Bisimulation for Markov Decision Processes Through Families of Functional Expressions

Norm Ferns[1] [?], Doina Precup[2], and Sophia Knight[3]

[1] ferns@di.ens.fr
Departement d'Informatique
Ecole Normale Superieure
45 rue d'Ulm, F-75230 Paris Cedex 05, France
[2] dprecup@cs.mcgill.ca
School of Computer Science
McGill University
Montreal, Canada, H3A 2A7
[3] sophia.knight@gmail.com
CNRS, LORIA
Universite de Lorraine
Nancy, France

Abstract. We transfer a notion of quantitative bisimilarity for labelled Markov processes [1] to Markov decision processes with continuous state spaces. This notion takes the form of a pseudometric on the system states, cast in terms of the equivalence of a family of functional expressions evaluated on those states and interpreted as a real-valued modal logic. Our proof amounts to a slight modi cation of previous techniques [2, 3] used to prove equivalence with a xed-point pseudometric on the state-space of a labelled Markov process and making heavy use of the Kantorovich probability metric. Indeed, we again demonstrate equivalence with a xed-point pseudometric de ned on Markov decision processes [4]; what is novel is that we recast this proof in terms of integral probability metrics [5] de ned through the family of functional expressions, shifting emphasis back to properties of such families. The hope is that a judicious choice of family might lead to something more computationally tractable than bisimilarity whilst maintaining its pleasing theoretical guarantees. Moreover, we use a trick from descriptive set theory to extend our results to MDPs with bounded measurable reward functions, dropping a previous continuity constraint on rewards and Markov kernels.

1 Introduction

Probabilistic bisimulation is a notion of state-equivalence for Markov transition systems, rst introduced by Larsen and Skou [6] based upon bisimulation for nondeterministic systems by Park and Milner [7, 8]. Roughly, states are deemed equivalent if they transition with the same probability to classes of equivalent states.

In the context of labelled Markov processes (LMPs), a robust quantitative notion of probabilistic bisimilarity has been devised in the form of a class of behavioural pseudometrics, dubbisimilarity metrics, de ned on the state space of a given process [9, 1, 10, 2, 11]. The de ning characteristic of these metrics is that the kernel of each is probabilistic bisimilarity, and otherwise each assigns a distance between 0 and 1 that measures the degree to which two states are bisimilar.

[?] Norm Ferns' contribution was partially supported by the AbstractCell ANR-Chair of Excellence.

Bisimilarity metrics were initially de ned in terms of a family of functional expressions interpreted as a real-valued logic on the states of an LMP by Desharnais et al. [9], building on ideas of Kozen [12] that logic could be generalized to handle probabilistic phenomena. Subsequently, van Breugel and Worrell [2] used category theory to de ne another class of bisimilarity metrics, and showed that their de nition was equivalent to a slightly modi ed version of the metrics of Desharnais et al. in terms of the family of functional expressions considered. Desharnais et al. [1] in turn reformulated this latter version solely in terms of order-theoretic xed-point theory. A crucial component in these formulations was the recognition that the initial version, when lifted to a metric on distributions, could be expressed as a Kantorovich metric.

For nite systems, the various formulations readily admit a variety of algorithms to estimate the distances [9, 3, 13]. In particular, the initial formulation in terms of a family of functional expressions led to an exponential-time algorithm based on choosing a suitable nite sub-family of functionals. This was vastly improved in [3] wherein an iterative polynomial-time algorithm exploited the fact that the Kantorovich metric is a specialized linear program.

In [14, 15, 4], the xed-point presentation of the LMP bisimilarity metric was adapted to nite Markov decision processes (MDPs) and MDPs with continuous state spaces, bounded continuous rewards, and (weakly) continuous Markov kernels. Insofar as nite systems are concerned, the addition of the reward parameter is minor; in fact, the iterative polynomial-time algorithm applies more or less directly [14]. Unfortunately, even for very simple toy systems the experimental space and time required is too great to be of practical use. This issue was explored in [16] where a Monte Carlo approach to estimating the Kantorovich metric, and the bisimilarity metric for MDPs in general, was devised and shown to outperform other approaches in an experimental setting. The Monte Carlo approach was even extended to MDPs in which the state space is a compact metric space [4]. However, this line of investigation is still very preliminary.

In this work, we seek to complete further the picture of bisimilarity metrics for MDPs by presenting a family of functional expressions that induce the xed-point bisimilarity metric, in analogy with the results of [2] for LMPs. We aim to shift the study of equivalences on MDP states to the study of such families and their properties. The right choice of family might lead to a more easily computable equivalence whilst maintaining some important theoretical guarantees. Additionally, we hope further to investigate Monte Carlo approaches to estimating similarity, for example, by sampling from a given family of functions. More speci cally, we carry out the following.

1. We adapt Proposition 2 of [2] to MDPs, showing that a class of functional expressions interpreted as a real-valued logic on the states of a given MDP is dense in the class of Lipschitz functions with respect to the pseudometric induced by the family. It is important to note that the proof here appears almost unchanged from [2]; what is important is that we recast the result in the terminology and conceptual framework of integral probability metrics and their generating classes of functions.
2. We remove the continuity constraints of Theorem 7, which establishes the xed-point bisimilarity metric for continuous MDPs, using techniques from descriptive set theory. This is, to the best of our knowledge, an original result.
3. We propose a preliminary Monte Carlo technique for estimating the bisimilarity metric by sampling from the family of functional expressions that encodes bisimilarity for MDPs. This too, appears to be an original result, but is based on a heuristic method and experimental evidence presented in [17].

The paper is organized as follows. In Section 2, we provide a brief summary on the relevant development of bisimilarity metrics and related results for LMPs and MDPs. In Section 3, we establish

a family of functional expressions that induces a metric equal to a previously-de ned bisimilarity metric for MDPs, and then generalize the applicability of this result by removing previous continuity constraints. Finally, in Section 4 and Section 5, we propose a Monte Carlo method for approximating the bisimilarity metric by sampling from a family of functional expressions, and conclude with suggestions for future work.

2 Background

The purpose of this section is to recall the development of pseudometrics capturing bisimilarity for labelled Markov processes, and set down what has already been carried over to Markov decision processes. In doing so, we x some basic terminology and notation for probabilistic systems. Unless otherwise stated, all the results of this section can be found in Prakash's book on labelled Markov processes [18].

2.1 Probability Measures on Polish Metric Spaces

Since we deal primarily with uncountably in nite state spaces, we must take into account the tension involved in imposing the right amount of structure on a space for general theoretical utility and imposing the right amount of structure for practical application. Much of the work on labelled Markov processes has been cast in the setting of Polish spaces and analytic spaces, which are general enough to include most practical systems of interest while structured enough to admit many useful theorems.

De nition 1.

1. A Polish metric space *is a complete, separable metric space.*
2. A Polish space *is a topological space that is homeomorphic to a Polish metric space.*
3. An analytic set *is a subset of a Polish space that is itself the continuous image of a Polish space.*
4. A standard Borel space *is a measurable space that is Borel isomorphic to a Polish space.*

If $(X; \ )$ is a topological space, then $C^b(X)$ is the set of continuous bounded real-valued functions on X. If $(X; B_X)$ is a standard Borel space then we let $S(X)$ and $P(X)$ denote the sets of subprobability measures and probability measures on X respectively, and remark that each is also a standard Borel space [19]. We will also assume that the reader is familiar with the theory of integration. If is a nite measure and f is an integrable function both de ned on $(X; B_X)$ then we denote the integral of f with respect to by (f).

Working at the level of standard Borel spaces allows us to use the rich structure of Polish metric spaces without necessarily having to x a metric beforehand. For example, when examining probability measures on such spaces, we can sometimes restrict to compact metric spaces, which in turn provide nite substructure for estimation schemes or over which convergence of certain functions can be made to be uniform. The following can be found in [20] and [21] and will be used to establish Theorem 9 in Section 3.

De nition 2. *Let P be a family of Borel probability measures on a metric space $(X; d)$.*

1. *P is said to be* uniformly tight *i for every > 0 there exists a compact subset K of X such that $P(X \ nK) <$ for every $P \ 2 \ P$:*

2. P is said to be relatively compact if every sequence of elements in P contains a weakly convergent subsequence.

Theorem 1 (Prohorov's Theorem). Suppose (X, d) is a Polish metric space. Let $P \subseteq \mathcal{P}(X)$. Then P is relatively compact if and only if it is uniformly tight.

Theorem 2 (Dini's Theorem). Suppose (K, τ) is a compact topological space, $(f_n)_{n \in \mathbb{N}}$ is a sequence of continuous real-valued functions on K, monotonically decreasing and converging pointwise to a continuous function f. Then $(f_n)_{n \in \mathbb{N}}$ converges to f uniformly on K.

2.2 Stochastic Kernels and Markov Processes

Definition 3. Let $(X, \mathcal{B}_X)$ and $(Y, \mathcal{B}_Y)$ be standard Borel spaces. A sub-Markov kernel[4] is a Borel measurable map from $(X, \mathcal{B}_X)$ to $(\mathcal{S}(Y), \mathcal{B}_{\mathcal{S}(Y)})$. A Markov kernel is a Borel measurable map from $(X, \mathcal{B}_X)$ to $(\mathcal{P}(Y), \mathcal{B}_{\mathcal{P}(Y)})$.

Equivalently, K is a sub-Markov (Markov) kernel from X to Y if

1. $K(x)$ is a sub-probability (probability) measure on $(Y, \mathcal{B}_Y)$ for each $x \in X$, and
2. $x \mapsto K(x)(B)$ is a measurable map for each $B \in \mathcal{B}_Y$.

We will simply write "K is a sub-Markov (Markov) kernel on X" when it is implicitly assumed that $Y = X$. Such stochastic kernels play the role of a transition relation in stochastic transition systems with continuous state spaces. The two Markov processes that we examine in detail are the labelled Markov process and the Markov decision process.

Definition 4. A labelled Markov process (LMP) is a tuple $(S, \mathcal{B}_S, A, \{K_a : a \in A\})$, where $(S, \mathcal{B}_S)$ is a standard Borel space, A is a finite set of labels, and for $a \in A$, K_a is a sub-Markov kernel on S.

Definition 5. A Markov decision process (MDP) is a tuple $(S, \mathcal{B}_S, A, \{P_a : a \in A\}, r)$, where $(S, \mathcal{B}_S)$ is a standard Borel space, A is a finite set of actions, $r : A \times S \to \mathbb{R}$ is a bounded measurable reward function, and for $a \in A$, P_a is a Markov kernel on S.

For each $a \in A$ we will denote by $r_a : S \to \mathbb{R}$ the function defined by $r_a(s) = r(a, s)$.

Remark 1. In [9, 22, 23, 10], Desharnais, Panangaden, et al. consider LMPs in which the state spaces are analytic sets; this is largely because the quotient of a Polish space may fail to be Polish but is always guaranteed to be analytic. We will not consider analytic sets in this work, but the interested reader should keep this in mind.

2.3 Bisimulation

We present bisimilarity for LMPs and MDPs as outlined in [10] and [4] and note that the latter amounts to little more than a mild extension through the addition of rewards to the definition of the former.

[4] This is also known as a stochastic relation, a stochastic transition kernel, or simply a stochastic kernel.

De nition 6. Given a relation R on a set S, a subset X of S is said to be R-closed if and only if the collection of all those elements of S that it is related to by R, $R(X) = \{s' \in S \mid \exists s \in X,\ sRs'\}$, is itself contained in X.

De nition 7. Given a relation R on a measurable space (S, Σ), we write $\Sigma(R)$ for the set of those Σ-measurable sets that are also R-closed, $\{X \in \Sigma \mid R(X) \subseteq X\}$.

When R is an equivalence relation then to say that a set X is R-closed is equivalent to saying that X is a union of R-equivalence classes. In this case $\Sigma(R)$ consists of those measurable sets that can be partitioned into R-equivalence classes.

De nition 8. Let $(S, \Sigma_S, A, \{K_a : a \in A\})$ be an LMP. An equivalence relation R on S is a bisimulation relation if and only if it satis es

$$sRs' \quad \text{for every } a \in A \text{ and for every } X \in \Sigma(R),\ K_a(s)(X) = K_a(s')(X).$$

Bisimilarity is the largest of the bisimulation relations.

De nition 9. Let $(S, \Sigma_S, A, \{P_a : a \in A\}, r)$ be an MDP. An equivalence relation R on S is a bisimulation relation if and only if it satis es

$$sRs' \quad \text{for every } a \in A,\ r_a(s) = r_a(s') \text{ and for every } X \in \Sigma(R),\ P_a(s)(X) = P_a(s')(X).$$

Bisimilarity is the largest of the bisimulation relations.

It turns out that bisimulation for LMPS and MDPs can be equivalently cast as the maximal xed-point of a monotone functional on a complete lattice. We present this here only in the context of MDPs; the statement for LMPs is analogous.

Theorem 3. Let $(S, \Sigma_S, A, \{P_a : a \in A\}, r)$ be an MDP, τ a Polish topology on S generating Σ_S and such that for each a in A, r_a and P_a are continuous with respect to τ, $\mathcal{P}(S)$ being endowed with the topology of weak convergence induced by τ. Assume that the image of r is contained in $[0, 1]$. De ne $F : Equ \to Equ$ by

$$sF(R)s' \iff \forall a \in A,\ r_a(s) = r_a(s') \text{ and } \forall X \in \Sigma(R),\ P_a(s)(X) = P_a(s')(X),$$

where Equ is the set of equivalence relations on S equipped with subset ordering. Then the greatest xed point of F is bisimilarity.

Lastly, we remark that bisimulation for LMPs has a logical characterization, and in turn, a characterization in terms of a real-valued modal logic. We omit the details for lack of space, but return to the latter idea in subsequent sections.

2.4 Probability Metrics

Metrizing bisimilarity for Markov processes essentially involves assigning a distance to their Markov kernels via a suitable probability metric. Gibbs and Su [24] survey a variety of such metrics. LMP bisimilarity was initially de ned in terms of an integral probability metric in [10], and later recast in terms of the Kantorovich metric in [2]. In order to present the Kantorovich metric, we rst recall the de nition of lower semicontinuity.

De nition 10. Let (X, τ) be a topological space and let $f : X \to \mathbb{R} \cup \{-\infty, +\infty\}$. Then f is lower semicontinuous if for each half-open interval of the form $(r, +\infty)$, the preimage $f^{-1}(r, +\infty) \in \tau$.

The Kantorovich Metric

De nition 11. Let S be a Polish space, h a bounded pseudometric on S that is lower semicontinuous on $S \times S$ with respect to the product topology, and $\mathrm{Lip}(h)$ be the set of all bounded functions $f : S \to \mathbb{R}$ that are measurable with respect to the Borel σ-algebra on S and that satisfy the Lipschitz condition $f(x) - f(y) \le h(x,y)$ for every $x, y \in S$. Let $P, Q \in \mathcal{P}(S)$. Then the Kantorovich metric $K(h)$ is de ned by

$$K(h)(P,Q) = \sup_{f \in \mathrm{Lip}(h)} (P(f) - Q(f)).$$

Lemma 1. Let S, h, P, and Q be as in De nition 11. Let $\Gamma(P,Q)$ consist of all measures on the product space $S \times S$ with marginals P and Q, i.e.,

$$\Gamma(P,Q) = \{\gamma \in \mathcal{P}(S \times S) : \gamma(E \times S) = P(E) \text{ and } \gamma(S \times E) = Q(E) \text{ for all } E \in \mathcal{B}_S\}. \quad (1)$$

Then the Kantorovich metric $K(h)$ satis es the inequality:

$$\sup_{f \in \mathrm{Lip}(h; C^b(S))} (P(f) - Q(f)) \le K(h)(P,Q) \le \inf_{\gamma \in \Gamma(P,Q)} \gamma(h) \quad (2)$$

where $\mathrm{Lip}(h; C^b(S))$ denotes functions on S that are continuous and bounded, 1-Lipschitz with respect to h, and have range $[0, \|h\|]$.

Note that h need not generate the topology on S, and so Lipschitz continuity with respect to h does not immediately imply continuity on S.

The leftmost and rightmost terms in inequality 2 are examples of in nite linear programs in duality. It is a highly nontrivial result that there is no duality gap in this case (see for example Theorem 1.3 and the proof of Theorem 1.14 in [25]).

Theorem 4 (Kantorovich-Rubinstein Duality Theorem). Assume the conditions of De nition 11 and Lemma 1. Then there is no duality gap in equation 2, that is,

$$K(h)(P,Q) = \sup_{f \in \mathrm{Lip}(h; C^b(S))} (P(f) - Q(f)) = \inf_{\gamma \in \Gamma(P,Q)} \gamma(h) \quad (3)$$

Note that for any point masses δ_x, δ_y, we have $K(h)(\delta_x, \delta_y) = h(x,y)$ since $\delta_{(x,y)}$ is the only measure with marginals δ_x and δ_y on the right-hand side of Equation 3. As a result, we obtain that any bounded lower semicontinuous pseudometric h can be expressed as $h(x,y) = \sup_{f \in F} (f(x) - f(y))$ for some family of continuous functions F.

Integral Probability Metrics The intuition behind the Kantorovich metric is that the quantitive di erence between two probability measures can be measured in terms of the maximal di erence between the expected values with respect to the two measures, of a class of test functions - in this case, the class of Lipschitz functions. For an arbitrary class of test functions, the induced metric is known as the integral probability metric generated by that class. All de nitions and results of this subsection are taken from [5].

De nition 12. Let F be a subset of bounded measurable real-valued functions on a Polish metric space (S,d). Then the integral probabilty metric associated with F is the probability metric $IPM(F)$ on $\mathcal{P}(S)$ de ned by

$$IPM(F)(P,Q) = \sup_{f \in F} |P(f) - Q(f)|$$

for probability measures P and Q.

For convenience, we will simply denote $IPM(F)$ by F. Remark that in general F is allowed to take on in nite values, though we will work with bounded sets of functions to avoid this. Additionally, we remark that this probability metric in turn induces a pseudometric on S via

$$F(x;y) := F(\ _x;\ _y):$$

Thus, as an abuse of notation we will use F to refer to a family of functions, the associated probability metric, and the induced pseudometric, with the intended meaning clear from the context.

De nition 13. Let F be a subset of bounded measurable real-valued functions on a Polish metric space $(S;d)$. The maximal generator of the integral probability metric associated to F is the set R_F of all bounded measurable real-valued functions on $(S;d)$, each of which satisi es the following: $g \ 2 \ R_F$ if and only if

$$jP(g) \quad Q(g)j \quad F(P;Q)$$

for every P and Q in $P(S)$.

It follows that R_F is the largest such family, and that $R_F(P;Q) = F(P;Q)$.

The following is Theorem 3.3 of [5].

Theorem 5. Let F be an arbitrary generator of R_F. Then

1. R_F contains the convex hull of F;
2. $f \ 2 \ R_F$ implies $f + \ 2 \ R_F$ for all $\ 2 [0;1]$ and $\ 2 \ R$;
3. If the sequence $(f_n)_{n \ 2 \ N}$ in R_F converges uniformly to f, then $f \ 2 \ R_F$.

In particular, for a given F, R_F is closed under uniform convergence.

2.5 Bisimulation Metrics

The metric analogue of bisimulation for LMPs was initially cast in terms of a family of functional expressions, interpreted as a real-valued logic over the states of a given Markov process [9]. A slightly modi ed version was then shown to be equivalent to a bisimultaion metric developed in the context of category theory in [2]. In [1], the authors in turn recast this latter metric fully using order-theoretic xed-point theory for discrete systems. Finally, this method was generalized to develop a bisimulation metric for MDPs with continuous state spaces in [4].

We present here the logic of [2] and the xed-point results of [4], as these are the results we will attempt to join in the subsequent sections.

De nition 14. For each $c \ 2 (0;1]$, F_c represents the family of functional expressions generated by the following grammar.

$$f ::= 1 \ j \ 1 \quad f \ j \ hai f \ j \ \max(f;f) \ j \ f \quad q \tag{4}$$

where $q \ 2 \ Q \setminus [0;1]$ and a belongs to a xed set of labels A.

Let $M = (S;B_S;A;fK_a : a \ 2 \ Ag)$ be an LMP. The interpretation of $f \ 2 \ F_c$, $f_M : S \ ! \ [0;1]$, is de ned inductively. Let $s \ 2 \ S$. Then

$$\begin{aligned}
1_M(s) &= 1 \\
(1 \quad f)_M(s) &= 1 \quad f_M(s) \\
(hai f)_M(s) &= c \ K_a(s)(f_M) \\
\max(f_1;f_2)_M(s) &= \max((f_1)_M(s);(f_2)_M(s)) \\
(f \quad q)_M(s) &= \max(f_M(s) \quad q;0);
\end{aligned}$$

Henceforth, we shall omit the subscript M and use f to refer both to an expression and its interpretation, with the difference clear from the context.

Remark 2. We may also add the expressions $\min(f, 1)$ and $f \dot{-} q$ as shorthand for the expressions $1 - \max(1 - f, 1 - f)$ and $1 - ((1 - f) - q)$. The operations $\wedge$ and $\dot{-}$ denoted truncated subtraction in the unit interval and truncated addition in the unit interval, respectively.

The relevance of such a formulation arises via a behavioural pseudometric.

The following is Theorem 3 of [2] and Theorem 8.2 of [18].

Theorem 6. Let $M = (S, B_S, A, \{K_a : a \in Ag\})$ be an LMP and for $c \in (0, 1]$, let F_c be the family of functional expressions defined in Definition 14. Define the map d_c on $S \times S$ as follows:

$$d_c(x, y) = \sup_{F_c} |f(x) - f(y)|. \tag{5}$$

Then d_c is a pseudometric on S whose kernel is bisimilarity.

As previously mentioned, the metric d_c can be formulated in terms of fixed-point theory, and indeed this construction has been carried over to MDPs, with the minor addition of taking into account reward differences. The following is Theorem 3.12 of [4].

Theorem 7. Let $M = (S, B_S, A, \{P_a : a \in Ag\}, r)$ be an MDP and let τ be a Polish topology on S that generates B_S. Assume that the image of r is contained in $[0, 1]$, and that for each a in A, r_a and P_a are continuous, $P(S)$ endowed with the weak topology induced by τ. Let $c \in (0, 1)$ be a discount factor, and lsc_m be be the set of bounded pseudometrics on S that are lower semicontinuous on $S \times S$ endowed with the product topology induced by τ. Define $F_c : \mathrm{lsc}_m \to \mathrm{lsc}_m$ by

$$F_c(h)(s, s') = \max_{a \in A} ((1 - c)|r_a(s) - r_a(s')| + c\, K(h)(P_a(s), P_a(s')))$$

where $K(h)$ is the Kantorovich metric induced by $h \in \mathrm{lsc}_m$. Then

1. F_c has a unique fixed point $\pi_c : S \times S \to [0, 1]$,
2. The kernel of π_c is bisimilarity,
3. for any $h_0 \in \mathrm{lsc}_m$, $\lim_{n \to \infty} F_c^n(h_0) = \pi_c$,
4. π_c is continuous on $S \times S$,
5. π_c is continuous in r and P, and
6. If MDP $M' = (S, B_S, A, \{P_a : a \in Ag\}, k\, r)$ for some $k \in [0, 1]$ then $\pi_{c;M'} = k\, \pi_{c;M}$.

Whereas the interest in finding small bisimilar systems for LMPs lies in being able to test properties of a system specified in a given logic, the interest in finding small bisimilar systems for MDPs concerns finding optimal planning strategies in terms of value functions. Given a discount factor $\gamma \in [0, 1)$, the optimal value function is the unique solution to the following Bellman optimality fixed-point equation.

$$v(s) = \max_{a \in A} (r_a(s) + \gamma P_a(s)(v)) \text{ for each } s \in S.$$

In general, such a v need not exist. Even if it does, there may not be a well-behaved, that is to say measurable, policy that is captured by it. Fortunately, there are several mild restrictions under which this is not the case. According to Theorem 8.3.6 and its preceding remarks in [26], if the state space is Polish and the reward function is uniformly bounded then there exists a unique solution v to the Bellman optimality equation and there exists a measurable optimal policy for it as well.

The following is Theorem 3.20 in [4].

Theorem 8. Let $M = (S, B_S, A, \{P_a : a \in A\}, r)$ be an MDP and let τ be a Polish topology on S that generates B_S. Assume that the image of r is contained in $[0,1]$, and that for each a in A, r_a and P_a are continuous, $P(S)$ endowed with the weak topology induced by τ. Let $c \in (0,1)$ be a discount factor. Let v^* be the optimal value function for the expected total discounted reward associated with M and discount factor $\gamma \in [0,1)$. Suppose $\gamma \leq c$. Then v^* is Lipschitz continuous with respect to ℓ_c with Lipschitz constant $\frac{1}{1-c}$, i.e., $|v^*(x) - v^*(y)| \leq (1-c)^{-1} \ell_c(x,y)$.

3 Bisimulation Metrics for MDPs Revisited

The goal of this section is two-fold. First, we establish a family of functional expressions as in Definition 14 that captures bisimulation for MDPs as defined in Theorem 7. This amounts to little more than Proposition 2 of [2] but using the terminology of generating classes for integral probability metrics. Second, we generalize the applicability of these results for MDPs by removing the continuity constraints in Theorem 7.

3.1 When is the Integral Probability Metric the Kantorovich Metric?

In this section we will show that under some very mild conditions, the maximal generator of a family of functional expressions is in fact the class of Lipschitz functions with respect to the distance induced by that family. In this case, the integral probability metric and the Kantorovich metric induced by the family coincide.

The following result is Lemma 4.6 of [1], itself adapted from Proposition 2 of [2], presented almost verbatim. The imposed Lipschitz condition makes measurability concerns almost an afterthought. What really matters here is the reframing of the result in terms of the integral probability metric and its maximal generator. Doing so will allow us to examine simpler grammars for bisimulation, as well as ways of approximating these.

Theorem 9. Suppose (S,d) is a Polish metric space and $\mathcal{F}$ is a family of real-valued functions on S that take values in the unit interval and are 1-Lipschitz continuous with respect to d. Suppose further that $\mathcal{F}$ contains the constant zero function and is closed under truncated addition with rationals in the unit interval, subtraction from the constant function 1, and taking the pointwise maximum of two functions. Let $R_{\mathcal{F}}$ be the maximal generator of $\mathcal{F}$ and $\mathrm{Lip}(\mathcal{F})$ be the set of real-valued measurable functions on S that are 1-Lipschitz with respect to the metric induced by $\mathcal{F}$. Then $R_{\mathcal{F}} = \mathrm{Lip}(\mathcal{F}) \cap C^b(S)$.

Proof. Firstly note that since by assumption all members of the family $\mathcal{F}$ are 1-Lipschitz continuous with respect to d, the induced pseudometric $\mathcal{F} \leq d$. Thus, $\mathrm{Lip}(\mathcal{F}) \supseteq \mathrm{Lip}(d) \supseteq C^b(S)$. From the definition of $R_{\mathcal{F}}$ applied to Dirac measures, it immediately follows that each of its members is 1-Lipschitz with respect to the pseudometric induced by $\mathcal{F}$. Thus, $R_{\mathcal{F}} \subseteq \mathrm{Lip}(\mathcal{F})$. In particular, every member of the maximal generator belongs to $C^b(S)$.

The reverse inclusion $\mathrm{Lip}(\mathcal{F}) \subseteq R_{\mathcal{F}}$ is somewhat more complicated to establish. By Theorem 5, we have that $R_{\mathcal{F}}$ is closed with respect to uniform convergence, and thus is also generated by $\overline{\mathcal{F}}$, the closure of $\mathcal{F}$ with respect to uniform convergence. In fact, we will show that $\mathcal{F}$ is dense in $\mathrm{Lip}(\mathcal{F})$ in the metric of uniform convergence; for then it follows that $\mathrm{Lip}(\mathcal{F}) = \overline{\mathcal{F}} \subseteq R_{\mathcal{F}}$. We do so in two steps. First we establish the result in the case where (S,d) is a compact metric space, as this allows us to replace pointwise convergence by uniform convergence at a certain point in the

proof. Finally, we extend this result to the general case of a Polish metric space by approximating it from within by a suitable compact subset.

Assume that (S,d) is a compact metric space. It is easily seen that $\overline{F}$ contains the constant zero function and remains closed under truncated addition with all constants in the unit interval, subtraction from 1, and taking maxima; in fact, it now follows that $\overline{F}$ is closed under countable suprema. To see this, suppose $(f_n)_{n\in\mathbb{N}}$ is a sequence in $\overline{F}$. Since $\overline{F}$ is uniformly bounded by 1 it follows that $f = \sup_{n\in\mathbb{N}} f_n$ exists and moreover it is continuous, as each f_n is 1-Lipschitz continuous with respect to d. Define $(h_n)_{n\in\mathbb{N}}$ in $\overline{F}$ by $h_n = \max_{1\le i\le n} f_i$. Then $(h_n)_{n\in\mathbb{N}}$ is monotonically increasing and converges pointwise to f. By Theorem 2, $(h_n)_{n\in\mathbb{N}}$ converges uniformly to f, and hence f belongs to $\overline{F}$. It now also follows that $\overline{F}$ is closed under truncated subtraction with constants in the unit interval, taking minima, and taking in ma.

Let $g \in \mathrm{Lip}(F)$. Without loss of generality, we assume the image of g belongs to $[0,1]$; for the Lipschitz property with respect to F implies that $\sup g - \inf g \le 1$ and we may replace g by $g^0 := g - \inf g$. It follows that if g^0 belongs to R_F then so does $g = g^0 + \inf g$.

Let $\varepsilon > 0$. Then there exists $f_{xy} \in F$ such that

$$g(x) - g(y) \le F(x,y) \le f_{xy}(x) - f_{xy}(y) + \varepsilon \tag{6}$$

Define $w_{xy} \in \overline{F}$ as follows:

$$w_{xy}(z) = \begin{cases} f_{xy}(z) & \text{if } f_{xy}(x) = g(x) \\ f_{xy}(z) \wedge_{[0,1]} (f_{xy}(x) - g(x)) & \text{if } f_{xy}(x) > g(x) \\ f_{xy}(z) \vee_{[0,1]} (g(x) - f_{xy}(x)) & \text{if } f_{xy}(x) < g(x) \end{cases} \tag{7}$$

Then $w_{xy}(x) = g(x)$ and $w_{xy}(y) \le g(y) + \varepsilon$.

Let $(u_n)_{n\in\mathbb{N}}$ be a dense sequence in (S,d). Define $(\varphi_{nm})_{n,m\in\mathbb{N}} \subset \overline{F}$ by $\varphi_{nm} = w_{u_n u_m}$ and define $(\varphi_n)_{n\in\mathbb{N}}$ by $\varphi_n = \inf_{m\in\mathbb{N}} \varphi_{nm}$. It follows that $(\varphi_n)_{n\in\mathbb{N}} \subset \overline{F}$. Moreover,

$$\varphi_n(u_n) = g(u_n) - g(u_n) + \varepsilon \quad \text{and for each } m \ne n, \quad \varphi_n(u_m) \le g(u_m) + \varepsilon.$$

Define $\varphi = \sup_{n\in\mathbb{N}} \varphi_n \in \overline{F}$. Then for any $n \in \mathbb{N}$,

$$g(u_n) \le \varphi(u_n) \le g(u_n) + \varepsilon. \tag{8}$$

Let $x \in S$. Then as the inequalities in line 8 hold for any subsequence of $(u_n)_{n\in\mathbb{N}}$ converging to x, and as both g and φ are continuous, it follows by taking limits that for any $x \in S$,

$$g(x) \le \varphi(x) \le g(x) + \varepsilon, \text{ or equivalently } \|\varphi - g\|_\infty < \varepsilon.$$

Define the sequence $(g_n)_{n\in\mathbb{N}}$ in $\overline{F}$ by $g_n = \varphi_{\frac{1}{n}}$. Then $(g_n)_{n\in\mathbb{N}}$ converges uniformly to g. Therefore, g belongs to $\overline{F} = R_F$, i.e. $\mathrm{Lip}(F) \subset R_F$.

Now suppose (S,d) is a general Polish metric space. Let $P, Q \in \mathcal{P}(S)$. Then $P = \{P, Q\}$ is finite, hence relatively compact. By Theorem 1, P is uniformly tight. Let $0 < \varepsilon < \frac{1}{2}$. Then there exists a compact subset K of S such that $P(S\setminus K) < \varepsilon$ and $Q(S\setminus K) < \varepsilon$.

Let G denote the functions of F restricted to K; for $f \in F$, we will write $f_K \in G$. Then as G still contains the constant zero function, and is closed under the same operations as F, and as (K,d) is a compact metric space, we have $R_G = \mathrm{Lip}(G)$. Let $g \in \mathrm{Lip}(F)$; as before, we assume

without loss of generality that the image of g is contained in [0;1]. Moreover, let g_K be g restricted to K and remark that g_K 2 Lip(G). Next we de ne $P_K ; Q_K$ 2 P(K) by

$$P_K(E) = \frac{P(E \setminus K)}{P(K)} \text{ and } Q_K(E) = \frac{Q(E \setminus K)}{Q(K)}:$$

Remark that $P(K) > 1 \quad > \frac{1}{2}$, and similarly for Q(K), so that each is well-de ned. Then as g_K 2 R_G,

$$jP_K(g_K) \quad Q_K(g_K)j \quad G(P_K ; Q_K):$$

Next for any 1-bounded measurable functionu on S and it's restriction u_K to K, we have

$$jP(u) \quad P_K(u_K)j = jP(u_{\quad K}) + P(u_{\quad SnK}) \quad P_K(u_K)j = j(P(u_{\quad K}) \quad P_K(u_K)) + P(u_{\quad SnK})j$$
$$j1 \quad \frac{1}{P(K)}j \; P(u_{\quad K}) + 1 \quad P(SnK) \quad \frac{1}{1} \quad 1 + \quad \frac{(2 \quad)}{1} \quad 3 ;$$

where $_K$ is the indicator function on K. Similarly $jQ(u) \quad Q_K(u_K)j \quad 3$: Finally,

$$\begin{aligned} jP(g) \quad Q(g)j \quad & jP(g) \quad P_K(g_K)j + jP_K(g_K) \quad Q_K(g_K)j + jQ_K(g_K) \quad Q(g)j \\ & 3 + G(P_K ; Q_K) + 3 \quad 6 + \sup_{f \; 2 F} jP_K(f_K) \quad Q_K(f_K)j \\ & 6 + \sup_{f \; 2 F} (jP_K(f_K) \quad P(f)j + jP(f) \quad Q(f)j + jQ(f) \quad Q_K(f_K)j) \\ & 12 + \sup_{f \; 2 F} (jP(f) \quad Q(f)j) \quad 12 + F(P;Q): \end{aligned}$$

As is arbitrary, it follows that $jP(g) \quad Q(g)j \quad F(P;Q)$ and g 2 R_F.

□

Corollary 1. *Suppose (S;d) is a Polish metric space and F is a family of real-valued functions on S that take values in the unit interval and are 1-Lipschitz continuous with respect to d. Suppose further that F contains the constant zero function and is closed under truncated addition with rationals in the unit interval, subtraction from the constant function 1, and taking the pointwise maximum of two functions. Then the integral probability metric associated to F is the Kantorovich metric of the pseudometric induced by F, i.e. F(P;Q) = K(F)(P;Q) for any P;Q 2 P(S).*

3.2 A Family of Functional Expressions for MDP Bisimulation

We now de ne a family of functional expressions as in De nition 14 that when evaluated on a given MDP, capture bisimilarity.

De nition 15. *For each c 2 (0;1], F_c represents the family of functional expressions generated by the following grammar.*

$$f ::= 0 \; j \; 1 \quad f \; j \; haif \; j \; max(f;f) \; j \; f \quad q \tag{9}$$

where q 2 Q \ [0;1] and a belongs to a xed set of labels A.

Let $M = (S, B_S, A, \{P_a : a \in A\}, r)$ be an MDP. The interpretation of $f \in F_c$, $f_M : S \to [0,1]$, is defined inductively. Let $s \in S$. Then

$$\begin{aligned}
0_M(s) &= 0 \\
(1 - f)_M(s) &= 1 - f_M(s) \\
(\langle a \rangle f)_M(s) &= r_a(s) + c \int P_a(s)(f_M) \\
\max(f_1, f_2)_M(s) &= \max((f_1)_M(s), (f_2)_M(s)) \\
(f \dot{-} q)_M(s) &= \min(f_M(s) + q, 1).
\end{aligned}$$

As before, we shall omit the subscript M when it is clear from the context, and remark that the family also contains the expressions $\min(f, f)$ and $f \dot{-} q$.

We now show that the integral probability metric generated by F_c agrees with the Kantorovich metric induced by the fixed-point bisimulation metric for MDPs. This is essentially the proof method in all of Section 4 of [1].

Theorem 10. *Suppose $M = (S, B_S, A, \{P_a : a \in A\}, r)$ is an MDP and let τ be a Polish topology on S that generates B_S. Assume that the image of r is contained in $[0,1]$, and that for each a in A, r_a and P_a are continuous, $\mathcal{P}(S)$ endowed with the weak topology induced by τ. Let $c \in (0,1)$ be a discount factor, and F_c be the family of functional expressions defined in Definition 15. Let G be a family of functional expressions such that $F_c \subseteq G \subseteq \mathrm{Lip}(F_c)$. Then the pseudometric induced by G coincides with the fixed-point metric $\mathfrak{s}_c$ given by Theorem 7.*

Proof. Let (S, d) be a Polish metric space such that d generates τ. Since $\mathfrak{s}_c$ is continuous, we can assume without loss of generality that $\mathfrak{s}_c \leq d$, as we can simply replace d by the equivalent metric $d + \mathfrak{s}_c$. By structural induction, $F_c \subseteq \mathfrak{s}_c$-$d$, and the range of each member of F_c is $[0,1]$. Therefore by Corollary 1, the integral probability metric and the kantorovich metric induced by F_c agree.

Notice that since F_c is closed under subtraction from the constant function 1, we have that for any $P, Q \in \mathcal{P}(S)$

$$\begin{aligned}
F_c(P,Q) &= \sup_{f \in F_c} |P(f) - Q(f)| = \sup_{f \in F_c} \max(P(f) - Q(f), Q(f) - P(f)) \\
&= \max(\sup_{f \in F_c} P(f) - Q(f), \sup_{f \in F_c} Q(f) - P(f)) \\
&= \max(\sup_{f \in F_c} P(f) - Q(f), \sup_{f \in F_c} P(1 - f) - Q(1 - f)) \\
&= \max(\sup_{f \in F_c} P(f) - Q(f), \sup_{f \in F_c} P(f) - Q(f)) \\
&= \sup_{f \in F_c} P(f) - Q(f)
\end{aligned}$$

which is not necessarily the case otherwise. A simple structural induction next shows that

$$F_c(x,y) = \sup_{a \in A, f \in F_c} |\langle a \rangle f(x) - \langle a \rangle f(y)|.$$

Therefore,

$$
\begin{aligned}
F_c(x,y) &= \sup_{a\in A, f\in F_c} \max(|hf(x) - hf(y)|, |hf(y) - hf(x)|) \\
&= \sup_{a\in A, f\in F_c} \max\ (1-c)(r_a(x) - r_a(y)) + c(P_a(x)(f) - P_a(y)(f)), \\
&\qquad (1-c)(r_a(y) - r_a(x)) + c(P_a(y)(f) - P_a(x)(f)) \\
&= \max_{a\in A} \max\ (1-c)(r_a(x) - r_a(y)) + c \sup_{f\in F_c}(P_a(x)(f) - P_a(y)(f)), \\
&\qquad (1-c)(r_a(y) - r_a(x)) + c \sup_{f\in F_c}(P_a(y)(f) - P_a(x)(f)) \\
&= \max_{a\in A} \max\ (1-c)(r_a(x) - r_a(y)) + c\ F_c(P_a(x), P_a(y)), \\
&\qquad (1-c)(r_a(y) - r_a(x)) + c\ F_c(P_a(x), P_a(y)) \\
&= \max_{a\in A} (1-c) \max\ (r_a(x) - r_a(y)), (r_a(y) - r_a(x)) + c\ F_c(P_a(x), P_a(y)) \\
&= \max_{a\in A} (1-c)|r_a(x) - r_a(y)| + c\ F_c(P_a(x), P_a(y)) \\
&= \max_{a\in A} (1-c)|r_a(x) - r_a(y)| + c\ K(F_c)(P_a(x), P_a(y)) \\
&= F_c(F_c)(x,y).
\end{aligned}
$$

The penultimate line follows from Corollary 1. Therefore, F_c is a fixed-point of the functional F_c defined in Theorem 7. As the fixed-point is unique, it follows that $\sim_c = F_c$. Finally, it follows from Theorem 9 and the definition of maximal generator that $G = F_c = \sim_c$. □

Remark 3. Theorem 10 provides another proof of Theorem 8. Consider the family G with the expression for the Bellman operator B for the MDP $M = (S, B_S, A, \{P_a : a \in A\}, (1-c)r)$ and discount factor c in $[0,1)$. Since $\mathrm{Lip}(F_c)$ is closed under B and the optimal value function scales with rewards, the result follows immediately. Otherwise, we obtain the result only for V_c since $B_c(f) = \max_{a\in A} hf$.

The usefulness of this theorem derives from our choice of G. On the one hand, we might attempt to see what is the minimal family, if one exists, that captures bisimilarity. On the other hand, we might consider explicitly adding operators, like the Bellman operator, that could help an estimation scheme converge faster. We will explore this further in Section 4.

Practical application is still hindered by the continuity constraints on the rewards and Markov kernels, as many interesting problems model discontinuous phenomena. In the next section, we will work to remove these constraints.

3.3 The General Case: Continuity from Measurability

We conclude this section with a neat little result from descriptive set theory that was first pointed out to the authors by Ernst-Erich Doberkat at the 2012 Bellairs Workshop on Probabilistic Systems organized by Prakash. In the most interesting reinforcement learning applications, continuity of the reward process cannot be guaranteed. Amazingly, we may remove the explicit assumption of continuity in [4] and the result still holds! We seek to establish the following.

Theorem 11. Let $(X; \tau)$ be a Polish space and $(P(X); \tau_{P(X)})$ be the space of probability measures on X equipped with the topology of weak convergence with respect to τ. Let $K : (X; \mathcal{B}_X) \to (P(X); \mathcal{B}_{P(X)})$ be a stochastic kernel. Then there exists a finer Polish topology τ' on X such that $\mathcal{B}(\tau') = \mathcal{B}_X$, $\mathcal{B}(\tau'_{P(X)}) = \mathcal{B}_{P(X)}$, and $K : (X; \tau') \to (P(X); \tau'_{P(X)})$ is continuous.

This result is a minor reworking of the following well-known measurability-to-continuity theorem, which is Corollary 3.2.6 in [27].

Theorem 12. Suppose $(X; \tau)$ is a Polish space, Y a separable metric space, and $f : X \to Y$ a Borel map. Then there is a finer Polish topology τ' on X generating the same Borel σ-algebra such that $f : (X; \tau') \to Y$ is continuous.

We will also make use of this characterization of Borel σ-algebra on the set of probability measures, which is Proposition 7.25 in [28].

Proposition 1. Let X be a separable metrizable space and $\mathcal{E}$ a collection of subsets of X which generates $\mathcal{B}_X$ and is closed under finite intersections. Then $\mathcal{B}_{P(X)}$ is the smallest σ-algebra with respect to which all functions of the form $\pi_E(p) = p(E)$, for $E \in \mathcal{E}$, are measurable from $P(X)$ to $[0; 1]$, i.e.,

$$\mathcal{B}_{P(X)} = \sigma[\{\pi_E^{-1}(B_R) : E \in \mathcal{E}\}].$$

For ease of exposition, we will divide the result into the following series of steps.

Lemma 2. Let $(X; \tau)$ and K be as in Theorem 11. Then there exists an increasing sequence $(\tau_n)_{n \in \mathbb{N}}$ of Polish topologies on X finer that τ such that $\mathcal{B}(\tau_n) = \mathcal{B}_X$ and $K : (X; \tau_{n+1}) \to (P(X); (\tau_n)_{P(X)})$ is continuous for all $n \in \mathbb{N}$.

Proof. By Proposition 1 for any Polish topology τ' generating $\mathcal{B}_X$, $\tau'_{P(X)}$ generates $\mathcal{B}_{P(X)}$. It is well known [29] that $\tau'_{P(X)}$ is also a Polish topology. Therefore, $K : (X; \tau') \to (P(X); \tau'_{P(X)})$ is a Borel map. By Theorem 12, there exists a finer Polish topology τ_0 such that $\mathcal{B}(\tau_0) = \mathcal{B}_X$ and $K : (X; \tau_0) \to (P(X); \tau_{P(X)})$ is continuous; but then $K : (X; \tau_0) \to (P(X); (\tau_0)_{P(X)})$ is Borel. Repeating this argument, there exists a finer topology τ_1 on X such that $\mathcal{B}(\tau_1) = \mathcal{B}_X$ and $K : (X; \tau_1) \to (P(X); (\tau_0)_{P(X)})$ is continuous. The result now easily follows for all $n \in \mathbb{N}$ by induction. □

Lemma 3. Let $(X; \tau)$, K, and $(\tau_n)_{n \in \mathbb{N}}$ be as in Lemma 2. Then the least upper bound topology $\tau_1 = \bigvee_{n \in \mathbb{N}} \tau_n$ exists and is Polish, $\mathcal{B}(\tau_1) = \mathcal{B}_X$, and $K : (X; \tau_1) \to (P(X); (\tau_n)_{P(X)})$ is continuous for all $n \in \mathbb{N}$.

Remark 4 ([27] Observation 2, pg. 93). Let $(\tau_n)_{n \in \mathbb{N}}$ be a sequence of Polish topologies on X such that for any two distinct elements $x; y$ of X, there exist disjoint sets $U; V \in \bigcup_{n \in \mathbb{N}} \tau_n$ such that $x \in U$ and $y \in V$. Then the topology τ_1 generated by $\bigcup_{n \in \mathbb{N}} \tau_n$ is Polish.

Proof. By Remark 4, τ_1 exists, is Polish, and is generated by $\bigcup_{n \in \mathbb{N}} \tau_n$. So $\bigcup_{n \in \mathbb{N}} \tau_n$ is a subbasis for τ_1. Let $O \in \tau_1$. Then O is an arbitrary union of finite intersections of elements of $\bigcup_{n \in \mathbb{N}} \tau_n$. So $O = \bigcup_{j \in J} (O_{j;1} \cap O_{j;2} \cap \cdots \cap O_{j;n_j})$ for some index set J. Let $i(j; k) = \min\{n \in \mathbb{N} : O_{j;k} \in \tau_n\}$ and $i(j) = \max\{i(j; k) : 1 \le k \le n_j\}$. Then $O_j = \bigcap_{1 \le k \le n_j} O_{j;k} \in \tau_{i(j)}$ because $(\tau_n)_{n \in \mathbb{N}}$ is increasing. So $O = \bigcup_{j \in J} O_j = \bigcup_{n \in \mathbb{N}} (\bigcup_{\{j : i(j) = n\}} O_j) = \bigcup_{n \in \mathbb{N}} O'_n$ where $O'_n = \bigcup_{\{j : i(j) = n\}} O_j \in \tau_n$. Therefore, each τ_1-open set is a countable union of open sets in $\bigcup_{n \in \mathbb{N}} \tau_n$. Since each $O'_n \in \tau_n = \mathcal{B}(\tau_n) = \mathcal{B}_X$,

$_1 \quad B_X$ and $(\ _1) \quad B_X$. On the other hand, $_0 \quad _1$ implies $B_X = (\ _0) \quad (\ _1)$. Therefore, $(\ _1) = B_X$.

Finally, continuity of $K : (X; \ _1) \ ! \ (P(X); (\ _n)_{P(X)})$ follows from that of $K : (X; \ _{n+1}) \ ! \ (P(X); (\ _n)_{P(X)})$, for all $n \ 2 \ N$. ∎

For the next result, we will need to appeal to the famous Portmanteau Theorem, as found for example in [20].

Theorem 13 (Portmanteau Theorem). Let P and $(P_n)_{n \ 2 N}$ be a sequence of probability measures on $(X; \ )$, a metric space with its Borel $\ $-algebra. Then the following $\ $ve conditions are equivalent:

1. $P_n \ ) \ P$
2. $\liminf_n \int f dP_n = \int f dP$ for all bounded, uniformly continuous real f.
3. $\limsup_n P_n(F) \quad P(F)$ for all closed F.
4. $\liminf_n P_n(G) \quad P(G)$ for all open G.
5. $\lim_n P_n(A) = P(A)$ for all P-continuity sets A.

Lemma 4. The least upper bound of the weak topologies $\__{n \ 2 N}(\ _n)_{P(X)}$ exists and

$$\__{n \ 2 N}(\ _n)_{P(X)} = (\ _1)_{P(X)}:$$

Proof. Again, $((\ _n)_{P(X)})_{n \ 2 N}$ is an increasing sequence of Polish spaces, and so $\__{n \ 2 N}(\ _n)_{P(X)}$ exists. Clearly, $\__{n \ 2 N}(\ _n)_{P(X)} \quad (\ _1)_{P(X)}$.

Suppose $P_n \ ) \ P$ in $\__{n \ 2 N}(\ _n)_{P(X)}$. Then $P_n \ ) \ P$ in $(\ _n)_{P(X)}$ for all $n \ 2 \ N$. Let O be a $\ _1$-open set. Then as in the proof of Lemma 3, $O = [_{n \ 2 N} O_n$ where each $O_n \ 2 \ _n$. Let $G_j = [^j_{n=1} O_n \ 2 \ _j \quad B_X$. Then $(G_j)_{j \ 2 N}$ increases to O. So $P_n(O) \quad P_n(G_j)$ for all $n; j \ 2 \ N$. So $\liminf_n P_n(O) \quad \liminf P_n(G_j) \quad P(G_j)$ for all $j \ 2 \ N$ by Theorem 13 in $(\ _j)_{P(X)}$. So $\liminf_n P_n(O) \quad \lim_j P(G_j) = P(O)$ by continuity from below. So $P_n \ ) \ P$ in $(\ _1)_{P(X)}$ by Theorem 13 in $(\ _1)_{P(X)}$. Therefore, $(\ _1)_{P(X)} \quad \__{n \ 2 N}(\ _n)_{P(X)}$ whence equality follows. ∎

We are now able to prove the main theorem of this section.

Proof (Theorem 11). By Lemmas 2 and 3, there exist Polish topologies $(_n)_{n \ 2 N}$ and $\ _1$ on X, $\ $ner than $\ $, such that $(\ _1) = B_X$ and $K : (X; \ _1) \ ! \ (P(X); (\ _n)_{P(X)})$ is continuous for all $n \ 2 \ N$. This is equivalent to continuity of $K : (X; \ _1) \ ! \ (P(X); \__{n \ 2 N}(\ _n)_{P(X)})$ since convergence in $\__{n \ 2 N}(\ _n)_{P(X)}$ is equivalent to convergence in $(_n)_{P(X)}$ for all $n \ 2 \ N$ (again this follows from $\__{n \ 2 N}(\ _n)_{P(X)}$-open sets being unions of $(_n)_{P(X)}$-open sets). But then $K : (X; \ _1) \ ! \ (P(X); (\ _1)_{P(X)})$ is continuous by Lemma 4. ∎

This argument easily extends to a countable family of stochastic kernels, so that we have the following.

Corollary 2. Let $(X; \ )$ be a Polish space and $(P(X); \ _{P(X)})$ be the space of probability measures on X equipped with the topology of weak convergence with respect to $\ $. Let $(K_n)_{n \ 2 N}$ be a sequence of stochastic kernels on X. Then there exists a $\ $ner Polish topology $\ ^0$ on X such that $(\ ^0) = B_X$, $(\ ^0_{P(X)}) = B_{P(X)}$, and each $K_n : (X; \ ^0) \ ! \ (P(X); \ ^0_{P(X)})$ is continuous.

Corollary 3. Let $M = (S; B_S; A; \{P_a : a \in Ag\}; r)$ be an MDP. Then there exists a Polish topology τ on S that generates B_S and makes r_a and P_a continuous for each a in A, where $P(S)$ is endowed with the topology of weak convergence induced by τ.

Thus, if r is bounded and measurable we may apply Theorem 7 to obtain a quantitative form of bisimilarity. It is important to keep in mind what is going on here from a practical point of view: if we begin with a modelling scenario in which the rewards are discontinuous with a given metric then this amounts to changing that metric to one with respect to which rewards are continuous. Therefore the usefulness of this result is contingent on the modelling problem at hand not crucially being dependent on any specific metric.

With that caveat in mind, we now come to the main result of this paper, a general version of Theorem 10.

Corollary 4. Suppose $M = (S; B_S; A; \{P_a : a \in Ag\}; r)$ is an MDP and that the image of r is contained in $[0; 1]$. Let $c \in (0; 1)$ be a discount factor, F_c be the family of functional expressions defined in Definition 15, and G be a family of functional expressions such that $F_c \cup G \subseteq Lip(F_c)$. Then the pseudometric induced by G is the unique fixed-point σ_c satisfying the equation

$$\sigma_c(x; y) = \max_{a \in A} ((1 - c)|r_a(x) - r_a(y)| + c\, K(\sigma_c)(P_a(x); P_a(y))) \text{ for all } x; y \in S$$

and whose kernel is bisimilarity.

4 Estimating Bisimulation

In this section, we discuss how focusing on families of functional expressions may make estimating bisimilarity more amenable in practice. Assume we are given an MDP $M = (S; B_S; A; \{P_a : a \in Ag\}; r)$ where the image of r is contained in $[0; 1]$. Computing a bisimilarity metric for a finite M has encompassed estimating the integral probability metric $F_c(P; Q)$, yielding an algorithm with exponential complexity [9], computing the Kantorovich metric, $K(\sigma)(P; Q)$, yielding an algorithm with polynomial complexity [3], and solving a linear program [30].

The major issue is that although computing the linear programming formulations of bisimilarity in the ideal case can be done in polynomial time, to do so in practice is highly inefficient; to understand why, one may remark that the linear programs for a given MDP are more complex than solving for the discounted value function for that MDP; although the latter is also known to be solvable in polynomial time by linear programming [31], in practice Monte Carlo techniques have been found to be much more successful. In fact, in [4], we focused on estimating the Kantorovich metric by replacing each P and Q by empirical measures; this idea is studied in better depth in [32]. We will not focus on that approach here.

Instead we focus on a heuristic approach implicitly used in [17] and Monte Carlo techniques used in [33]. In the former, the problem at hand is, given a distribution over MDPs with a common state space, to try to find a policy that optimizes the expected total geometrically-discounted sum of rewards achieved at each state, where the average is taken over a number of sample runs performed on a number of MDPs drawn according to the given MDP distribution. The authors attack this problem by generating a family of functional expressions according to some distribution, and using these to estimate optimal planning strategies - the so-called *formula-based exploration / exploitation strategies*. In [33], the authors solve the problem of trying to compute the infimum over a given set by instead sampling and then estimating the essential infimum. Since in our case we are interested in suprema, let us recall the definition of essential supremum.

De nition 16. Let $(X; B_X;)$ be a measure space. The essential supremumof a bounded measurable function $f : (X; B_X) ! (R; B_R)$ is given by the following.

$$\text{ess sup}f = \inf f \quad 2 R : \quad (fx 2 X : f(x) > \quad g) = 0 g:$$

In other words, ess supf is the least real number that is an upper bound onf except for a set of -measure zero. It follows that in general, ess supf supf . Suppose further that B_X is a Borel -algebra, f is continuous, and is a strictly positive measure, i.e. every non-empty open subset of X has strictly positive -measure. Then sincefx 2 X : f(x) > g = $f^{1}(; 1)$ is open, it follows that it has -measure zero if and only if it is the empty set; in this case, the essential supremum and the supremum agree. We will use this in conjunction with Lemma 2 from [33], restated here in terms of the essential supremum in place of the essential in mum.

Lemma 5. Let (; ; P) and $(X; B_X;)$ be probability spaces and assume that we can sample random variables $X_1; X_2; : : : ; X_n$ mapping to X , independently and identically distributed according to . Then if $f : X ! R$ is bounded and measurable we have

$$\max_{1 i n} f(X_i) ! \text{ess sup}f \text{ in } \text{-probability as } n ! 1: \tag{10}$$

This allows for another Monte Carlo technique for (under)approximating the Kantorovich metric for bisimilarity in an MDP.

Proposition 2. Let $M = (S; B_S; A; fP_a : a 2 Ag; r)$ be an MDP where the image of r is contained in [0; 1]. Let c 2 (0; 1) be a discount factor and F_c be the family of functional expressions de ned in De nition 15 and interpreted over M . Let $\overline{F_c}$ be the closure of F_c with respect to uniform convergence. Let 2 $P(\overline{F_c})$ be strictly positive. Suppose $f_1; f_2; : : : ; f_n$ are independent, identically distributed samples drawn according to . Then

$$\max_{1 i n} jP(f_i) \quad Q(f_i)j ! F_c(P; Q) \text{ in } \text{-probability as } n ! 1: \tag{11}$$

Proof. SinceS is Hausdor , $C^b(S)$ with the uniform norm is a Banach space. Therefore, $\overline{F_c}$, as a closed subset of $C^b(S)$, is itself a measurable subspace when equipped with the Borel sets given by the uniform norm. For a given P; Q 2 P(S) let $g : \overline{F_c} ! R$ be de ned by g(f) = jP(f) Q(f)j. Then g is continuous and bounded by 1. The result now follows from Lemma 5 and the preceding remarks, and Corollary 4. □

Remark in particular that

$$\max_{1 i n} jf_i(x) \quad f_i(y)j ! \ _c(x; y) \text{ in } \text{-probability as } n ! 1: \tag{12}$$

To turn this into a proper algorithm is beyond the scope of this work - one needs to x a particular measure and provide sample complexity results, among other things. However, we remark that being able to sample from a much smaller class than the class of all Lipschitz functions should improve performance regardless of how other parameters are set.

5 Conclusions

We have shown, with slight modi cation, that the family of functional expressions developed in [9, 2] to capture quantitative bisimilarity for LMPs does the same for MDPs with continuous state spaces and bounded measurable reward functions. We have used the same techniques as in these previous works - in particular, a density result in Proposition 2 of [2] - reworded in the terminology of generating classes for integral probability metrics. The hope is that by focusing on these generating classes of functions, we may nd better practical algorithms for assessing equivalence between states in a Markov process - either by under or over-approximating a particular class, or by sampling from it in some manner.

Moreover, we have used a trick from descriptive set theory to remove a previous continuity constraint on the rewards and Markov kernels in Theorem 7, thereby widening its applicability.

5.1 Related Work

The notion of bisimilarity metrics, both in terms of logical expressions and in terms of how to compute them using linear programming formulations, really derives from the work of [9] and [2] for LMPs. In [9], the emphasis was on developing a robust theoretical notion of quantitative bisimilarity and establishing a decision procedure for it, albeit with exponential complexity. In [2], the emphasis was again on establishing a robust notion of quantitative bisimilarity while at the same time yielding a theoretical polynomial complexity bound by means of the Kantorovich metric. Complexity results in general are discussed in [30]. However, in none of these are more than a few toy examples worked through, and the idea of Monte Carlo techniques for more e cient practical implementations is not broached.

The idea of examining the relationship between probability measures by studying generating classes of functions was explored in [5, 34] for integral probability metrics and stochastic orders. Müller takes the point of view of looking at maximal generators for such orders, and demonstrates that in general, minimal orders may not exist.

To the best of our knowledge, the only practical work to exploit optimality based on functional expression occurs in [17]. Here, the goal is to determine an optimal planning strategy on average, when one is acting on an unknown MDP but given a distribution over its reward and transition parameters. The advantage of the functional expression approach here is that it is independent of the particulars of a given model.

5.2 Future Work

The point of view of this work is that one should focus on families of functional expressions for quantitative bisimilarity as we suspect this may be more advantageous in practice. Thus, an immediate concern is to turn Proposition 2 into a full- edged Monte Carlo algorithm. Among the necessities are choosing the right class of functional expressions from which to sample, as small as possible a subset of F_c, constructing a strictly positive probability measure with which to sample the class of functionals, and most importantly, a sample complexity bound to inform us of how many samples should be required for a given level of con dence.

From the theoretical side, we are interested in nding minimal classes that generate the same bisimilarity metric, and equivalences obtained from using other classes. In both cases, it might be fruitful to consider only non-empty closed subsets of $C^b(S)$ with the uniform norm. We can order

this space, and add in the empty set, to get a complete lattice; moreover, we can equip it with the Hausdor metric, and the resulting Borel -algebra, known as the E ros Borel space, will be a standard Borel space provided $(S; B_S)$ is as well ([27], pg. 97). Doing so may allow us to relate the di erences between the equivalences induced by two families of functional expressions in terms of their quantitative di erence in E ros Borel space. In particular, we are interested in coarser more easily computable equivalences, and how to relate these to the theoretical guarantees given by bisimilarity.

In statistical parlance, the interpreted class of functional expressions is just a family of random variables; and testing whether or not two states are bisimilar amounts to testing how their Markov kernels di er on this test set of random variables. Conceptually, this ts in with Prakash's view that Markov processes should be viewed as transformers of random variables [35]. As (real-valued) stochastic kernels subsume both random variables and subprobability measures, we may complete this conceptual picture by viewing a Markov process - itself a family of kernels - as a transformer of families of kernels. It remains to be seen if this point of view in general can lead to better algorithms in practice.

Acknowledgements. This work is dedicated with love to Prakash Panangaden. Absolutely none of it would have been possible without him. The authors would also like to thank Ernst-Erich Doberkat and Jean Goubault-Larrecq for suggesting Theorem 11, as well as Igor Khavkine and Jérôme Feret for verifying the results of Subsection 3.3.

References

1. Desharnais, J., Jagadeesan, R., Gupta, V., Panangaden, P.: The Metric Analogue of Weak Bisimulation for Probabilistic Processes. In: LICS '02: Proceedings of the 17th Annual IEEE Symposium on Logic in Computer Science, Copenhagen, Denmark, 22-25 July 2002, Washington, DC, USA, IEEE Computer Society (2002) 413{422
2. van Breugel, F., Worrell, J.: Towards Quantitative Veri cation of Probabilistic Transition Systems. In: ICALP '01: Proceedings of the 28th International Colloquium on Automata, Languages and Programming,, London, UK, Springer-Verlag (2001a) 421{432
3. van Breugel, F., Worrell, J.: An Algorithm for Quantitative Veri cation of Probabilistic Transition Systems. In: CONCUR '01: Proceedings of the 12th International Conference on Concurrency Theory, London, UK, Springer-Verlag (2001b) 336{350
4. Ferns, N., Panangaden, P., Precup, D.: Bisimulation Metrics for Continuous Markov Decision Processes. SIAM Journal on Computing 40(6) (2011) 1662{1714
5. Muller, A.: Integral Probability Metrics and Their Generating Classes of Functions. Advances in Applied Probability 29 (1997) 429{443
6. Larsen, K.G., Skou, A.: Bisimulation Through Probabilistic Testing. Information and Computation 94(1) (1991) 1{28
7. Milner, R.: A Calculus of Communicating Systems. Volume 92 of Lecture Notes in Computer Science. Springer-Verlag, New York, NY (1980)
8. Park, D.: Concurrency and Automata on In nite Sequences. In: Proceedings of the 5th GI-Conference on Theoretical Computer Science, London, UK, Springer-Verlag (1981) 167{183
9. Desharnais, J., Gupta, V., Jagadeesan, R., Panangaden, P.: Metrics for Labeled Markov Systems. In: CONCUR '99: Proceedings of the 10th International Conference on Concurrency Theory, London, UK, Springer-Verlag (1999) 258{273
10. Desharnais, J., Gupta, V., Jagadeesan, R., Panangaden, P.: Metrics for Labelled Markov Processes. Theor. Comput. Sci. 318(3) (2004) 323{354

11. van Breugel, F., Hermida, C., Makkai, M., Worrell, J.: Recursively De ned Metric Spaces Without Contraction. Theoretical Computer Science 380(1-2) (July 2007) 143{163
12. Kozen, D.: A Probabilistic PDL. In: STOC '83: Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing, New York, NY, USA, ACM (1983) 291{297
13. van Breugel, F., Sharma, B., Worrell, J.: Approximating a Behavioural Pseudometric Without Discount for Probabilistic Systems. In Seidl, H., ed.: Foundations of Software Science and Computational Structures, 10th International Conference, FOSSACS 2007, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2007, Braga, Portugal, March 24-April 1, 2007, Proceedings. Volume 4423 of Lecture Notes in Computer Science., Springer (2007) 123{137
14. Ferns, N., Panangaden, P., Precup, D.: Metrics for Finite Markov Decision Processes. In: AUAI '04: Proceedings of the 20th Annual Conference on Uncertainty in Arti cial Intelligence, Arlington, Virginia, United States, AUAI Press (2004) 162{169
15. Ferns, N., Panangaden, P., Precup, D.: Metrics for Markov Decision Processes with In nite State Spaces. In: Proceedings of the 21th Annual Conference on Uncertainty in Arti cial Intelligence (UAI-05), Arlington, Virginia, AUAI Press (2005) 201{208
16. Ferns, N., Castro, P.S., Precup, D., Panangaden, P.: Methods for Computing State Similarity in Markov Decision Processes. In: Proceedings of the 22nd Annual Conference on Uncertainty in Arti cial Intelligence (UAI-06), Arlington, Virginia, AUAI Press (2006)
17. M. Castronovo, F. Maes, R.F., Ernst., D.: Learning Exploration/Exploitation Strategies for Single Trajectory Reinforcement Learning. In: In Proceedings of the 10th European Workshop on Reinforcement Learning (EWRL 2012), Edinburgh, Scotland, June 30-July 1 2012. Volume 24. (2012) 1{10
18. Panangaden, P.: Labelled Markov Processes. Imperial College Press (2009)
19. Giry, M.: A Categorical Approach to Probability Theory. Categorical Aspects of Topology and Analysis (1982) 68{85
20. Billingsley, P.: Convergence of Probability Measures. Wiley (1968)
21. Dudley, R.M.: Real Analysis and Probability. Cambridge University Press (August 2002)
22. Desharnais, J.: Labelled Markov Processes. PhD thesis, McGill University (2000)
23. Desharnais, J., Edalat, A., Panangaden, P.: Bisimulation for Labeled Markov Processes. Information and Computation 179(2) (Dec 2002) 163{193
24. Gibbs, A.L., Su, F.E.: On Choosing and Bounding Probability Metrics. International Statistical Review 70 (2002) 419{435
25. Villani, C.: Topics in Optimal Transportation (Graduate Studies in Mathematics, Vol. 58). American Mathematical Society (2003)
26. Hernandez-Lerma, O., Lasserre, J.B.: Further Topics on Discrete-Time Markov Control Processes. Applications of Mathematics. Springer, New York (1999)
27. Srivastava, S.M.: A Course on Borel Sets. Volume 180 of Graduate texts in mathematics. Springer (2008)
28. Bertsekas, D.P., Shreve, S.E.: Stochastic Optimal Control: The Discrete-Time Case. Athena Scienti c (2007)
29. Parthasarathy, K.R.: Probability Measures on Metric Spaces. Academic, New York (1967)
30. Chen, D., van Breugel, F., Worrell, J.: On the Complexity of Computing Probabilistic Bisimilarity. In Birkedal, L., ed.: FoSSaCS. Volume 7213 of Lecture Notes in Computer Science., Springer (2012) 437{451
31. Puterman, M.L.: Markov Decision Processes: Discrete Stochastic Dynamic Programming. John Wiley & Sons, Inc., New York, NY, USA (1994)
32. Sriperumbudur, B.K., Fukumizu, K., Gretton, A., Scholkopf, B., Lanckriet, G.R.G.: On the Empirical Estimation of Integral Probability Metrics. Electronic Journal of Statistics 6 (2012) 1550{1599
33. Bouchard-Côte, A., Ferns, N., Panangaden, P., Precup, D.: An Approximation Algorithm for Labelled Markov Processes: Towards Realistic Approximation. In: QEST '05: Proceedings of the Second International Conference on the Quantitative Evaluation of Systems (QEST'05) on The Quantitative Evaluation of Systems, Washington, DC, USA, IEEE Computer Society (2005) 54{61

34. Muller, A.: Stochastic Orders Generated by Integrals: A Uni ed Study. Advances in Applied Probability 29 (1997) 414{428
35. Chaput, P., Danos, V., Panangaden, P., Plotkin, G.D.: Approximating Markov Processes by Averaging. In Albers, S., Marchetti-Spaccamela, A., Matias, Y., Nikoletseas, S.E., Thomas, W., eds.: Automata, Languages and Programming, 36th Internatilonal Collogquium, ICALP 2009, Rhodes, greece, July 5-12, 2009, Proceedings, Part II. Volume 5556 of Lecture Notes in Computer Science., Springer (2009) 127{138