Lax Probabilistic Bisimulation

Jonathan Taylor

Master of Science

School of Computer Science

McGill University

Montreal,Quebec

2008-08-29

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

DEDICATION

This thesis is dedicated to my parents, Margaret and Gray Taylor, who never gave up on me. For this, I am forever thankful.

ACKNOWLEDGEMENTS

I must first thank Prakash Panangaden, the coolest supervisor in the world, for guiding my research so skillfully, for being so patient with me, for being so witty, and of course for all the great wine parties.

I would like to also thank Doina Precup who joined our ranks when my research moved towards AI. Thank you for all your advice and work on the papers we submitted.

Thank you Michael Langer for convincing me to come to McGill and giving me so much invaluable advice. Thank you Luc Devroye for your constant support and for both challenging me and making me laugh from day one.

I would also like to thank my great friend and lab mate Jordan Frank who did his masters by my side; for knowing what I was going through, for making it fun, for submitting this document for me and countless other things.

I would also like to thank Eric, Mike, Patrick and Vlad for always being there, listening to me complain, and for being all around the bestest friends.

Merci Stéphane, de m'avoir aidé à traduire l'abrégé en français.

Thanks to everyone who came out for beers. I will not try to enumerate you all here for I would invariably forget one of you. This is the nature of beer.

Last but far from least, I would like to thank my supportive and loving girlfriend Jennie. In the end, impressing you was exactly the motivation I needed. Tack. Du är perfekt och jag älskar dig.

ABSTRACT

Probabilistic bisimulation is a widely studied equivalence relation for stochastic systems. However, it requires the behavior of the states to match on actions with matching labels. This does not allow bisimulation to capture symmetries in the system. In this thesis we define lax probabilistic bisimulation, in which actions are only required to match within given action equivalence classes. We provide a logical characterization and an algorithm for computing this equivalence relation for finite systems. We also specify a metric on states which assigns distance 0 to lax-bisimilar states. We end by examining the use of lax bisimulation for analyzing Markov Decision Processes (MDPs) and show that it corresponds to the notion of a MDP homomorphism, introduced by Ravindran & Barto. Our metric provides an algorithm for generating an approximate MDP homomorphism and provides bounds on the quality of the best control policy that can be computed using this approximation.

ABRÉGÉ

La bisimulation probabiliste est une relation d'équivalence pour système stochastique grandement étudiée. Toutefois, il demande que le comportement des états soit équivalent pour les actions portant le même nom. Ceci ne permet pas la bisimulation de capturer les symétries dans le système. Dans cette thèse, nous définissons la bisimulation lax probabiliste dans laquelle les actions sont seulement requises d'être équivalente sous une classe d'équivalence donnée. Nous proposons une caractérisation logique et un algorithme pour calculer cette relation d'équivalence pour les systèmes finis. Nous spécifions aussi une métrique sur les états qui assigne une distance de 0 aux états lax-bisimilaires. Nous terminons en examinant l'utilité de la bisimulation lax pour l'analyse des processus de décisions markoviens (PDM) et démontrons que la bisimulation lax correspond à la notion d'homomorphisme dans les PDMs, introduites par Ravindran & Barto. Notre métric fournit un algorithme pour générer un PDM homomorphique approximatif et fournit des bornes sur la qualité de la meilleure politique calculée à partir de cette approximation.

TABLE OF CONTENTS

LIST OF TABLES

Table page

LIST OF FIGURES

CHAPTER 1
Introduction

The formal analysis of large stochastic systems often requires reducing the state space of the system, by grouping together states that exhibit similar behavior. Probabilistic bisimulation [LS91, KS60] is an equivalence relation for such systems that captures naturally the notion of behavioral similarity between states. Much recent research has been devoted to this topic [Bai96, BK97, DG97, DDLP05, DEP02, DPW06, PLS00, DGJP02b]. However, in the presence of real numbers, like probabilities or times, the notion of equivalence relation is too exact, because it requires exact matching of real numbers. If these numbers are acquired from data, or represented with limited precision (as is always the case in a digital computer), exact matching is very hard to achieve. Metrics are the ideal substitute in this case and there has been considerable interest in metric analogues of probabilistic bisimulation [DGJP99, DGJP04, vBW01b, vBW01a, DGJP02a, FPP04, FPP05]. This relaxation of the notion of probabilistic bisimulation opens the door to approximate reasoning algorithms for probabilistic processes.

However, both bisimulation and its metric analogue require the behavior of states to match on the same actions (or labels). This can be very restrictive, for example if the action space is continuous. Moreover, there are many situations in which one wants to match actions that are not *named* identically but which are closely related. A very common case is when there is a symmetry in the underlying system.

Our work is motivated by applications from robotics and artificial intelligence. For example, consider a robot navigating in a square grid. Suppose that the goal is to reach the centre of the region. In this case, a move to the right at the left end of the room is essentially identical to a move to the left at the right end of the room, from the point of view of bringing the robot closer to its target. One could exploit this symmetry for extra compression in the state space. However, the usual notion of bisimulation does not allow such symmetries to be taken into account: because the action of going right has very different outcomes on the right and left side of the grid, these states will all be considered different, and no aggregation is possible. Intuitively, what we would like is to allow *different* actions to match each others' effects in different parts of the state space. The goal of this thesis is to relax the notion of bisimulation to allow different actions to match in different states.

1.1 Contributions

We study a version of strong probabilistic bisimulation which we call *lax bisimulation* where we match actions that are at zero distance to each other in a metric – denoted α – rather than just matching identical actions. A natural idea is to think that one can introduce names for these distance zero equivalence classes and proceed with ordinary bisimulation on these "lumped" sets of actions. However, this is not correct, as we will show by example below. Essentially, the reason is that if we just lumped the equivalence classes we would introduce non-determinism whereas our theory preserves the fully probabilistic nature of the systems with which we work.

Our work is carried out in the context of probabilistic transition systems with continuous state spaces. We are contributing a new general theory of bisimulation that includes:

- a definition of lax bisimulation
- a logical characterization of lax bisimulation,
- an algorithm to compute distinguishing formulas for states that are not lax bisimilar and
- a metric analogue to lax bisimulation.

As well, by applying this theory to the domain of Markov Decision Processes we include the following contributions:

- a metric that indicates the degree of symmetry between MDP states
- bounds on the performance loss caused by aggregating states in an MDP
- an algorithm to do aggregation.

1.2 Outline

The work will begin with a background chapter that accomplishes two things. First, the appropriate mathematical structures are introduced necessary to the theory of bisimulation. Next, bisimulation and its metric analogue are shown in their classic unlaxed form.

Chapter 3 focuses on relaxing these definitions and demonstrating that the theory can be generalized. Chapter 4 provides a logical characterization of this unlaxed bisimulation as well as a corresponding algorithm that can compute the necessary characterizing formulas for a system.

Chapter 5 is a case study in which we analyze a specific type of probabilistic system called a Markov Decision Process (MDP). In this case, we show that the notion of bisimulation corresponds exactly to that of symmetry as described by the theory of MDP homomorphisms. This theory allows states that can behave similarly by performing symmetric actions to be lumped together to create an aggregate MDP. The resulting bisimulation metric actually indicates the level of symmetry between two states and we show how this can be used as a recipe to form an approximate MDP homomorphism. We conclude by providing tight bounds on the performance of the approximate MDP.

In chapter 6 we outline the contributions of the thesis and discuss potential avenues of future work.

CHAPTER 2
Background

Our mission is to formalize a theory that can describe the behavioral similarity of different parts of a system. Mathematically we think of these parts as different states (or elements) of a state space S. In this thesis, we assume that S is an analytic space. This is hardly restrictive as it is inclusive of all discrete spaces and closed subspaces of $\mathcal{R}^n$ [DEP02]. We now proceed to examine the spatial structures that will be used to describe the similarity between states. References for this material are available in many classic analysis texts [Rud66].

2.0.1 Relations

One way to indicate state similarity is to directly equate states through a relation. A *relation* B is a subset of $S \times S$ in which $(s,t) \in R$ indicates that s is related to t by B and this is often denoted simply as sBt. The type of relations that are most useful to us will satisfy a few basic properties.

1. Reflexivity: for any x, xBx.
2. Symmetry: for any x and y where xBy, yBx also.
3. Transitivity: for any $x, y, z \in S$, xBz and zBy means that xBy.

If a relation B satisfies the above properties that we say that it is an *equivalence relation*. An equivalence relation divides the state space of a system into partitions. Bisimulation is an example of such a relation in which the states within each partition behave the same way. We will make this precise later.

2.0.2 Metrics

Equivalence relations are completely binary; states are either related or not. One can think of many examples in which states should be related to varying degrees. Indeed, in the case of behavioural similarity, we may want to indicate to what degree states are similar even if they do not have the exact same behaviour. The obvious way to do this is to indicate similarity through a numerical value. The appropriate mathematical structure required here is that of a metric.

A metric can be thought of as a distance function that assigns a numerical value to two points denoting how far apart they are. What these numerical values are and how they are assigned varies but the rules that they must follow are inspired by our conventional notion of spatial distance. To start, the distance from one point to itself is zero. Also, the distance from one point to a second must be equal to the distance from the second to the first. Lastly, the distance from one point to a second detouring through a third point cannot be less than the distance to the third directly. Indeed, these rules are quantitative analogs to the rules of an equivalence relation.

We say that a (1-bounded pseudo) metric on a space X is a map $d : X \times X \longrightarrow [0, 1]$ such that for $x, y, z \in X$ we have that

1. $x = y \implies d(x, y) = 0$
2. $d(x, y) = d(y, x)$
3. $d(x, y) \leq d(x, z) + d(z, y)$

Given a metric d we let $Rel(d)$ denote the equivalence relation that equates all pairs of points (x, y) in which $d(x, y) = 0$.

Let us now define the specific class of metrics that we use in this thesis:

Definition 2.0.1 *A function $h : X \longrightarrow \mathcal{R}$ is lower semicontinuous (lsc) if whenever $\{x_n\} \longrightarrow x$ we have that* $\liminf h(x_n) \geq h(x)$.

For a given set X we let $\mathcal{M}$ denote the set of lsc metrics on X. If X is not specified, it can be assumed we are talking about the set of states S in a probabilistic system. It will be shown later that under an appropriate ordering, this space is an example of a complete lattice.

2.1 Fixed Points

Bisimulation will be defined later in a manner that resembles a recursive refinement of the state space. It will be seen that this form facilitates significant analysis through the use of fixed point theory [DP02] on lattices. A set L is said to be partially ordered if it is endowed with an ordering $\leq$. We say such a set is a complete lattice if arbitrary subsets of L have least upper bounds and greatest lower bounds. A point $x \in L$ in which $f(x) \leq x$, $f(x) = x$ or $x \leq f(x)$ are respectively said to be prefixed, fixed and postfixed points.

Theorem 2.1.1 *(Knaster-Tarski Fixed Point Theorem). Let L be a complete lattice, and suppose $f : L \longrightarrow L$ is monotone. Then f has a least fixed point, which is also its least prefixed point, and f has a greatest fixed point, which is also its greatest postfixed point.*

In this thesis, we will use this theorem twice. First, we will define bisimulation in terms of operators on REL and EQU. These spaces will denote respectively, the space of binary relations and the space of topologically-closed equivalence relations

on an analytic space S. Both spaces under the subset ordering form complete lattices [FPP05].

Second, we will define an analogous metric on the space of lsc metrics. Indeed, given an arbitrary set of lsc functions, the pointwise supremum is also lsc. As well, the pointwise supremum of a metric is also a metric. Thus, this space forms a complete lattice under pointwise ordering.

2.2 Probability

The sort of systems that will be investigating are probabilistic. We have attempted to address a set of systems as diverse as possible. In order to include systems with continuous state spaces it will be necessary to recall a few basic definitions from measure theory [Bil95]. Although it is always helpful to look to discrete systems for insight such systems will be special cases of this more general theory.

A collection Σ of subsets of S is a $\sigma-$-algebra if

1. Σ contains both the empty set $\emptyset$ and the entire set S.
2. Σ is closed under complements.
3. Σ is closed under countable unions.

The smallest σ-algebra containing a collection of sets $\mathcal{B}$ is called the σ-algebra generated by $\mathcal{B}$ and is denoted $\sigma(\mathcal{B})$. The tuple (S, Σ) is then said to be a measurable space but we will often just say S without making Σ explicit.

Within such a space we are concerned with a specific set of functions known as subprobability measures. Such a function $\mu : \Sigma \longrightarrow [0, 1)$ has the properties that

1. $\mu(\emptyset) = 0$

2. for any countable pairwise disjoint collection of sets $E_1, E_2, \ldots$, $\mu(\cup_i E_i) = \sum_i E_i$.

In addition to measuring a set E directly through $\mu(E)$, we can measure a set through a certain type of function called a measurable function. A function f between measurable spaces (X, Σ_X) and (Y, Σ_Y) is said to be *measurable* if $\forall E \in \Sigma_Y$ it is true that $f^{-1}(E) \in \Sigma_X$. Here, we are mostly concerned with the case where Y is the real numbers and Σ_Y is the Borel σ-algebra. This is the smallest σ-algebra containing all the open intervals.

The following proposition [Bil95] will be useful to us.

Proposition 2.2.1 *Let $\mathcal{B}$ be a non empty collection of sets where $S \in \mathcal{B}$ and $T \in \mathcal{B}$ means that $S \cap T \in \mathcal{B}$. If $\sigma(\mathcal{B})$ is the σ-algebra generated by $\mathcal{B}$ and two probability measures P and Q agree on $\mathcal{B}$ then they agree on all of $\sigma(\mathcal{B})$.*

2.3 Probability Metrics

Later we will analyze the similarity of states by looking at how the states behave when they attempt to perform various actions. As this behaviour is described by probability distributions, we will need a way to measure the degree of similarity between two distributions.

The **Total Variation Metric** between two probability measures P and Q is defined as

$$TV(P, Q) = \sup_{X \in \Sigma} |P(X) - Q(X)|$$

This metric is easy to understand and computationally efficient but it is not always appropriate as it treats each state in the state space distinctly. As will be

seen, we may want to indicate that some state pairs are less distinct than others and take this into account when similarity is computed. Of course, the best way to describe how distinct states are is by using a metric. We then want to have a way to lift metrics on the state space to metrics on probability distributions over the state space.

The construction given below is inspired by work of Kantorovich [Kan40]. The lifting is given by a linear program; the duality theory of linear programs is crucial in the theory.

Definition 2.3.1 *Given a metric $d \in \mathcal{M}$ and probability distributions P and Q over the state space S the* ***Kantorovich Metric*** *$K(d)(P, Q)$ is defined to by*

$$K(d) = \sup_f (P(f) - Q(f))$$

where $P(f)$ denotes the integral of f with respect to the probability measure P and the supremum is taken over all bounded measurable $f : s \longrightarrow \mathcal{R}$ satisfying the Lipchitz condition that $\forall x, y \in S$

$$f(x) - f(y) \leq d(x, y)$$

in the case of discrete systems, the linear program is as follows:

$$\max_{u_i} \sum_{i=1}^{|S|} (P(s_i) - Q(s_i))u_i$$

$$\textit{subject to } \forall i, j.u_i - u_j \leq d(s_i, s_j)$$

$$\forall i.0 \leq u_i \leq 1$$

which has the following equivalent dual program:

$$\min_{\lambda_{kj}} \sum_{k,j=1}^{|S|} \lambda_{kj} d(s_k, s_j)$$

$$\textit{subject to } \forall k. \sum_j \lambda_{kj} = P(s_k)$$

$$\forall j. \sum_k \lambda_{kj} = Q(s_j)$$

$$\forall k, j.\lambda_{kj} \geq 0$$

2.4 Labelled Markov Processes

Labelled Markov processes are probabilistic versions of labelled transition systems. A Markov process is defined for each label. The transition probability is given by a *stochastic kernel* (Feller's terminology [Fel71]), also commonly called a *Markov kernel*. Hence, the lack of determinism has two sources: the "choice" of the labels (no probabilities are attributed to this at all) and the probabilistic transitions made by the process. This is the "reactive" model studied by Larsen and Skou [LS91] who used it only in a discrete state-space setting.

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

Definition 2.4.1 *A **transition (sub-)probability function** on a measurable space (S, Σ) is a function $\tau : S \times \Sigma \longrightarrow [0, 1]$ such that for each fixed $s \in S$, the set function $\tau(s, \cdot)$ is a (sub-)probability measure, and for each fixed $X \in \Sigma$ the function $\tau(\cdot, X)$ is a measurable function.*

One interprets $\tau(s, X)$ as the probability of the process that starts in state s to make a transition to one of the states in X. In general, the transition probabilities could depend on time (in the sense that the transition probability could be different at every step), but they must be independent of the past history. We will always consider the time-independent case.

We will work with *sub-probability* functions; i.e. with functions where $\tau(s, S) \leq 1$ rather than $\tau(s, S) = 1$. We view processes where the transition functions are only sub-probabilities as being *partially defined*. All the theory extends immediately to the case of full probabilities.

Definition 2.4.2 *A **partial labeled Markov process (LMP) $\mathcal{S}$** with label set $\mathcal{A}$ is a structure $(S, i, \Sigma, \{\tau_a \mid a \in \mathcal{A}\})$, where S is the set of states, Σ is the Borel σ-field on S, and*

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

is a transition sub-probability function.

For simplicity, we will fix the label set to be $\mathcal{A}$ (this does not restrict the theory). Hence, we will write (S, Σ, τ) for partial labelled Markov processes, instead of the more precise $(S, \Sigma, \{\tau_a \mid a \in \mathcal{A}\})$.

2.5 Probabilistic Bisimulation

The fundamental process equivalence that we consider is *strong probabilistic bisimulation*. Probabilistic bisimulation means matching the moves and probabilities *exactly*. Thus, each system must be able to make the same transitions with the same probabilities as the other.

Let B be a binary relation on a set S. We say a set $X \subseteq S$ is B-*closed* if $B(X) := \{t | \exists s \in X, sBt\}$ is a subset of X. If B is reflexive, then clearly this condition is equivalent to requiring $B(X) = X$. If B is an equivalence relation, a set is B-closed if and only if it is a union of equivalence classes. We write $\Sigma(B)$ for those Σ-measurable sets that are also B-closed.

Definition 2.5.1 *Let $\mathcal{S} = (S, \Sigma, \tau)$ be a labelled Markov process. An equivalence relation B on S is a* ***bisimulation*** *if whenever sBs', with $s, s' \in S$, we have that for all $a \in \mathcal{A}$ and every B-closed measurable set $X \in \Sigma$, $\tau_a(s, X) = \tau_a(s', X)$. Two states are bisimilar if they are related by a bisimulation relation.*

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

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

In either case it is clear that bisimulation is actually an equivalence relation [DEP02].

Proposition 2.5.2 *Bisimulation is an equivalence relation.*

It is not always clear which states of a system are bisimilar and even less clear which ones are not. In order to make such analysis easier, one can define a simple modal logic and prove that two states are bisimilar if and only if they satisfy exactly the same formulas. Indeed, for finite-state processes one can decide whether two states are bisimilar and effectively construct a distinguishing formula in case they are not [DGJP02a]. The logic is called $\mathcal{L}$ and has the following syntax:

$$\mathsf{T} \mid \phi_1 \wedge \phi_2 \mid \langle a \rangle_q \phi$$

where a is an action and q is a rational number.

Given a labelled Markov process $\mathcal{S} = (S, \Sigma, \tau)$ we denote by $s \models \phi$ the fact that the state s satisfies the formula ϕ. The definition of the relation $\models$ is given by induction on formulas. The definition is obvious for the propositional constant T and conjunction. We say $s \models \langle a \rangle_q \phi$ if and only if $\exists X \in \Sigma.(\forall s' \in X.s' \models \phi) \wedge (\tau_a(s, X) > q)$. In other words, the process in state s can make an a-move to some state that satisfies ϕ, with probability strictly greater than q.

The following important theorem, proved in [DEP98, DGJP02a], relates logic $\mathcal{L}$ and bisimulation.

Theorem 2.5.3 *Let (S, Σ, τ) be a labelled Markov process. Two states $s, s' \in S$ are bisimilar if and only if they satisfy the same formulas of $\mathcal{L}$.*

CHAPTER 3
Lax Probabilistic Bisimulation

Bisimulation requires bisimilar states to behave similarly by matching identical actions. This thesis investigates the possibility of relaxing this condition so that non identical actions can be matched. Such a relaxation will relate more states allowing for more compression.

3.1 Formulation

To facilitate this "relaxed" matching we make the assumption that the action set A is finite and fix a labelled Markov process and consider a fixed metric $\alpha : S \times A \longrightarrow [0, 1]$ that quantifies how similar state action pairs are. Within this "relaxed" notion of bisimulation, we allow states to match actions with zero distance in α. In this way, metrics which "relate" more actions will allow more states to be related.

Definition 3.1.1 *A relation B is a lax probabilistic bisimulation relation if whenever sRt we have that*

(1) $\forall a\ \exists b$ such that $\alpha((s, a), (t, b)) = 0$ and for all B-closed sets X we have that $\tau_a(s, X) = \tau_b(t, X)$

(2) $\forall b\ \exists a$ such that $\alpha((s, a), (t, b)) = 0$ and for all B-closed sets X we have that $\tau_b(t, X) = \tau_a(s, X)$

Note that from here on, when we say lax bisimulation or simply bisimulation, we are implicitly referring to lax probabilistic bisimulation unless otherwise explicitly stated.

As with regular bisimulation, the goal is to group states together and thus any sane relation of states should be an equivalence relation. The definition above, combined with the fact that $Rel(\alpha)$ itself is an equivalence relation, can easily be seen to only admit equivalence relations.

Theorem 3.1.2 *A bisimulation relation is an equivalence relation.*

If we have two different bisimulation relations, then the finite additivity of the probability function assures us that their union is also a bisimulation relation.

Theorem 3.1.3 *The union of two bisimulation relations is a bisimulation relation.*

This is a very useful property as we would like to group as many states as possible to get the simplest abstraction of the underlying system. To this end, we consider merging all bisimulation relations into one.

Definition 3.1.4 *The union of all lax bisimulation relations is called the bisimilarity relation and is denoted by $\sim$. We say that two states s and t are bisimilar if $s \sim t$.*

In order to show that $\sim$ itself is a bisimulation relation it is useful to look at an alternate definition of bisimulation in terms of an operator on relations. Indeed, let REL and EQU be the complete lattices of binary relations and topologically closed equivalence relations over S.

Definition 3.1.5 *Define $\mathcal{F} : REL \longrightarrow REL$ so that $s\mathcal{F}(B)s$ such that*

$$\forall a, \exists b \text{ where } \alpha((s,a),(t,b)) = 0 \text{ and } \forall X \in \Sigma(B_{rst}), \tau_a(s,X) = \tau_b(t,X)$$

and

$$\forall b, \exists a \text{ where } \alpha((s,a),(t,b) = 0 \text{ and } \forall X \in \Sigma(B_{rst}), \tau_a(s,X) = \tau_b(t,X)$$

Where B_{rst} is the smallest equivalence relation containing B.

Theorem 3.1.6 *$\sim$ is a bisimulation relation.*

Proof . Clearly $\mathcal{F}$ is monotonic and thus by theorem 2.1.1 it has a greatest fixed point. Also, one can see that $\mathcal{F}(B) = B$ if and only if B is a lax bisimulation relation, and thus the greatest fixed point is a bisimulation relation and so it must be contained in $\sim$. As every lax bisimulation relation is contained in the greatest fixed point, we actually have that $\sim$ must be the greatest fixed point and thus a bisimulation relation also. ∎

This relation $\sim$ provides the highest degree of compression under the constraints provided by α. When the state space is finite the theorem above yields an algorithm for calculating $\sim$ by locating the greatest fixed point of $\mathcal{F}$.

3.2 A Metric Analogue

As was explained in the introduction, an exact equivalence is inappropriate when one is dealing with systems with quantitative parameters. In this chapter we develop a metric analogue of α-lax bisimulation. Ferns et al.[FPP04, FPP05] developed a metric theory for MDPs, of course one with probabilistic bisimulation as its kernel rather than lax bisimulation. Here we develop the lax version for LMPs and to this end, we first define the specific type of metric in which we are generally interested in.

Similar ideas for defining metrics were proposed by van Breugel and Worrell [vBW01a].

The following two lemmas are given in Ferns et al [Fer07].

Lemma 3.2.1 *Given a metric $d \in \mathcal{M}$ and actions a, b, the map that takes $d(s,t)$ $\longrightarrow K(d)(\tau_s^a, \tau_t^b)$ is lower semi continuous in the state pair.*

Lemma 3.2.2 *Given a metric $d \in \mathcal{M}$ we have that*

$$K(d)(P,Q) = 0 \iff P(X) = Q(X), \forall X \in \Sigma(Rel(d))$$

In the case of the lax metric we have to compare sets of actions. Here we take advantage of a "natural" metric between compact subsets of a metric space called the *Hausdorff metric*. Given a metric, the Hausdorff metric measures the distance between two (compact) sets. It is a tight bound on the largest distance between a point in one set and a point of the other set. The Hausdorff metric is thus zero exactly when the two point sets coincide.

Definition 3.2.3 *Given a finite 1-bounded metric space $(\mathcal{S}, d)$, let $\mathcal{P}(\mathcal{S})$ be the powerset of $\mathcal{S}$, then the Hausdorff metric $H(d) : \mathcal{P}(\mathcal{S}) \times \mathcal{P}(\mathcal{S}) \longrightarrow [0,1]$ is given by*

$$H(d)(X,Y) = \max(\max_{x \in X} \min_{y \in Y} d(x,y), \max_{y \in Y} \min_{x \in X} d(x,y))$$

In the setting of defining a metric analogue to lax bisimulation this arises as follows. Given two states that we are comparing we have to look not just at the distance between the probability distributions *for the same action*, but, rather, at the metric between *sets of distributions* corresponding to the different possible actions in the equivalence classes. Since the sets of actions are finite we need not worry about compactness. In this vain, let us define the following operator on the space of lsc metrics.

Definition 3.2.4 *Given $d \in \mathcal{M}$ and any $c \in (0,1)$ we define $\delta(d)$ as follows*

$$\delta(d)((s,a)(t,b)) = (1-c)\alpha((s,a),(t,b)) + cK(d)(\tau_s^a, \tau_t^b)$$

$$F(d)(s,t) = H(\delta(d))(s,t)$$

Note that for simplicity we do not make the dependence of δ on c explicit.

We would like to show that the fixed point of this operator is a lsc metric and relate it to lax bisimulation. For this, we will first need the following lemma.

Lemma 3.2.5 *The lax bisimilarity relation $\sim$ is a closed subset of $S \times S$.*

Proof . Let $E \in EQU$, then $\mathcal{F}(E)$ is clearly an equivalence relation due to the fact that $Rel(\alpha)$ is an equivalence relation.

Let $\{(x_n, y_n)\}$ be a sequence in $\mathcal{F}(E)$ converging to some pair of states (x,y). Let $a \in A$, then for every n there exists some b in which $\alpha((x_n,a),(y_n,b)) = 0$. This means that for every E-closed measurable set X we have that $\tau_{x_n}^a(X) = \tau_{y_n}^{b_n}(X)$. Now as the action space is finite, there must be an infinite subsequence $\{(x'_n, y'_n)\}$ in which the b_n's are all the same action, say b. Thus we have that $\tau_{x'_n}^a(X) = \tau_{y'_n}^b(X)$. Let γ be the discrete metric assigning distance 1 to points if and only if they are not related by E. Since E is closed, we have that γ is lsc so that $K(\gamma)$ is defined. So we have that $K(\gamma)(\tau_{x'_n}^a, \tau_{y'_n}^b) = 0$ by the lemma 3.2.2. Hence, by lemma 3.2.1 $K(\gamma)(\tau_s^a, \tau_t^b)$ is lsc in terms of s and t we have that $K(\gamma)(\tau_x^a, \tau_x^b) = 0$ so that $\tau_x^a(X) = \tau_y^b(X)$. As well α is lsc so $\alpha((x,a),(y,b)) = 0$ and thus $(x,y) \in \mathcal{F}(E)$ and $\mathcal{F}(E)$ is closed.

∎

Now we can proceed with the main result of this section.

Theorem 3.2.6 *F is monotonic and has a least fixed point d_{fix} in which $Rel(d_{fix}) = \sim$.*

Proof . To see that $F(d) \in \mathcal{M}$, note that by Lemma 3.2.1, the map taking (s,t) to $K(d)(\tau_s^a, \tau_t^b)$ is lsc. As well the sum of two lsc metrics is lsc also. Indeed because all the max's and min's are over finite sets F itself is lsc.

The monotonicity of the Hausdorff and Kantorovich metrics together imply monotonicity of F. Thus F has a least fixed point d_{fix}.

By the lemma, $Rel(d_{fix}) = \mathcal{F}(Rel(d_{fix}))$ and is thus contained in bisimulation. For the other direction, consider the $d_\sim$ metric that assigns 0 distance to all bisimilar states and 1 otherwise. This is lsc as $\sim$ is closed (by the previous Lemma). So $\sim = \mathcal{F}(\sim) = \mathcal{F}(Rel(d_\sim)) = Rel(F(d_\sim))$, which implies that $F(d_\sim) \leq d_\sim$. Since d_{fix} is the least prefixed point of F, we must have that $d_{fix} \leq l \leq d_\sim$ so $\sim \subseteq Rel(d_{fix})$.

∎

With the previous theorem we have shown that d_{fix} is in some sense an analogue to the bisimilarity relation $\sim$. The metric d_{fix} completely captures the information in the relation as $Rel(d_{fix}) = \sim$ but for states that are not bisimilar, it will also give a numerical value indicating how different they actually are. In this sense, the metric can be used to describe the similarity between states in a system in a much less brittle manner. This will become very evident when we find that *good* approximate MDPs can be made by lumping together states that are close in this metric.

CHAPTER 4
Logical Characterization

4.1 Formulation and Explanation

As previously explained in the case of unlaxed bisimulation, it is not obvious how to show that two states are bisimilar. The characterization of bisimulation through a modal logic assists us in proving such properties. We would like to define a logic that characterizes the concept of lax bisimulation in such a way that states are lax bisimilar exactly when they satisfy the same formulas in this logic. For simplicity, we assume that the similarity in actions is independent of state, or more formally, that $\forall s, t, s', t'.\alpha((s, a), (t, b)) = \alpha((s', a), (t', b))$. This allows us the slight abuse of notation in which we use $Rel(\alpha)$ as an equivalence relation over actions and simply denote it by α. To be clear this simplifies the original definition of bisimulation as follows.

Definition 4.1.1 *A relation B is a lax probabilistic bisimulation relation if whenever sRt we have that*

(1) $\forall a$ $\exists b$ such that $a\alpha b$ and for all B-closed sets X we have that $\tau_a(s, X) = \tau_b(t, X)$

(2) $\forall b$ $\exists a$ such that $a\alpha b$ and for all B-closed sets X we have that $\tau_b(t, X) = \tau_a(s, X)$

In [DEP98, DEP02] the logical characterization for unlaxed bisimulation was shown using a very simple logic with no negative constructs. It turns out that the

logic does not quite work for the case of lax bisimulation. The one modal operator appearing in the logic of [DEP98, DEP02] is $\langle a \rangle_q$ where a is an action and q is a rational number. For lax bisimulation one has to modify the logic so as to capture the fact that several different formulas must be satisfied simultaneously, because different actions can be matched. Interestingly, this cannot just be done by conjunction.

Definition 4.1.2 *The logic $\mathcal{L}$ has the syntax*

$$T|\phi_1 \wedge \phi_2|\langle a \rangle(\phi_1, ..., \phi_k; l_1, ..., l_k; u_1, ..., u_k)$$

where T and conjunction have the obvious meaning and

$s \models \langle a \rangle(\phi_1, ..., \phi_k; l_1, ..., l_k; u_1, ..., u_k)$ means that $\exists b \in A$ with $b \alpha a$ and $\exists X_1, \ldots X_k \in \Sigma$ so $\tau_b(s, X_i) \in (l_i, u_i), \forall i \in \{1 \ldots k\}$ and when $t \in X_i$, we have that $t \models \phi_i$.

One can immediately notice that the last construct has evolved into something significantly more complicated. In particular:

- Any α-related action can now be used.
- Lower bound constraints ($> q_i$) have been replaced by interval constraints ($\in (l_k, u_k)$).
- Multiple formulas must be satisfied simultaneously.

The first modification seems reasonable as it allows equating states that use related actions to accomplish the same behaviour. This, however, introduces a subtlety: it makes it necessary to use interval constraints instead of lower bounds. To see this, consider the system depicted in Figure 4–1, in which s and t both make a transitions to x with probability 1, and can also make b transitions with probability $\frac{1}{2}$ and $\frac{1}{5}$ respectively. Intuitively, even if a and b are α-related, s and t are still not

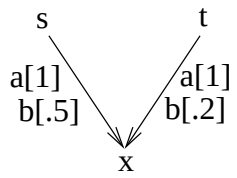

Figure 4–1: A simple LMP

bisimilar, because their transition probability functions are not properly matched. However, they would satisfy all the same lower bound formulas because the a transitions are the highest probability, and they are equal. Thus, it is necessary to use intervals. In this case, the two states can be differentiated by a formula such as $\langle a \rangle (T, 0.4, 0.6)$ which s satisfies but t does not.

Is it really necessary to have the $\langle a \rangle$ construct include multiple formulas? This is a very subtle point. At first sight it looks just like conjunction. Maybe we could simply write

$$\langle a \rangle (\phi_1, l_1, u_1) \wedge \ldots \wedge \langle a \rangle (\phi_k, l_k, u_k)$$

instead of

$$\langle a \rangle (\phi_1, ..., \phi_k; l_1, ..., l_k; u_1, \ldots, u_k).$$

These are, in fact, quite different: in the first formula each term can be matched by a *different* α-equivalent action while in the second formula, the *same* α-equivalent action must be used for each formula.

To illustrate this, consider Figure 4–2 where the α classes are $\{a, b, c\}, \{d\}, \{e\}$ and all transitions have probability 0.5. Clearly s and t are not lax bisimilar because for any bisimulation B, we have that x and y are in their own singleton classes, and there is no α-related action t can use to match s's a actions to both these B-closed

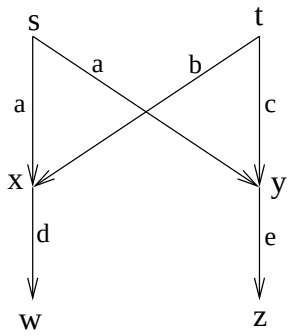

Figure 4–2: Labelled Transition System where the α classes are $\{a, b, c\}, \{d\}, \{e\}$.

sets. The formula needed to distinguish these two states is:

$$\phi = \langle a \rangle (\langle d \rangle (T; l; u), \langle e \rangle (T; l; u); l, l, u, u)$$

where, for example, $l = 0.4$ and $u = 0.6$. For this formula, we have that $s \models \phi$ but $t \not\models \phi$. The proposed replacement formula using conjunction would be satisfied by both s and t, and thus the logic would be too weak.

4.2 Characterization

We now proceed to show that $\mathcal{L}$ indeed characterizes lax bisimulation.

Lemma 4.2.1 *For any formula ϕ in $\mathcal{L}$ the set $[\phi]$ is measurable.*

Proof . We proceed by structural induction on ϕ. First, $s \models T$, $\forall s \in S$ and $S \in \Sigma$, so the base case is true. The set $[\phi_1 \wedge \phi_2] = [\phi_1] \cap [\phi_2] \in \Sigma$ because Σ is a σ-field. Since $A = [a]_\alpha$ is finite, we also have that the set

$$\langle a \rangle (\phi_1, \ldots \phi_k; l_1, \ldots l_k; u_1, \ldots u_k) \ = \cup_{b \in \mathcal{A}} \cap_{i=1}^{k} \tau_b(\cdot, [\phi_i])^{-1}((l_i, u_i))$$

is in Σ. This is due to the fact that each term is measurable (since the τ_b's are Borel measurable) and we are taking a finite union of finite intersections, which is also measurable (because Σ is a σ-algebra). ∎

Proposition 4.2.2 *If B is a bisimulation and sBt then s and t satisfy the same formulas.*

Proof . The cases for T and conjunction are trivial, so let us assume the statement is true for some formulas ϕ_i i.e., any pair of B-related states either both satisfy ϕ_i, or neither of them does. This means that $[\phi_i]$ is B-closed, and by the previous lemma it is also measurable. Now if $s \models \langle a \rangle(\phi_1, \ldots \phi_k; l_1, \ldots l_k; u_1, \ldots u_k)$ then we know there is some $b \in [a]_\alpha$ such that for $1 \leq i \leq k$, $\tau_b(s, [\phi_i]) \in (l_i, u_i)$. Since s and t are bisimilar, there exists a $c \in [b]_\alpha = [a]_\alpha$ with $\tau_c(t, [\phi_i]) = \tau_b(s, [\phi_i]) \in (l_i, u_i)$ and thus $t \models \langle a \rangle(\phi_1, \ldots \phi_k; l_1, \ldots l_k; u_1, \ldots u_k)$ as well. ∎

The reverse direction of the proof follows more or less the same pattern of the corresponding proof from [DGJP00, DGJP03]; it relies on properties of analytic spaces. However, the result here is not just a corollary of the result in [DGJP03], it has to be redone for the new logic.

Definition 4.2.3 *We say that $s \approx t$ if $\forall \phi \in \mathcal{L}$ we have that $s \models \phi \iff t \models \phi$. It is clear that $\approx$ is an equivalence relation and thus we define $q : S \longrightarrow S/\approx$ by $q(s) = [s]_\approx$. When (S, Σ) is measurable, this defines the quotient $(S/\approx, \Sigma_\approx)$ such that a subset $E \subseteq S/\approx$ is in $\Sigma_\approx$ if $q^{-1}(E) \in \Sigma$.*

Lemma 4.2.4 *If $s \approx t$ then $\forall a \in \mathcal{A}$, $\exists b \in \mathcal{A}$ such that $a \alpha b$ and $\forall \phi \in \mathcal{L}$ we have that $\tau_a(s, [\phi]) = \tau_b(t, [\phi])$.*

Proof . Suppose the statement is not true. Then, for some $a \in \mathcal{A}$ we have that $\forall b \in \mathcal{A}$ where $b \alpha a$ there is a formula ϕ_b in which $\tau_a(s, [\phi_b]) \neq \tau_b(t, [\phi_b])$. Because there is a finite number of α classes, there are finitely many such formulas. Thus, we can find intervals $(l_{b_1}, u_{b_1}), \ldots (l_{b_k}, u_{b_k}))$ such that

$$s \models \langle a \rangle (\phi_1, \ldots \phi_k; l_1, \ldots l_k; u_1, \ldots u_k)]$$

but

$$t \not\models \langle a \rangle (\phi_1, \ldots \phi_k; l_1, \ldots l_k; u_1, \ldots u_k)]$$

contradicting $s \approx t$. ∎

The following two lemmas are the crucial facts about analytic spaces that one needs. The first lemma is a very strong structural property which says that there cannot be very many sub-σ-algebras of an analytic space: if the sub-σ-algebra is not too large (countably generated) and not too small (separates points) then it *is* the ambient σ-algebra. This is called the "unique structure theorem." The result appears as Theorem 3.3.5 of "Invitation to C-$*$ algebras" by Arverson [Arv76] and one of its corollaries.

Lemma 4.2.5 *Let $(X, \mathcal{B})$ be an analytic space and let $\mathcal{B}_0$ be a countably generated sub-σ-field of $\mathcal{B}$ which separates points in X. Then $\mathcal{B}_0 = \mathcal{B}$.*

The second lemma shows that quotienting under a suitable equivalence relation preserves the property of being an analytic space.

Lemma 4.2.6 *Let X be an analytic space and let $\sim$ be an equivalence relation on X. Assume there is a sequence $f_1, f_2, \ldots$ of real valued Borel functions on X such*

that for any pair of points $x, y \in X$, $x \sim y$ if and only if $f_n(x) = f_n(y)$ for all n. Then $X/\sim$ is an analytic space.

The main theorem can now be stated.

Theorem 4.2.7 *The relation $\approx$ is an α-lax probabilistic bisimulation.*

Proof . We first show that $S/\approx$ is analytic. Let $\{\phi_i : i \in \mathbb{N}\}$ be an indexing of all formulas. We know that $[\phi_i]$ is a measurable set, so the characteristic function $\chi_{\phi_i} : s \longrightarrow \{0, 1\}$ is also measurable. Then we have that $x \approx y$ if and only if $\forall i \in \mathbb{N}$, $\chi_{\phi_i}(x) = \chi_{\phi_i}(y)$. Thus, by Lemma 4.2.6 it follows that $S/\approx$ is an analytic space.

Let $\mathcal{B} = \{q([\phi_i]_S) : i \in \mathbb{N}\}$. Then for any $q([\phi_i]) \in \mathcal{B}$, it is clear that $q^{-1}(q([\phi_i])) = [\phi_i] \in \Sigma$ as shown above. Now $\sigma(\mathcal{B})$ separates points $x, y \in S/\approx$, $x \neq y$ if there is a formula ϕ so that for all states $x' \in q^{-1}(x)$ and $y' \in q^{-1}(y)$ we have that $x' \in [\phi]$ and $y' \notin [\phi]$. This implies $x \in q([\phi])$ and $y \notin q([\phi])$. Thus, since $\sigma(\mathcal{B})$ is countably generated, from Lemma 4.2.5 we have that $\sigma(\mathcal{B}) = \Sigma_{\approx}$.

Let $s \approx t$ and let $a \in \mathcal{A}$. For any $B \in \sigma(\mathcal{B})$, we define $\rho(B) = \tau_a(s, q^{-1}(B))$. Then by Lemma 4.2.4 we know that there is a $b \in \mathcal{A}$ such that $a \alpha b$ and for any $[\phi]$ we have $\tau_a(s, [\phi]) = \tau_b(t, [\phi])$. Note that for any $B \in \mathcal{B}$ we also have that $q^{-1}(B) = [\phi]$ for some ϕ and thus $\rho(B) = \tau_b(t, q^{-1}(B))$. Now $\mathcal{B}$ is closed under finite intersections, so so from proposition 2.2.1, we have that $\rho(B) = \tau_b(t, q^{-1}(B))$ also for any $B \in \Sigma_{\approx}$. Now if $X \in \Sigma$ is $\approx$-closed then we have that $X = q^{-1}(q(X))$ and hence $q(X) \in \Sigma_{\approx}$. This implies $\tau_a(s, X) = \rho(q(X)) = \tau_b(t, X)$. The other direction follows similarly and thus $\approx$ is a bisimulation. ∎

Theorem 4.2.8 *Given two states $s, t \in S$ we have that $s \sim t$ if and only if $\forall \phi \in \mathcal{L}$, $s \models \phi \iff t \models \phi$.*

Proof . By Proposition 4.2.2, if $s \sim t$ then they both satisfy the same formulas. Conversely, if s and t satisfy the same formulas, then by Definition 4.2.3 $s \approx t$. By Theorem 4.2.7 $\approx$ is a bisimulation, so $\approx \subseteq \sim$ which implies $s \sim t$. ∎

We have, as a corollary, another avenue to show that $\sim$ is a bisimulation relation, at least in this limited context in which α is treated as an equivalence relation.

Corollary 4.2.9 *The bisimilarity relation $\sim$ is itself a bisimulation relation.*

The characterization of lax bisimulation can be exploited to more easily demonstrate whether two states are bisimilar or not. For example, if one wanted to demonstrate that two states s and t are not bisimilar one can simply present a specific formula ϕ that is satisfied by s and not t. Without a logical characterization, a complex proof by contradiction would have to be presented. Furthermore, for a fixed finite system, it should be intuitively obvious that the maximum "depth" of formula needed to differentiate non bisimilar states is also finite. This observation motivates an algorithm that breaks the state space of a finite system into bisimilarity classes.

4.3 Computing Distinguishing Formulas

In this section we give an algorithm that given a system with a finite state spaces, finds the bisimilarity classes, and for each pair of non-bisimilar states provides a formula on which the states disagree. This follows the general ideas of the algorithm from [DEP02] but the fact that there is a new kind of formula means that the details and proof are different.

Due to the discreteness of the action and state spaces, we can assume that the minimum difference in transition probabilities is strictly greater than ϵ. It is then convenient to consider formulas of the form.

$$\langle a \rangle(\phi_1, \ldots, \phi_k; q_1 - \epsilon, \ldots, q_k; q_1 + \epsilon, \ldots, q_k + \epsilon)$$

which we simply denote by $\langle a \rangle(\vec{\phi}, \vec{q})$.

The algorithm iteratively refines a partition of the state space S until the partition represents the equivalence classes of the lax bisimilarity relation. D_1 is initialized with to contain only the set of all states. At each step it is split by looking at all possible transitions to sets already existing in D_1. Through this process each set B in D_1 gets assigned a formula $F(B)$ so that B contains all the states that represent B. This process continues until D_1 remains unchanged.

Theorem 4.3.1 *Two states satisfy the same formulas iff they belong to the same sets in D_1 after executing the previous algorithm.*

Proof . The loops will terminate the first time D_1 does not change. As D_1 can only increase, the loop is executed a maximum of $2^{|S|}$ times and thus the algorithm will terminate.

We show necessity by showing that every set in $D_1 \cup D_2$ is represented by a formula. The whole set S is represented by the formula T. So, suppose that after k iterations of the inside for loop, every element of $D_1 \cup D_2$ represents a formula. We must then show that each of the sets returned by split(B,$[a]$,$\vec{C}$) represents a formula. Now we are assuming that B and $\vec{C}$ represent formulas ϕ and $\vec{\psi}$. A constructed set B_1 is of the form $\{s \in B : \exists b \in [a] \wedge \tau_b(s, C_1) \in (q_1 - \epsilon, q_1 + \epsilon) \wedge \ldots \wedge \tau_b(s, C_n) \in (q_n - \epsilon, q_n + \epsilon)\}$ which represents the formula $\phi \wedge \langle a \rangle(\vec{\psi}, \vec{q})$ thus indeed all the sets added to $D_1 \cup D_2$ represent formulas. This means that if two states satisfy the same formulas then they must be in the same sets of $D_1 \cup D_2$.

Function *bisim()*
$F(S) = T$
$D_1 = \{S\}$
$D_2 = \emptyset$
while $D_1 \neq D_2$ **do**
  foreach $[a] \in Act_\alpha$ *and* $\vec{C} \in \mathcal{P}(D_1)^1$ **do**
    $D_2 = D_1$
    $D_1 = \emptyset$
    foreach $B \in D_2$ **do**
      $D_1 = D_1 \cup \text{split}(B, [a], \vec{C})$
    end
  end
end
Function *split(B,[a],$\vec{C}$)*
$(C_1, ..., C_n) = \vec{C}$;
$D = \{B\}$
$Q = \{(\tau_b(t, C_1), \ldots, \tau_b(t, C_n) : t \in B, b \in [a]\}$
foreach $(q_1, ..., q_n) \in Q$ **do**
  $B_1 = \{s \in B : \exists b \in [a] \wedge$
    $\tau_b(s, C_1) \in (q_1 - \epsilon, q_1 + \epsilon) \wedge ... \wedge \tau_b(s, C_n) \in (q_n - \epsilon, q_n + \epsilon)\}$
  $F(B_1) = F(B) \wedge \langle a \rangle (F(C_1), ..., F(C_n); \vec{q})$
  $D = D \cup \{B_1\}$
end
return D

To show sufficiency we will show that every element $B \in D_1 \cup D_2$ represents a formula ϕ in the restricted logic

$$T|\phi_1 \wedge \phi_2|\langle a \rangle(\vec{\phi}, \vec{q})$$

To see why this works, assume a formula ϕ in the original logic separates two states s and t in such a way that $s \models \phi$ but $t \not\models \phi$. Such a formula can contain arbitrary intervals, but each of these can be shrunk to the size of 2ϵ in such a way that they are centered around the exact probabilities s needs to satisfy ϕ. Thus if two states s and t do not satisfy the same formulas in the original logic, then they will not satisfy the same formulas in this reduced logic. By showing that all formulas are represented by sets in the reduced logic we will then have shown that they are not in the same sets in D_1.

We do this by structural induction of the formulas in the restricted logic. First, the whole set S represents the formula T.

Now let $\phi_1, ..., \phi_k$ be formulas represented in the sets in D_1 by sets $C_1, ..., C_k$. Assume that $\langle a \rangle(\phi_1, ..., \phi_k, q_1, ..., q_k)$ is not represented for some action a and rationals $q_1, ..., q_k$ so necessarily there is some state s that satisfies it. Then D_1 will be modified by the call split(S,[a],$\vec{C}$) which contradicts the algorithm being finished and thus formulas of this form must be represented in D_1.

Now assume that ϕ and ψ are both represented in D_1. By the construction of ϕ in the algorithm we can see that it is the result of some finite number of calls to split and thus can be written as

$$\phi = \langle a_1 \rangle(\vec{\phi_1}, \vec{q_1}) \wedge \cdots \wedge \langle a_n \rangle(\vec{\phi_n}, \vec{q_n})$$

where $\vec{\phi_i}$ is represented by a vector of sets from D_1 which we call C_i. Now we will show that $\phi \wedge \psi$ is represented in D_1 by induction on n. Indeed assume that $\zeta \wedge \psi$ is represented in D_1 when ζ is a conjunction of k formulas. Now, for some formula ϕ that is a conjunction of $k+1$ such terms we can write

$$\phi = \zeta \wedge \langle a_{k+1} \rangle(\vec{\phi_{k+1}}, \vec{q_{k+1}})$$

then by our induction hypothesis we assume that $\zeta \wedge \phi$ is represented by the set $C' \in D_1$. Then a set representing $\phi \wedge \psi$ will be created in the call to split(C, a_{k+1}, C') contradicting the algorithm being finished.

∎

CHAPTER 5
Case Study: Markov Decision Processes

In this chapter, we apply lax bisimulation to a specific type of probabilistic system called the Markov Decision Process. Markov Decision Processes (MDPs) are similar to LMPs except that numeric rewards are associated with each choice of action in each state. Maximizing the long run rewards by choosing appropriate actions in each state is a very popular formalism for decision making under uncertainty [Put94]. A significant problem is computing this optimal strategy when the state and action space are very large and/or continuous. A popular approach is *state abstraction*, in which states are grouped together in partitions, or aggregates, and the optimal policy ie and/or continuous. Li et al (2006) provide a nice comparative survey of approaches to state abstraction. The work we present here bridges two such methods: bisimulation-based approaches and methods based on MDP homomorphisms.

Bisimulation has been specialized for MDPs by Givan et al (2003). Indeed, one can simply use the difference in rewards for state-action pairs as the defining metric. However, equivalence notions are very brittle for probabilistic systems, as a small change in the values of the transition probability distributions (or rewards) for different states can drastically change the equivalence relation between states. Thus, it is desirable to have a notion that captures the degree in which two states differ. In recent work, Ferns et al (2004, 2005, 2006) has extended the unlaxed bisimulation metric to MDPs.

One of the disadvantages of bisimulation and the corresponding metrics is that they require that the behavior matches for exactly the same actions. However, in many cases of practical interest, actions with the exact same label may not match, but the environment may contain symmetries and other types of special structure, which may allow correspondences between states by matching their behavior with *different* actions. This idea was formalized by [RB04] with the concept of MDP homomorphisms. MDP homomorphisms specify a map matching equivalent states as well as equivalent actions in such states. This matching can then be used to transfer policies between different MDPs. Because MDP homomorphisms specify equivalence relations, they are also brittle. Hence [RB04] proposed using *approximate homomorphisms*, which allow aggregation of states which are not exactly equivalent. They define an MDP over these partitions and quantify the approximate loss resulting from using this MDP, compared to the original system. As expected, the bound depends on the quality of the partition. Subsequent work, e.g. [WB06], constructs such partitions heuristically.

In this chapter, we attempt to construct provably good, approximate MDP homomorphisms from first principles. First, we relate the notion of MDP homomorphisms to the concept of lax bisimulation. This allows us to define the lax bisimulation metric on states, similarly to existing bisimulation metrics. Interestingly, this approach works both for discrete and for continuous actions. We show that the difference in the optimal value function of two states is bounded above by this metric. This allows us to provide a state aggregation algorithm with provable

approximation guarantees. Empirical illustrations show that this approach provides much better state space compression than the use of bisimilarity metrics.

5.1 Background

A finite Markov decision process (MDP) is a tuple $\langle S, A, P, R \rangle$, where S is a finite set of states, A is a set of actions, $P : S \times A \times S \longrightarrow [0, 1]$ is the transition model, with $P(s, a, s')$ being the probability of transition from state s to s' under action a, and $r : S \times A \longrightarrow \mathbf{R}$ is the reward function with $r(s, a)$ being the reward for performing action a in state s. For the purpose of this chapter, the state space S is assumed to be finite, but the action set A could be finite or infinite (as will be detailed later). We assume without loss of generality that rewards are bounded in $[0, 1]$. We will use $P^a_{ss'}$ and r^a_s as shorthand for $P(s, a, s')$ and $r(s, a)$ respectively.

5.1.1 Policies and their Values

A deterministic policy $\pi : S \longrightarrow A$ specifies which action should be taken in every state. By following policy π from state s, an agent can expect to receive a value $V^\pi(s) = E(\sum_{t=0}^{\infty} \gamma^t r_t | s_0 = s, \pi)$ where $\gamma \in (0, 1)$ is a discount factor and r_t is the reward received at time t. In a finite MDP, the optimal value function V^* is unique and satisfies the following formulas, known as the Bellman optimality equations:

$$V^*(s) = \max_{a \in A} (r^a_s + \gamma \sum_{s'} P^a_{ss'} V^*(s')), \forall s \in S$$

If the action space is continuous, we will assume that it is compact, so the max can be taken and the above results still hold [Put94]. Given the optimal value function, an optimal policy is easily inferred by simply taking at every state greedy action with respect to the one-step-lookahead value. It is well known that the optimal value

function can be computed through a sequence of iterates: $V_0(s) = 0$ and

$$V_{n+1}(s) = \max_a (r_s^a + \gamma \sum_{s'} P_{ss'}^a V_n(s')),$$

which converges to V^* uniformly.

5.1.2 MDP Bisimulation

Ideally, if the state space is very large, "similar" states should be grouped together in order to speed up this type of computation. Bisimulation can be adapted to MDPs [GDG03] as a notion of equivalence between states. The traditional definition is as follows: A relation $B \subseteq S \times S$ is a *bisimulation relation* if

$$sBt \Leftrightarrow \forall a.(r_s^a = r_t^a \text{ and } \forall X \in S/B.P(s, a, X) = P(t, a, X))$$

.

From this definition, it follows that bisimilar states can match each others' actions to achieve the same rewards. Hence, bisimilar states have the same optimal value [GDG03] (and indeed, will have the same value under any policy). However, bisimulation is not robust to small changes in the rewards or the transition probabilities so again we are better off moving to a metric.

To this end, Ferns et al (2004) proposed a *bisimulation metric*, defined as the least fixed point e_{fix} of the following operator on the lattice of 1-bounded metrics :

$$G(e)(s, t) = \max_a ((1 - c)|r_s^a - r_t^a| + cK(e)(P_s^a, P_t^a)) \tag{5.1}$$

Again, this is simply the unlaxed bisimulation metric using the difference in rewards to define the underlying metric space.

5.2 Lax MDP Bisimulation

In many MDPs of practical interest, actions with the exact same label may not match, but the environment may contain symmetries and other types of special structure, which may allow correspondences between *different* actions at certain states. For example, consider the MDP in Figure 5–1, where a reward of 1 is obtained at state 3, and the transition dynamics are deterministic for all navigation actions. In this case, states 1 and 4 have a bisimulation distance of 1. However, the action of going down in state 1 can be matched perfectly (in terms of effects) by the action of going up in state 4. Indeed, this is the exact type of matching laxed bisimulation performs but our definition of lax bisimulation relies on the presence of a metric between state-action pairs. The formulation of the unlaxed metric in the previous section gives a hint in how this should be defined. Indeed if we define the underlying metric as:

$$\alpha((s,a),(t,b)) = |r_s^a - r_t^b|$$

then states should only be equated if they can match each others rewards immediately and transition with equal probabilities to states where the same holds.

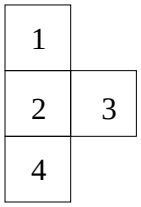

Figure 5–1: A small grid world.

This notion is very closely related to the idea of MDP homomorphisms (Ravindran & Barto, 2003). We now formally establish this connection.

Definition 5.2.1 *(Ravindran & Barto, 2003) A* MDP homomorphism h *from* $M = \langle S, A, P, R\rangle$ *to* $M' = \langle S', A', P', R'\rangle$ *is a tuple of surjections* $\langle f, \{g_s : s \in S\}\rangle$ *with* $h(s,a) = (f(s), g_s(a))$, *where* $f : S \longrightarrow S'$ *and* $g_s : A_s \longrightarrow A'_{f(s)}$ *such that* $R(s,a) = R'(f(s), g_s(a))$ *and* $P(s,a,f^{-1}(f(s'))) = P'(f(s), g_s(a), f(s'))$

Hence, a homomorphism puts states into a correspondence, and has a state-dependent mapping between actions as well. We now show that homomorphisms are identical to lax probabilistic bisimulation.

Lemma 5.2.2 *Let* h *be a MDP homomorphism and define the relation* B *such that* sBt *iff* $f(s) = f(t)$. *Then* B *is a lax probabilistic bisimulation.*

Proof: Let sBt and let $a \in A$ (so $g_s(a) \in A$). By the hypothesis, $f(s) = f(t)$ and since g_t is a surjection to $A_{f(t)} = A_{f(s)}$, there must be some $b \in A_t$ with $g_t(b) = g_s(a)$. Hence,

$$R(s,a) = R'(f(s), g_s(a)) = R'(f(t), g_t(b)) = R(t,b)$$

Let X be a non-empty B-closed set such that $f^{-1}(f(s')) = X$ for some s'. Then:

$$\begin{aligned} P(s,a,X) &= P'(f(s), g_s(a), f(s')) = P'(f(t), g_t(b), f(s')) \\ &= P(t,b,X) \end{aligned}$$

which concludes the proof. $\diamond$

Analogously, we can construct an MDP homomorphism from a lax bisimulation.

Lemma 5.2.3 *Let* B *be a lax bisimulation relation. Then there exists a MDP homomorphism in which* $sBt \implies f(s) = f(t)$.

Proof: Consider the partition S/B induced by the equivalence relation B on set S. For each equivalence class $X \in S/B$, we choose a representative state $s_X \in X$ and define $f(s_X) = s_X$ and $g_{s_X}(a) = a, \forall a \in A$. Then, for any $s \sim s_X$, we define $f(s) = s_X$. From the definition of lax bisimulation , we have that $\forall a \exists b, P(s, a, X') = P(s_X, b, X'), \forall X' \in S/B$. Hence, we set $g_s(a) = b$. Then, we have:

$$\begin{aligned} P'(f(s), g_s(a), f(s')) &= P'(f(s_X), b', f^{-1}(f(s')) \\ &= P(s_X, b, f^{-1}(f(s')) \\ &= P(s, a, f^{-1}(f(s')) \end{aligned}$$

Also, $R'(f(s), g_s(a)) = R'(f(s_X), b) = R(s_X, a)$. $\diamond$

Putting these two lemmas together gives us the following theorem.

Theorem 5.2.4 *Two states s and t are bisimilar if and only if they are related by some MDP homomorphism $f, g_s : s \in S$ in the sense that $f(s) = f(t)$.*

By specifying α above, we have implicitly defined a lax MDP bisimulation metric d_{fix} analogous to the unlaxed MDP bisimulation metric e_{fix}. As both e_{fix} and d_{fix} quantify the difference in behaviour between states, it is not surprising to see that they constrain the difference in optimal value. Indeed, the bound below has previously been shown [FPP04] for e_{fix}, but we also show that our laxed metric d_{fix} is even tighter.

Theorem 5.2.5 *Let e_{fix} be the metric defined in [FPP04]. Then if $\gamma \leq c$, we have the following bounds.*

$$(1 - c)|V^*(s) - V^*(t)| \leq d_{fix}(s, t) \leq e_{fix}(s, t)$$

Proof . We show via induction on n that

$$(1-c)|V_n(s) - V_n(t)| \leq d_{fix}(s,t) \leq e_{fix}(s,t)$$

and then the result follows by merely taking limits.

For the base case note that

$$(1-c)|V_0(s) - V_0(t)| = d_0(s,t) = e_0(s,t) = 0$$

Now, assume this holds for n. By the monotonicity of F, we have that

$$F(d_n)(s,t) \leq F(e_n)(s,t)$$

Now, for any a, it is clear that

$$\delta(e_n)((s,a),(t,a)) \leq G(e_n)(s,t)$$

but then it is easy to see that

$$\begin{aligned}
&F(e_n)(s,t) \\
&\leq \max(\max_a \delta(e_n)((s,a),(t,a)), \max_b \delta(e_n)((s,b),(t,b)) \\
&\leq \max(\max_a G(e_n)(s,t), G(e_n)(s,t)) \\
&= G(e_n)(s,t)
\end{aligned}$$

and so $d_{n+1} \leq e_{n+1}$

Without loss of generality, assume that $V_{n+1}(s) > V_{n+1}(t)$. Then

$$
\begin{aligned}
&(1-c)|V_{n+1}(s) - V_{n+1}(t)| \\
&= (1-c)|\max_a(r_s^a + \gamma \sum_u P_{su}^a V_n(u)) - \max_b(r_t^b + \gamma \sum_u P_{tu}^b V_n(u))| \\
&= (1-c)|(r_s^{a'} + \gamma \sum_u P_{su}^{a'} V_n(u)) - (r_t^{b'} + \gamma \sum_u P_{tu}^{b'} V_n(u))| \\
&= (1-c)\min_b |(r_s^{a'} + \gamma \sum_u P_{su}^{a'} V_n(u)) - (r_t^b + \gamma \sum_u P_{tu}^b V_n(u))| \\
&\leq (1-c)\max_a \min_b |(r_s^a + \gamma \sum_u P_{su}^a V_n(u)) - (r_t^b + \gamma \sum_u P_{tu}^b V_n(u))| \\
&\leq \max_a \min_b((1-c)|r_s^a - r_t^b| + c|\sum_u (P_{su}^a - P_{tu}^b)\frac{(1-c)\gamma}{c} V_n(u)|)
\end{aligned}
$$

Now since $\gamma \leq c$,

$$0 \leq \frac{(1-c)\gamma}{c} V_i(u) \leq \frac{(1-c)\gamma}{c(1-\gamma)} \leq 1$$

and by the induction hypothesis

$$\frac{(1-c)\gamma}{c} V_n(s) - \frac{(1-c)\gamma}{c} V_n(t) \leq (1-c)|V_n(s) - V_n(t)| \leq d_n(s,t)$$

So $\{\frac{(1-c)\gamma}{c} V_n(u) : u \in S\}$ is a feasible solution to the LP for $K(d_n)(P_s^a, P_t^b)$. We then continue the inequality:

$$
\begin{aligned}
&(1-c)|V_{n+1}(s) - V_{n+1}(t)| \\
&\leq \max_a \min_b((1-c)|r_s^a - r_t^b| + cT_k(d_n)(P_s^a, P_t^b)) \\
&= F(d_n)(s,t) = d_{n+1}(s,t)
\end{aligned}
$$

5.3 State Aggregation

A MDP homomorphism allows one to aggregate states into Partitions by defining a relabeling of each state's actions in the Partition so that each state in the Partition can accumulate long run rewards in the same manner. That said, sensible models can be defined [FPP04] in which Partitioned states are not homomorphic by averaging the transition probabilities within each Partition. Here, we merge these two ideas by defining a relabeling of each state's actions and then taking averages.

Definition 5.3.1 *Given a MDP M, an aggregated MDP M' is given by $(S', A, \{P^a_{CD} : a \in A; C, D \in S'\}, \{r^a_C : a \in A, C \in S'\}, \rho, g_s : s \in S)$ where S' is a partition of S, $\rho : S \longrightarrow S'$ maps states to their aggregates, each $g_s : A \longrightarrow A$ (we say g^a_s as shorthand for $g_s(a)$) relabels the action set and we have that $\forall C, D \in S'$ and $a \in A$ that*

$$P^a_{CD} = \frac{1}{|C|} \sum_{s \in C} P^{g^a_s}_s(D) \text{ and } r^a_C = \frac{1}{|C|} \sum_{s \in C} r^{g^a_s}_s$$

When a MDP homomorphism defines a homomorphism, all the states in a partition have actions that are relabeled specifically so they can exactly match each others behaviour. Thus a policy in the aggregate MDP can be lifted to the original MDP by choosing the relabeled actions.

Definition 5.3.2 *If M' is an aggregation of a MDP M and π' is a policy in M' then the lifted policy is defined by $\pi(s) = g_s(\pi'(a))$.*

Using a lax bisimulation metric, it is possible to choose appropriate relabellings so that states within a Partition can approximately match each others actions.

Definition 5.3.3 *Given a lax bisimulation metric d and a MDP M, we say that an aggregated MDP M' is d-consistent if each aggregated class C has a state $s \in C$ which we call the representative of C in which $\forall t \in C$ we have that*

$$\delta(d)((s, g_s^a), (s, g_t^a)) \leq F(d)(s, t)$$

When the relabellings are chosen in this way, we can solve for the maximum value function of the aggregated MDP and be assured that the original state's maximum value is quite close to the maximum value of the Partition it is contained in.

Theorem 5.3.4 *If $\gamma \leq c$ and M' is a d_ζ-consistent aggregation of a MDP M and $n \leq \zeta$ then $\forall s \in S$ we have that*

$$(1 - c)|V_n(\rho(s)) - V_n(s)| \leq m(\rho(s)) + M \sum_{k=1}^{n-1} \gamma^{n-k}$$

and if π' is any policy in M' and π is the lifted policy to M then

$$(1 - c)|V_n^{\pi'}(\rho(s)) - V_n^{\pi}(s)| \leq m(\rho(s)) + M \sum_{k=1}^{n-1} \gamma^{n-k}$$

where $m(C) = 2 \max_{t \in C} d_\zeta(s', t)$ such that s' is the representative state of C and $M = \max_C m(C)$.

Proof .

$$
\begin{aligned}
&|V_{n+1}(\rho(s)) - V_{n+1}(s)| \\
&= |\max_a (r^a_{\rho(s)} + \gamma \sum_{D \in S'} P^a_{\rho(s)D} V_n(D)) - \max_a (r^a_s + \gamma \sum_u P^a_{su} V_n(u))| \\
&\leq \frac{1}{|\rho(s)|} \sum_{t \in \rho(s)} \max_a (|r^{g^a_t}_t - r^{g^a_s}_s| \\
&+ \gamma |\sum_{D \in S'} \sum_{u \in D} P^{g^a_t}_{tu} V_n(D) - \sum_u P^{g^a_s}_{su} V_n(u)|) \\
&\leq \frac{1}{|\rho(s)|} \sum_{t \in \rho(s)} \max_a (|r^{g^a_t}_t - r^{g^a_s}_s| + \gamma |\sum_u P^{g^a_t}_{tu} V_n(\rho(u)) - P^{g^a_s}_{su} V_n(u)|) \\
&\leq \frac{1}{|\rho(s)|} \sum_{t \in \rho(s)} \max_a (|r^{g^a_t}_t - r^{g^a_s}_s| + \gamma |\sum_u (P^{g^a_t}_{tu} - P^{g^a_s}_{su}) V_n(u) \\
&+ \gamma |\sum_u P^{g^a_t}_{tu} (V_n(\rho(u)) - V_n(u))|) \\
&\leq \frac{1}{(1-c)|\rho(s)|} \sum_{t \in \rho(s)} \max_a ((1-c)|r^{g^a_s}_s - r^{g^a_t}_t| + c|\sum_u (P^{g^a_t}_{tu} - P^{g^a_s}_{su}) \frac{(1-c)\gamma}{c} V_n(u)|) \\
&+ \frac{\gamma}{|\rho(s)|} \sum_{t \in \rho(s)} \max_a \sum_u P^{g^a_t}_{tu} |V_n(\rho(u)) - V_n(u)|
\end{aligned}
$$

Now from the previous theorem we know that $\{\frac{(1-c)\gamma}{c} V_n(u) : u \in S\}$ is a feasible solution to the primal LP for $K(d_n)(P^{g^a_s}_s, P^{g^a_t}_t)$. So, let z be the representative used for $\rho(s)$ then we have

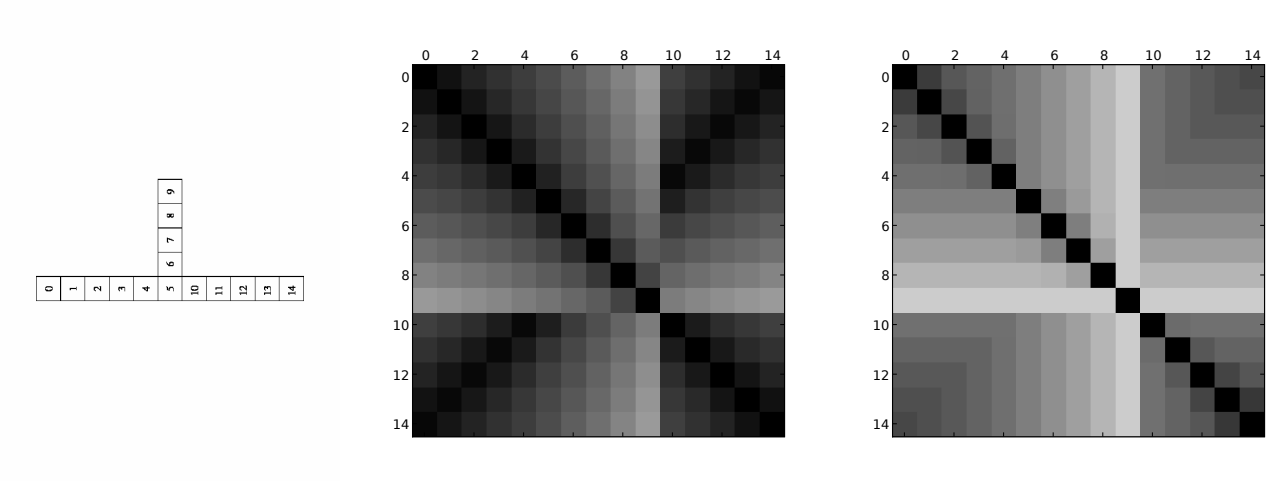

Figure 5–2: Computed Metrics after 100 Iterations. Left: The MDP structure. Center: The lax metric. Right: The unlaxed metric.

$$
\begin{aligned}
&\leq (1-c)|r_s^{g_s^a} - r_t^{g_t^a}| + cK(d_n)(P_s^{g_s^a}, P_t^{g_t^a}) \\
&\leq (1-c)|r_s^{g_s^a} - r_t^{g_t^a}| + cK(d_\zeta)(P_s^{g_s^a}, P_t^{g_t^a}) \\
&\leq (1-c)|r_s^{g_s^a} - r_z^{g_z^a}| + cK(d_\zeta)(P_s^{g_s^a}, P_z^{g_z^a}) \\
&+ (1-c)|r_z^{g_z^a} - r_t^{g_t^a}| + cK(d_\zeta)(P_z^{g_z^a}, P_t^{g_t^a}) \\
&= d_\zeta(s,z) + d_\zeta(z,t) \leq m(\rho(s))
\end{aligned}
$$

We continue with the original inequality using these two results.

$$
\begin{aligned}
&\leq \frac{1}{(1-c)} \sum_{t \in \rho(s)} ((1-c)|r_s^{g_s^a} - r_t^{g_t^a}| + cK(d_n)(P_s^{g_s^a}, P_t^{g_t^a})) \\
&+ \frac{\gamma}{|\rho(s)|} \sum_{t \in \rho(s)} \max_a \sum_u P_{tu}^{g_t^a} \max_u |V_n(\rho(u)) - V_n(u)| \\
&\leq \frac{1}{(1-c)|\rho(s)|} \sum_{t \in \rho(s)} m(\rho(s)) + \gamma \max_u |V_n(\rho(u)) - V_n(u)| \\
&\leq \frac{m(\rho(s))}{(1-c)} + \gamma \max_u (\frac{m(\rho(s))}{(1-c)} + M \sum_{k=1}^{n-1} \gamma^{n-k}) \\
&\leq \frac{1}{(1-c)} (m(\rho(s)) + \gamma \max_u m(\rho(u)) + M \sum_{k=1}^{n-1} \gamma^{n+1-k}) \\
&\leq \frac{1}{(1-c)} (m(\rho(s)) + M \sum_{k=1}^{n} \gamma^{(n+1)-k})
\end{aligned}
$$

The second proof is nearly identical except that instead of max'ing over actions the action selected by the policy, $a = \pi'(\rho(s))$, and lifted policy $g_s^a = \pi(s)$ is used.

∎

By taking limits we get the following theorem.

Theorem 5.3.5 *If $\gamma \leq c$ and M' is an d_{fix}-consistent aggregation of a MDP M, then $\forall s \in S$ we have that*

$$(1-c)|V^*(\rho(s)) - V^*(s)| \leq m(\rho(s)) + \frac{\gamma}{1-\gamma} M$$

and if π' is any policy in M' and π is the lifted policy to M then

$$(1-c)|V^{\pi'}(\rho(s)) - V^{\pi}(s)| \leq m(\rho(s)) + \frac{\gamma}{1-\gamma} M$$

where $m(C) = 2\max_{t \in C} d_{fix}(s', t)$ *such that* s' *is the representative state of* C *and* $M = \max_C m(C)$.

One appropriate way to aggregate states is to choose some desired error bound $\epsilon > 0$ and ensure that the states within each partition are within an ϵ-ball. A simple way to do this is given in the following algorithm.

$Partitions = \emptyset$;

while $S \neq \emptyset$ **do**

$s = RandomElementOf(S)$;

$Partition = \{t \in S : d(s, t) \leq \epsilon\}$;

$S = S - Partition$;

$Partitions = Partitions \cup \{Partition\}$;

end

It has been noted that when the above condition holds, then under the unlaxed bisimulation metric e_{fix}, we can be assured that for each state s, that $|V^*(\rho(s)) - V(s)|$ is bounded by $\frac{2\epsilon}{(1-c)(1-\gamma)}$. The theorem above shows that under the lax bisimulation metric d_{fix} this difference is actually bounded by $\frac{4\epsilon}{(1-c)(1-\gamma)}$. Despite this, we will later illustrate that a massive reduction in state space can be achieved by moving from e_{fix} to d_{fix} even when moving from ϵ to $\epsilon' = \frac{\epsilon}{2}$.

Indeed the two sets of policies that yield optimal value in the original and aggregated MDP are not explicitly related. That said there is an obvious way [RB04] to construct a policy for the original MDP from a policy for the aggregated MDP.

For large systems, it might not be feasible to compute the metric e_{fix} in the original MDP. In this case, we might want to use some sort of heuristic or prior knowledge to create an aggregation. In this case, it was recently shown by [RB04]

using work from [Whi78] that a bound the difference in values between an optimal policy in the aggregated MDP and the lifted policy in the original MDP can be obtained. We will now derive a result in which this bound is a simple corollary to.

Theorem 5.3.6 *If M' is an aggregation of a MDP M, π' is an optimal policy in M', π is the policy lifted from π' to M and d'_{fix} corresponds to our metric computed on M' then*

$$
\begin{aligned}
&|V^{\pi}(s) - V^{\pi'}(\rho(s))| \\
&\leq \frac{2}{1-\gamma} \max_{s,a} |r_s^{g_s^a} - r^a_{\rho(s)}| + \frac{\gamma}{(1-c)} \max_{s,a} K(d'_{fix})(P_s^{g_s^a}, P^a_{\rho(s)})
\end{aligned}
$$

Proof .

$$
\begin{aligned}
&|V^{\pi}(s) - V^{\pi'}(\rho(s))| \\
&\leq \frac{2}{1-\gamma} \max_{s,a} |r_s^{g_s^a} - r^a_{\rho(s)} + \gamma \sum_C (P_{sC}^{g_s^a} - P^a_{\rho(s)C}) V^{\pi'}(C)| \\
&\leq \frac{2}{1-\gamma} \max_{s,a} |r_s^{g_s^a} - r^a_{\rho(s)}| + \max_{s,a} \gamma | \sum_C (P_{sC}^{g_s^a} - P^a_{\rho(s)C}) V^{\pi'}(C)| \\
&\leq \frac{2}{1-\gamma} \max_{s,a} |r_s^{g_s^a} - r^a_{\rho(s)}| + \max_{s,a} \frac{\gamma}{(1-c)} K(d'_{fix})(P_s^{g_s^a}, P^a_{\rho(s)})
\end{aligned}
$$

The first inequality originally comes from [Whi78] and is applied to MDPs in [RB04]. The last inequality holds since π' is an optimal policy and thus by the proof of theorem 5.2.5 we know that $\{\frac{V^{\pi'}(C)}{(1-c)} : C \in S'\}$ is a feasible solution. ∎

As a corollary, we can get the same bound as in [RB04] by bounding the Kantorovich by the total variation metric.

Corollary 5.3.7 *Let* $\Delta = \max_{C,a} r_C^a - \min_{C,a} r_C^a$ *be the maximum difference in rewards in the aggregated MDP then*

$$\begin{aligned}
&|V^\pi(s) - V^\pi(\rho(s))| \\
&\leq \frac{2}{1-\gamma}\left(\max_{s,a}|r_s^{g_s^a} - r_{\rho(s)}^a| + \frac{\gamma}{1-\gamma}\Delta \cdot TV(P_s^{g_s^a}, P_{\rho(s)}^a)\right)
\end{aligned}$$

Proof . This follows from the fact that

$$\begin{aligned}
&\max_{C,D} d'_{fix}(C,D) \\
&\leq (1-c)\Delta + c\max_{C,D} d'_{fix}(C,D) \\
&\cdots \\
&\leq \frac{(1-c)\Delta}{1-c} \\
&\leq \frac{(1-c)\Delta}{1-\gamma}
\end{aligned}$$

and using the total variation as an approximation [GS01] then

$$K(d'_{fix})(P_s^{g_s^a}, P_{\rho(s)}^a) \leq \max_{C,D} d'_{fix}(C,D) \cdot TV(P_s^{g_s^a}, P_{\rho(s)}^a)$$

∎

However, it is better to be able to make some guarantees just from the input MDP in order to avoid enumerating the exponential number of possible MDP aggregations.

5.4 Illustration

We illustrate how the lax metric d_{fix} differs from the non-lax metric e_{fix}. We have made the claim that while the unlaxed metric does capture behavioural bisimilarity, it does not capture performance similarity. We illustrate this on a small MDP consisting of 15 states as illustrated in Figure 5–2. There is a small reward in state 10 but nowhere else. There are the four obvious navigational actions $\{U, D, L, R\}$. For each state and each of these actions we sampled a value p from the uniform distribution over $[0.85, 0.95]$ and assigned p as the probability that the action will succeed and $1 - p$ as the probability that the agent stays in the current state. The most interesting state pair in this metric is 0 and 14. The convergence of the metric for this state pair is illustrated in Figure 5–3.

Also displayed in Figure 5–2 are representations of the two metrics after convergence has been established. The values on one axis range from 0 to 14 and represent the state. Thus the i, j'th entry in the matrix represents the distance from state i to state j. White would indicate a distance of 1 and black a distance of 0.

One should draw their attention towards the bottom left and bottom right of each plot. In the lax metric, the bottom left mirrors the bottom right much better than in the unlaxed metric. This is because in the unlax metric, there is a divergence between the two halves of the vertical hallway. In the lax version of the metric, up and down actions are being matched together and thus the only divergence is due to the noise we added.

Now let us consider the utility of these metrics in aggregating MDPs. Consider the cross MDP displayed in Figure 5–4. There is a reward of 1 in the center and

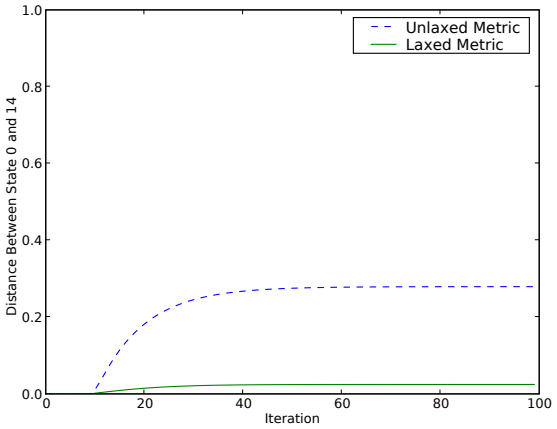

Figure 5–3: Convergence of the Metrics

the probability of succeeding in movement has some noise as described before. For a given ϵ, we used the random partitioning algorithm outlined earlier to create a state aggregation. In Figure 5–5 you can see the size of the aggregated MDPs plotted against ϵ. In the case of the lax metric, we used $\epsilon' = \frac{\epsilon}{2}$ to compensate for the factor of 2 difference in the error bound. It is very revealing that the number of partitions drops very quickly and levels at around 6 or 7 for our algorithm. This is because the MDP is collapsing to a state space close to the natural choice of $\{\{C\}\} \cup \{\{Ni, Si, Wi, Ei\} : i \in \{1, 2, 3, 4, 5, 6\}\}$. Under the unlaxed metric, this is not likely to occur, and thus the first states to be partitioned together are the ones neighbouring each other (which can actually have quite different behaviours).

5.5 Discussion

Although the metric is potentially quite expensive to compute initially there are many domains in which having an accurate aggregation is worth it. This is especially true when one has large computing resources available initially to compute the metric

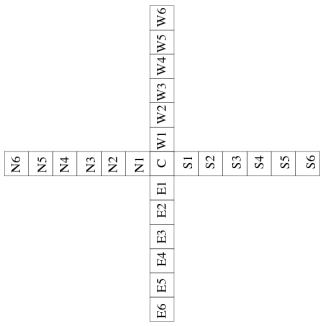

Figure 5–4: Cross MDP

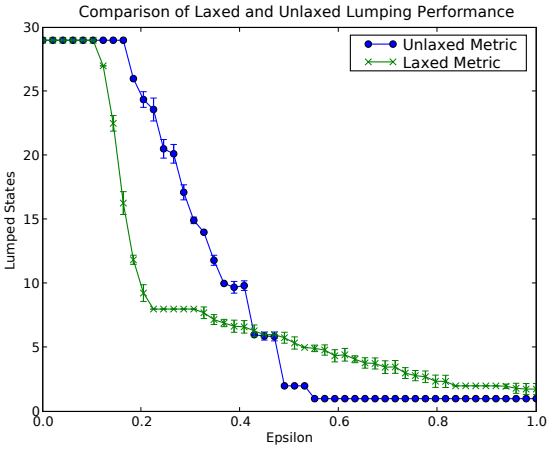

Figure 5–5: Aggregation Performance of Metrics: Average Number of Partitions over 100 Runs vs. $\epsilon/2$

and collapse the MDP into a close approximation. The smaller aggregation might then be small enough to fit on a small mobile device. Also, a policy optimized to yield any value function can be calculated quickly and lifted to yield actions in the original space. This will be of great benefit when the goal is to achieve a certain value and this value changes over time.

The metric can also be used to find subtasks in a larger problem that can be solved using controllers from a pre-supplied library. For example, if a controller is available to navigate single rooms, the metric might be used to lump states in a building schematic into "rooms". The aggregate MDP can then be used to solve the high level navigational task using the controller to navigate specific rooms.

CHAPTER 6
Conclusion

In this thesis we addressed some of the limitations of the standard notion of probabilistic bisimulation and associated metrics. The standard theory allows states to be equated if they behave similarly by matching the same actions. We relax this requirement by considering what happens when matching is allowed amongst related actions.

The theory follows the same general pattern as the usual theory but includes the following crucial differences:

1. The suitability of action matches was specified by a metric over actions; a distance of zero between actions is all that is necessary to satisfy the relaxed version of bisimulation;
2. The logical characterization needs new types of formulas; the logic used to characterize bisimulation does not correctly characterize lax bisimulation.
3. The metric analogue lifts the action metric to the level of states by looking at the worst case action matches; the Hausdorff metric is leveraged to accomplish this.

Lax bisimulation and its corresponding metric provide a sound theoretical basis for analyzing the symmetry in very general probabilistic systems. As a case study we considered measuring the similarity of state-action pairs in a Markov Decision Process and used it in an algorithm for constructing approximate MDP homomorphisms.

Our approach works significantly better in practice than the bisimulation metrics of Ferns et al. The theoretical bound on the error in the value function presented in (Ravindran & Barto, 2004) can be derived using our metric.

6.1 Future Work

The work in this thesis has been of a theoretical nature. Algorithms were developed mostly to hint towards some of the possibilities such a theory might yield. Thus, a main avenue for future work is reducing the computational complexity of these algorithms. Two sources of complexity include the quadratic dependence on the number of actions, and the evaluation of the Kantorovich metric.

The first issue can be addressed by various approximations. For example, one might use prior knowledge to prune action matchings that are believed to be bad performers. Alternatively, greater flexibility could be achieved using a randomized approach in which a heuristic was used to sample pairs of actions, rather than considering all possibilities.

It is also worth investigating the possibility of replacing the Kantorovich metric (which is very convenient from the theoretical point of view) with a more practical approximation. There has already been work [FCPP06] on approximating the unlaxed metric centered around sampling either the Kantorovich metric directly or the Total Variation metric as an approximation. The results are encouraging and these methods should be adapted to the unlaxed case.

Finally, an extension to continuous actions is very important. Under some mild assumptions a large portion of the results go through. Unfortunately, some of the proofs explicitly count actions to arrive at contradictions. It is quite conceivable that

these issues can be worked out and the metric can be used to explore discretizing fully continuous systems.

REFERENCES

[Arv76] W. Arveson. *An Invitation to C^*-Algebra*. Springer-Verlag, 1976.

[Bai96] C. Baier. Polynomial time algorithms for testing probabilistic bisimulation and simulation. In *Proceedings of the 8th International Conference on Computer Aided Verification (CAV'96)*, number 1102 in Lecture Notes in Computer Science, pages 38–49, 1996.

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

[BK97] C. Baier and M. Kwiatkowska. Domain equations for probabilistic processes. *Electronic Notes in TCS*, 7, July 1997. Extended Abstract.

[DDLP05] Vincent Danos, Josée Desharnais, François Laviolette, and Prakash Panangaden. Bisimulation and cocongruence for probabilistic systems. *Information & Computation*, 2005. To appear.

[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] Josée Desharnais, Abbas Edalat, and Prakash Panangaden. Bisimulation for labelled markov processes. *Inf. Comput.*, 179(2):163–193, 2002.

[DG97] Tom G. Dean and Robert Givan. Model minimization for Markov Decision Processes. In *Proceedings of AAAI'97*, pages 106–111, 1997.

[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.

[DGJP02a] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. The metric analogue of weak bisimulation for labelled Markov processes. In *Proceedings of the Seventeenth Annual IEEE Symposium On Logic In Computer Science*, pages 413–422, July 2002.

[DGJP02b] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Weak bisimulation is sound and complete for *pctl*$*$. In L. Brim, P. Jancar, M. Kretinsky, and A. Kucera, editors, *Proceedings of 13th International Conference on Concurrency Theory, CONCUR02,*, number 2421 in Lecture Notes In Computer Science, pages 355–370. Springer-Verlag, 2002.

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

[DGJP04] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for labelled markov processes, 2004.

[DP02] B. A. Davey and H. A. Priestley. *Introduction to Lattices and Order*. Cambridge University Press, 2002.

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

[FCPP06] Norm Ferns, Pablo Samuel Castro, Doina Precup, and Prakash Panangaden. Methods for computing state similarity in markov decision processes. In *UAI*, 2006.

[Fel71] W. Feller. *An Introduction to Probability Theory and its Applications II*. John Wiley and Sons, 2nd edition, 1971.

[Fer07] Norm Ferns. State-similarity metrics for continuous markov decision processes. 2007.

[FPP04] Norm Ferns, Prakash Panangaden, and Doina Precup. Metrics for finite markov decision processes. In *AUAI '04: Proceedings of the 20th conference on Uncertainty in artificial intelligence*, pages 162–169, Arlington, Virginia, United States, 2004. AUAI Press.

[FPP05] Norman Ferns, Prakash Panangaden, and Doina Precup. Metrics for markov decision processes with infinite state spaces. In *Proceedings of the 21th Annual Conference on Uncertainty in Artificial Intelligence (UAI-05)*, page 201, Arlington, Virginia, 2005. AUAI Press.

[GDG03] Robert Givan, Thomas Dean, and Matthew Greig. Equivalence notions and model minimization in Markov decision processes. *Artificial Intelligence*, 147(1-2):163–223, 2003.

[GS01] Alison L. Gibbs and Francis Edward Su. On choosing and bounding probability metrics. *International Statistical Review*, 70:419–435, 2001.

[Kan40] L. V. Kantorovich. A new method for solving some classes of extremal problems. *Comptes Rendus (Doklady) Acad. Sci. USSR*, 28:211–214, 1940.

[KS60] J. G. Kemeny and J. L. Snell. *Finite Markov Chains*. Van Nostrand, 1960.

[LS91] Kim G. Larsen and Arne Skou. Bisimulation through probabilistic testing. *Inf. Comput.*, 94(1):1–28, 1991.

[LWL06] Lihong Li, Thomas J. Walsh, and Michael L. Littman. Towards a unified theory of state abstractions for mdps. In *In Proceedings of the Ninth International Symposium on Artificial Intelligence and Mathematics*, pages 531–539, 2006.

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

[Put94] Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley, 1994.

[RB04] B. Ravindran and A. Barto. Approximate homomorphisms: A framework for non-exact minimization in Markov decision processes. In *Proceedings of the Fifth International Conference on Knowledge Based Computer Systems (KBCS 04)*, 2004.

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

[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.

[WB06] Alicia P. Wolfe and Andrew G. Barto. Decision tree methods for finding reusable mdp homomorphisms. In *AAAI*, 2006.

[Whi78] Ward Whitt. Approximations of dynamic programs, i. *Mathematics of Operations Research*, 3(3):231–243, 1978.