightarrow L_\infty^+(Y,q)$, as follows: $\mathbb{E}_\alpha(f) = \frac{\mathrm{d}M_\alpha(f \rhd p)}{\mathrm{d}q}$. As α is in $\mathbf{Rad}_\infty$, we have that $M_\alpha(p) \leq Kq$, for some constant K, which is equivalent to $M_\alpha(f \rhd p) \leq K_f\, q$ for all $f \in L_\infty^+(X,p)$ (where obviously K_f depends on f); thus the Radon-Nikodym derivative is defined and is in $L_\infty^+(X,p)$. That is, the following diagram commutes by definition:

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

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

$\mathbb{E}_{(-)}$ can similarly be defined on $\mathbf{Rad}_1$ as follows:

$$\begin{array}{ccc}
L_1^+(X,p) & \xrightarrow[\rhd p]{} & \mathcal{M}^{\ll p}(X) \\
\Big\downarrow{\scriptstyle \mathbb{E}_\alpha} & & \Big\downarrow{\scriptstyle M_\alpha(-)} \\
L_1^+(Y,q) & \xleftarrow{\frac{\mathrm{d}}{\mathrm{d}q}} & \mathcal{M}^{\ll q}(Y)
\end{array}$$

From the definition of $\mathbb{E}_{(-)}$, we obtain that $\mathbb{E}_{\alpha\circ\beta} = \mathbb{E}_\alpha \circ \mathbb{E}_\beta$, which can be seen from this diagram:

$$\begin{array}{ccc}
L_1^+(X,p) & \underset{\rhd p}{\overset{\frac{\mathrm{d}}{\mathrm{d}p}}{\leftrightarrows}} & \mathcal{M}^{\ll p}(X) \\
\Big\downarrow{\scriptstyle \mathbb{E}_\alpha} & & \Big\downarrow{\scriptstyle M_\alpha(-)} \\
L_1^+(Y,q) & \underset{\rhd q}{\overset{\frac{\mathrm{d}}{\mathrm{d}q}}{\leftrightarrows}} & \mathcal{M}^{\ll q}(Y) \\
\Big\downarrow{\scriptstyle \mathbb{E}_\beta} & & \Big\downarrow{\scriptstyle M_\beta(-)} \\
L_1^+(Z,r) & \underset{\rhd r}{\overset{\frac{\mathrm{d}}{\mathrm{d}r}}{\leftrightarrows}} & \mathcal{M}^{\ll r}(Z)
\end{array}$$

The two squares commute by definition of $\mathbb{E}_{(-)}$, and the pairs of arrows coming from the Radon-Nikodym theorem are mutual inverses. Thus $\mathbb{E}_{(-)}$ is compatible with composition by functoriality of the image measure.

To show that $\mathbb{E}_{(-)}$ is a functor, we need to show that it is an order-continuous linear map of cones. But this is trivial as it is the composition of three order-continuous linear maps of cones.

The following diagram combines the information we have until now.

Proposition 3.3.1 *Given $\alpha \in \mathbf{Rad}_\infty$, the following commutes:*

$$\begin{array}{ccccc}
L_\infty^+(X,p) & \underset{\rhd p}{\overset{\frac{\mathrm{d}}{\mathrm{d}p}}{\leftrightarrows}} & \mathcal{M}^{\leq Kp}(X) & \underset{\phi^{-1}}{\overset{\phi}{\leftrightarrows}} & L_1^{+,*}(X,p) \\
\Big\downarrow{\scriptstyle \mathbb{E}_\alpha} & & \Big\downarrow{\scriptstyle (M_\alpha(-))^*} & & \Big\downarrow{\scriptstyle ((-)\circ\alpha)^*} \\
L_\infty^+(Y,q) & \underset{\rhd q}{\overset{\frac{\mathrm{d}}{\mathrm{d}q}}{\leftrightarrows}} & \mathcal{M}^{\leq Kq}(Y) & \underset{\phi}{\overset{\phi^{-1}}{\rightleftarrows}} & L_1^{+,*}(Y,q)
\end{array}$$

Thus the arrows corresponding to the Radon-Nikodym theorem act as a natural isomorphism between the functors $\mathbb{E}_{(-)}$ and $M_{(-)}$. We can summarize this as a theorem.

Theorem 3.3.2 *The three functors $\mathbb{E}_{(-)}$, $M_{(-)}$, and the dual of the precomposition functor, are all naturally isomorphic.*

This allows us to situate the operation of conditional expectation in a categorical setting in a pleasing fashion.

We state an important identity about the conditional expectation functor:

Proposition 3.3.3 *Consider $\alpha \in \mathbf{Rad}_\infty$ and $\mathbb{E}_\alpha : L^+_\infty(X,p) \longrightarrow L^+_\infty(Y,q)$. Then for all f in $L^+_\infty(X,p)$, we have that*

$$\int \mathbb{E}_\alpha(f) \cdot (-)\, \mathrm{d}q = \int f \cdot (- \circ \alpha)\, \mathrm{d}p$$

as functionals in $L^{+,*}_1(Y,q)$.

Alternatively, for all f in $L^+_\infty(X,p)$ and u in $L^+_1(Y,q)$, we have that

$$\langle \mathbb{E}_\alpha(f), u \rangle = \langle f, u \circ \alpha \rangle$$

and thus $\mathbb{E}_\alpha$ is the adjoint of the precomposition map $- \circ \alpha$ from $L^+_\infty(Y,q)$ to $L^+_\infty(X,p)$

The proof lies in the commutativity of the outside of the diagram of proposition 3.3.1

As above, we can slightly modify these functors to map the category $\mathbf{Rad}_1$ to $\omega\mathbf{CC}$, and the following commutes as well:

$$\begin{array}{ccccc}
L^+_1(X,p) & \underset{\triangleright p}{\overset{\frac{\mathrm{d}}{\mathrm{d}p}}{\leftrightarrows}} & \mathcal{M}^{\ll p}(X) & \underset{\phi^{-1}}{\overset{\phi}{\leftrightarrows}} & L^{+,*}_\infty(X,p) \\
\downarrow{\scriptstyle \mathbb{E}_\alpha} & & \downarrow{\scriptstyle M_\alpha(-)} & & \downarrow{\scriptstyle ((-)\circ\alpha)^*} \\
L^+_1(Y,q) & \underset{\triangleright q}{\overset{\frac{\mathrm{d}}{\mathrm{d}q}}{\leftrightarrows}} & \mathcal{M}^{\ll q}(Y) & \underset{\phi}{\overset{\phi^{-1}}{\rightleftarrows}} & L^{+,*}_\infty(Y,q)
\end{array}$$

3.4 Some operator norms

One last detail that needs to be tied up is the calculation of the norm of some operators. We begin with some preliminary results.

Lemma 3.4.1 *For all* $\alpha : (X, p) \longrightarrow (Y, q),\ f \in L_1^+(Y, q)$,

$$M_\alpha((f \circ \alpha) \triangleright p) = f \triangleright (M_\alpha(p))$$

Proof . Given $B \in \Sigma_Y$, we have

$$\begin{aligned}
M_\alpha((f \circ \alpha) \triangleright p)(B) &= ((f \circ \alpha) \triangleright p)(\alpha^{-1}(B)) && \text{(by definition)} \\
&= \int (f \circ \alpha) \cdot \mathbf{1}_{\alpha^{-1}(B)} \, \mathrm{d}p && \text{(again by definition)} \\
&= \int (f \cdot \mathbf{1}_B) \circ \alpha \, \mathrm{d}p && \text{(by lemma 3.2.4)} \\
&= \int f \cdot \mathbf{1}_B \, \mathrm{d}(M_\alpha(p)) && \text{(by change of variable)} \\
&= (f \triangleright M_\alpha(p))(B) && \text{(by definition)}
\end{aligned}$$

∎

Proposition 3.4.2 *Given* f *in* $L_1(Y, q)$, $\mathbb{E}_\alpha(f \circ \alpha) = d(\alpha) \cdot f$

Proof . Using above lemma,

$$\begin{aligned}
\mathbb{E}_\alpha(f \circ \alpha) &= \frac{\mathrm{d}M_\alpha(f \circ \alpha \triangleright p)}{\mathrm{d}q} \\
&= \frac{\mathrm{d}f \triangleright M_\alpha(p)}{\mathrm{d}q} \\
&= \frac{\mathrm{d}f \triangleright M_\alpha(p)}{\mathrm{d}M_\alpha(p)} \cdot \frac{\mathrm{d}M_\alpha(p)}{\mathrm{d}q} \\
&= f \cdot \frac{\mathrm{d}M_\alpha(p)}{\mathrm{d}q}
\end{aligned}$$

∎

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

Lemma 3.4.3 $\mathbb{E}_\alpha : L_\infty^+(X) \longrightarrow L_\infty^+(Y)$ *has norm* $\|d(\alpha)\|_\infty$.

Proof . Let $f \in L^+_\infty(X)$. Then $f \leq_p \|f\|_\infty \cdot \mathbf{1}_X$, and so by monotonicity of $\mathbb{E}_\alpha$, $\mathbb{E}_\alpha(f) \leq_q \|f\|_\infty d(\alpha)$, and so $\|\mathbb{E}_\alpha(f)\|_\infty \leq \|f\|_\infty \cdot \|d(\alpha)\|_\infty$, and so $\|\mathbb{E}_\alpha\| \leq \|d(\alpha)\|_\infty$.

On the other hand, we have that $\mathbb{E}_\alpha(\mathbf{1}_X) = d(\alpha)$ and thus the norm of $\mathbb{E}_\alpha$ is exactly $\|d(\alpha)\|_\infty$. ∎

Lemma 3.4.4 *The map $(-) \circ \alpha : L^+_\infty(Y, q) \longrightarrow L^+_\infty(X, p)$ has norm 1.*

Proof . Let $g \in L^+_\infty(Y)$. Suppose $\|g\|_\infty = M$. Let $N > M$. Then we have

$$p(g \circ \alpha > N) = p(\alpha^{-1}(g^{-1}(N, \infty))) = M_\alpha(p)(g^{-1}(N, \infty))$$

But we have that $M_\alpha(p) \leq Kq$ for some constant K, and so

$$M_\alpha(p)(g^{-1}(N, \infty)) \leq Kq(g^{-1}(N, \infty)) = 0$$

since $N > \|g\|_\infty$. Thus we have that $\|g \circ \alpha\|_\infty \leq \|g\|_\infty$, and so $\|(-) \circ \alpha\| \leq 1$. But then $\mathbf{1}_Y \circ \alpha = \mathbf{1}_X$ and so the norm is exactly 1. ∎

3.5 The Approximation Map

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

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

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

The map $(-) \circ \alpha$ can be considered an abstract Markov kernel; the map $\mathbb{E}_\alpha$ is an abstract Markov kernel if $\|d(\alpha)\|_\infty \leq 1$. This is actually a very restrictive condition on α, as the following lemma shows:

Lemma 3.5.1 *If* $\alpha : (X, p) \longrightarrow (Y, q)$ *is in* $\mathbf{Rad}_\infty$ *and* $\|d(\alpha)\|_\infty \leq 1$*, then* $d(\alpha) =_q \mathbf{1}_Y$*.*

Proof . We have that

$$\int_Y d(\alpha)\,\mathrm{d}q = \int_Y \mathrm{d}M_\alpha(q) = \int_X \mathbf{1}_Y \circ \alpha\,\mathrm{d}p = \int_X \mathbf{1}_X\,\mathrm{d}p = p(X) = 1$$

As $\|d(\alpha)\|_\infty \leq 1$, $\mathbf{1}_Y - d(\alpha) \geq_q 0$. Also,

$$\int_Y \mathbf{1}_Y - d(\alpha)\,\mathrm{d}q = q(Y) - 1 = 0$$

And thus $\mathbf{1}_Y - d(\alpha) =_q 0$ ∎

And thus if α is used in the context of AMPs, we have that $M_\alpha(p) = q$.

CHAPTER 4
Bisimulation

As with labelled transition systems or labelled Markov processes, we must define a notion of bisimulation for abstract Markov processes. The strategy is to examine the definitions that we already have for labelled Markov processes and to translate them in the context of AMPs. Before proceeding with this, we will first show that any labelled Markov process can be cast into a labelled abstract Markov process, as it ensures that our abstract setting is worth studying.

4.1 From LMPs to AMPs

In order to take a labelled Markov process and change it into a labelled abstract Markov process, we need to define a measure on the state space such that every Markov kernel in the LMP is nonsingular.

We begin with a few definitions.

Definition 4.1.1 *Given a set $\mathcal{A}$, we define $\mathcal{A}^*$ to be the free monoid on $\mathcal{A}$.*

So $\mathcal{A}^*$ is nothing but the set of finite words over $\mathcal{A}$. In our case, the set $\mathcal{A}$ is finite or countable set, as it is the set of actions; thus $\mathcal{A}^*$ is countable as well. If w and v are two words in $\mathcal{A}^*$, we let $w \cdot v$ be their concatenation. We let ϵ be the empty word. We now define a certain class of measures on $\mathcal{A}^*$, on which we put the powerset σ-algebra. Let us first define, for all words $v \in \mathcal{A}^*$, a function $f_v : \mathcal{A}^* \longrightarrow \mathcal{A}^*$ as $f_v(w) = w \cdot v$. Thus f_v just appends the word v to its argument.

Definition 4.1.2 *A measure m on $\mathcal{A}^*$ is* forward-closed *if $M_{f_v}(m) \ll m$ for every $v \in \mathcal{A}^*$. In other words, m is forward-closed if and only if for all words w such that $m(w) > 0$, we have that $m(w \cdot v) > 0$ for all $v \in \mathcal{A}^*$.*

Thus a measure is forward-closed if, for any word w to which we give positive measure, any word with w as a prefix will have positive measure. Note that since $\mathcal{A}^*$ is countable, one can have a finite measure on $\mathcal{A}^*$ such that every word has nonzero measure.

Let (X, Σ, τ_a) be a LMP. Recall from section 2.5 that the operator $\bar{T}_{\tau_a}$ is an operator on the finite measures on (X, Σ) which transforms measures "forwards in time". Let $w = v \cdot a$ be a word in $\mathcal{A}^*$, with $a \in \mathcal{A}$. We recursively define operators on the space $\mathcal{M}(X)$ of finite measures on X:

$$\bar{T}_\epsilon = id$$

$$\bar{T}_w = \bar{T}_{v \cdot a} = \bar{T}_{\tau_a} \circ \bar{T}_v$$

where id is the identity operator on $\mathcal{M}(X)$. Thus $\bar{T}_w(\mu)$ is just the measure μ transformed through the actions in the word w in the usual left-to-right order.

Proposition 4.1.3 *Let (X, Σ, τ_a) be a LMP. Let m be a finite forward-closed measure on $\mathcal{A}^*$. Let μ be a finite measure on X.*

We define a measure $K_{\mu,m}$ on X as follows. If $B \in \Sigma$, we let

$$K_{\mu,m}(B) = \int_{\mathcal{A}^*} \bar{T}_w(\mu)(B) \,\mathrm{d}m(w) = \sum_{w \in \mathcal{A}^*} m(w) \bar{T}_w(\mu)(B)$$

Then $K_{\mu,m}$ is a finite measure on X, and τ_a is nonsingular with respect to $K_{\mu,m}$ for every $a \in \mathcal{A}$

Proof . First of all, the finiteness of $K_{\mu,m}$ is immediate. Indeed, it is clear from the definition of $\bar{T}_{\tau_a}$ that $\bar{T}_{\tau_a}(\mu)(X) \leq \mu(X)$. Thus we have that $K_{\mu,m}(B) \leq \mu(X)m(\mathcal{A}^*)$.

To show that τ_a is nonsingular, we must show that, for all $A \in \Sigma$ and $a \in \mathcal{A}$,

$$K_{\mu,m}(A) = 0 \Rightarrow \tau_a(x, A) =_{K_{\mu,m}} 0$$

It is easier to show the contrapositive. Suppose we have $A \in \Sigma$ and an action $a \in \mathcal{A}$ such that $\tau_a(x, A) \neq_{K_{\mu,m}} 0$. Then there is a real number $\delta > 0$ and a set $B_\delta \in \Sigma$ such that $\tau_a(x, A) > \delta$ for $x \in B_\delta$, and such that $K_{\mu,m}(B_\delta) > 0$. This is clear as we can approximate $\tau_a(x, A)$ by simple functions.

We want to show that $K_{\mu,m}(A) > 0$. As $K_{\mu,m}(B_\delta) > 0$, we must have that $m(w)\bar{T}_w(\mu)(B_\delta) > 0$ for some word $w \in \mathcal{A}^*$, by the definition of $K_{\mu,m}$. Hence $m(w) > 0$ and $\bar{T}_w(\mu)(B_\delta) > 0$. But then, we have that

$$\begin{aligned}
\bar{T}_{w \cdot a}(\mu)(A) &= \left(\bar{T}_{\tau_a} \circ \bar{T}_w(\mu)\right)(A) \\
&= \bar{T}_{\tau_a}\left(\bar{T}_w(\mu)\right)(A) \\
&= \int_X \tau_a(x, A)\, \mathrm{d}\bar{T}_w(\mu)(x) \\
&\geq \int_{B_\delta} \tau_a(x, A)\, \mathrm{d}\bar{T}_w(\mu)(x) \\
&\geq \int_{B_\delta} \delta\, \mathrm{d}\bar{T}_w(\mu)(x) \\
&\geq \delta \int_{B_\delta} \mathrm{d}\bar{T}_w(\mu)(x) \\
&= \delta\, \bar{T}_w(\mu)(B_\delta) \\
&> 0
\end{aligned}$$

Furthermore, as $m(w) > 0$, we have that $m(w \cdot a) > 0$ as m is forward-closed, and so $m(w \cdot a)\bar{T}_{w \cdot a}(\mu)(A) > 0$. This expression is a term in the defining sum for $K_{\mu,m}(A)$ and thus $K_{\mu,m}(A) > 0$ and we are done ∎

The measure $K_{\mu,m}$ can be interpreted as some weighing of all the "reachable states" starting from a measure μ, with a weighing m defined on the possible sequences of actions the user can pick.

Corollary 4.1.4 *Given any LMP (X, Σ, τ_a) and measure μ on X, there exists a probability measure p_μ on X making this LMP into an AMP. Also, one can ensure that $\mu \ll p_\mu$.*

Proof . Take any finite forward-closed measure m on $\mathcal{A}^*$. Define p_μ as the measure $K_{\mu,m}$ normalized to be a probability measure. By the above theorem, τ_a is nonsingular on the probability space $(X, \Sigma, K_{\mu,m})$ and thus the Markov kernels correspond uniquely to abstract Markov kernels on X.

If one wants to impose $\mu \ll p_\mu$, pick m such that $m(\epsilon) > 0$. Thus every word, including the empty word, has nonzero measure in $\mathcal{A}^*$. Thus, for any $A \in \Sigma$, the term $\bar{T}_\epsilon(\mu)(A)$ is a term in the defining sum for $K_{\mu,m}(A)$; but $\bar{T}_\epsilon$ is the identity transformation on measures, and so $\bar{T}_\epsilon(\mu)(A) = \mu(A)$. Hence if $\mu(A) > 0$, we have that $K_{\mu,m}(A) > 0$, and thus $p_\mu(A) > 0$. ∎

4.2 Event bisimulation and zigzags

Recall that for a given LMP $(X, \Sigma, \hat{\tau})$ (with a single action), a sub-σ-algebra $\Lambda \subseteq \Sigma$ is an event-bisimulation if $\hat{\tau}(x, B)$ is Λ-measurable for all B in Λ. In the language of abstract Markov processes, if we have an AMP (X, Σ, p, τ), this definition implies that $\tau : L^+_\infty(X, \Sigma, p) \longrightarrow L^+_\infty(X, \Sigma, p)$ sends the subspace $L^+_\infty(X, \Lambda, p)$ to itself, so that the following commutes:

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

Indeed, the definition of event-bisimulation precisely says that $\tau(\mathbf{1}_B)$ is Λ-measurable for all $B \in \Lambda$. As any function in $L^+_\infty(X, \Lambda)$ can be is the increasing pointwise limit of simple functions, and thanks to the order-continuity of τ, we get the above conclusion.

A generalization to the above would be a map $\alpha : (X, \Sigma, p) \longrightarrow (Y, \Lambda, q)$ in the category $\mathbf{Rad}_\infty$, with X and Y respectively equipped with AMPs τ

and ρ, such that the following commutes:

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

Suppose that the abstract Markov kernels ρ and τ come from nonsingular Markov kernels $\hat{\rho}$ and $\hat{\tau}$, respectively. Using the above diagram, we get, for $B \in \Lambda$:

$$\begin{aligned} \hat{\tau}\left(x, \alpha^{-1}(B)\right) &= \tau\left(\mathbf{1}_{\alpha-1(B)}\right)(x) \\ &= \tau\left(\mathbf{1}_B \circ \alpha\right)(x) \\ &= \left(\rho\left(\mathbf{1}_B\right) \circ \alpha\right)(x) \\ &= \rho\left(\mathbf{1}_B\right)(\alpha(x)) \\ &= \hat{\rho}\left((\alpha(x)), B\right) \end{aligned}$$

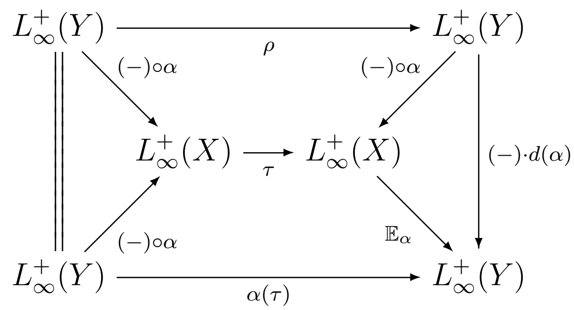
And thus the function α is a zigzag in the sense of definition 2.3.4. We can thus make the following definition:

Definition 4.2.1 *Given two labelled AMPs* (X, Σ, p, τ_a) *and* (Y, Λ, q, ρ_a), *a map* $\alpha : (X, \Sigma, p) \longrightarrow (Y, \Lambda, q)$ *in the category* $\mathbf{Rad}_\infty$ *is a* zigzag *if the following diagram commutes for every action a:*

$$\begin{array}{ccc} L_\infty^+(X,p) & \xrightarrow[\tau_a]{} & L_\infty^+(X,p) \\ \uparrow{\scriptstyle (-)\circ\alpha} & & \uparrow{\scriptstyle (-)\circ\alpha} \\ L_\infty^+(Y,q) & \xrightarrow[\rho_a]{} & L_\infty^+(Y,q) \end{array}$$

We shall now consider AMPs with a single action for simplicity.

Note that if there is a zigzag from (X, Σ, p, τ) to (Y, Λ, q, ρ), then ρ is very closely related to the projection of τ onto Y via α (which we defined in the previous chapter). Indeed, we have the following diagram:



We have that $\mathbb{E}_\alpha(f \circ \alpha) = f \cdot d(\alpha)$ from lemma 3.4.2. This implies that $\alpha(\tau) = \rho \cdot d(\alpha)$. In particular, if $d(\alpha) = \mathbf{1}_Y$ — which happens if $M_\alpha(p) = q$ — then ρ is equal to $\alpha(\tau)$, the projection of τ onto Y. Note that the condition $M_\alpha(p) = q$ means that the image measure is precisely the measure in the codomain of α. However, we have already shown that for the projection of an AMP by a map α to make sense, we required that $\|d(\alpha)\|_\infty \leq 1$, which implied, by lemma 3.5.1, that $d(\alpha) = \mathbf{1}_Y$. We will now presuppose this condition in the forthcoming analysis.

4.3 The category AMP

In the previous section, we saw that zigzags and projections coincided perfectly given that the map between the state spaces was particularly well-behaved.

Definition 4.3.1 *A map $\alpha : (X, p) \longrightarrow (Y, q)$ in* **Prb** *is said to be* measure-preserving *if $M_\alpha(p) = q$.*

Let **MMPM** be the category of probability spaces and measure-preserving maps. It is a subcategory of **Prb** and of $\mathbf{Rad}_\infty$.

In effect, this ensures that the map α is essentially surjective. However, there is no reason why we would consider essentially surjective maps which are not surjective in the usual sense, except maybe to keep an artificial and unnecessary sense of generality. Furthermore, requiring surjectivity makes many of the forthcoming mathematical arguments much easier to follow.

The surjective measure-preserving maps form a further subcategory of **MMPM**. We will augment this category with additional structure relevant to our situation.

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

4.4 Bisimulation defined on AMP

As we have discussed when we introduced zigzags for LMPs, it should be noticed that surjective measure-preserving maps between probability spaces typically involve information loss. Thus, we define a preorder on **AMP** as follows: given two AMPs (X, Σ, p, τ) and (Y, Λ, q, ρ), we say that $Y \preceq X$ if there is an arrow $\alpha : (X, \Sigma, p, \tau) \longrightarrow (Y, \Lambda, q, \rho)$ in **AMP**. This preorder will allow us to formalize a few concepts.

Reconsider example 2.3.6. We may consider two arbitrarily complicated AMPs (X, Σ, p, τ) and (Y, Λ, q, ρ) such that $\tau(\mathbf{1}_X) = \mathbf{1}_X$ and $\rho(\mathbf{1}_Y) = \mathbf{1}_Y$. The dynamics on the state spaces may be very complex, but the above conditions on τ and ρ imply that the action is never disabled. Thus, in the spirit of example 2.3.6, they should both be bisimilar to a one-point space with a trivial AMP defined on it. However, it appears ludicrous and counterintuitive that one should be able to construct a span of zigzags as in definition 2.3.5; this would require the construction of a third state space U which would somehow weave together the complex behaviors of the AMPs X and Y. However, both AMPs X and Y have a zigzag to a one-point space, and this one-point space truly reflects their dynamics with respect to the user. It thus appears to us that the correct definition of a bisimulation is in terms of *cospans* of zigzags morphisms. Such a view of bisimulation was already present in [DDLP06], where LMPs were constructed as coalgebras of an appropriate functor. We thus define bisimulation of AMPs as follows:

Definition 4.4.1 *We say that two objects of* **AMP***,* (X, Σ, p, τ) *and* (Y, Λ, q, ρ)*, are* bisimilar *if there is a third object* (Z, Γ, r, π) *with a pair of zigzags*

$$\alpha : (X, \Sigma, p, \tau) \longrightarrow (Z, \Gamma, r, \pi)$$

$$\beta : (Y, \Lambda, q, \rho) \longrightarrow (Z, \Gamma, r, \pi)$$

making a cospan diagram

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

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

4.5 Bisimulation as an equivalence relation

Ideally, bisimulation would be an equivalence relation on the objects of **AMP**. The following theorem shows that it is the case.

Theorem 4.5.1 *Let*

$$\alpha : (X, \Sigma, p, \tau) \longrightarrow (Y, \Lambda, q, \rho)$$

$$\beta : (X, \Sigma, p, \tau) \longrightarrow (Z, \Gamma, r, \kappa)$$

be a span of zigzags. Then the pushout $(W, \Omega, \lambda, \pi)$ *exists and the pushout maps* $\delta : Y \longrightarrow W$ *and* $\gamma : Z \longrightarrow W$ *are zigzags.*

The proof requires several lemmas.

Lemma 4.5.2 *Let* $\alpha : (X, \Sigma, p) \longrightarrow (Y, \Lambda, q)$ *be a measure-preserving map of probability spaces. Then for all* $h \in L_\infty(X)$, $\mathbb{E}_\alpha(h) \circ \alpha = h \Leftrightarrow h$ *is* $\alpha^{-1}(\Lambda)$*-measurable.*

Proof . The right implication is obvious. For the left implication, note that the function $\mathbb{E}_\alpha(h) \circ \alpha$ is also $\alpha^{-1}(\Lambda)$-measurable. Thus, to show the equality, we must show that for all $B \in \Lambda$,

$$\int_{\alpha^{-1}(B)} \mathbb{E}_\alpha(h) \circ \alpha \, \mathrm{d}p = \int_{\alpha^{-1}(B)} h \, \mathrm{d}p$$

which we can show by the following equality of measures:

$$\begin{aligned} M_\alpha\left((\mathbb{E}_\alpha h \circ \alpha) \rhd p\right) &= \mathbb{E}_\alpha h \rhd M_\alpha(p) && \text{(by lemma 3.4.1)} \\ &= \mathbb{E}_\alpha h \rhd q && \text{(α is measure-preserving)} \\ &= M_\alpha\left(h \rhd p\right) && \text{(by definition of $\mathbb{E}_\alpha$)} \end{aligned}$$

The first and last measures being equal is precisely equivalent to the above equality of integrals. ∎

Lemma 4.5.3 *Let* $\alpha : (X, \Sigma, p, \tau) \longrightarrow (Y, \Lambda, q, \rho)$ *be an arrow in* **AMP**. *Then* α *is a zigzag if and only if* $\mathbb{E}_\alpha(-) \circ \alpha|_{Im(\tau(-\circ\alpha))} = id$, *i.e. if and only if for all* $f \in L^+_\infty(Y)$, $\mathbb{E}_\alpha(\tau(f \circ \alpha)) \circ \alpha = \tau(f \circ \alpha)$.

Proof . If α is a zig-zag, the following diagram commutes:

$$\begin{array}{ccc} L_\infty(X) & \xrightarrow[\tau]{} & L_\infty(X) \\ \uparrow {\scriptstyle (-)\circ\alpha} & & \downarrow {\scriptstyle \mathbb{E}_\alpha} \\ L_\infty(Y) & \xrightarrow[\rho]{} & L_\infty(Y) \\ \downarrow {\scriptstyle (-)\circ\alpha} & & \downarrow {\scriptstyle (-)\circ\alpha} \\ L_\infty(X) & \xrightarrow[\tau]{} & L_\infty(X) \end{array}$$

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

Corollary 4.5.4 *$\alpha : (X, \Sigma, p, \tau) \longrightarrow (Y, \Lambda, q, \rho)$ 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 4.5.5 *If $\alpha : (X, \Sigma, p, \tau) \longrightarrow (Y, \Lambda, q, \rho)$ in* **AMP** *is a zigzag, $\beta : (Y, \Lambda, q, \rho) \longrightarrow (Z, \Gamma, r, \kappa)$ is a map in* **AMP***, and $\gamma = \beta \circ \alpha$ is a zigzag, then β is a zigzag.*

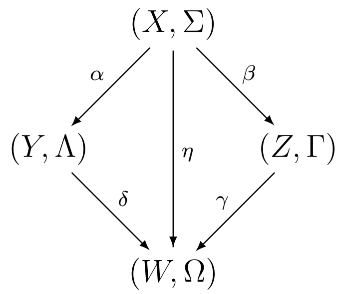
Proof .

$$\begin{aligned} \kappa(f) \circ \beta \circ \alpha &= \kappa(f) \circ \gamma \\ &= \tau(f \circ \gamma) && (\gamma \text{ is a zigzag}) \\ &= \tau(f \circ \beta \circ \alpha) \\ &= \rho(f \circ \beta) \circ \alpha && (\alpha \text{ is a zigzag}) \end{aligned}$$

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

We are now ready to prove the above theorem.

Proof of theorem 4.5.1. It is well-known 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:



Note here that, of course, $\eta = \delta \circ \alpha = \gamma \circ \beta$.

We have to construct a measure on W such that the maps δ and γ are measure preserving (we know that they are surjective by the construction of the pushout in **Set**). Recall that the probability measures on X, Y and Z were called p, q and r, respectively. Let us define on (W, Ω) the measure λ in the obvious way, that is, $\lambda = M_\eta(p)$. Note that by the definition of η, the fact that α is measure-preserving, and the functoriality of M_-, we have

$$\lambda = M_\eta(p) = M_\delta M_\alpha(p) = M_\delta(q)$$

and so we automatically have that δ is measure-preserving. The same can be done with γ.

Finally, we have to construct an AMP on (W, Ω, λ). We take $\pi = \eta(\tau)$, which is the projection of the AMP on X through η. Thus, for all f in $L^+_\infty(W)$, we have $\pi(f) = \mathbb{E}_\eta(\tau(f \circ \eta))$. Note that as $\mathbb{E}_-$ is a functor and α is an arrow in **AMP**, we have $\pi(f) = \mathbb{E}_\delta \mathbb{E}_\alpha(\tau((f \circ \delta) \circ \alpha)) = \mathbb{E}_\delta(\rho(f \circ \delta)) = \delta(\rho)$, and thus δ is an arrow in **AMP** as well. The same argument works for γ.

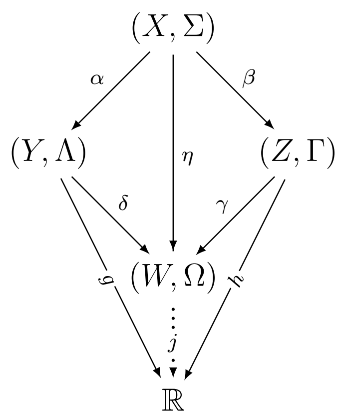
We have:

$$\begin{aligned} \tau(f \circ \eta) &= \tau(f \circ \delta \circ \alpha) \\ &= \rho(f \circ \delta) \circ \alpha && \text{(as } \alpha \text{ is a zigzag)} \end{aligned}$$

and similarly, as β is a zigzag, we have

$$\tau(f \circ \eta) = \kappa(f \circ \gamma) \circ \beta$$

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

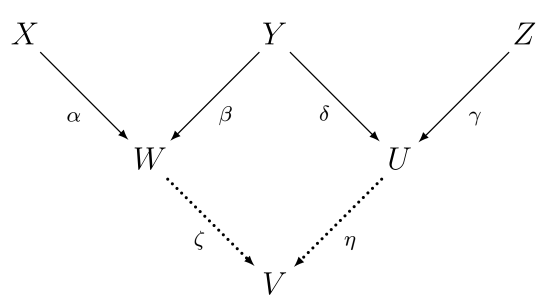


As W is a pushout, there is a unique map $j : W \longrightarrow \mathbb{R}$ such that $\tau(f \circ \eta) = j \circ \eta$. Thus $\tau(f \circ \eta)$ is $\eta^{-1}(\Omega)$-measurable, and so η is a zigzag by corollary 4.5.4.

Finally, by Lemma 4.5.5, δ and γ are zigzags. ∎

Corollary 4.5.6 *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 their state space for clarity. 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:



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. ∎

Note that by defining bisimulation as a cospan, we obtain a coarser relation that if we had defined bisimulation as a span. Indeed, if there is a span of zigzags $f : U \longrightarrow X$ and $g : U \longrightarrow Y$, then by the above theorem, we can construct a pushout with the morphisms being zigzags, thus constructing a cospan of zigzags between X and Y; thus, our definition includes, but is not quite the same as, the previous definitions of bisimulation. However, we believe that our definition is more mathematically pleasing, and easier to work with.

4.6 The smallest bisimulation and logical characterization

Given an AMP (X, Σ, p, τ), one question one may ask is whether there is a "smallest" object $\left(\tilde{X}, \Xi, r, \xi\right)$ in **AMP** such that, for every zigzag from X to another AMP (Y, Λ, q, ρ), there is a zigzag from (Y, Λ, q, ρ) to $\left(\tilde{X}, \Xi, r, \xi\right)$. It can be shown that such an object exists, by generalizing theorem 4.5.1.

Proposition 4.6.1 *Let $\{\alpha_i : (X, \Sigma, p, \tau) \longrightarrow (Y_i, \Lambda_i, q_i, \rho_i)\}$ be the set of all zigzags in* **AMP** *with domain (X, Σ, p, τ). This yields a generalized pushout diagram, and as in Theorem 4.5.1, the pushout $\left(\tilde{X}, \Xi, r, \xi\right)$ exists and the pushout maps are zigzags. We thus obtain zigzags β_i from $(Y_i, \Lambda_i, q_i, \rho_i)$ to $\left(\tilde{X}, \Xi, r, \xi\right)$ and a single zigzag η from (X, Σ, p, τ) to $\left(\tilde{X}, \Xi, r, \xi\right)$.*

Proof . The proof is exactly the same as for Theorem 4.5.1; we use the fact that the category **Mes** is cocomplete. ∎

If one considers the subcategory $\mathbf{ZZ}_X$ of **AMP** consisting of all objects (Y, Λ, q, ρ) such that there is a zigzag from X to Y, together with the zigzag maps between these objects, then $\left(\tilde{X}, \Xi, r, \xi\right)$ is a bottom element of $\mathbf{ZZ}_X$

with respect to the preorder $\preceq$. This object has important uniqueness properties.

Corollary 4.6.2 *Up to isomorphism, the object* $\left(\tilde{X}, \Xi, r, \xi\right)$ *is the unique bottom element of* $\mathbf{ZZ}_X$*. That is, if* (W, Ω, q, ρ) *is another AMP such that there is a zigzag* μ *from* $\tilde{X}$ *to* W*, then* μ *is an isomorphism.*

Proof . Let η be the map from (X, Σ, p, τ) to $\left(\tilde{X}, \Xi, r, \xi\right)$ obtained from the generalized pushout. Then $\mu \circ \eta$ is a zigzag from X to W. Hence $\mu \circ \eta : X \longrightarrow W$ is part of the generalized pushout diagram from X, and so there is a pushout map $\epsilon : W \longrightarrow \hat{X}$ such that $\epsilon \circ \mu \circ \eta = \eta$.

Now η is surjective, hence right-cancellable, and so we have $\epsilon \circ \mu = \mathrm{id}|_{\hat{X}}$. Hence μ is monic, thus injective, as a map of measure spaces. But μ is also surjective, and so $\epsilon = \mu^{-1}$. Thus μ is an isomorphism of measure spaces which trivially extends to an isomorphism of AMPs. ∎

Thus, we can say that $\left(\tilde{X}, \Xi, r, \xi\right)$ is the meet (or infimum) of all objects $(Y_i, \Lambda_i, q_i, \rho_i)$ which are bisimilar to the AMP X, with respect to the preorder $\preceq$. However, this "smallest" object is given in an abstract way; is it possible to capture it constructively?

The following few lemmas indicate the way to answer this question.

Lemma 4.6.3 *Suppose* $\alpha : (X, \Sigma, p, \tau) \longrightarrow (Y, \Lambda, q, \rho)$ *is a map in* **AMP** *such that* $\alpha^{-1}(\Lambda) = \Sigma$*. Then* α *is a zigzag.*

Proof . This is a direct consequence of corollary 4.5.4. 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 4.6.4 *Let* $\alpha : (X, \Sigma, p, \tau) \longrightarrow (Y, \Lambda, q, \rho)$ *be a zigzag. Then* α *factors into two zigzags as follows:* $i_\alpha : (X, \Sigma, p, \tau) \longrightarrow (X, \alpha^{-1}(\Lambda), p, \tau)$*, which is the identity on* X*, reducing the* σ*-algebra; and* $\hat{\alpha} : (X, \alpha^{-1}(\Lambda), p, \tau) \longrightarrow (Y, \Lambda, q, \rho)$

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 4.5.4. ∎

Thus every zigzag from an object (X, Σ, p, τ) in **AMP** yields a sub-σ-algebra Λ of Σ. Also, the bottom element $\tilde{X}$ defined above yields the smallest such sub-σ-algebra. These sub-σ-algebras are event-bisimulations in the sense of [DDLP06]. Recall that in this paper, it was shown that the smallest event-bisimulation can be obtained by a logical characterization. The result can be directly recast in the context of AMPs. Indeed, given an AMP (X, Σ, p, τ), one can consider the nonsingular Markov kernel $\hat{\tau}$ from which the abstract Markov kernel τ comes; the logical characterization result of [DDLP06] applies and can then be brought back into the context of AMPs. We first modify the semantics of $\mathcal{L}$ for AMPs:

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

We shall also write, for a measurable set A, $\left[\!\!\left[\langle a \rangle_q A\right]\!\!\right] = \{s : \tau_a(\mathbf{1}_A) > q\}$

We restate the logical characterization result in our formulation:

Theorem 4.6.5 *Given a labelled AMP (X, Σ, p, 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 $i : (X, \Sigma, p, \tau_a) \longrightarrow (X, \sigma([\![\mathcal{L}]\!]), p, \tau_a)$ 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, given an AMP (X, Σ, p, τ) and a zigzag morphism η to its smallest bisimulation $\left(\tilde{X}, \Xi, r, \xi\right)$, we have that $\eta^{-1}(\Xi) = \sigma([\![\mathcal{L}]\!])$.

CHAPTER 5
Approximations of AMPs

In this section, if the measurable map $i_\Lambda : (X, \Sigma) \longrightarrow (X, \Lambda)$ is the identity on the set X, then the resulting AMP morphism shall be denoted $i_\Lambda : (X, \Sigma, p, \tau) \longrightarrow (X, \Lambda, p, \Lambda(\tau))$, as p is just restricted on a smaller σ-algebra.

We will also elide the measure when describing an object in **AMP**, for readability, if the measure is clear.

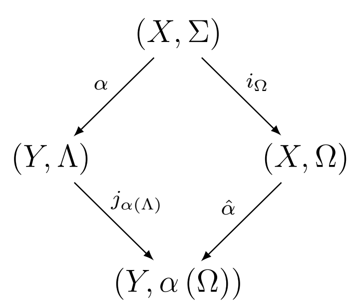
5.1 Some lemmas on measure spaces

We will need some results on measurable spaces.

Lemma 5.1.1 *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.

Lemma 5.1.2 *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:*



Where 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 . 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**, the pushout Y is then equipped with the largest σ-algebra making the maps measurable, which is the case here as $j_{\alpha(\Lambda)}$ is an isomorphism of σ-algebras by Lemma 5.1.1. ∎

The point of the preceding lemma is that the map α is an isomorphism of σ-algebras while the state space diminishes, while i_Ω reduces the σ-algebra without altering the state space; these two operations can be combined into the pushout space $(Y, \alpha(\Omega))$, and it turns out that one can do these two operations in any order.

5.2 Finite approximations

Given an arbitrary AMP, it may be very difficult to study its behavior if its state space is very large or uncountable. It is therefore sensible to devise a way to reduce the state space to a manageable size. How reliable are such approximations?

Let (X, Σ, p, τ_a) be a labelled AMP. Let $\mathcal{P} = 0 < q_1 < q_2 < \ldots < q_n \leq 1$ be a finite partition of the unit interval with each q_i a rational number. We shall call these *rational partitions.* We define a family of finite π-systems [Bil95], subsets of Σ, as follows:

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

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

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

We begin with a very important result:

Proposition 5.2.1 *Given any labelled AMP (X, Σ, p, τ_a), the σ-algebra $\sigma(\bigcup \Lambda_{\mathcal{P},M})$, 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. ∎

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

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

Now since $\Lambda_{\mathcal{P},M}$ is finite, it is atomic, and so it partitions our state space X, yielding an equivalence relation. Quotienting by this equivalence relation gives a map

$$\pi_{\mathcal{P},M} : (X, \Lambda_{\mathcal{P},M}, p) \longrightarrow \left(\hat{X}_{\mathcal{P},M}, \Omega_{\mathcal{P},M}, p_{\mathcal{P},M}\right)$$

where $\hat{X}_{\mathcal{P},M}$ is the (finite!) set of atoms of $\Lambda_{\mathcal{P},M}$ and $\Omega_{\mathcal{P},M}$ is just the powerset of $\hat{X}_{\mathcal{P},M}$. The measure $p_{\mathcal{P},M}$ is defined in the obvious way as the image measure through $\pi_{\mathcal{P},M}$; thus $\pi_{\mathcal{P},M}$ is measure-preserving as well.

As the σ-algebra on $\hat{X}_{\mathcal{P},M}$ is its powerset, we will often refrain from writing $\Omega_{\mathcal{P},M}$ when involving a finite approximants.

We thus have measure-preserving approximation maps $\phi_{\mathcal{P},M} = \pi_{\mathcal{P},M} \circ i_{\Lambda_{\mathcal{P},M}}$ from our original state space to a finite state space. We can easily extend these to morphisms of AMP's: as all of the above spaces are defined using maps with (X, Σ, p, τ_a) as domain, we define AMP's on the finite approximants spaces by letting $\rho_{(\mathcal{P},M),a} = \phi_{\mathcal{P},M}(\tau_a)$.

5.3 A projective system of finite approximants

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

Thus, given two approximation pairs such that $(\mathcal{P}, M) \leq (\mathcal{Q}, N)$, we have a measurable measure-preserving map

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

which is the identity on points, and which is well defined by the inclusion $\Lambda_{\mathcal{P},M} \subseteq \Lambda_{\mathcal{Q},N} \subseteq \Sigma$. We therefore have a projective system of such maps indexed by our poset of approximation pairs. A consequence of Lemma 5.1.2 is that these maps induce a map on the finite approximation spaces

$$j_{(\mathcal{P},M),(\mathcal{Q},N)} : \hat{X}_{\mathcal{Q},N} \longrightarrow \hat{X}_{\mathcal{P},M}$$

such that the following commutes:

$$\begin{array}{ccc} (X, \Lambda_{\mathcal{Q},N}) & \xrightarrow{i_{(\mathcal{P},M),(\mathcal{Q},N)}} & (X, \Lambda_{\mathcal{P},M}) \\ \downarrow{\scriptstyle \pi_{\mathcal{Q},N}} & & \downarrow{\scriptstyle \pi_{\mathcal{P},M}} \\ \hat{X}_{\mathcal{Q},N} & \xrightarrow{j_{(\mathcal{P},M),(\mathcal{Q},N)}} & \hat{X}_{\mathcal{P},M} \end{array}$$

Therefore, the approximation map $\phi_{(\mathcal{P},M)}$ factors through the approximation map $\phi_{(\mathcal{Q},N)}$ as $\phi_{(\mathcal{P},M)} = j_{(\mathcal{P},M),(\mathcal{Q},N)} \circ \phi_{(\mathcal{Q},N)}$. Hence, the maps $j_{(\mathcal{P},M),(\mathcal{Q},N)}$ along with the approximants $\hat{X}_{(\mathcal{P},M)}$ also form a projective system of surjective measure-preserving maps with respect to our poset of approximation pairs.

5.4 Existence of the projective limit

Each finite approximant $\hat{X}_{\mathcal{P},M}$ can be considered as a topological space; indeed, one can put the discrete topology on $\hat{X}_{\mathcal{P},M}$, as it is a finite set. This gives a compact Hausdorff space. Thus we have a projective system of measure-preserving maps of probability spaces where each σ-algebra is generated by a compact Hausdorff topology. These topological considerations allow us to use a result by Choksi [Cho58]:

Proposition 5.4.1 *(From [Cho58]) Let (X_i, Σ_i, m_i) be a projective system of measure-preserving maps indexed by a directed set I, with maps f_{ij} if $i \leq j$. Suppose that X_i is a compact Hausdorff topological space and m_i is inner regular. Then the projective limit $(X_\infty, \Sigma_\infty, m_\infty)$ exists, and the measure m_∞ is inner regular with respect to the compact Hausdorff topology induced by the projective system on X_∞.*

We skip the details of the proof but concentrate on the important parts. How Choksi proves this is by first constructing the projective limit in **Mes**, which always exists, as discussed in section 2.2. The main problem comes from defining a measure on X_∞. In general, it may only be possible to define a finitely additive set function m on the algebra of sets $\mathcal{M} = \bigcup_{i \in I} f_{i\infty}^{-1}(\Sigma_i)$. Given $A \in \Sigma_i$, we define $m\left(f_{i\infty}^{-1}(A)\right) = m_i(A)$. It is well-defined as I is a directed set and the maps are measure-preserving. We need to extend this function to a measure on $\Sigma_\infty = \sigma(\mathcal{M})$, which is what Choksi succeeds in doing. It is then immediate that the functions $f_{i\infty}$ are measure-preserving.

One point that Choksi does not address in his proof is the universality of the projective limit. That is, given another probability space (Y, Λ, λ) with measure-preserving maps $g_i : (Y, \Lambda, \lambda) \longrightarrow (X_i, \Sigma_i, m_i)$, is there a unique measure-preserving map $u : (Y, \Lambda, \lambda) \longrightarrow (X_\infty, \Sigma_\infty, m_\infty)$ through which the maps g_i factor? We know that such a map exists in **Mes**, and thus we only need to show that u is measure-preserving.

Lemma 5.4.2 *The universal map* $u : (Y, \Lambda, \lambda) \longrightarrow (X_\infty, \Sigma_\infty, m_\infty)$ *is measure-preserving.*

Proof . We only need to show that m_∞ and $M_u(\lambda)$ coincide on the algebra of sets $\mathcal{M}$. Indeed, $\mathcal{M}$ is a π-system which generates Σ_∞. By the uniqueness of extension theorem [Bil95], this is enough to show that m_∞ and $M_u(\lambda)$ coincide on Σ_∞. Any element of $\mathcal{M}$ is the preimage of measurable set A in some probability space (X_i, Σ_i, m_i). Pick any such A; we must then show that $m_\infty\left(f_{i\infty}^{-1}(A)\right) = M_u(\lambda)\left(f_{i\infty}^{-1}(A)\right)$. But we have that

$$\begin{aligned} M_u(\lambda)\left(f_{i\infty}^{-1}(A)\right) &= \left(M_{f_{i\infty}} M_u(\lambda)\right)(A) && \text{(def. of image measure)} \\ &= M_{g_i}(\lambda)(A) && (M_{-} \text{ is functorial}) \\ &= m_i(A) && (g_i \text{ is measure-preserving}) \end{aligned}$$

Furthermore, as m_∞ is an extension of m, we have

$$m_\infty\left(f_{i\infty}^{-1}(A)\right) = m_i(A)$$

Thus m_∞ and $M_u(\lambda)$ coincide on $\mathcal{M}$ and we are done. ∎

We restate Choksi's result for the case of our approximation spaces.

Corollary 5.4.3 *The projective system* $\left(\hat{X}_{\mathcal{P},M}, \Omega_{\mathcal{P},M}, p_{\mathcal{P},M}\right)$ *of finite approximants of an AMP* (X, Σ, p, τ_a)*, indexed by the approximation pairs, has a projective limit* $\left(\hat{X}_\infty, \Omega_\infty, p_\infty\right)$ *in* **MMPM**. *Furthermore, as the maps* $j_{(\mathcal{P},M),(\mathcal{Q},N)}$ *of the projective system are surjective, the limit maps are surjective as well.*

Thus we have the limit maps, $\psi_{\mathcal{P},M} : \hat{X}_\infty \longrightarrow \hat{X}_{\mathcal{P},M}$ for every approximation pair. We also have the measure-preserving maps $\phi_{\mathcal{P},M} : (X, \Sigma, p) \longrightarrow \left(\hat{X}_{\mathcal{P},M}, p_{\mathcal{P},M}\right)$ from the underlying probability space of our AMP to the finite approximants. By the above lemma (5.4.2), we have a unique measure-preserving map $\kappa : X \longrightarrow \hat{X}_\infty$ such that $\psi_{\mathcal{P},M} \circ \kappa = \phi_{\mathcal{P},M}$, i.e., the approximation maps from X factor through κ. Note that κ is surjective as well.

Proposition 5.4.4 *The σ-algebra $\kappa^{-1}(\Omega_\infty)$ is precisely equal to $\sigma\, [\![\mathcal{L}]\!]$*

Proof . Recall that the σ-algebra Ω_∞ is generated by the inverse images of the limit maps $\psi_{\mathcal{P},M}$; we have $\Omega_\infty = \sigma\left(\bigcup {\psi_{\mathcal{P},M}}^{-1}(\Omega_{\mathcal{P},M})\right)$, where the union is over all approximation pairs. Now we know that

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

and so, by well-behavedness of preimages, we have

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

∎

Finally, we define the AMP ζ_a on $\left(\hat{X}_\infty, \Omega_\infty, p_\infty\right)$ in the obvious way; that is, as the projection of the AMP τ_a through κ. Then the projection of ζ_a onto the finite approximants through $\psi_{\mathcal{P},M}$ is precisely equal to $\rho_{(\mathcal{P},M),a}$ as they were previously defined, since $\psi_{\mathcal{P},M} \circ \kappa = \phi_{\mathcal{P},M}$.

Thus, the projective limit of measure spaces can be extended to a projective limit of AMP's. Note that the AMP structure is, for now, quite superfluous, as all the AMPs we have defined come from the projection of the original AMP on X. We thus need to relate the projected AMPs to the original AMP.

5.5 Convergence

Proposition 5.5.1 *The universal map κ obtained from the projective limit is a zigzag.*

Proof . As $\kappa^{-1}(\Omega_\infty) = \sigma(\llbracket\mathcal{L}\rrbracket)$, we invoke Lemma 4.6.4 and proposition 5.4.4 to factor κ as $\hat{\kappa} \circ i_\kappa$, where

$$i_\kappa : (X, \Sigma, p, \tau_a) \longrightarrow (X, \sigma(\llbracket\mathcal{L}\rrbracket), p, \tau_a)$$
$$\hat{\kappa} : (X, \sigma(\llbracket\mathcal{L}\rrbracket), p, \tau_a) \longrightarrow \left(\hat{X}_\infty, \Omega_\infty, p_\infty, \zeta_a\right)$$

i_κ is a zigzag as $\sigma(\llbracket\mathcal{L}\rrbracket)$ is an event bisimulation; $\hat{\kappa}$ is a zigzag as it preserves the σ-algebras. Thus κ is a zigzag. ■

Thus, if we let $\left(\tilde{X}, \Xi, r, \xi_a\right)$ be the smallest bisimulation obtained as in proposition 4.6.1, we have a zigzag $\omega : \left(\hat{X}_\infty, \Omega_\infty, p_\infty, \zeta_a\right) \longrightarrow \left(\tilde{X}, \Xi, r, \xi_a\right)$. This zigzag must be an isomorphism of σ-algebras as Ξ is the smallest possible σ-algebra on $\tilde{X}$. We will now show that there is a zigzag going in the other direction.

Proposition 5.5.2 *Let $\alpha : (X, \Sigma, p, \tau_a) \longrightarrow (Y, \Theta, q, \rho_a)$ be a zigzag. Then these two AMPs have the same finite approximants.*

Corollary 5.5.3 *Two bisimilar AMPs have the same finite approximants.*

We will first need to show some lemmas. Let $\llbracket\phi\rrbracket_X$ be the set associated to formula ϕ in the labelled AMP on X.

Lemma 5.5.4 *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*

$$\alpha^{-1}\left(\left\llbracket \langle a\rangle_q A\right\rrbracket_Y\right) = \left\llbracket \langle a\rangle_q \alpha^{-1}(A)\right\rrbracket_X$$

Proof .

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

∎

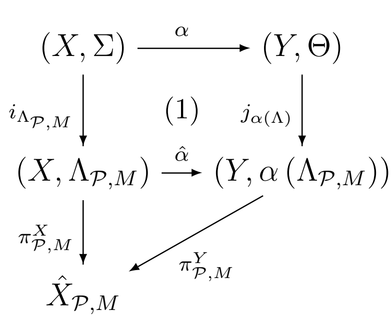
Lemma 5.5.5 *Let (X,Σ,p,τ_a) be a labelled AMP and $\Omega\subseteq\Sigma$ be an event-bisimulation. Then (X,Ω,p,τ_a) and (X,Σ,p,τ_a) have the same finite approximants.*

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

$$\Lambda_{\mathcal{P},M}\subseteq\sigma\left(\llbracket\mathcal{L}\rrbracket\right)\subseteq\Omega\subseteq\Sigma$$

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 5.5.2. The following diagram of AMPs will be referred to during the proof:



First, by Lemma 5.5.5 and the factoring property of zigzags (lemma 4.6.4), 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 5.1.1, α is an isomorphism of σ-algebras. Let $\Lambda_{\mathcal{P},M} \subseteq \Sigma$ be an approximating σ-algebra on X.

By Lemma 5.1.2, the square (1) in the above diagram commutes and is a pushout. The measures and AMPs are defined in the usual way.

We need only show that $\alpha\left(\Lambda_{\mathcal{P},M}\right)$ is precisely the approximating σ-algebra obtained on Y by the approximation pair $(\mathcal{P}, M)$. Lemma 5.5.4 guarantees that this is the case, as sets of the form $\left[\!\left[\langle a\rangle_q A\right]\!\right]_Y$ generate the approximating σ-algebras.

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 AMPs, and so the bottom triangle of the above diagram commutes; thus the two original AMPs (X, Σ, p, τ_a) and (Y, Θ, q, ρ_a) have the same finite approximations. ∎

We conclude with the main result.

Theorem 5.5.6 *Given an AMP (X, Σ, p, τ_a), the projective limit of its finite approximants $\left(\hat{X}_\infty, \Omega_\infty, p_\infty, \zeta_a\right)$ is isomorphic to its smallest bisimulation $\left(\tilde{X}, \Xi, r, \xi_a\right)$.*

Proof . As X and $\tilde{X}$ are bisimilar, they have the same approximants, and thus the projective limits of these approximants $\left(\hat{X}_\infty, \Omega_\infty, p_\infty, \zeta_a\right)$ is the same. Therefore, by Proposition 5.5.1 there is a zigzag

$$\epsilon : \left(\tilde{X}, \Xi, r, \xi_a\right) \longrightarrow \left(\hat{X}_\infty, \Omega_\infty, p_\infty, \zeta_a\right)$$

Hence, by Corollary 4.6.2, ϵ is an isomorphism of AMPs. ∎

CHAPTER 6
Related Work

6.1 Labelled Markov processes

As extensively discussed in this thesis, bisimulation truly is a notion of equality among different processes. However, equality, being a true/false statement, is very brittle and unable to tell when two processes are, in some sense, close. Thus, pseudometrics were developed in order to capture such a notion[DGJP99, DGJP04, vBW01b, vBW01a]. One important property of these pseudometrics is that the kernel precisely corresponds to bisimulation. Furthermore, these metrics allow a precise quantification of the convergence rate of approximation schemes. For example, the finite approximants defined by a hierarchy of levels, discussed in chapter 2, converge to the process they approximate with respect to the metric of [DGJP04].

It is also of interest to algorithmically compute the finite approximations to a labelled Markov process. This is typically quite difficult, as it requires computing sets defined by the preimage of the Markov kernels τ. In order to circumvent this difficulty, a Monte Carlo algorithm was devised by Bouchard-Côté et al. in order to, in effect, approximate the finite approximations [BCFPP05]. Convergence of the Monte Carlo scheme was shown using the metric above.

6.2 The functor $\mathbb{E}_{-}$

It appears that our development of the functor $\mathbb{E}_{-}$ is original. Of course, the concept has been around; an obvious special case of this functor is the usual conditional expectation operator on a sub-σ-algebra Λ, where the map α is in fact a map $i_\Lambda : (X, \Sigma) \longrightarrow (X, \Lambda)$ which is the identity on points.

A special case of the functor $\mathbb{E}_-$ is well known in the dynamical systems literature. Consider a map $\alpha : (X, \Sigma, p) \longrightarrow (X, \Sigma, p)$ in $\mathbf{Rad}_1$. Then the operator $\mathbb{E}_\alpha$ can be iterated on $L_1^+(X)$. Furthermore, it can be shown that this operator has norm 1, and is thus a Markov operator on X as per definition 2.5.3. This Markov operator is precisely the one associated to the Markov kernel of example 2.3.2 using α, and is called the Frobenius-Perron operator [LM94]. Recall that $\mathbb{E}_\alpha$ is the adjoint of the pre-composition functor, as demonstrated in proposition 3.3.3. In this special case of α, the arrow $(-) \circ \alpha$ obtained from the precomposition functor is called the Koopman operator. Note that in this case, the Frobenius-Perron operator is a Markov operator and is thus interpreted to transform measures "forwards in time"; in our case, we used the operator $\mathbb{E}_\alpha$ as an abstract Markov kernel (as we considered maps $\alpha \in \mathbf{Rad}_\infty$) and was thus used as a likelihood transformer.

The only reference we have found to the general case of our functor $\mathbb{E}_-$ is in a paper by Scheffer [Sch69]; the only restriction is that the maps α be measure-preserving. The author noted the adjunction with precomposition, but did not expose the operation as a functor.

6.3 Markov operators

As far as we know, this is the first time that probabilistic systems, in the sense of this work, are studied using Markov operators. Nevertheless, Markov operators have been extensively studied, and a detailed overview of this body of work would be beyond the scope of this thesis, as most of the work we are aware of is about ergodicity (e.g. [Hop54, Fog69, Haw06]) or oriented towards functional analysis (e.g. spectrum of Markov operators, see [Sch74]).

One recent work of interest to us is that of Ding et al. ([DLZ02]). In this article, the authors postcompose a Markov operator with a conditional expectation operator in order to obtain a finite approximation. However, the state space of the Markov operator is restricted to be $[0, 1]$, and the finite σ-algebras used for the projection are not obtained in any systematic way. Finally, the main objective of the authors is to compute invariant measures of a Markov operator by taking a limit of the invariant measures of the finite approximants; the dynamics of the approximants themselves are not studied.

CHAPTER 7
Conclusions

7.1 Contributions

We review the main contributions of this work.

- We view Markov processes in a "dual" point of view; instead of looking at state or measure transitions "forwards in time", we view Markov processes as transformers of functional predicates, which operate "backwards in time". This allows us to consider a new class of probabilistic systems, intimately connected to ordinary labelled Markov processes, that we called labelled abstract Markov processes.
- We generalized conditional expectation into a functor that encompasses connected concepts which were until now handled separately. Furthermore, we showed that this functor is naturally isomorphic, and its arrows have precisely the same norm as, two other well-known functors, one of which is the image measure functor. Using this functor allowed us to construct a category of AMPs where the arrows behave as projections.
- We defined zigzags and bisimulation for this category, and showed that our definition for bisimulation was transitive. We also showed that given any AMP, there exists a minimal AMP to which it is bisimilar. We imported previous results showing that a modal logic generated the event-bisimulation induced by the minimal bisimilar AMP.
- We devised a scheme to produce a family of finite approximants to any AMP; the scheme uses the modal logic. This family of finite approximants converges, in a categorical sense, to the minimal

bisimilar AMP. In other words, given the finite approximants to a given AMP, one can reconstruct a bisimilar process which is minimal in a precise sense.

7.2 Future work

The theory of abstract Markov processes is by no means complete, and many more results are expected. Of particular interest is the development of metrics. As we are using linear operators, it is very tempting to use the operator norm in order to define a metric. However, we have not yet successfully defined a suitable metric characterizing bisimulation, although we have many ideas which we wish to explore. Indeed, one metric defined in [DGJP04] involves a logic of functional expressions, where real-valued functions are defined on the state space of a LMP; two points are bisimilar if and only if every functional expression, interpreted in this LMP, yields the same value at both points. As the functional expressions use the operator T_τ for a given Markov kernel τ, it appears that our linear operator point of view will streamline such a definition.

Another important question is to investigate the rôle of the probability measure on the state spaces we consider. Indeed, if we have, on a measurable space, two probability measures p and q such that $p \ll q$ and $q \ll p$, it is easy to show that any Markov kernel which is nonsingular with respect to one is nonsingular with respect to another. It thus appears that imposing a probability measure on our measurable spaces is a proxy for a structure of "negligible sets" (as two measures as p and q above share precisely the same sets of zero measure). It may then be possible to let go of measures altogether. This idea was already explored in [DDLP06].

We also wish to extend our framework to include stochastic observations. Intuitively speaking, beyond the knowledge of whether an action in enabled or not, the user will be able to obtain partial information about the state space, given stochastically. For example, one may imagine that the user can measure the "temperature" at a particular state, which yields extra information. Such an observation can be considered an abstract Markov kernel, and one can extend the definition of the category **AMP**, and the definition of a zigzag, to include this new structure.

We also have written a Python program which approximates our finite approximants using Monte Carlo methods, in the spirit of [BCFPP05]. However, as we do not have metrics, it is difficult to prove convergence of these approximants, and thus we have not discussed this algorithm.

One very interesting problem which we are working on is that of writing an explicit representation of the state space underlying the "smallest bisimulation". Indeed, it was shown with labelled Markov processes that one could construct a "universal" labelled Markov process as the solution to a domain equation [DGJP03]. We believe that constructing an appropriate structure on the set of formulas of $\mathcal{L}$ would yield a set whose σ-algebra would precisely be $\sigma(\llbracket \mathcal{L} \rrbracket)$, the underlying σ-algebra of the "smallest bisimulation" of every AMP.

Furthermore, we wish to give an explicit categorical definition of the objects of **AMP** or $\mathbf{AMP}_{\mathcal{A}}$. Indeed, given the set of action $\mathcal{A}$, one may consider the monoid $\mathcal{A}^*$ consisting of all words with $\mathcal{A}$ as an alphabet. This monoid may be considered as a one-object category, and a functor from this category to $\omega\mathbf{CC}$, where the single object is mapped to $L^+_\infty(X)$ and the arrows are linear operators with norm less than 1, has all of the information we need about a labelled AMP; furthermore, a natural transformation of such functors is very much like a zigzag.

In a more general setting, it appears to us that the definition of bisimulation as a span of morphism is mistaken; indeed, the definition in terms of cospans appears to be much more natural. Thus, we wish to formalize this impression and to show that bisimulation in terms of cospans should be the preferred definition, whether the system is probabilistic or nondeterministic.

Finally, even though this thesis was done by viewing probabilistic systems as interactive, our approximation scheme for abstract Markov kernels applies in any situation; it may thus be possible to use our ideas in any context where abstract Markov kernels or Markov operators appear.

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.

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

[CPDP09] Philippe Chaput, Prakash Panangaden, Vincent Danos, and Gordon Plotkin. Approximating Markov processes by averaging. In *Proceedings of ICALP 2009*, 2009. to be published.

[DD03] Vincent Danos and Josée Desharnais. Labeled Markov Processes: Stronger and faster approximations. In *Proceedings of the 18th 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. Seventh Workshop on Coalgebraic Methods in Computer Science 2004.

[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.

[DGJP99] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for labeled Markov systems. In *Proceedings of CONCUR99*, number 1664 in Lecture Notes in Computer Science. Springer-Verlag, 1999.

[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.

[DLZ02] Jiu Ding, Tien Yien Li, and Aihui Zhou. Finite approximations of Markov operators. *Journal of Computational and Applied Mathematics*, 147(1):137 – 152, 2002.

[DP06] Vincent Danos and Gordon Plotkin. Abstract lmps and compression. Private communication, January 2006.

[Fog69] Shaul R. Foguel. *The Ergodic theory of Markov processes.* Van Nostrand Reinhold Co., 1969.

[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.

[LM94] A. Lasota and M. Mackey. *Chaos, Fractals, and Noise.* Springer, 1994. Stochastic Aspects of Dynamics.

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

[Mac98] S. Mac Lane. *Categories for the Working Mathematician.* Number 5 in Graduate Texts in Mathematics. Springer-Verlag, 1998. Second Edition.

[Mil80] R. Milner. *A Calculus for Communicating Systems*, volume 92 of *Lecture Notes in Computer Science.* Springer-Verlag, 1980.

[Par81] David Park. Title unknown. Slides for Bad Honnef Workshop on Semantics of Concurrency, 1981.

[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.

[Sch69] Carel L. Scheffer. Limits of directed projective systems of probability spaces. *Probability Theory and Related Fields*, 13(1):60–80, March 1969.

[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*, number 33 in TUCS General Publication, pages 127–143. Turku Centre for Computer Science, June 2004. Available from `www.mathstat.dal.ca/~selinger/papers.html`.

[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.

[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.