Master's Thesis

Quantifying Uncertainty in Uncertain Markov Chains

Sander Suverkropp

February 25, 2026

Supervisor:
dr. Jurriaan Rot

Second reader:
dr. Sebastian Junges

Radboud University

Abstract

Uncertain Markov decision processes (uMDPs) are probabilistic models used for sequential decision-making. These models are a variation on Markov decision processes where the fixed transition probabilities are replaced by an uncertainty set of possible transition probabilities. They can be more or less uncertain, based on the size of the uncertainty sets, as well as a variety of other factors. We do not have a way of quantifying the uncertainty in uMDPs. In this thesis, we work towards this goal by quantifying the uncertainty in uncertain Markov chains (uMCs), or more precisely uncertain Markov reward models. These are similar to uMDPs, but do not have actions. We introduce the notion of uncertainty functions, which are functions that map uMCs to a number representing their uncertainty. In this thesis, they are defined using the maximum distance between elements of the uncertainty set under different pseudometrics. The pseudometrics used consist of the existing bisimilarity metric, as well as a three new pseudometrics based on the distribution over traces, the distribution over values and the expected value. From these pseudometrics, we derive four different uncertainty functions. We only have a method to calculate the uncertainty function based on expected value. This makes this the only function that is currently usable to quantify uncertainty.

Contents

1 Introduction

Markov decision processes (MDPs) are a type of mathematical model for sequential decision-making in a probabilistic environment. They have a wide range of practical applications. For instance, they have been used to set salmon fishing quotas in such a way that the long term yield is maximized [32]. In hydroelectric dams, they have been used to determine how much energy to produce based on the current energy prices and the amount of water in the reservoir [32]. Smart overnight charging of EVs can be used to compensate for the reduced flexibility of green energy production compared to fossil fuels. MDPs can be used in this process by planning the charging over the entire night [21]. They have also been used to analyse the optimal strategy for the ambulance dispatching problem, where you have to decide which ambulance to send to an incident [19].

MDPs can be formulated as a type of transition system, where the transitions are labelled by an action and a reward. In each state we can take one of the available actions. The combination of state and action determines the probability distribution over successor states, as well as how much reward we receive. For an MDP, a policy is a function which selects one of the available actions in each state. The goal of an MDP is to find a policy that maximizes the expected reward.

When we model real world problems as MDPs, the exact transition probabilities are often not known. There are two ways that the transition probabilities are commonly derived. They can be estimated by experts, or approximated from collected data.

Small differences between the probabilities in the model and the probabilities in the real world can result in a suboptimal policy. This is because the optimal policy calculated from an MDP can be sensitive to small changes or inaccuracies in the transition probabilities [24].

To obtain a policy robust to slight inaccuracies in the transition probabilities, we can use uncertain MDPs [27]. In these models, the transition probabilities are not fixed, but are specified to be within a certain "uncertainty set" of possible transition probabilities. One common way to do this is specifying an interval within which the probability has to lie. For an uncertain MDP, we can look at the worst-case expected reward, that is, the expected reward using the transition probabilities for which the policy performs worst. This gives us a lower bound for the expected reward. The policy with the highest worst-case reward is called the robust policy.

Not all uncertain MDPs have the same amount of uncertainty. This leads us to the question of how we can quantify the uncertainty in uncertain MDPs. Obviously, the larger the uncertainty set, the larger the uncertainty of the model. However, there are also other factors that can be taken into account. For instance, if a state is unlikely to be reached, the uncertainty in its transitions could be considered to contribute less to the overall uncertainty of the model than if that state is likely to be reached. Another factor that can be taken into account is the difference between the possible successor states in an uncertain transition. If they are very similar, the overall uncertainty is less than if they are very different.

Quantifying the uncertainty in uncertain MDPs would be useful for several reasons. For example, there are different methods for gathering data to refine the probabilities in the uncertain MDP [29]. These could be compared by examining which method most efficiently reduces the uncertainty. Another motivation is to see whether the range of probabilities in the uncertain MDP is narrow enough to derive a good policy. If the uncertainty is still high, it might be worth investigating the probabilities more finely, while if it is already low, this would be wasted effort.

In this thesis, we use uncertain Markov chains as a stepping stone to quantifying uncertainty in uncertain MDPs. The Markov chains used in this thesis are like MDPs, but without actions, also known as Markov reward models. Uncertain Markov chains have an uncertainty set of possible transition probabilities similar to uncertain MDPs. We can view them as a subset of uncertain MDPs, namely those with a single action.

To quantify uncertainty in uncertain Markov chains, we first need to define the properties which such a quantification needs to have. We have captured this in the concept of an uncertainty function, which maps uncertain Markov chains to a number representing their uncertainty. There are two basic properties we want these uncertainty functions to have. First, if the uncertain Markov chain has only one possible transition function, there is no uncertainty. The uncertainty function should be zero to reflect this. Second, suppose we have two uncertain Markov chains $\mathcal{M}$ and $\mathcal{N}$, and that for every element in the uncertainty set of $\mathcal{M}$, there is a corresponding one in the uncertainty set of $\mathcal{N}$ that results in the same behaviour. Then we want an uncertainty function to give a $\mathcal{N}$ a value that is at least as large as that of $\mathcal{M}$. To formalize this, we use bisimilarity to express that Markov chains have the same behaviour.

In this thesis, we define uncertainty functions by using the maximum distance between elements of the uncertainty set given some distance function. This is a natural fit for the problem, because the uncertainty in uncertain Markov chains does not have any probability measure associated with it. Instead, the uncertainty set forms a set of possibilities, each of which needs to be taken into account. The distance functions that we use for this are pseudometrics. That means they can assign a distance of zero to two different Markov chains. One case where this is useful is for Markov chains that are bisimilar.

In particular, we look at the existing bisimilarity metric as well as three pseudometrics defined in this thesis. The bisimilarity metric generalizes bisimilarity to a continuous function. The other three pseudometrics are: the trace metric, the value distribution metric and the expected value metric. These compare the distribution over traces, the distribution over values and the expected value respectively.

In this thesis, we introduce the concept of an uncertainty function, which is a function measuring the uncertainty in an uncertain Markov chain. We also show that uncertainty functions can be derived from bounded pseudometrics on Markov chains that assign distance zero to Markov chains that are bisimilar. Then we introduce three new pseudometrics on Markov chains: the trace metric, the value distribution metric and the expected value metric. We compare each of these to the existing bisimilarity metric, thus showing that they can be used to derive an uncertainty function. We also discuss the computational complexity of these pseudometrics, as well as their associated uncertainty functions. Lastly, we give several ideas for possible extensions of uncertainty functions to uncertain MDPs.

1.1 Outline

In Section 2 we define several important concepts, including uncertain Markov chains. In Section 3, we introduced the concept of an uncertainty function, which quantifies the amount of uncertainty in an uncertain Markov chain. Next, in Section 4, four pseudometrics on Markov chains are defined. We compare these metrics and prove they can be strictly ordered in Section 5. This also allows us to conclude that each of them can be used to define an uncertainty function. In Section 6, we discuss algorithms to compute the pseudometrics and their associated uncertainty functions. After that, in Section 7, we sketch how the concept of uncertainty functions and the specific metrics and associated uncertainty functions introduced in this thesis could be extended to uncertain MDPs. In Section 8, we discuss related work. And finally, the conclusion is Section 9.

Acknowledgements

I am grateful to everyone who helped me during the process of writing this thesis. In particular, I would like to thank:

- my supervisor Jurriaan Rot who was very patient throughout this process, even when I made little progress.
- Nils Jansen and Marnix Suilen for their help and the discussions during the beginning of this project.
- my parents and sister for their emotional support.
- Anna, who helped me get access to many papers I would not otherwise have been able to read.

2 Preliminaries

In this section, we recall background on several topics, including pseudometrics and (uncertain) Markov chains.

2.1 Pseudometrics

Definition 2.1 (Couplings)**.** A probability distribution $\lambda \in \mathbb{P}(X \times Y)$ is a *coupling* of probability distributions $P \in \mathbb{P}(X)$ and $Q \in \mathbb{P}(Y)$ if and only if for all measurable subsets $A \subseteq X$ and $B \subseteq Y$,

$$\lambda(A \times Y) = P(A) \quad \text{and} \quad \lambda(X \times B) = Q(B).$$

The notation $\Lambda(P, Q)$ denotes the set of all couplings between P and Q.

Definition 2.2 (Pseudometric)**.** A function $d\colon X \times X \to \mathbb{R}_{\geq 0}$ is a *pseudometric* if for all $x, y, z \in X$,

$$d(x, x) = 0$$

$$d(x, y) = d(y, x)$$
$$d(x, z) \leq d(x, y) + d(y, z)$$

Lemma 2.3. *Pseudometric have the following properties:*

(a) If d_Y is a pseudometric, and $f\colon X \to Y$, then $d_X(x, y) = d_Y(f(x), f(y))$ is also a pseudometric.

(b) For any two pseudometrics d, d', and a constant c, $d + cd'$ is also a pseudometric.

(c) If each element of a sequence is a pseudometric, then the limit of the sequence will also be a pseudometric.

These pseudometrics can be ordered by a partial ordering by saying that $d \leq d'$ if

$$\forall \mathcal{M}, \mathcal{N} d(\mathcal{M}, \mathcal{N}) \leq d'(\mathcal{M}, \mathcal{N}).$$

2.1.1 Kantorovich metric

In this subsection we will define the Kantorovich metric[1]. These definitions can be found in Chapter 1 and Section 7.1 of [31].

The Kantorovich metric is defined in the context of Polish spaces. A Polish space is a separable completely metrizable topological space. Relevant for this thesis is that a finite set (such as the state space of a Markov chain) with the discrete topology is a Polish space.

Definition 2.4 (Kantorovich metric)**.** Let X and Y be Polish spaces. Given a function $f\colon X \times Y \to \mathbb{R}$, the Kantorovich metric $\mathcal{K}(f)\colon \mathbb{P}(X) \times \mathbb{P}(Y) \to \mathbb{R}$ is defined by

$$\mathcal{K}(f)(P, Q) = \inf_{\lambda \in \Lambda(P,Q)} \int_{X \times Y} f(x, y) \lambda(dx, dy). \tag{1}$$

[1] This metric is also known by the names Monge-Kantorovich, Kantorovich-Rubinstein, Hutchinson, Mallows, Wasserstein, Vasserstein, Earth Mover's Distance, Fortet-Mourier, and Dudley [15].

Note that while it is called the Kantorovich metric, it is not always a metric when applied to an arbitrary function. It is, however, a pseudometric when applied to a pseudometric. In Theorem 7.4 of [31], Villani proves that given a metric d, $\mathcal{K}(d)$ is always a metric. This can easily be adapted to show that if d is instead a pseudometric, $\mathcal{K}(d)$ is also a pseudometric.

Lemma 2.5. *The operator $\mathcal{K}(-)$ preserves pseudometrics. That is, if d is a pseudometric on X, then $\mathcal{K}(d)$ is a pseudometric on probabilities over X.*

From Theorem 1.3 of [31], we also get the following:

Lemma 2.6. *The minimizing coupling of a Kantorovich metric always exists. In other words, the infimum in* (1) *is a minimum.*

Now we will prove a lemma about the minimizing coupling in the case of a finite set X. For such sets, we can convert the integral into a sum:

$$\mathcal{K}(f)(P,Q) = \inf_{\lambda \in \Lambda(P,Q)} \sum_{x,y \in X} f(x,y)\lambda(x,y)$$

The following lemma shows that if the distance between two points is zero, there is always the coupling minimizing the Kantorovich metric that assigns them the maximum possible probability.

Lemma 2.7. *Let d be a pseudometric over a finite set X, and $P, Q \in \mathbb{P}(X)$. Furthermore, let $x_0, y_0 \in X$ such that $d(x_0, y_0) = 0$. Then there exists a coupling λ between P and Q that minimizes the sum in the Kantorovich metric, and $\lambda(x_0, y_0) = \min(P(x_0), Q(y_0))$.*

Proof. Without loss of generality, we assume that $P(x_0) \le Q(y_0)$, such that $\min(P(x_0), Q(y_0)) = P(x_0)$.

By Lemma 2.6 there exists at least one coupling that minimizes the sum. Call the couplings for which the sum is minimized Λ_0.

We define Λ_1 to be the set of couplings $\lambda \in \Lambda_0$ for which the number $\#\{y \in X \setminus \{y_0\} \mid \lambda(x_0, y) > 0\}$ is minimal. Note that if this minimum is 0, then it follows that $\lambda(x_0, y_0) = P(x_0)$. Because of this, we can assume this minimum is greater than 0.

Now we can define Λ_2 to be the set of couplings $\lambda \in \Lambda_1$ for which the number of $x \in X \setminus \{x_0\}$ such that $\lambda(x, y_0) > 0$ is minimal. Note that this minimum must be greater than zero, because for all $\lambda \in \Lambda_1$, $\lambda(x_0, y_0) < P(x_0) \le Q(y_0)$.

Let λ be a coupling in Λ_2. Then we can choose $y_1 \ne y_0$ such that $\lambda(x_0, y_1) > 0$, and $x_1 \ne x_0$ such that $\lambda(x_1, y_0) > 0$. Let $r = \min(\lambda(x_1, y_0), \lambda(x_0, y_1)$. Now we can define a new coupling λ' as follows

$$\lambda'(x,y) = \begin{cases} \lambda(x_0,y_1) - r & \text{if } x = x_0 \wedge y = y_1 \\ \lambda(x_1,y_0) - r & \text{if } x = x_1 \wedge y = y_0 \\ \lambda(x_0,y_0) + r & \text{if } x = x_0 \wedge y = y_0 \\ \lambda(x_1,y_1) + r & \text{if } x = x_1 \wedge y = y_1 \\ \lambda(x,y) & \text{otherwise} \end{cases}$$

The sum of the Kantorovich metric is still minimized by λ', by the triangle inequality:

$$\sum_{x,y \in X} \lambda'(x,y)d(x,y) - \sum_{x,y \in X} \lambda(x,y)d(x,y) = r(d(x_1,y_1) - d(x_1,y_0) - d(x_0,y_1)) \le r \cdot 0$$

So $\lambda' \in \Lambda_0$. Now there are two cases:

- If $r = \lambda(x_0, y_1)$, then $\lambda'(x_0, y_1) = 0$. But then there are fewer $y \in X \setminus \{y_0\}$ for which $\lambda'(x_0, y)$ is positive then for which $\lambda(x_0, y)$ is positive. But then λ cannot be in Λ_1, a contradiction.
- If $r = \lambda(x_1, y_0)$, then $\lambda'(x_1, y_0) = 0$. But then there are fewer $x \in X \setminus \{x_0\}$ for which $\lambda'(x, y_0)$ is positive then for which $\lambda(x, y_0)$ is positive. But then λ cannot be in Λ_2, a contradiction.

□

2.2 Banach fixed point theorem

The Banach fixed point theorem shows that a function that reduces the distance between points to which it is applied has a unique fixed point [5]. This theorem and a proof can also be found in [4] as Theorem 1.34. In this thesis, we use it for the definition of the bisimilarity metric and later in a proof relating the bisimilarity metric to the trace metric.

Definition 2.8 (Contraction mapping)**.** Given a metric space (X, d), a function $T\colon X \to X$ is a *contraction mapping* if there exists a $q \in [0, 1)$ such that for all $x, y \in X$,

$$d(T(x), T(y)) \leq q d(x, y).$$

Theorem 2.9 (Banach fixed point theorem)**.** *Let (X, d) be a non-empty complete metric space, and let $T\colon X \to X$ be a contraction mapping. Then T has a unique fixed point x^*. In addition, for any point $x_0 \in X$,*

$$\lim_{n \to \infty} T^n(x_0) = x^*.$$

2.3 Markov chains

Markov chains are a type of transition system where the probability of the next state only depends on the current state. A Markov chain consists of a number of states, one of which is the initial state, as well as a probability distribution for each state representing the probability of transitioning to a given other state. In this thesis, we use Markov chains as a simplified version of MDPs. As such, we associate a reward with each state in a Markov chain. Markov chains with such a reward function are also called Markov reward models. For more background on Markov chains, see [22]. Particularly, Markov chains are defined in Section 2.1 of [22], and the discounted value is discussed in Section 4.8.1 of the same source. Our definition corresponds to time-homogeneous discrete time Markov chains with a reward function in this source. In Figure 1, an example of a Markov chain is given.

Definition 2.10 (Markov chain)**.** A *Markov chain* (MC) M is a tuple (S, s_{init}, P, R), where S is a finite set of states, $s_{init} \in S$ is the initial state, $P = (P_s)_{s \in S}\colon S \to \mathbb{P}(S)$ is the transition function, i.e. $P_s \in \mathbb{P}(S)$, and $R\colon S \to [0, 1]$. Let MC denote the class of all MCs.

The semantics of a Markov chain is its expected value. This is based on the expected reward in each time step. To ensure that this value is finite, even if the Markov chain keeps having reward, the value is discounted by a discount factor $\gamma \in (0, 1)$ for each time step in the future that the reward is received. As a result, the maximum value a Markov chain can have is $\sum_{i=0}^{\infty} \gamma^i = \frac{1}{1-\gamma}$.

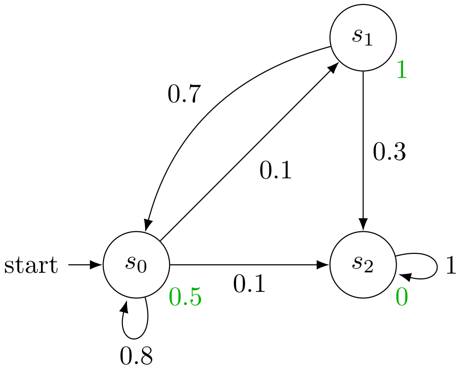

Figure 1: An example of a Markov chain. The label next to a state represents the reward of the transitions, while the number next to the arrow represents the probability of the transitions.

Definition 2.11 (Expected discounted value)**.** Given a discount factor $\gamma \in (0, 1)$, the *expected γ-discounted value* of a state s in a Markov chain M, or just *value* for short, is defined as the least fixed point of

$$V^M(s) = R(s) + \gamma \sum_{s' \in S} P_s(s')V(s').$$

We define the value of an entire Markov chain as the value of its initial state:

$$V(M) = V^M(s^M_{init})$$

Now we will give an example by calculating the value of the Markov chain M in Figure 1.

Example 2.12. For this example we will use a discount factor of $\gamma = 0.9$. The state s_2 has only one outgoing transition, and this transition is to itself. This means we can calculate its value before calculating the others.

$$V(s_2) = R(s_2) + \gamma \sum_{s' \in S} P_{s_2}(s')V(s') = 0 + 0.9(P_{s_2}(s_2)V(s_2)) = 0.9V(s_2)$$

From this, we can conclude that $V(s_2) = 0$. The values of the other two states depend on each other. For s_0 we have

$$\begin{aligned} V(s_0) &= R(s_0) + \gamma \sum_{s' \in S} P_{s_0}(s')V(s') \\ &= 0.5 + 0.9\,(0.8V(s_0) + 0.1V(s_1) + 0.1V(s_2)) \\ &= 0.5 + 0.72V(s_0) + 0.09V(s_1). \end{aligned}$$

For s_1 we have

$$\begin{aligned} V(s_1) &= R(s_1) + \gamma \sum_{s' \in S} P_{s_1}(s')V(s') \\ &= 1 + 0.9\,(0.7V(s_0) + 0.3V(s_2)) \\ &= 1 + 0.63V(s_0). \end{aligned}$$

Substituting the formula for $V(s_1)$ into the one for $V(s_0)$, we get

$$V(s_0) = 0.5 + 0.72V(s_0) + 0.09(1 + 0.63V(s_0)) = 0.59 + 0.7767V(s_0).$$

Now we can rewrite this to get the value of the state s_0, which is also the value of the Markov chain as a whole.

$$V(M) = V(s_0) = \frac{0.59}{1 - 0.7767} = \frac{0.59}{0.2233} \approx 2.642$$

2.3.1 Bisimilarity

Bisimilarity is a notion of behavioural equivalence. Bisimulation for Markov chains with reward can be seen as a simplified version of bisimulation for MDPs. This was introduced by Larsen and Skou for probabilistic transition systems [23]. Givan, Dean and Greig adapted it to MDPs by incorporating reward into the definition [16].

Definition 2.13 (Bisimilarity)**.** A relation on states of Markov chains $T \subseteq S \times S$ is a *bisimulation relation* if for all pairs of states s, t that are related by T,

1. $R(s) = R(t)$, and
2. There exists a coupling λ between P_s and P_t such that each pair of states s', t' in the support of λ is related by T. That is,

$$\forall s', t'.\lambda(s', t') > 0 \implies (s', t') \in T.$$

When two states s and t are related by any bisimulation relation, we call them *bisimilar*, which is denoted by $s \sim t$. This bisimilarity relation $\sim$ forms the maximal bisimulation relation.

2.4 Traces

An infinite sequence of states is called a *trace*. We will denote a trace consisting of states $s_0, s_1, s_2, \ldots$ by $\mathbf{s}$. Given a trace, we can calculate its discounted value similarly to how the value of a state or Markov chain is calculated.

Definition 2.14 (Discounted value)**.** Given a discount factor $\gamma \in (0, 1)$, the γ*-discounted value* of a trace $V\colon S^\omega \to \mathbb{R}$ is defined as

$$V^M(\mathbf{s}) = \sum_{i=0}^{\infty} \gamma^i R(s_i).$$

Markov chains induce a probability measure on the set of traces. To define this measure, we first need to define the σ-algebra associated with a Markov chain. This is from Definition 10.10 of [3].

The *Cylinder set* of $s_0 \ldots s_n \in S^{n+1}$ is defined by

$$\text{Cyl}(s_0 \ldots s_n) = \{\mathbf{t} \in S^\omega \mid \forall i \leq n.t_i = s_i\}$$

The σ-algebra associated with Markov chain M is the smallest σ-algebra that contains $\text{Cyl}(s_0 \ldots s_n)$ for all finite paths $s_0 \ldots s_n \in S^{n+1}$.

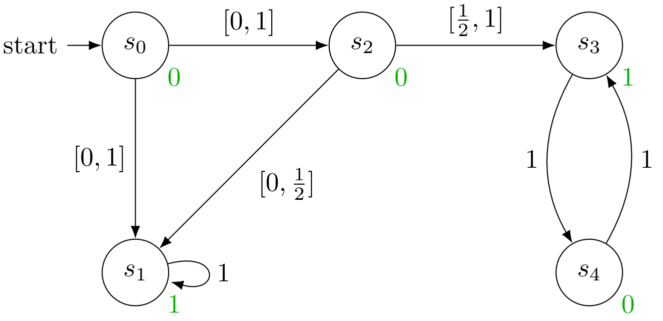

Figure 2: An example of a uMC.

The probability measure $\Pr^M$ associated with M is defined by

$$\Pr^M(\mathrm{Cyl}(s_0 \dots s_n)) = \delta(s_0, s_{init}^M) \prod_{i=0}^{n-1} P_{s_i}(s_{i+1})$$

where $\delta(s,t) = \begin{cases} 1 & \text{if } s = t \\ 0 & \text{if } s \neq t \end{cases}$.

If we define the space of traces as a topological space, with a topology generated by the cylinder sets, this σ-algebra is also the Borel σ-algebra. In addition, the space is then a Polish space. It is separable because taking one element from each cylinder set forms a countable dense subset. We can define a metric $d(\mathbf{s}, \mathbf{t}) = \frac{1}{n}$, where n is the first element where $\mathbf{s}$ and $\mathbf{t}$ differ. With this metric, the space is also complete.

As we would expect, the value of a Markov chain is equal to the expected value of a trace from the corresponding probability distribution over traces.

$$V(M) = \mathop{\mathbb{E}}_{\mathbf{s} \sim \Pr^M} [V(\mathbf{s})]$$

2.5 Uncertain Markov chains

Uncertain Markov chains are similar to Markov chains, but instead of one specific transition function, there is an uncertainty set of possible transition functions.

Definition 2.15 (Uncertain Markov chain)**.** An *uncertain Markov chain* (uMC) $\mathcal{M}$ is a tuple $(S, s_{init}, \mathcal{P}, R)$, where S is a finite set of states, $s_{init} \in S$ is the initial state, $\mathcal{P}$ is the uncertainty set, a non-empty set of transition functions, i.e. $\forall P \in \mathcal{P}$, $P \colon S \to \mathbb{P}(S)$, and $R \colon S \to [0,1]$. Let UMC denote the class of all uncertain MCs.

An example of a uMC is shown in Figure 2. In this example, the uncertainty is denoted by intervals. The possible probability distributions are all distributions such that the probability of each transition is within its respective interval.

An uncertain Markov chain is s-rectangular if its set of transition functions can be written as

$$\mathcal{P} = \prod_{s \in S} \mathcal{P}_s$$

where $\mathcal{P}_s$ is a set of probability distributions over states. This means that the possible transition probabilities in each state are independent of the transition probabilities in other states. The sets $\mathcal{P}_s$ are called local uncertainty sets. From here on, I will assume all uncertain Markov chains are s-rectangular.

An uncertain Markov chain can equivalently be seen as a set of Markov chains. We call this set the uncertainty set. Note that this is not the same as the set of possible transition functions $\mathcal{P}$, which is also called an uncertainty set. It will be clear from context which of the two is the intended meaning.

Definition 2.16 (Uncertainty set)**.** The *uncertainty set* of a uMC $\mathcal{M} = (S, s_{init}, \mathcal{P}, R)$, is

$$[\![\mathcal{M}]\!] = \{(S, s_{init}, (P_s)_{s \in S}, R) \mid \forall s \in S, P_s \in \mathcal{P}_s\}.$$

We can determine two values for an uncertain Markov chain, the robust value and the optimistic value. These look at the worst and best case respectively.

Definition 2.17. Given a discount factor $\gamma \in (0, 1)$, the *optimistic value* of an uncertain Markov chain is defined by

$$\overline{V}(\mathcal{M}) = \sup_{M \in [\![\mathcal{M}]\!]} V(M).$$

Similarly, the *robust*, or *pessimistic*, *value* of an uncertain Markov chain is defined by

$$\underline{V}(\mathcal{M}) = \inf_{M \in [\![\mathcal{M}]\!]} V(M).$$

For s-rectangular uMCs, the transition probabilities in each state are independent. Then there is one Markov chain in the uncertainty set that maximizes the value in all states, yielding the optimistic value, and similarly for the robust value. This means we can also define the optimistic and robust value for each state, and calculate them as a least fixed point. The optimistic value of a state s is the least fixed point of

$$\overline{V}(s) = R(s) + \gamma \sup_{P_s \in \mathcal{P}_s} \sum_{s' \in S} P_s(s')\overline{V}(s').$$

Analogously, the robust value is the least fixed point of

$$\underline{V}(s) = R(s) + \gamma \inf_{P_s \in \mathcal{P}_s} \sum_{s' \in S} P_s(s')\underline{V}(s').$$

Then the optimistic (or robust) value of an uncertain Markov chain is equal to the optimistic (or robust) value of its initial state.

As an example, we will calculate the optimistic and robust values of the uncertain Markov chain in Figure 2, which we will call $\mathcal{M}$.

Example 2.18. We will start with the optimistic value. For the states s_1, s_3 and s_4, there is only one successor state, and no uncertainty. We have

$$\begin{aligned}
\overline{V}(s_1) &= R(s_1) + \gamma\overline{V}(s_1) = 1 + \gamma\overline{V}(s_1) \\
\overline{V}(s_3) &= R(s_3) + \gamma\overline{V}(s_4) = 1 + \gamma\overline{V}(s_4) \\
\overline{V}(s_4) &= R(s_4) + \gamma\overline{V}(s_3) = \gamma\overline{V}(s_3)
\end{aligned}$$

Rewriting these equations gives us the following values.

$$\overline{V}(s_1) = \frac{1}{1-\gamma} \qquad \overline{V}(s_3) = \frac{1}{1-\gamma^2} \qquad \overline{V}(s_4) = \frac{\gamma}{1-\gamma^2}$$

With these values, we can look at the value for state s_2. Its transition probabilities can be described a single parameter $p \in [0, \frac{1}{2}]$, with the probability of s_1 being p, and the probability of s_3 being $1-p$. That is, $\mathcal{P}_{s_2} = \{P_p \mid p \in [0, \frac{1}{2}]\}$.

$$\begin{aligned}
\overline{V}(s_2) &= R(s_2) + \gamma \sup_{P \in \mathcal{P}_{s_2}} \sum_{s' \in S} P(s')\overline{V}(s') \\
&= \gamma \sup_{p \in [0, \frac{1}{2}]} p\overline{V}(s_1) + (1-p)\overline{V}(s_3) \\
&= \gamma \sup_{p \in [0, \frac{1}{2}]} p\frac{1}{1-\gamma} + (1-p)\frac{1}{1-\gamma^2} = \gamma\left(\frac{0.5}{1-\gamma} + \frac{0.5}{1-\gamma^2}\right)
\end{aligned}$$

Finally, we can determine the optimistic value of the initial state, which is also the optimistic value of $\mathcal{M}$. Note that $\overline{V}(s_1) > \overline{V}(s_2)$. This means that the transition probabilities that would maximize the value of s_0 are $P_{s_0}(s_1) = 1$ and $P_{s_0}(s_2) = 0$.

$$\overline{V}(\mathcal{M}) = \overline{V}(s_0) = R(s_0) + \gamma\overline{V}(s_1) = \gamma\frac{1}{1-\gamma}$$

The robust value of the states s_1, s_3 and s_4 is equal to their optimistic value, because it is impossible to reach a state with uncertain transitions from them. For s_2, we again parametrize the possible transition probabilities by $p \in [0, \frac{1}{2}]$.

$$\begin{aligned}
\underline{V}(s_2) &= R(s_2) + \gamma \inf_{P_{s,p} \in \mathcal{P}_s} \sum_{s' \in S} P_{s,p}(s')\underline{V}(s') \\
&= \gamma \inf_{p \in [0, \frac{1}{2}]} p\underline{V}(s_1) + (1-p)\underline{V}(s_3) \\
&= \gamma \inf_{p \in [0, \frac{1}{2}]} p\frac{1}{1-\gamma} + (1-p)\frac{1}{1-\gamma^2} = \gamma\frac{1}{1-\gamma^2}
\end{aligned}$$

Now we can calculate the robust value of the initial state and of $\mathcal{M}$.

$$\begin{aligned}
\underline{V}(\mathcal{M}) = \underline{V}(s_0) &= R(s_0) + \gamma \inf_{p \in [0,1]} p\underline{V}(s_1) + (1-p)\underline{V}(s_2) \\
&= \gamma \inf_{p \in [0,1]} p\frac{1}{1-\gamma} + (1-p)\gamma\frac{1}{1-\gamma^2} = \gamma^2\frac{1}{1-\gamma^2}
\end{aligned}$$

3 Uncertainty on Markov chains

In this section we introduce the notion of uncertainty functions on Markov chains. This notion is designed to encompass functions that quantify the amount of uncertainty in uncertain Markov chains. Formally, the definition is as follows.

Definition 3.1 (Uncertainty function)**.** An *uncertainty function on Markov chains* is a function $U\colon \text{UMC} \to \mathbb{R}$ such that for all uncertain MCs $\mathcal{M}, \mathcal{N}$,

$$|[\![\mathcal{M}]\!]| = 1 \implies U(\mathcal{M}) = 0 \qquad \text{(certainty)}$$
$$\forall M \in [\![\mathcal{M}]\!]\ \exists N \in [\![\mathcal{N}]\!]. M \sim N \implies U(\mathcal{M}) \leq U(\mathcal{N}) \qquad \text{(monotonicity under bisimilarity)}$$

Both properties of uncertainty functions use the uncertainty set as an indicator for the level of uncertainty. The first property, *certainty*, expresses that if the uncertainty set contains only one element, there is no uncertainty, so the uncertainty function should return zero.

The second property, monotonicity under bisimilarity, is about comparing the uncertainty in different uncertain Markov chains. One way doing this would be to look at uncertain Markov chains where the uncertainty set of one is included in the uncertainty set of the other. Then the uncertainty of the first would be less than or equal to the uncertainty of the second. However, when quantifying uncertainty, we are interested not in the exact representation of Markov chains in the uncertainty set but in their behaviour. For this reason, we generalize this notion, by considering bisimilar Markov chains as equal. This results in the property of monotonicity under bisimilarity.

Note that even when $[\![\mathcal{M}]\!]$ is a strict subset of $[\![\mathcal{N}]\!]$, with $[\![\mathcal{N}]\!]$ containing Markov chains that are not bisimilar to any Markov chain in $[\![\mathcal{M}]\!]$, $U(\mathcal{N})$ does not have to be strictly larger than $U(\mathcal{M})$. This is because the Markov chains that are only in the larger uncertainty set might not be meaningfully different from the Markov chains that are in both uncertainty sets. What meaningfully different means in this context can depend on the type of uncertainty that the uncertainty function expresses. For example, two Markov chains might be considered equivalent if they have the same expected value. Another example is the trivial uncertainty function: $\forall \mathcal{M} \in \text{UMC}\,.U(\mathcal{M}) = 0$. Thus, a larger uncertainty set does not necessarily mean that there is meaningfully more uncertainty. The reason we use bisimilarity and not expected discounted value in the definition of uncertainty functions, is that we want to uncertainty functions to express something about the behaviour of uncertain Markov chains.

From these properties, we can also derive that if all Markov chains in the uncertainty set are bisimilar to each other, the uncertainty function must return 0. This makes sense, because it means that there is no uncertainty about the behaviour of the uncertain Markov chain.

Lemma 3.2. *If U is an uncertainty function, and $\mathcal{M}$ an uncertain Markov chain such that for all Markov chains $M, N \in [\![\mathcal{M}]\!]$ we have $M \sim N$, then $U(\mathcal{M}) = 0$.*

Proof. We take any Markov chain $N \in [\![\mathcal{M}]\!]$, and look at the uncertain Markov chain $\mathcal{N}$ whose uncertainty set is $\{N\}$. Then, by certainty, $U(\mathcal{N}) = 0$. Furthermore, we know that for all $M \in [\![\mathcal{M}]\!]$, $N \sim M$. By monotonicity under bisimilarity, it follows that $U(\mathcal{M}) \leq U(\mathcal{N}) = 0$. □

We can alternatively characterize monotonicity under bisimilarity as a combination of two other properties: monotonicity and what we call stability under bisimilarity. To do this, we will first give an auxiliary definition that lifts the notion of bisimilarity to sets of Markov chains.

Definition 3.3 (equality up to bisimilarity)**.** We call two sets of Markov chains S, T *equal up to bisimilarity* if both the following statements hold.

$$\forall M \in S\ \exists N \in T. \quad M \sim N$$
$$\forall N \in T\ \exists M \in S. \quad M \sim N$$

We denote this $S \sim T$.

Lemma 3.4. *A function $U\colon$ UMC $\to \mathbb{R}$ is monotonic under bisimilarity if and only if the following two properties hold for all uncertain Markov chains $\mathcal{M}, \mathcal{N}$*

1. *Monotonicity:*
$$[\![\mathcal{M}]\!] \subseteq [\![\mathcal{N}]\!] \implies U(\mathcal{M}) \leq U(\mathcal{N})$$

2. *Stability under bisimilarity:*
$$[\![\mathcal{M}]\!] \sim [\![\mathcal{N}]\!] \implies U(\mathcal{M}) = U(\mathcal{N})$$

Proof. First we will show that these properties follow from monotonicity under bisimilarity. If $[\![\mathcal{M}]\!] \subseteq [\![\mathcal{N}]\!]$, then it follows that for each $M \in [\![\mathcal{M}]\!]$, we have $M \in [\![\mathcal{N}]\!]$, such that $M \sim M$. As a result, by monotonicity under bisimilarity, it follows that $U(\mathcal{M}) \leq U(\mathcal{N})$.

If $[\![\mathcal{M}]\!] \sim [\![\mathcal{N}]\!]$, then by monotonicity under bisimilarity, we have both $U(\mathcal{M}) \leq U(\mathcal{N})$ and $U(\mathcal{N}) \leq U(\mathcal{M})$. It follows that $U(\mathcal{M}) = U(\mathcal{N})$.

Now we will show that stability under bisimilarity and monotonicity together imply monotonicity under bisimilarity. Suppose that we have uMCs $\mathcal{M}, \mathcal{N}$ such that

$$\forall M \in [\![\mathcal{M}]\!]\ \exists N \in [\![\mathcal{N}]\!]. M \sim N.$$

We will construct uncertain Markov chains $\mathcal{M}'$ and $\mathcal{N}'$, such that

$$[\![\mathcal{M}]\!] \sim [\![\mathcal{M}']\!]$$
$$[\![\mathcal{N}]\!] \sim [\![\mathcal{N}']\!]$$
$$[\![\mathcal{M}']\!] \subseteq [\![\mathcal{N}']\!].$$

If these properties hold, then by stability under bisimilarity and monotonicity, we have $U(\mathcal{M}) = U(\mathcal{M}') \leq U(\mathcal{N}') = U(\mathcal{N})$.

The idea is that both $\mathcal{M}'$ and $\mathcal{N}'$ will have all states of $\mathcal{M}$ and $\mathcal{N}$ apart from $s_{init}^{\mathcal{M}}$. For $\mathcal{M}'$, the initial state will have the transitions of $s_{init}^{\mathcal{M}}$, while for $\mathcal{N}'$, it will have the transitions of both $s_{init}^{\mathcal{M}}$ and $s_{init}^{\mathcal{N}}$ as possible transitions.

More formally, we define $\mathcal{M}'$ by changing $\mathcal{M}$ in two ways. Firstly we add all states of $\mathcal{N}$, without any transitions to them, so they are unreachable. Secondly, we replace the initial state with $s_{init}^{\mathcal{N}}$, and move all the transitions that first went to $s_{init}^{\mathcal{M}}$ to $s_{init}^{\mathcal{N}}$ instead. To do this, we introduce a function $\phi : S \to S$ that maps $s_{init}^{\mathcal{N}}$ to s_{init}^{M}.

$$\phi(s) = \begin{cases} s_{init}^{\mathcal{M}} & \text{if } s = s_{init}^{\mathcal{N}} \\ s & \text{otherwise} \end{cases} \tag{2}$$

If we apply this before some probability distribution $P \in \mathcal{P}^{\mathcal{M}}_s$ for some s, this has the desired effect. More formally, $\mathcal{M}' = (S, s^{\mathcal{N}}_{init}, \mathcal{P}^{\mathcal{M}'}, R)$, where

$$S = S^{\mathcal{M}} \setminus \{s^{\mathcal{M}}_{init}\} \uplus S^{\mathcal{N}}$$

$$R(s) = \begin{cases} R^{\mathcal{M}}(s) & \text{if } s \in S^{\mathcal{M}} \\ R^{\mathcal{N}}(s) & \text{if } s \in S^{\mathcal{N}} \end{cases}.$$

$$\mathcal{P}^{\mathcal{M}'}_s = \begin{cases} \{P \circ \phi \mid P \in \mathcal{P}^{\mathcal{M}}_s\} & \text{if } s \in S^{\mathcal{M}} \\ \mathcal{P}^{\mathcal{N}}_s & \text{if } s \in S^{\mathcal{N}} \setminus \{s^{\mathcal{N}}_{init}\} \\ \{P \circ \phi \mid P \in \mathcal{P}^{\mathcal{M}}_{s^{\mathcal{M}}_{init}}\} & \text{if } s = s^{\mathcal{N}}_{init} \end{cases}$$

Note that since each $M \in [\![\mathcal{M}]\!]$ is bisimilar to some $N \in [\![\mathcal{N}]\!]$, it follows that the reward at the initial state must be equal, i.e. $R^{\mathcal{M}}(s^{\mathcal{M}}_{init}) = R^{\mathcal{N}}(s^{\mathcal{N}}_{init})$. With this, it follows that $[\![\mathcal{M}]\!] \sim [\![\mathcal{M}']\!]$.

Since we want $[\![\mathcal{M}']\!] \subseteq [\![\mathcal{N}']\!]$, we want both uMCs to have the same states, rewards and initial state. The transition function of $\mathcal{N}'$ is the same as that of $\mathcal{M}'$, except for the initial state. The transition set for the initial state is the union of two sets: the transitions of the initial state in $\mathcal{M}'$ and the transitions of the initial state in $\mathcal{N}$. More formally, we have $\mathcal{N}' = (S, s^{\mathcal{N}}_{init}, \mathcal{P}^{\mathcal{N}'}, R)$, where

$$\mathcal{P}^{\mathcal{N}'}_s = \begin{cases} \{P \circ \phi \mid P \in \mathcal{P}^{\mathcal{M}}_s\} & \text{if } s \in S^{\mathcal{M}} \\ \mathcal{P}^{\mathcal{N}}_s & \text{if } s \in S^{\mathcal{N}} \setminus \{s^{\mathcal{N}}_{init}\} \\ \mathcal{P}^{\mathcal{M}}_{s^{\mathcal{N}}_{init}} \cup \{P \circ \phi \mid P \in \mathcal{P}^{\mathcal{M}}_{s^{\mathcal{M}}_{init}}\} & \text{if } s = s^{\mathcal{M}}_{init} \end{cases}$$

From the definitions, we can see that $\mathcal{P}^{\mathcal{M}'} \subseteq \mathcal{P}^{\mathcal{N}'}$, and as a result, $[\![\mathcal{M}']\!] \subseteq [\![\mathcal{N}']\!]$.

It remains to show that $[\![\mathcal{N}]\!] \sim [\![\mathcal{N}']\!]$. Let $N \in [\![\mathcal{N}]\!]$. To construct $N' \in [\![\mathcal{N}']\!]$, we need to pick transition probability distributions for each state in $s \in S$. For $s \in S^{\mathcal{N}}$, we choose $P^{N'}_s = P^{N}_s$. The transitions for the other states can then be arbitrarily chosen, since those states cannot be reached. Then it follows that N and N' are bisimilar.

Now let $N' \in [\![\mathcal{N}']\!]$. There are two cases here:

1. $P^{N'}_{s^{N}_{init}} = P_0 \circ \phi$ for some $P_0 \in \mathcal{P}^{\mathcal{M}}_{s^{\mathcal{M}}_{init}}$ In this case, we know that $N' \in [\![\mathcal{M}']\!]$ as well. That means it is bisimilar to some $M' \in [\![\mathcal{M}]\!]$. By our assumption, this M' must be bisimilar to some $N \in [\![\mathcal{N}]\!]$. As a result, we have $N' \sim M' \sim N$.

2. $P^{N'}_{s^{N}_{init}} \in \mathcal{P}^{\mathcal{N}}_{s^{\mathcal{N}}_{init}}$: Then the states in $S^{\mathcal{M}}$ are unreachable. We can construct an $N \in [\![\mathcal{N}]\!]$ by choosing the same transition functions as in N'. Then it is clear that N is bisimilar to N'.

□

Note that monotonicity is a necessary, but not sufficient condition to be monotonic under bisimilarity. For instance, take the following function.

$$U(\mathcal{M}) = \begin{cases} 0 & \text{if } |[\![\mathcal{M}]\!]| = 1 \\ 1 & \text{otherwise} \end{cases}$$

This clearly fulfils certainty and monotonicity. Now we consider the uncertain Markov chains $\mathcal{M}, \mathcal{N}$ from Figure 3. Then the uncertainty sets of $\mathcal{M}$ and $\mathcal{N}$ are equivalent up to bisimilarity, since they both give zero reward at each transition. However, $U(\mathcal{M}) = 1 \neq 0 = U(\mathcal{N})$. As a result, it is neither stable under bisimilarity, nor monotonic under bisimilarity.

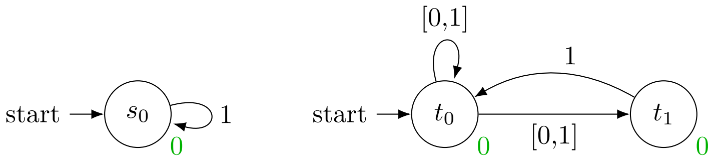

Figure 3: Two uncertain Markov chains $\mathcal{M}$ (left) and $\mathcal{N}$ (right).

4 Distances on Markov chains

In this section we will introduce a method to construct an uncertainty function out of a distance function on Markov chains. We do this by looking at the maximum distance between two elements of the uncertainty set of the uncertain Markov chain. For this construction to work, we need the distance function to be a bounded pseudometric that assigns a distance of 0 to Markov chains that are bisimilar.

Lemma 4.1. *Let $d\colon MC \times MC \to \mathbb{R}$ be a bounded pseudometric, such that*

$$\forall M, N \in MC.M \sim N \implies d(M, N) = 0. \tag{3}$$

Then the function

$$U_d(\mathcal{M}) = \sup_{M,N \in [\![\mathcal{M}]\!]} d(M, N), \tag{4}$$

is an uncertainty function.

Proof. Note that the supremum in this definition is defined because the pseudometric has an upper bound, and uncertain Markov chains have a non-empty uncertainty set by definition, so we are not taking the supremum of an empty set.

If there is only one Markov chain M in the uncertainty set of some uncertain Markov chain $\mathcal{M}$, then the induced uncertainty function works out to be

$$U_d(\mathcal{M}) = d(M, M) = 0$$

To show that this definition satisfies monotonicity under bisimilarity, let $\mathcal{M}, \mathcal{N}$ be two uncertain Markov chains, such that $\forall M \in [\![\mathcal{M}]\!] \exists N \in [\![\mathcal{N}]\!].M \sim N$. Firstly, we note that for Markov chains M, N, K, if $M \sim N$, then $d(M, N) = 0$ by assumption (3). Then we have

$$d(M, K) \leq d(M, N) + d(N, K) = d(N, K)$$

$$d(N, K) \leq d(N, M) + d(M, K) = d(M, K)$$

It follows that $d(M, K) = d(N, K)$. Now consider a pair $M_0, M_1 \in [\![\mathcal{M}]\!]$. By our assumption, we have N_0, N_1 in $[\![\mathcal{N}]\!]$ such that $M_0 \sim N_0$ and $M_1 \sim N_1$. Then we have

$$d(M_0, M_1) = d(N_0, M_1) = d(N_0, N_1)$$

This means that the distances between Markov chains in the uncertainty set of $\mathcal{M}$ are a subset of the distances between Markov chains in the uncertainty set of $\mathcal{N}$.

$$\{d(M_0, M_1) \mid M_0, M_1 \in [\![\mathcal{M}]\!]\} \subseteq \{d(N_0, N_1) \mid N_0, N_1 \in [\![\mathcal{N}]\!]\}$$

As a result, we have

$$U_d(\mathcal{M}) = \sup_{M_0, M_1 \in [\![\mathcal{M}]\!]} d(M_0, M_1) \leq \sup_{N_0, N_1 \in [\![\mathcal{N}]\!]} d(N_0, N_1) = U_d(\mathcal{N}).$$

□

We call the function from (4) the *induced uncertainty function* of d.

In the rest of this section, we will examine four different pseudometrics on Markov chains, proving that they are pseudometrics and giving examples to show how they differ.

4.1 Bisimilarity metric

One metric on Markov chains is the bisimilarity metric. It is a quantitative extension of the bisimilarity relation. That is, the distance between two Markov chains is zero if and only if they are bisimilar, and if we change two bisimilar Markov chains a little bit, the resulting Markov chains will be assigned a small distance by the bisimilarity metric.

This metric is an adapted version of the bisimilarity metric for MDPs, defined in [9, 10, 11]. As a helper function, we define $\theta_\gamma(s,t) = (1-\gamma)|R(s) - R(t)|$.

Definition 4.2 (Bisimilarity metric)**.** Given two Markov chains M, N, and a discount factor $\gamma \in (0,1)$, define $\rho^*\colon S^M \times S^N \to \mathbb{R}$ as limit of the sequence ρ_n, where $\rho_0(s,t) = 0$ and for $n > 0$,

$$\rho_n(s,t) = \theta_\gamma(s,t) + \gamma\mathcal{K}(\rho_{n-1})(P_s, P_t).$$

Then the *bisimilarity metric* between M and N is $d_B(M,N) = \rho^*(s^M_{init}, s^N_{init})$.

Another way of characterizing the bisimilarity metric is as the fixpoint of an equation.

Lemma 4.3. *The bisimilarity metric is the unique fixpoint of* $\Phi\colon [0,1]^{S^M \times S^N} \to [0,1]^{S^M \times S^N}$, *defined as*

$$\Phi(\rho)(s,t) = \theta_\gamma(s,t) + \gamma\mathcal{K}(\rho)(P_s, P_t). \tag{5}$$

Proof. First of all, we need to show that Φ is well-defined. That is, for any function $\rho\colon S^M \times S^N \to [0,1]$, and any states $s \in S^M$ and $t \in S^N$, $\Phi(\rho)(s,t) \in [0,1]$. We note that $\Phi(\rho)(s,t)$ must be positive, because all values of ρ are positive. For the upper bound, we note that $|R(s) - R(t)| \leq 1$. With this, we have

$$\Phi(\rho)(s,t) = \theta_\gamma(s,t) + \gamma\mathcal{K}(\rho)(P_s, P_t) \leq (1-\gamma) + \gamma\mathcal{K}(\rho)(P_s, P_t)$$

Since the Kantorovich metric is a weighted sum of values of ρ, with the weights summing to 1, we can conclude that $\mathcal{K}(\rho)(P_s, P_t) \leq 1$. Then if follows that $\Phi(\rho(s,t) \leq 1$.

We can prove that Φ has a unique fixpoint equal to the bisimilarity metric by using the Banach fixed point theorem (Theorem 2.9). As a closed subset of $\mathbb{R}^n$, for some $n \in \mathbb{N}$, $[0,1]^{S^M \times S^N}$ is a complete metric space. What remains is to show that Φ is a contraction mapping. That is, for all $\rho_1, \rho_2\colon S^M \times S^N \to \mathbb{R}$, and some $q \in [0,1)$, we have to prove

$$\|\Phi(\rho_1) - \Phi(\rho_2)\|_\infty \leq \delta\|\rho_1 - \rho_2\|_\infty.$$

We take $q = \gamma$. Writing out the definition of Φ, we get

$$\begin{aligned}
&\|\Phi(\rho_1) - \Phi(\rho_2)\|_\infty \\
=& \min_{(s,t)\in S^M \times S^N} |\Phi(\rho_1) - \Phi(\rho_2)| \\
=& \min_{(s,t)\in S^M \times S^N} |\theta_\gamma(s,t) + \gamma\mathcal{K}(\rho_1)(P_s, P_t) - \theta_\gamma(s,t) - \gamma\mathcal{K}(\rho_2)(P_s, P_t)| \\
=&\gamma \min_{(s,t)\in S^M \times S^N} |\mathcal{K}(\rho_1)(P_s, P_t) - \mathcal{K}(\rho_2)(P_s, P_t)|.
\end{aligned}$$

Now, we fix the states s, t that minimize this expression. Without loss of generality, we can assume that $\mathcal{K}(\rho_1)(P_s, P_t) \geq \mathcal{K}(\rho_2)(P_s, P_t)$, so that the expression in the absolute value bars is non-negative. Let λ_i be the coupling between P_s and P_t that minimizes the sum $\sum_{s',t'} \lambda_i(s',t')\rho_i(s',t')$, for $i \in \{1, 2\}$. Applying all this, we get

$$\gamma \min_{(s,t)\in S^M \times S^N} |\mathcal{K}(\rho_1)(P_s, P_t) - \mathcal{K}(\rho_2)(P_s, P_t)| = \gamma \left(\sum_{s',t'} \lambda_1(s',t')\rho_1(s',t') - \sum_{s',t'} \lambda_2(s',t')\rho_2(s',t') \right).$$

Since λ_1 minimizes this sum, replacing it by λ_2 will increase the sum and thus the whole expression.

$$\begin{aligned}
&\gamma \left(\sum_{s',t'} \lambda_1(s',t')\rho_1(s',t') - \sum_{s',t'} \lambda_2(s',t')\rho_2(s',t') \right) \\
&\leq \gamma \left(\sum_{s',t'} \lambda_2(s',t')\rho_1(s',t') - \sum_{s',t'} \lambda_2(s',t')\rho_2(s',t') \right) \\
&= \gamma \sum_{s',t'} \lambda_2(s',t') \left(\rho_1(s',t') - \rho_2(s',t')\right) \\
&\leq \gamma \sum_{s',t'} \lambda_2(s',t') |\rho_1(s',t') - \rho_2(s',t')|
\end{aligned}$$

Note that the last part is an absolute difference between function values of ρ_1 and ρ_2. By definition, this must be less than or equal to $\|\rho_1 - \rho_2\|_\infty$.

$$\gamma \sum_{s',t'} \lambda_2(s',t') |\rho_1(s',t') - \rho_2(s',t')| \leq \gamma \sum_{s',t'} \lambda_2(s',t') \|\rho_1 - \rho_2\|_\infty = \gamma \|\rho_1 - \rho_2\|_\infty$$

With this, we have shown that Φ is a contraction mapping, and thus that it has a unique fixed point. In addition, this fixed point is the limit $\lim_{n\to\infty} \Phi^n(\rho)$ for any function $\rho\colon S^M \times S^N \to [0, 1]$. The sequence ρ_n from Definition 4.2 is exactly equal to this sequence if we take the zero function ρ_0 for the initial function ρ: $\rho_n = \Phi^n(\rho_0)$. Then it follows that the limit ρ^* is equal to the unique fixed point of Φ. □

Now we can show that this is a pseudometric. This can also be found in [10, 11].

Lemma 4.4. *For any Markov chain, ρ^* forms a bounded pseudometric on states, and the bisimilarity metric d_B forms a bounded pseudometric on Markov chains as a result.*

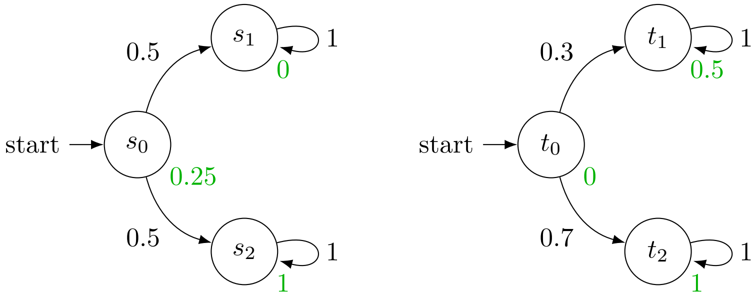

Figure 4: Two Markov chains M (left) and N (right).

Proof. We will first show that ρ^* is a pseudometric. By Lemma 2.3 (c), it suffices to show that ρ_n is a pseudometric for all n. We show this by induction on n. For the base case, for any states s, t, u, we have

$$\rho_0(s, s) = 0 \tag{6}$$

$$\rho_0(s, t) = 0 = \rho_0(t, s) \tag{7}$$

$$\rho_0(s, u) = 0 \leq 0 + 0 = \rho_0(s, t) + \rho_0(t, u) \tag{8}$$

For $n > 0$, we know that ρ_{n-1} is a pseudometric by the induction hypothesis. By Lemma 2.5 and 2.3 (a), it follows that $s, t \mapsto \mathcal{K}(\rho_{n-1})(P_s, P_t)$ is also a pseudometric. In addition, θ_γ is a pseudometric. Finally, by Lemma 2.3 (b), it follows that ρ_n is a pseudometric.

By 2.3 (a), it now follows that the bisimilarity metric is a bounded pseudometric as well. The alternative definition of ρ^* in Lemma 4.3 shows that its function values lie within $[0, 1]$, so ρ^* and the bisimilarity metric are bounded. □

To illustrate the bisimilarity metric, we can look at some simple examples. If M and N have the same reward for every transition, a and b respectively, then $|R(s) - R(t)| = |a - b|$ for any pair of states s, t from M and N. Then the smallest fixpoint ρ^* is

$$\forall s, t. \rho^*(s, t) = \sum_{i=0}^{\infty} \gamma^i (1 - \gamma)|a - b| = |a - b|.$$

So in this case, the bisimilarity metric is

$$d_B(M, N) = |a - b|.$$

For a more complex example, consider the Markov chains in Figure 4. To find the bisimilarity metric, we first need to calculate the ρ^* function. We start with the states s_1, s_2, t_1, t_2.

$$\begin{aligned}
\rho^*(s_1, t_1) &= \sum_{i=0}^{\infty} \gamma^i (1-\gamma)|0 - 0.5| = 0.5 \\
\rho^*(s_1, t_2) &= \sum_{i=0}^{\infty} \gamma^i (1-\gamma)|0 - 1| = 1 \\
\rho^*(s_2, t_1) &= \sum_{i=0}^{\infty} \gamma^i (1-\gamma)|1 - 0.5| = 0.5 \\
\rho^*(s_2, t_2) &= \sum_{i=0}^{\infty} \gamma^i (1-\gamma)|1 - 1| = 0
\end{aligned}$$

Then we can calculate the ρ^* function for the initial states. The coupling that minimizes the distance is

$$\begin{aligned}
\lambda(s_1, t_1) &= 0.3 \\
\lambda(s_1, t_2) &= 0.2 \\
\lambda(s_2, t_2) &= 0.5
\end{aligned}$$

Then we can now calculate $\rho^*(s_0, t_0)$, which is the bisimilarity metric.

$$d_B(M, N) = (1-\gamma)0.25 + \gamma(0.3 \cdot 0.5 + 0.2 \cdot 1 + 0.5 \cdot 0) = 0.25 + 0.1\gamma$$

4.2 Trace metric

Instead of executing two Markov chains in parallel, like for bisimilarity, we can look at the distribution of traces of two Markov chains. This leads us to the trace metric. To define this new distance on Markov chains, we first define a distance on infinite traces of states.

Definition 4.5 (Trace metric)**.** Given two Markov chains M and N, define $F \colon S_M^\omega \times S_N^\omega \to \mathbb{R}$ by

$$F(\mathbf{s}, \mathbf{t}) = \sum_{i=0}^{\infty} \gamma^i \theta_\gamma(s_i, t_i)$$

Note that the infinite sum is well-defined since $\theta_\gamma(s, t)$ is bounded. Then the *trace metric* between M and N is defined as

$$d_T(M, N) = \mathcal{K}(F)(\mathrm{Pr}^M, \mathrm{Pr}^N)$$

where $\mathcal{K}(F)$ is the Kantorovich distance over F. Recall that Pr^M is the distribution over traces induced by the Markov chain M.

Now we can show that this is a pseudometric.

Lemma 4.6. *F is a pseudometric on traces, and d_T is a bounded pseudometric on Markov chains.*

Proof. The Kantorovich operator lifts a bounded pseudometric on a set to a bounded pseudometric on distributions over that set. As a result, we merely need to show that F is a bounded pseudometric on traces. Because $R(s) \in [0, 1]$ for any state s, F is bounded by $\sum_{i=0}^{\infty} \gamma^i (1-\gamma) = 1$.

Now it remains to show that F is a pseudometric. First of all, for any trace $\mathbf{s}$, we have

$$F(\mathbf{s},\mathbf{s}) = \sum_{i=0}^{\infty} \gamma^i \theta_\gamma(s_i, s_i) = \sum_{i=0}^{\infty} \gamma^i (1-\gamma) \cdot 0 = 0.$$

Symmetry follows from the symmetry of θ_γ. For all traces $\mathbf{s}, \mathbf{t}$, we have

$$F(\mathbf{s},\mathbf{t}) = \sum_{i=0}^{\infty} \gamma^i \theta_\gamma(s_i, t_i) = \sum_{i=0}^{\infty} \gamma^i \theta_\gamma(t_i, s_i) = F(\mathbf{t},\mathbf{s}).$$

And finally, the triangle inequality follows from the triangle inequality on absolute differences. For all traces $\mathbf{s}, \mathbf{t}, \mathbf{u}$, we have

$$\begin{aligned} F(\mathbf{s},\mathbf{t}) + F(\mathbf{t},\mathbf{u}) &= \sum_{i=0}^{\infty} \gamma^i \theta_\gamma(s_i, t_i) + \sum_{i=0}^{\infty} \gamma^i \theta_\gamma(t_i, u_i) \\ &= \sum_{i=0}^{\infty} \gamma^i (1-\gamma)(|R(s_i) - R(t_i)| + |R(t_i) - R(u_i)|) \\ &\geq \sum_{i=0}^{\infty} \gamma^i (1-\gamma)|R(s_i) - R(u_i)| = \sum_{i=0}^{\infty} \gamma^i \theta_\gamma(s_i, u_i) = F(\mathbf{s},\mathbf{u}). \end{aligned}$$

□

Now we can look at some examples to show that the trace metric is different from the bisimilarity metric. The core idea is that for the trace metric, it does not matter at which step two traces start to differ, only what the probability of each trace is, and the reward at each step. For the bisimilarity metric, this does matter. If we look at the uncertain Markov chains in Figure 5, we can see that they both have identical traces. As such, they are identical for the trace metric. However, the bisimilarity metric does differentiate between these uncertain Markov chains. This is because the bisimilarity metric effectively executes the Markov chains in parallel, and the split between the two traces happens at a different time step in each Markov chain.

Lemma 4.7. *There exist Markov chains M, N such that*

$$d_B(M, N) > d_T(M, N).$$

Proof. Two such Markov chains are shown in Figure 5. For M, there are two traces, both with probability 0.5: $s_0 s_1 s_2 \ldots$ and $s_0 s_3 s_4 \ldots$. The rewards per step for these are $0, 0, 0, \ldots$ and $0, 0, 1, \ldots$ respectively. For N, there are also two traces with probability 0.5: $t_0 t_1 t_2 \ldots$ and $t_0 t_1 t_3 \ldots$. Their rewards per step are also $0, 0, 0, \ldots$ and $0, 0, 1, \ldots$ respectively. To calculate the trace metric, we first calculate the distances between the traces.

$$\begin{aligned} F(s_0 s_1 s_2 \ldots, t_0 t_1 t_2 \ldots) &= 0 \\ F(s_0 s_1 s_2 \ldots, t_0 t_1 t_3 \ldots) &= \sum_{i=2}^{\infty} \gamma^i (1-\gamma) = \gamma^2 \\ F(s_0 s_3 s_4 \ldots, t_0 t_1 t_2 \ldots) &= \sum_{i=2}^{\infty} \gamma^i (1-\gamma) = \gamma^2 \\ F(s_0 s_3 s_4 \ldots, t_0 t_1 t_3 \ldots) &= 0 \end{aligned}$$

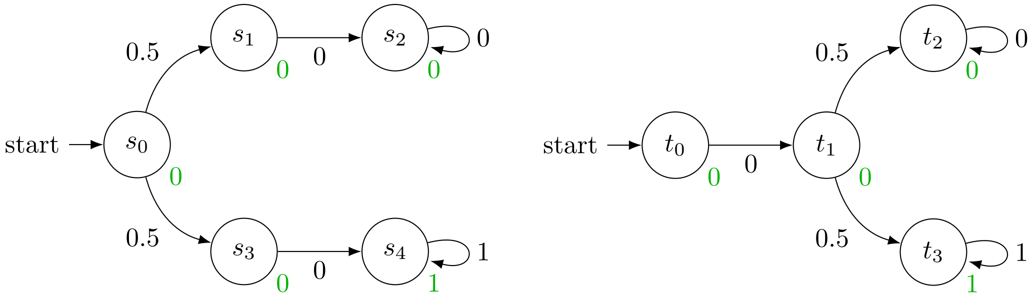

Figure 5: Two Markov chains M (left) and N (right).

From this, we can see that the coupling that minimizes the trace metric is

$$\lambda(s_0 s_1 s_2 \ldots, t_0 t_1 t_2 \ldots) = 0.5$$
$$\lambda(s_0 s_3 s_4 \ldots, t_0 t_1 t_3 \ldots) = 0.5.$$

The trace metric is then

$$d_T(M, N) = 0.5F(s_0 s_1 s_2 \ldots, t_0 t_1 t_2 \ldots) + 0.5F(s_0 s_3 s_4 \ldots, t_0 t_1 t_3 \ldots) = 0.$$

For the bisimilarity metric, the relevant parts of ρ^* function are as follows:

$$\rho^*(s_2, t_2) = \rho^*(s_4, t_3) = 0$$
$$\rho^*(s_2, t_3) = \rho^*(s_4, t_2) = 1$$
$$\rho^*(s_1, t_1) = \rho^*(s_3, t_1) = 0.5\gamma$$
$$\rho^*(s_0, t_0) = 0.5\gamma^2$$

So we have $d_B(M, N) = 0.5\gamma^2 > 0 = d_T(M, N)$. □

4.3 Expected value metric

Another way of defining at the distance between Markov chains is to look at the difference in the expected discounted reward.

Definition 4.8 (Expected value metric)**.** The *expected value metric* is defined as

$$d_V(M, N) = (1 - \gamma)|V(M) - V(N)|$$

Lemma 4.9. *The expected value metric is a bounded pseudometric on Markov chains.*

Proof. The absolute difference is a pseudometric on $\mathbb{R}$, so by Lemma 2.3 (a), it follows that the expected value metric is a pseudometric.

Boundedness follows from the fact that the expected value is bounded by $\sum_{i=0}^{\infty} \gamma^i = \frac{1}{1-\gamma}$. Then the expected value metric is bounded by 1. □

To illustrate the expected value metric, we can look at the Markov chains in Figure 4. The value in s_1 is 0, and the value in s_2 is $\frac{1}{1-\gamma}$. Then the value of M is

$$V(M) = 0.25 + \gamma(0.5V(s_1) + 0.5V(s_2)) = 0.25 + \gamma\left(0.5 \cdot 0 + 0.5 \cdot \frac{1}{1-\gamma}\right) = 0.25 + \frac{0.5\gamma}{1-\gamma}$$

For N, the value in t_1 is $\frac{0.5}{1-\gamma}$, and the value in t_2 is $\frac{1}{1-\gamma}$. The expected value of N is

$$V(N) = 0 + \gamma\left(0.3 \cdot V(t_1) + 0.7 \cdot V(t_2)\right) = \gamma\left(0.3 \cdot \frac{0.5}{1-\gamma} + 0.7 \cdot \frac{1}{1-\gamma}\right) = \frac{0.85\gamma}{1-\gamma}$$

Then the expected value metric is

$$\begin{aligned} d_V(M,N) &= (1-\gamma)|V(M) - V(N)| \\ &= (1-\gamma)\left|0.25 + \frac{0.5\gamma}{1-\gamma} - \frac{0.85\gamma}{1-\gamma}\right| \\ &= |0.25(1-\gamma) + 0.5\gamma - 0.85\gamma| \\ &= |0.25 - 0.6\gamma| \end{aligned}$$

4.4 Value distribution metric

Instead of looking at the average value, we can look at the value distribution.

Definition 4.10 (Value distribution metric)**.** Define

$$G(\mathbf{s}, \mathbf{t}) = (1-\gamma)|V(\mathbf{s}) - V(\mathbf{t})|.$$

The *value distribution metric* is defined as

$$d_D(M,N) = \mathcal{K}(G)(\Pr^M, \Pr^N).$$

Lemma 4.11. *The value distribution metric is a bounded pseudometric.*

Proof. The absolute difference is a metric on real numbers, and the discounted value is bounded by $\sum_{i=0}^{\infty} \gamma^i = \frac{1}{1-\gamma}$. As a result, G is a bounded pseudometric on traces. Because the Kantorovich operation preserves pseudometrics and boundedness, it follows that d_D is a bounded pseudometric. □

The value distribution metric distinguishes more Markov chains than the expected value metric, because it looks at the distribution of the possible values, and not just their average. We will illustrate this with the following example.

Lemma 4.12. *There exists Markov chains M, N such that*

$$d_V(M,N) < d_D(M,N)$$

Proof. Two such Markov chains are shown in Figure 6. The Markov chain M is deterministic, and always gives a discounted value of $\frac{0.5}{1-\gamma}$. The Markov chain N has two possible traces, both with

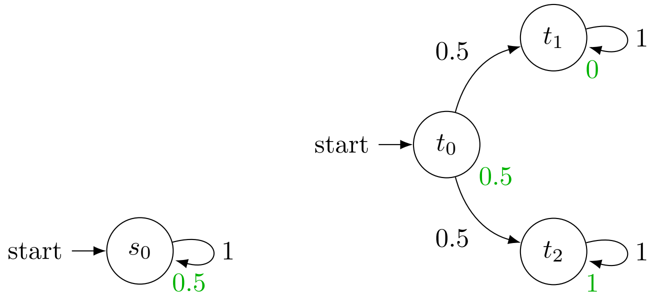

Figure 6: Two Markov chains M (top) and N (bottom).

probability 0.5: $t_0 t_1 \ldots$ and $t_0 t_2 \ldots$. They have values of 0.5 and $0.5 + \frac{\gamma}{1-\gamma}$ respectively. That means that $V(N) = 0.5 + \frac{0.5\gamma}{1-\gamma} = \frac{0.5}{1-\gamma}$. Since the values of M and N are the same,

$$d_V(M, N) = (1-\gamma)|V(M) - V(N)| = (1-\gamma)\left|\frac{0.5}{1-\gamma} - \frac{0.5}{1-\gamma}\right| = 0$$

For the value distribution metric, we note that there is only one possible coupling between $\Pr^M$ and $\Pr^N$, with $\lambda(\mathbf{s}, \mathbf{t}) = 0.5$ for either trace $\mathbf{t}$ in N. Then we have

$$\begin{aligned}
d_D(M, N) &= \mathcal{K}(G)(\Pr^M, \Pr^N) \\
&= 0.5G(s_0 \ldots, t_0 t_1 \ldots) + 0.5G(s_0 \ldots, t_0 t_2 \ldots) \\
&= 0.5(1-\gamma)|V(s_0 \ldots) - V(t_0 t_1 \ldots)| + 0.5(1-\gamma)|V(s_0 \ldots) - V(t_0 t_2 \ldots)| \\
&= 0.5(1-\gamma)\left|\frac{0.5}{1-\gamma} - 0.5\right| + 0.5(1-\gamma)\left|\frac{0.5}{1-\gamma} - 0.5 - \frac{\gamma}{1-\gamma}\right| \\
&= 0.25\gamma + 0.25\gamma \\
&= 0.5\gamma > 0 = d_V(M, N).
\end{aligned}$$

□

The value distribution metric also differs from the trace metric. Specifically, the value distribution metric looks at the total discounted reward for each trace, while the trace metric also looks at which specific step the reward is obtained.

Lemma 4.13. *There exists Markov chains M, N such that*

$$d_D(M, N) < d_T(M, N)$$

Proof. We will look at the two Markov chains in Figure 7. Both of them are deterministic, so they have only one trace, which we will call $\mathbf{s}$ and $\mathbf{t}$ respectively. This means there is only one coupling between these distributions, with $\lambda(\mathbf{s}, \mathbf{t}) = 1$. As a result, the trace and value distribution metrics work out to be $F(\mathbf{s}, \mathbf{t})$ and $G(\mathbf{s}, \mathbf{t})$ respectively. For the trace metric, this means that

$$d_T(M, N) = F(\mathbf{s}, \mathbf{t}) = \sum_{i=0}^{\infty} \gamma^i (1-\gamma)|R(s_i), R(t_i)| = \sum_{i=0}^{\infty} \gamma^i (1-\gamma) 0.5 = 0.5.$$

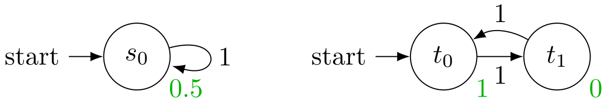

Figure 7: Two Markov chains M (left) and N (right).

For the value distribution metric, we have

$$d_D(M,N) = G(\mathbf{s},\mathbf{t}) = (1-\gamma)|V(\mathbf{s}) - V(\mathbf{t})| = (1-\gamma)\left|\frac{0.5}{1-\gamma} - \frac{1}{1-\gamma^2}\right| = \left|0.5 - \frac{1}{1+\gamma}\right|$$

Now it remains to show that $0.5 > |0.5 - \frac{1}{1+\gamma}|$. We have

$$\left|0.5 - \frac{1}{1+\gamma}\right| = \left|\frac{0.5\gamma - 0.5}{1+\gamma}\right| = \frac{0.5 - 0.5\gamma}{1+\gamma} < 0.5.$$

□

5 Comparison of metrics

We have introduced four pseudometrics on Markov chains: the expected value metric d_V, the value distribution metric d_D, the trace metric d_T, and the bisimilarity metric d_B. In the previous section we have already seen some examples showing that they are different. In this section we will show that there exists a strict ordering on these four pseudometrics.

Theorem 5.1. *The following inequalities hold:*

$$d_V < d_D < d_T < d_B$$

To show this, we will prove the three inequalities in order. The first inequality that we will prove is that the trace metric is less than or equal to the bisimilarity metric. The idea is to construct a coupling between the distribution of traces out of the couplings for each state pair between the transition distributions. This gives us an upper bound for the trace metric, that we can prove is equal to the bisimilarity metric.

Lemma 5.2.

$$d_T \leq d_B$$

Proof. Let M, N be uncertain Markov chains. For each pair $(s,t) \in S^M \times S^N$, let $\lambda_{s,t} \in \Lambda(P_s, P_t)$ be a coupling that minimizes

$$\sum_{s',t'} \lambda_{s,t}(s,t)\rho^*(s',t').$$

Note that this is the minimizing coupling of $\mathcal{K}(\rho^*)(P_s,P_t)$, so it exists by Lemma 2.6. Then we can define a coupling

$$\lambda^*(\mathrm{Cyl}(s_0 \dots s_n) \times \mathrm{Cyl}(t_0 \dots t_n)) = \begin{cases} \prod_{i=0}^{n-1} \lambda_{s_i,t_i}(s_{i+1},t_{i+1}) & \text{if } (s_0,t_0) = (s^M_{init}, s^N_{init}) \\ 0 & \text{otherwise} \end{cases}. \qquad (9)$$

First we need to show that this is indeed a coupling between Pr^M and Pr^N. Let $s_0 \dots s_n$ be some finite trace of states. Then we have

$$\lambda^*(\mathrm{Cyl}(s_0 \dots s_n) \times S^M) = \sum_{t_0 \dots t_n \in S_N^{n+1}} \lambda^*(\mathrm{Cyl}(s_0 \dots s_n) \times \mathrm{Cyl}(t_0 \dots t_n)).$$

If $s_0 \neq s^M_{init}$, then this is 0, just like $\mathrm{Pr}^M(s_0 \dots s_n)$. So we will now look at the case $s_0 = s^M_{init}$. Also, we fix $t_0 = s^N_{init}$ in the sum, since all other terms are 0. Then we can write out the definition of λ^*.

$$\begin{aligned} &\sum_{t_1 \dots t_n \in S_N^n} \lambda^*(\mathrm{Cyl}(s_0 \dots s_n) \times \mathrm{Cyl}(t_0 \dots t_n)) \\ &= \sum_{t_1 \dots t_n \in S_N^n} \prod_{i=0}^{n-1} \lambda_{s_i,t_i}(s_{i+1},t_{i+1}) \\ &= \sum_{t_1 \in S_N} \lambda_{s_0,t_0}(s_1,t_1) \cdots \sum_{t_{n-1} \in S_N} \lambda_{s_{n-2},t_{n-2}}(s_{n-1},t_{n-1}) \sum_{t_n \in S_N} \lambda_{s_{n-1},t_{n-1}}(s_n,t_n) \end{aligned}$$

Now, because λ_{s_i,t_i} is a coupling between P_{s_i} and P_{t_i}, we have $\sum_{t_{i+1} \in S_N} \lambda_{s_i,t_i}(s_{i+1}, t_{i+1}) = P_{s_i}(s_{i+1})$, so we can rewrite this to:

$$\begin{aligned}
&\sum_{t_1 \in S_N} \lambda_{s_0,t_0}(s_1, t_1) \cdots \sum_{t_{n-1} \in S_N} \lambda_{s_{n-2},t_{n-2}}(s_{n-1}, t_{n-1}) \sum_{t_n \in S_N} \lambda_{s_{n-1},t_{n-1}}(s_n, t_n) \\
&= \sum_{t_1 \in S_N} \lambda_{s_0,t_0}(s_1, t_1) \cdots \sum_{t_{n-1} \in S_N} \lambda_{s_{n-2},t_{n-2}}(s_{n-1}, t_{n-1}) P_{s_{n-1}}(s_n) \\
&= \prod_{i=0}^{n-1} P_{s_i}(s_{i+1}) \\
&= \mathrm{Pr}^M(\mathrm{Cyl}(s_0 \dots s_n))
\end{aligned}$$

This shows that λ^* is indeed a coupling between Pr^M and Pr^N. Using this, we now have an upper bound for the trace metric:

$$d_T(M, N) = \mathcal{K}(F)(\mathrm{Pr}^M, \mathrm{Pr}^N) = \min_{\lambda \in \Lambda(\mathrm{Pr}^M, \mathrm{Pr}^N)} \int_{S^\omega \times S^\omega} F(\mathbf{s}, \mathbf{t}) \lambda(d\mathbf{s}, d\mathbf{t}) \leq \int_{S^\omega \times S^\omega} F(\mathbf{s}, \mathbf{t}) \lambda^*(d\mathbf{s}, d\mathbf{t})$$

Writing out the definition of F, we get

$$\int_{S^\omega \times S^\omega} F(\mathbf{s}, \mathbf{t}) \lambda^*(d\mathbf{s}, d\mathbf{t}) = \int_{S^\omega \times S^\omega} \sum_{n=0}^{\infty} \gamma^n \theta_\gamma(s_n, t_n) \lambda^*(d\mathbf{s}, d\mathbf{t}).$$

In this sum, all the terms are positive. Also, λ^* is a probability metric, so it is σ-finite. By Tonelli's theorem (252G and 252H in [14]), we can then swap the integral and sum:

$$\int_{S^\omega \times S^\omega} \sum_{n=0}^{\infty} \gamma^n \theta_\gamma(s_n, t_n) \lambda^*(d\mathbf{s}, d\mathbf{t}) = \sum_{n=0}^{\infty} \gamma^n \int_{S^\omega \times S^\omega} \theta_\gamma(s_n, t_n) \lambda^*(d\mathbf{s}, d\mathbf{t})$$

Since the part within the integral only depends on the states s_n and t_n, we can simplify this by rewriting the integral to a sum over the finitely many sequences of length $n + 1$. If $s_0 \neq s^M_{init}$ or $t_0 \neq s^N_{init}$, then $\lambda^*(\mathrm{Cyl}(s_0 \dots s_n), \mathrm{Cyl}(t_0 \dots t_n)) = 0$, so we can restrict the sum to sequences starting with s^M_{init} and s^N_{init}. We write $\mathrm{Seq}^{n+1}(s) = \{s_0 \dots s_n \in S^{n+1} \mid s_0 = s\}$.

$$\begin{aligned}
&\sum_{n=0}^{\infty} \gamma^n \int_{S^\omega \times S^\omega} \theta_\gamma(s_n, t_n) \lambda^*(d\mathbf{s}, d\mathbf{t}) \\
&= \sum_{n=0}^{\infty} \gamma^n \sum_{\substack{s_0 \dots s_n \in \mathrm{Seq}^{n+1}(s^M_{init}) \\ t_0 \dots t_n \in \mathrm{Seq}^{n+1}(s^N_{init})}} \theta_\gamma(s_n, t_n) \lambda^*(\mathrm{Cyl}(s_0 \dots s_n) \times \mathrm{Cyl}(t_0 \dots t_n)) \\
&= \sum_{n=0}^{\infty} \gamma^n \sum_{\substack{s_0 \dots s_n \in \mathrm{Seq}^{n+1}(s^M_{init}) \\ t_0 \dots t_n \in \mathrm{Seq}^{n+1}(s^N_{init})}} \theta_\gamma(s_n, t_n) \prod_{i=0}^{n-1} \lambda_{s_i,t_i}(s_{i+1}, t_{i+1})
\end{aligned}$$

Now we can define the sequence of finite sums that tends to this infinite sum.

$$\delta_k(s, t) = \sum_{n=0}^{k-1} \gamma^n \sum_{\substack{s_0 \dots s_n \in \mathrm{Seq}^{n+1}(s) \\ t_0 \dots t_n \in \mathrm{Seq}^{n+1}(t)}} \theta_\gamma(s_n, t_n) \prod_{i=0}^{n-1} \lambda_{s_i,t_i}(s_{i+1}, t_{i+1})$$

Then we have now shown that

$$\int_{S^\omega \times S^\omega} F(\mathbf{s}, \mathbf{t}) \lambda^*(d\mathbf{s}, d\mathbf{t}) = \lim_{k \to \infty} \delta_k(s^M_{init}, s^N_{init}).$$

This limit is equal to the bisimilarity metric. To show this, we will introduce an alternative equation for the bisimilarity metric. Consider the following equation.

$$\tilde{\rho}(s,t) = \theta_\gamma(s,t) + \gamma \sum_{s',t'} \lambda_{s,t}(s',t') \tilde{\rho}(s',t')$$

This equation has ρ^* as a fixed point. Analogously to Φ from Lemma 4.3, this is a contraction mapping on $[0,1]^{S^M \times S^N}$. As a result, by Banach fixed point theorem (Theorem 2.9), ρ^* is also the limit of the following sequence:

$$\begin{aligned} \tilde{\rho}_0(s,t) &= 0 \\ \tilde{\rho}_{k+1}(s,t) &= \theta_\gamma(s,t) + \gamma \sum_{s',t'} \lambda_{s,t}(s',t') \tilde{\rho}_k(s',t') \end{aligned}$$

Now, we will show by induction that $\tilde{\rho}_k$ is equal to δ_k. The base case is trivial:

$$\tilde{\rho}_0(s,t) = 0 = \delta_0(s,t)$$

For the induction step, we first apply the induction hypothesis:

$$\begin{aligned} &\tilde{\rho}_{k+1}(s,t) \\ &= \theta_\gamma(s,t) + \gamma \sum_{s',t' \in S} \lambda_{s,t}(s',t') \tilde{\rho}_k(s',t') \\ &= \theta_\gamma(s,t) + \gamma \sum_{s',t' \in S} \lambda_{s,t}(s',t') \delta_k(s',t') \\ &= \theta_\gamma(s,t) + \gamma \sum_{s',t' \in S} \lambda_{s,t}(s',t') \sum_{n=0}^{k-1} \gamma^n \sum_{\substack{s_0 \dots s_n \in \mathrm{Seq}^{n+1}(s') \\ t_0 \dots t_n \in \mathrm{Seq}^{n+1}(t')}} \theta_\gamma(s_n,t_n) \prod_{i=0}^{n-1} \lambda_{s_i,t_i}(s_{i+1},t_{i+1}) \end{aligned}$$

We can incorporate the sum over s', t' in the sum over the state sequences. Then s, t become the new s_0, t_0, so $\lambda_{s,t}(s',t') = \lambda_{s_0,t_0}(s_1,t_1)$. This becomes the new term for $i = 0$ of the product.

$$\begin{aligned} &\theta_\gamma(s,t) + \gamma \sum_{s',t' \in S} \lambda_{s,t}(s',t') \sum_{n=0}^{k-1} \gamma^n \sum_{\substack{s_0 \dots s_n \in \mathrm{Seq}^{n+1}(s') \\ t_0 \dots t_n \in \mathrm{Seq}^{n+1}(t')}} \theta_\gamma(s_n,t_n) \prod_{i=0}^{n-1} \lambda_{s_i,t_i}(s_{i+1},t_{i+1}) \\ &= \theta_\gamma(s,t) + \sum_{n=0}^{k-1} \gamma^{n+1} \sum_{\substack{s_0 \dots s_{n+1} \in \mathrm{Seq}^{n+2}(s) \\ t_0 \dots t_{n+1} \in \mathrm{Seq}^{n+2}(t)}} \theta_\gamma(s_{n+1},t_{n+1}) \prod_{i=0}^{n} \lambda_{s_i,t_i}(s_{i+1},t_{i+1}) \end{aligned}$$

We can now renumber the sum to start from $n = 1$. Then the missing term $n = 0$ is exactly equal to $|R(s) - R(t)|$.

$$\begin{aligned}
&\theta_\gamma(s,t) + \sum_{n=0}^{k-1} \gamma^{n+1} \sum_{\substack{s_0 \ldots s_{n+1} \in \mathrm{Seq}^{n+2}(s) \\ t_0 \ldots t_{n+1} \in \mathrm{Seq}^{n+2}(t)}} \theta_\gamma(s_{n+1}, t_{n+1}) \prod_{i=0}^{n} \lambda_{s_i,t_i}(s_{i+1}, t_{i+1}) \\
&= \theta_\gamma(s,t) + \sum_{n=1}^{k} \gamma^{n} \sum_{\substack{s_0 \ldots s_{n} \in \mathrm{Seq}^{n+1}(s) \\ t_0 \ldots t_{n} \in \mathrm{Seq}^{n+1}(t)}} \theta_\gamma(s_{n}, t_{n}) \prod_{i=0}^{n-1} \lambda_{s_i,t_i}(s_{i+1}, t_{i+1}) \\
&= \sum_{n=0}^{k} \gamma^{n} \sum_{\substack{s_0 \ldots s_{n} \in \mathrm{Seq}^{n+1}(s) \\ t_0 \ldots t_{n} \in \mathrm{Seq}^{n+1}(t)}} \theta_\gamma(s_{n}, t_{n}) \prod_{i=0}^{n-1} \lambda_{s_i,t_i}(s_{i+1}, t_{i+1}) \\
&= \delta_{k+1}(s,t)
\end{aligned}$$

We finish the proof by concluding that

$$\begin{aligned}
d_T(M,N) = \mathcal{K}(F)(\mathrm{Pr}^M, \mathrm{Pr}^N) \leq \int_{S^\omega \times S^\omega} &F(\mathbf{s},\mathbf{t}) \lambda^*(d\mathbf{s}, d\mathbf{t}) \\
&= \lim_{k\to\infty} \delta_k(s^M_{init}, s^N_{init}) = \lim_{k\to\infty} \tilde{\rho}_k(s^M_{init}, s^N_{init}) = d_B(M,N).
\end{aligned}$$

□

Now we can easily show that the value distribution metric is less than or equal to the trace metric by moving the absolute value inside the sum.

Lemma 5.3.

$$d_D \leq d_T$$

Proof. For two traces $\mathbf{s}, \mathbf{t}$, we have

$$\begin{aligned}
G(\mathbf{s},\mathbf{t}) = (1-\gamma)|V(\mathbf{s}) - V(\mathbf{t})| = (1-\gamma)\left|\sum_{i=0}^{\infty} \gamma^i R(s_i) - \sum_{i=0}^{\infty} \gamma^i R(t_i)\right| &= (1-\gamma)\left|\sum_{i=0}^{\infty} \gamma^i (R(s_i) - R(t_i))\right| \\
&\leq \sum_{i=0}^{\infty} \gamma^i \theta_\gamma(s_i, t_i) = F(\mathbf{s},\mathbf{t}).
\end{aligned}$$

Because the Kantorovich operator is monotonic, this proves that $d_D \leq d_T$. □

For the expected value metric and the value distribution metric, we can prove the inequality by moving the absolute value through the integral.

Lemma 5.4.

$$d_V \leq d_D$$

Proof. Let M, N be Markov chains. Then

$$d_D(M,N) = \mathcal{K}(G)(\mathrm{Pr}^M, \mathrm{Pr}^N) = \min_{\lambda \in \Lambda(\mathrm{Pr}^M, \mathrm{Pr}^N)} \int_{S^\omega \times S^\omega} G(\mathbf{s}, \mathbf{t}) \lambda(d\mathbf{s}, d\mathbf{t})$$

We let λ_{min} be the coupling that minimizes this expression. Then we can write

$$\begin{aligned} d_D(M,N) &= \int_{S^\omega \times S^\omega} G(\mathbf{s}, \mathbf{t}) \lambda_{min}(d\mathbf{s}, d\mathbf{t}) \\ &= \int_{S^\omega \times S^\omega} (1-\gamma)|V(\mathbf{s}) - V(\mathbf{t})| \lambda_{min}(d\mathbf{s}, d\mathbf{t}) \\ &= (1-\gamma) \int_{S^\omega \times S^\omega} |V(\mathbf{s}) - V(\mathbf{t})| \lambda_{min}(d\mathbf{s}, d\mathbf{t}) \\ &\geq (1-\gamma) \left| \int_{S^\omega \times S^\omega} V(\mathbf{s}) - V(\mathbf{t}) \lambda_{min}(d\mathbf{s}, d\mathbf{t}) \right|. \end{aligned}$$

V^M and V^N are measurable functions, since they are the limit of the sequence $\sum_{i=0}^{n} \gamma^i R(s_i)$, and each of those functions is measurable. According to Section 1.1.1 from [31], we can then split the integral into

$$(1-\gamma)\left| \int_{S^\omega} V(\mathbf{s}) \, \mathrm{Pr}^M(d\mathbf{s}) - \int_{S^\omega} V(\mathbf{t}) \, \mathrm{Pr}^N(d\mathbf{t}) \right| = (1-\gamma)|V(M) - V(N)| = d_V(M,N)$$

□

Proof of the theorem. The inequalities are shown in the Lemmas 5.2, 5.3 and 5.4. In addition, we know that the metrics are not equal by Lemmas 4.7, 4.12 and 4.13 in the previous section. As a result, we can conclude that

$$d_V < d_D < d_T < d_B.$$

□

Using this, we can now show that all four pseudometrics fulfil the assumptions of Lemma 4.1. As a result, for each of these pseudometrics, the maximal distance between elements of the uncertainty set forms an uncertainty function.

Corollary 5.5. *For each of the four pseudometrics d_V, d_D, d_T, d_B,*

$$U_d(\mathcal{M}) = \sup_{M,N \in [\![\mathcal{M}]\!]} d(M,N)$$

is an uncertainty function.

Proof. Let M, N be two bisimilar Markov chains. Then by the construction of the bisimilarity metric, we have $d_B(M,N) = 0$. As a result of Theorem 5.1, we also have

$$d_V(M,N) = d_D(M,N) = d_T(M,N) = d_B(M,N) = 0.$$

In addition, we have shown that each of these functions are bounded pseudometrics in Lemmas 4.4, 4.6, 4.9 and 4.11. By Lemma 4.1, it follows that for each of these pseudometrics, U_d is an uncertainty function. □

6 Computation of the induced uncertainty functions

In the previous section, we have determined that the four pseudometrics can be used to construct induced uncertainty functions of the form $U_d(\mathcal{M}) = \sup_{M,N \in [\![\mathcal{M}]\!]} d(M, N)$. In this section we will look at algorithms for calculating each of the metrics, as well as their induced uncertainty functions.

6.1 Expected value metric

The induced uncertainty function of the expected value metric is defined by

$$U_{d_V}(\mathcal{M}) = \sup_{M,N \in [\![\mathcal{M}]\!]} (1-\gamma)|V(M) - V(N)|.$$

This largest difference in expected reward can be calculated by finding MDPs in the uncertainty set with the largest and smallest possible expected reward. These are equal to the optimistic and pessimistic value of the uncertain Markov chain, respectively. Thus, the induced uncertainty function is equal to $(1-\gamma)\left(\overline{V}(\mathcal{M}) - \underline{V}(\mathcal{M})\right)$.

Computing the optimistic or pessimistic value of an uncertain Markov chain is a specific case of computing those values for uncertain MDPs, which add actions to Markov chains. Computing the optimistic and pessimistic, or robust, values of MDPs is a well-studied problem, for example see [24, 18, 33, 17] Assuming s-rectangularity and convex local uncertainty sets, the complexity to compute the optimistic or pessimistic value of an uncertain Markov chain with a state set S to an accuracy of ε is $\mathcal{O}(|S|^2 \log(\frac{1}{\varepsilon})^2)$ [24].

6.2 Bisimilarity metric

To compute the bisimilarity metric d_B on Markov chains, we can look at them as the special case of MDPs where there is only a single action. For an MDP with state set S and action set A, and a discount factor γ, we can compute the bisimilarity metric to a degree of accuracy $\varepsilon \in \mathbb{R}$ in $O(|A||S|^4 \log|S| \frac{\ln \varepsilon}{\ln \gamma})$ operations [10, 11]. Applying this to Markov chains, we get a complexity of $O(|S|^4 \log|S| \frac{\ln \varepsilon}{\ln \gamma})$. In [13], a comparison is made between different calculations of the bisimilarity metric.

Additionally, Ferns and Precup [12] showed that the bisimilarity metric between two MDPs is equal to the optimal value of an MDP. This transformation could be used in combination with algorithms for computing the value of an MDP to find the bisimilarity distance between MDPs, and thus Markov chains.

The induced uncertainty function U_{d_B} is harder to compute. The problem is that we need to look at a possibly infinite number of pairs of Markov chains. One way we could try to handle this by only looking at the extreme points of the uncertainty set in terms of the transition probabilities. This is inspired by linear programming, where an optimum can be found at an extreme point of the set defined by the constraints [20]. We can define extreme points of an uncertainty set as follows:

Definition 6.1. Let $\mathcal{M}$ be an uncertain Markov chain. A Markov chain $M \in [\![\mathcal{M}]\!]$ *lies in between* Markov chains $N_0, N_1 \in [\![\mathcal{M}]\!]$ if there exists an $r \in [0,1]$, such that for any states s, t, $P_s^M(t) = rP_s^{N_0}(t) + (1-r)P_s^{N_1}(t)$. A Markov chain $M \in [\![\mathcal{M}]\!]$ is an extreme point of $[\![\mathcal{M}]\!]$ if and only if it does not lie between two other points in $[\![\mathcal{M}]\!]$.

We will first look at an example in which this method would indeed work. The details of this calculation can be found in Appendix A.

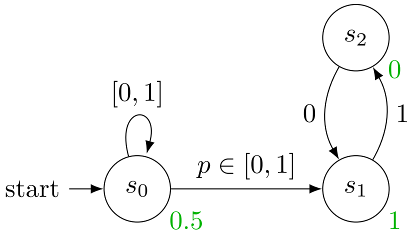

Figure 8: An uncertain Markov chain $\mathcal{M}$.

Example 6.2. As an example, we will compute the induced uncertainty function of the bisimilarity metric on the uMC $\mathcal{M}$ in Figure 8. To determine the bisimilarity uncertainty of $\mathcal{M}$, we will compute the bisimilarity metric on two Markov chains $M, N \in [\![\mathcal{M}]\!]$. Let p and q be the probabilities of the transition from s_0 to s_1 in M and N respectively. For s_1 and s_2, we will omit the superscript, because s_1^M is bisimilar to s_1^N, and s_2^M to s_2^N. To find $d_B(M, N)$, we will compute ρ^* using Lemma 4.3.

We can find the distance between s_1 and s_2 by simply writing out the definition: $\rho^*(s_1, s_2) = 1$. The distances $\rho^*(s_0^M, s_1)$ and $\rho^*(s_0^M, s_2)$ depend on each other. The first works out to be

$$\rho^*(s_0^M, s_1) = \frac{\gamma p}{1 - \gamma^2(1-p)^2} + \frac{(1-\gamma)0.5}{1 - \gamma(1-p)}.$$

Now we can calculate $\rho^*(s_0^M, s_0^N)$. Assuming that $p \leq q$, the coupling λ between $P_{s_0^M}$ and $P_{s_0^N}$ that minimizes the sum $\sum_{s,t} \rho^*(s,t)\lambda(s,t)$ is as follows:

$$\begin{aligned}
\lambda(s_0^M, s_0^N) &= 1 - q \\
\lambda(s_0^M, s_1) &= q - p \\
\lambda(s_1, s_0^N) &= 0 \\
\lambda(s_1, s_1) &= p
\end{aligned}$$

With this, we can work out the distance $\rho^*(s_0^M, s_0^N)$.

$$\rho^*(s_0^M, s_0^N) = \frac{\gamma^2(q-p)p}{(1-\gamma^2(1-p)^2)(1-\gamma(1-q))} + \frac{0.5(1-\gamma)\gamma(q-p)}{(1-\gamma(1-p))(1-\gamma(1-q))}$$

Using the partial derivatives, we can determine that the maximum value is achieved with $p = 0$ and $q = 1$. These are the extreme points in the uncertainty set. If we fill in these values, we get

$$U_{d_B}(\mathcal{M}) = 0.5\gamma.$$

In this example, the maximum distance between two Markov chains in the uMC was equal to the distance between the two extreme points of $[\![\mathcal{M}]\!]$, where $p = 0$ and $p = 1$ respectively. However, as we will show next, this is not always the case.

6.3 Extreme points are insufficient to compute the bisimilarity uncertainty function

In this section we will show that the Markov chains in an uncertainty set with the maximum bisimilarity distance are not always extreme points. An example of this is the uMC $\mathcal{M}$ in Figure 9.

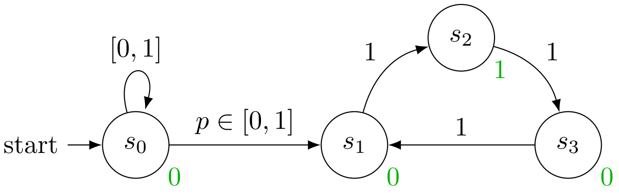

Figure 9: An uncertain Markov chain $\mathcal{M}$.

Intuitively, the idea is that for a high enough discount factor γ, the contribution of the eventual behaviour of the Markov chain to the bisimilarity metric will be much larger than the contribution of the first few cycles. Thus, we can look at the average behaviour in the long term for estimating bisimilarity metric. In addition, we will look at cases where at least one of the two Markov chains is deterministic. This means that the couplings in the bisimilarity metric are all trivial, and that its value is equal to the discounted average difference in reward. Combining these ideas, we estimate the bisimilarity metric with the eventual average difference in reward.

For $p = 1$, the reward will be 1 in every third step, while for $p = 0$, the reward is always zero. So between the extreme points $p = 0$ and $p = 1$, the average difference in reward is $\frac{1}{3}$.

However, if we look at the distance between $p = 1$ and some p between 0 and 1, the distance could be higher. This be because for $p \in (0, 1)$, we will eventually end up in the cycle s_1, s_2, s_3 with probability 1. This means a reward of 1 at every third time step. However, this may not be on the same time steps as in the case $p = 1$. Depending on the exact value of p, the distribution over the time steps in which we get reward may be different. For $p = 1$, the rewards will fall in the time steps $3k$ for $k \in \mathcal{N}$. For large values of $p < 1$, the probability of getting reward at time steps $\{3k \mid k \in \mathcal{N}\}$ will be much larger than the probability of getting reward at the time steps $\{3k + 1 \mid k \in \mathcal{N}\}$ or $\{3k + 2 \mid k \in \mathcal{N}\}$. However, for small values of p, the probabilities for all three will be approximately equal. If we assume them to be perfectly equal, the probability of receiving reward at time steps other than $\{3k \mid k \in \mathcal{N}\}$, and thus other than the case $p = 1$, is $\frac{2}{3}$. If this is the case, the difference in reward with the case $p = 1$ will be 1 in two out of every three time steps. As a result, the average difference in reward would be $\frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9} > \frac{1}{3}$. This indicates that for large enough γ, the largest distance will probably not be between the extreme points of the uncertainty set.

Now we can apply this same reasoning and calculation to the uncertain Markov chain we looked at in Example 6.2 (see Figure 8). In we compare the cases $p = 0$ and $p = 1$, we get a difference in reward of $\frac{1}{2}$ in every time step after the first. For the cases $p = 1$ and $p \in (0, 1)$, making the same assumptions as for the uMC in Figure 9, we also get an average difference in reward of $\frac{1}{2}$. This means that the Markov chains with higher difference in the initial states will have a higher bisimilarity distance. As a result, and as we have seen in the example, the maximum distance is achieved at the extreme points.

The argument above gives an informal intuition why, for the uncertain Markov chain from Figure 9, we might expect the bisimilarity metric between Markov chains within the uncertainty set $[\![\mathcal{M}]\!]$ than the bisimilarity metric between the extreme points of $[\![\mathcal{M}]\!]$. We will now prove this formally by calculating the distances.

Theorem 6.3. *There exists an uncertain Markov chain $\mathcal{M}$ for which the maximal bisimilarity* *distance between Markov chains in $[\![\mathcal{M}]\!]$ is not attained at the extreme points of the space of transition probabilities.*

Proof. To show that the uncertain Markov chain $\mathcal{M}$ from Figure 9 is such an uncertain Markov chain, we will calculate the distance between Markov chains in its uncertainty set. We let M_p denote the Markov chain in $[\![\mathcal{M}]\!]$, for which the probability of the transition from s_0 to s_1 is p. Now, we will compute the bisimilarity metric on M_p and M_q for $p, q \in [0,1]$. We will start by calculating ρ^* on s_1, s_2 and s_3.

$$\begin{aligned}
\rho^*(s_1, s_2) &= (1-\gamma) + \gamma\rho^*(s_2, s_3) \\
\rho^*(s_1, s_3) &= \gamma\rho^*(s_2, s_1) \\
\rho^*(s_2, s_3) &= (1-\gamma) + \gamma\rho^*(s_3, s_1)
\end{aligned}$$

Now we can substitute these in the formula for $\rho^*(s_1, s_2)$.

$$\rho^*(s_1, s_2) = (1-\gamma) + \gamma((1-\gamma) + \gamma^2\rho^*(s_1, s_2)) = 1 - \gamma^2 + \gamma^3\rho^*(s_1, s_2)$$

From this, we get both $\rho^*(s_1, s_2)$ and $\rho^*(s_1, s_3)$, which are the values we actually need.

$$\begin{aligned}
\rho^*(s_1, s_2) &= \frac{1-\gamma^2}{1-\gamma^3} \\
\rho^*(s_1, s_3) &= \gamma\frac{1-\gamma^2}{1-\gamma^3}
\end{aligned}$$

Next, we will look at the distance between s_0^p on the one hand and s_1, s_2 and s_3 on the other.

$$\begin{aligned}
\rho^*(s_0^p, s_1) &= (1-\gamma)|0-0| + \gamma\mathcal{K}(\rho^*)(P_{s_0^p}, P_{s_1}) = \gamma((1-p)\rho^*(s_0^p, s_2) + p\rho^*(s_1, s_2)) \\
\rho^*(s_0^p, s_2) &= (1-\gamma)|0-1| + \gamma\mathcal{K}(\rho^*)(P_{s_0^p}, P_{s_2}) = (1-\gamma) + \gamma((1-p)\rho^*(s_0^p, s_3) + p\rho^*(s_1, s_3)) \\
\rho^*(s_0^p, s_3) &= (1-\gamma)|0-0| + \gamma\mathcal{K}(\rho^*)(P_{s_0^p}, P_{s_3}) = \gamma(1-p)\rho^*(s_0^p, s_1)
\end{aligned}$$

Again, we can make substitutions to calculate this. We will make the substitutions in the formula for the value of $\rho(s_0^p, s_1)$.

$$\begin{aligned}
\rho^*(s_0^p, s_1) &= \gamma((1-p)\rho^*(s_0^p, s_2) + p\rho^*(s_1, s_2)) \\
&= \gamma\left((1-p)((1-\gamma) + \gamma((1-p)\rho^*(s_0^p, s_3) + p\rho^*(s_1, s_3))) + p\frac{1-\gamma^2}{1-\gamma^3}\right) \\
&= \gamma\left((1-p)\left((1-\gamma) + \gamma\left((1-p)\gamma(1-p)\rho^*(s_0^p, s_1) + p\gamma\frac{1-\gamma^2}{1-\gamma^3}\right)\right) + p\frac{1-\gamma^2}{1-\gamma^3}\right) \\
&= \gamma(1-p)(1-\gamma) + \gamma^3(1-p)^3\rho^*(s_0^p, s_1) + \gamma^3(1-p)p\frac{1-\gamma^2}{1-\gamma^3} + \gamma p\frac{1-\gamma^2}{1-\gamma^3}
\end{aligned}$$

From this, we get

$$\rho^*(s_0^p, s_1) = \frac{\gamma(1-p)(1-\gamma) + \gamma^3(1-p)p\frac{1-\gamma^2}{1-\gamma^3} + \gamma p\frac{1-\gamma^2}{1-\gamma^3}}{1-\gamma^3(1-p)^3}.$$

Now we can compute the crucial distance $\rho^*(s_0^p, s_0^q)$. We will assume that $p \leq q$. To do this, we will need to find a coupling λ between $P_{s_0^p}$ and $P_{s_0^q}$ that minimizes the weighted sum of distances

$\sum_{s,t} \rho^*(s,t)\lambda(s,t)$. The probability distribution $P_{s_0^p}$ consist of a probability p of the state s_1, and a probability $1-p$ of the state s_0^p. Just like in the previous example, we can maximize the probability weight on (s_1, s_1) by Lemma 2.7, and then the rest of the coupling follows.

$$\begin{aligned}\lambda(s_1, s_1) &= p\\ \lambda(s_0^p, s_1) &= q - p\\ \lambda(s_0^p, s_0^q) &= 1 - q\end{aligned}$$

With this coupling, we can calculate the distance.

$$\begin{aligned}\rho^*(s_0^p, s_0^q) &= (1-\gamma)|0-0| + \gamma\mathcal{K}(\rho^*)(P_{s_0^p}, P_{s_0^q})\\ &= \gamma\left((q-p)\rho^*(s_0^p, s_1) + (1-q)\rho^*(s_0^p, s_0^q)\right)\\ &= \gamma(q-p)\frac{\gamma(1-p)(1-\gamma) + \gamma^3(1-p)p\frac{1-\gamma^2}{1-\gamma^3} + \gamma p\frac{1-\gamma^2}{1-\gamma^3}}{1-\gamma^3(1-p)^3} + \gamma(1-q)\rho^*(s_0^p, s_0^q)\end{aligned}$$

By moving the $\rho^*(s_0^p, s_0^q)$ to the left-hand side, we get the distance.

$$\rho^*(s_0^p, s_0^q) = \frac{\gamma(q-p)}{1-\gamma(1-q)} \cdot \frac{\gamma(1-p)(1-\gamma) + \gamma^3(1-p)p\frac{1-\gamma^2}{1-\gamma^3} + \gamma p\frac{1-\gamma^2}{1-\gamma^3}}{1-\gamma^3(1-p)^3}$$

Now we can calculate the bisimilarity distance between M_0 and M_1.

$$d_B(M_0, M_1) = \frac{\gamma(1-0)}{1-\gamma(1-1)} \cdot \frac{\gamma(1-0)(1-\gamma) + \gamma^3(1-0)0\frac{1-\gamma^2}{1-\gamma^3} + \gamma 0\frac{1-\gamma^2}{1-\gamma^3}}{1-\gamma^3(1-0)^3} = \frac{\gamma^2(1-\gamma)}{1-\gamma^3}$$

If we take $\gamma = 0.9$, this works out to approximately 0.299. We can compare this to the distance between $M_{0.1}$ and M_1:

$$d_B(M_{0.1}, M_1) = \frac{\gamma(1-0.1)}{1-\gamma(1-1)} \cdot \frac{\gamma(1-0.1)(1-\gamma) + \gamma^3(1-0.1)0.1\frac{1-\gamma^2}{1-\gamma^3} + \gamma 0.1\frac{1-\gamma^2}{1-\gamma^3}}{1-\gamma^3(1-0.1)^3} \approx 0.329$$

This shows that the maximum distance is not always the distance between extreme points in the space of transition probabilities. □

6.4 Trace metric

The trace metric between Markov chains can be approximated by computing all possible traces and their probabilities up to a certain number of steps. Let the approximation to n steps of the distance F be F_n. Then the error of the approximation F_n is

$$|F(\mathbf{s},\mathbf{t}) - F_n(\mathbf{s},\mathbf{t})| = \sum_{i=n}^{\infty} \gamma^i \theta_\gamma(s_i, t_i) \leq (1-\gamma)\sum_{i=n}^{\infty} \gamma^i = \gamma^n.$$

Because the Kantorovich metric is a weighted average of distances, using this approximation to calculate the trace metric $d_T = \mathcal{K}(F)$ will also have an error of at most γ^n.

To get to an error of at most ε, n needs to be at least $\log_\gamma(\varepsilon)$. If we let B be the maximum number of possible successor state for a single state, this means there will be at most $B^{\log_\gamma(\varepsilon)}$ traces. Given probability distributions over T different traces, the Kantorovich metric can be computed in $\mathcal{O}(T^2 \log(T))$ steps as a linear program [25, 10]. If we take values of $\varepsilon = 0.01$ and $\gamma = 0.9$, then this becomes approximately $B^{87.4} \log(B^{43.7})$, which is not feasible.

To compute the induced uncertainty function, we need to find maximal distance between two Markov chains in the uncertainty set. This is difficult, because the uncertainty set can have infinitely many elements.

Just like for the bisimilarity metric, it is insufficient to examine the extreme points. From Theorem 6.3, we know that the points with the maximal bisimilarity distance are not always the extreme points in the uncertainty set. The same uncertain Markov chain, from Figure 9, also works to prove this for the trace metric. Because the $p = 1$ case in the example is deterministic, the value of the bisimilarity metric and the trace metric are identical in this case.

Corollary 6.4. *There exists an uncertain Markov chain $\mathcal{M}$ for which the maximal trace metric between Markov chains in $[\![\mathcal{M}]\!]$ is not attained at the extreme points of the space of transition probabilities.*

Proof. The uncertain Markov chain $\mathcal{M}$ in Figure 9 also works as a counterexample here. The relevant cases are $p = 0$ and $p = 1$ on one hand and $p = 0.1$ and $p = 1$ on the other. The case $p = 1$ is deterministic, so it has only one possible trace. Recall that the trace metric is defined as the Kantorovich metric over the function F of the probability distributions over traces. In the proof of Lemma 5.2, we have shown a that the integral of the values of F over the coupling λ^*, as defined in (9), is equal to the bisimilarity metric. Because there is only one trace for M_1, there is also only one coupling between the $\Pr^{M_1}$ and $\Pr^{M_0}$. As a result, λ^* must be the unique coupling between these probability distributions, and the trace metric and the bisimilarity metric must coincide in this case. The same is true for $\Pr^{M_1}$ and $\Pr^{M_{0.1}}$. Therefore, with a discount factor of $\gamma = 0.9$,

$$d_T(M_0, M_1) \approx 0.299 < 0.329 \approx d_T(M_{0.1}, M_1).$$

□

6.5 Value distribution metric

Like the trace metric, the value distribution metric can be approximated by calculating the traces up to a certain number of steps. Let G_n be the approximation of G to n steps, and let V_n be the total discounted value of a trace up to n steps. Then the error of this approximation is at most

$$\begin{aligned}
&|G(\mathbf{s},\mathbf{t}) - G_n(\mathbf{s},\mathbf{t})| \\
&= (1-\gamma)||V(\mathbf{s}) - V(\mathbf{t})| - |V_n(\mathbf{s}) - V_n(\mathbf{t})|| \\
&= (1-\gamma)\left|\left|R_n(\mathbf{s}) + \sum_{i=n}^{\infty}\gamma^i R(s_i) - \left(R_n(\mathbf{t}) + \sum_{i=n}^{\infty}\gamma^i R(t_i)\right)\right| - |R_n(\mathbf{s}) - R_n(\mathbf{t})|\right| \\
&\leq (1-\gamma)\left|\sum_{i=n}^{\infty}\gamma^i (R(s_i) - R(t_i))\right| \\
&\leq (1-\gamma)\sum_{i=n}^{\infty}\gamma^i = \gamma^n.
\end{aligned}$$

For the trace metric, it is necessary to keep track of each trace along with its probability. For the value distribution metric, however, we only need to keep track of the probability of being in a given state with a given amount of accumulated reward. This means that we can ignore the path taken to a state as long as the accumulated reward is the same.

If we round the values to a precision of δ at each step, we can further reduce the number of state-reward pairs to evaluate. This results in an error of at most $n\frac{\delta}{2}$ in the calculated values in the reward distributions. Because the distributions of both Markov chains can have this error in opposite directions, the total error from rounding will be at most $n\delta$.

Combining these two ideas, we can formulate an algorithm to approximate the value distribution metric. For each of the n time steps that we simulate, we will calculate a probability distribution H_i over pairs of a state and (approximate) accumulated reward (s, r_{acc}). The distribution before the first time step, H_0, will have probability 1 to be in the initial state with 0 accumulated reward. To calculate H_{i+1}, we initialize it at probability 0 for all pairs. Then we go through each of the pairs with positive probability in H_i. For the pair (s, r_{acc}), we calculate the new accumulated reward

$$r'_{acc} = r_{acc} + (1 - \gamma)\gamma^i R(s)$$

and round it to the closest multiple of δ. We then go through each possible transition from s. For a transition to a successor state s', we add $H_{i-1}(s, r_{acc}) \cdot P_s(s')$ to the probability for the pair (s', r'_{acc}) in H_i. To get a probability distribution of the (approximate) value, we add up the probabilities in H_n for the same accumulated reward in different states.

To calculate the value distribution metric between two Markov chains, we first calculate the value distributions of both Markov chains as described above. Then, we calculate the Kantorovich metric (with the absolute difference) on these distributions. To do this, it suffices to compute L_1-distance of cumulative histograms [26].

To find the complexity of this algorithm, we will go through the required steps. In each of the n time steps, we need to go through all states in S. For each state, there are at most $\frac{1}{\delta}$ possible values for the accumulated reward. For each of those, we need to go through at most B possible successor states, add the new total reward, and multiply the probability by the probability of this transition. All in all, this results in $\mathcal{O}(n|S|B\frac{1}{\delta})$ operations, where B is again the maximum number of possible successor states that a state can have.

If we calculate the value distributions in this way, there will be at most $\frac{1}{\delta}$ different total reward values at the end. Because these are distributions on $\mathbb{R}$, a one-dimensional space, this is an identical to the L_1-distance between the cumulative histograms [26]. As a result, the distance can be calculated in linear time. The total complexity for this step is thus $\mathcal{O}(\frac{1}{\delta})$.

Combining this, the total error will be $\varepsilon = \gamma^n + n\delta$. If we rewrite this, we have $\delta = \frac{\varepsilon - \gamma^n}{n}$. As a result, the complexity is $\mathcal{O}(|S|B\frac{n^2}{\varepsilon - \gamma^n})$. To further refine this result, it would be necessary to express n in terms of ε directly.

If we take values of $\varepsilon = 0.01$ and $\gamma = 0.9$, then minimizing $\frac{n}{\delta}$ gives values of $n = 57$ iterations and $\delta \approx 0.00013$. This means there would be approximately $880000 \cdot |S|B$ steps in the calculation of the value distribution metric.

To calculate the induced uncertainty function, it would be necessary to find the Markov chains in the uncertainty set at which the value distribution metric is maximal. I did not find an obvious method to do this. It also remains unclear whether this maximal distance is necessarily achieved at extreme points of the space of transition probabilities.

7 Discussion on the extension to MDPs

In this section, we will sketch possible future work about extending uncertainty functions to uncertain MDPs. This adds actions to the uncertain Markov chains that we have looked at in the previous sections. We will look at extending both the general definitions of uncertainty functions, and specific uncertainty functions. We will also look at some additional considerations that come with quantifying uncertainty in uncertain MDPs.

7.1 Markov decision processes

MDPs are defined similarly to Markov chains, but with an extra set of actions, and transition functions that are also indexed by actions.

Definition 7.1 (Markov Decision Process)**.** A *Markov Decision Process* (MDP) is a tuple $M = (S, A, s_{init}, (P_{s,a})_{(s,a)\in S\times A}, R)$, where S is a finite set of states, A is a finite set of actions, $s_{init} \in S$ is the initial state, $P_{s,a} \in \mathbb{P}(S)$ is the transition function, and $R\colon S \times A \to [0,1]$.

Uncertain MDPs are defined analogously to uncertain Markov chains.

Definition 7.2 (Uncertain MDP)**.** An *uncertain Markov Decision Process* (uncertain MDP) is a tuple $\mathcal{M} = (S, A, s_{init}, \mathcal{P}, R)$, where S is a finite set of states, A is a finite set of actions, $s_{init} \in S$ is the initial state, $\mathcal{P} \subseteq \mathbb{P}(S)^{S\times A}$ is a non-empty set of transition functions and $R\colon S \times A \to [0,1]$.

The class of all uncertain MDPs is denoted by UMDP. The uncertainty set of an uncertain MDP $\mathcal{M}$ is defined as

$$[\![\mathcal{M}]\!] = \{(S, A, s_{init}, P, R) \mid P \in \mathcal{P}^{\mathcal{M}}\}.$$

7.2 Uncertainty functions on MDPs

Uncertainty functions on MDPs can be defined analogously to uncertainty functions on Markov chains.

Definition 7.3 (Uncertainty function)**.** An *uncertainty function on uncertain MDPs* is a function $U\colon$ UMDP $\to \mathbb{R}$ such that for all uncertain MDPs $\mathcal{M}, \mathcal{N}$,

$$\begin{aligned} |[\![\mathcal{M}]\!]| = 1 &\implies U(\mathcal{M}) = 0 && \text{(certainty)} \\ [\![\mathcal{M}]\!] \subseteq [\![\mathcal{N}]\!] &\implies U(\mathcal{M}) \leq U(\mathcal{N}) && \text{(monotonicity under bisimilarity)} \end{aligned}$$

We expect that the second requirement, monotonicity under bisimilarity, can be proven to be equivalent to the combination of stability under bisimilarity and monotonicity. This could be done by adapting the proof for Lemma 3.4 to MDPs. However, a straightforward adaptation of this proof would only hold for certain types of uncertain MDPs. In the proof for Markov chains, we construct an uncertain Markov chain whose possible transitions for the initial state are the union of the transitions of the initial states of two different uncertain Markov chains. As a result, this proof does not work for functions that are only defined on interval Markov chains, because the union of two intervals is not necessarily an interval. If we apply a similar construction to uncertain MDPs, the resulting uncertain MDP is not necessarily s, a-rectangular, even if the two original uncertain MDPs are. That is, the possible transition probabilities of different actions in the same state may depend on each other. As a result, a straightforward adaption of this proof would not apply to functions that are only defined on s, a-rectangular uncertain MDPs.

For uncertainty on uncertain MDPs, there is an additional property that might sometimes be useful. Since the end goal of learning an uncertain MDP is often to find the actions that would optimize the reward gained, there will be certain actions that are not relevant. In this context, there are two questions that are most relevant: Which actions give the most reward, and how much reward they give. A quantification of uncertainty can help us understand whether we need more data to answer these questions. Note that for both of these questions, only certain state-action pairs are relevant: those which could be part of an optimal policy of some MDP in the uncertainty set. This leaves some actions for which it is trivial to see that they will always give less expected value than some other actions, which we might call *bad actions.* The uncertainty of the value of bad actions is not relevant to our questions, and thus it would be a useful property if an uncertainty function is not dependent on their value.

7.3 Lifting uncertainty functions

There are a few different approaches to lifting uncertainty functions on Markov chains to MDPs. For any uncertainty function, we can choose a particular policy, and compute the uncertainty function on the Markov chain induced by that policy. This can be used if there is one specific policy that we are interested in. For the bisimilarity metric and the value metric specifically, there are also analogues for MDPs that can be used to construct uncertainty functions.

7.3.1 Evaluating uncertainty on one policy

In a context where we are simultaneously learning the uncertain MDP and a policy for it, the current policy could be used to evaluate the uncertainty. A drawback of this is that when the policy changes, the new values of the uncertainty function may not be meaningfully comparable to the old values.

Another option is to always use the robust policy. However, this has the downside that the resulting function is not monotonic. This is because when the uncertainty is decreased, the robust policy might change from a low-uncertainty path to a high-uncertainty path. The value of the function then increases, while the uncertainty in the MDP decreased.

If we want to preserve the property of monotonicity, we can also choose to use the policy which maximizes the uncertainty. However, this comes with its own difficulties. Firstly, it might be difficult to compute the policy maximizing the uncertainty. Secondly, the maximizing policy might use bad actions, as described earlier. As a result, it could give an overly pessimistic view of the uncertainty.

7.3.2 Value uncertainty

We can also use two different policies for the two different MDPs we are comparing. For the expected value metric, a natural choice for the policies would be the robust and optimistic policies. That is, we can define the value uncertainty for MDPs as the difference in the value of the optimistic policy and the value of the robust policy. This would give us an upper bound on the value realized by the robust policy and the potential optimal policy in any given MDP in the uncertainty set.

To apply this same idea to other metrics, we need to be careful. If we simply try to look at the robust and optimistic policies with another metric, we may get a function that is not monotonic.

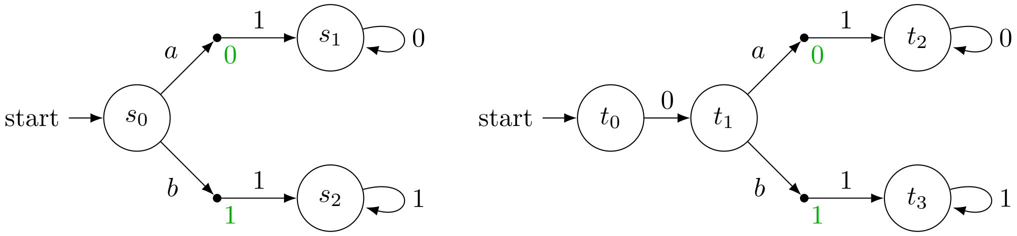

Figure 10: Two Markov decision processes M (left) and N (right).

In order to get a monotonic function, we need to choose the policies in such a way that the distance between the induced Markov chains always decreases when the uncertainty decreases. In addition, to ensure that the uncertainty function is well-defined, we need to ensure that the two chosen policies coincide when the uncertainty is zero. For the other metrics, it is not immediately clear how to do this.

7.3.3 Bisimilarity metric on MDPs

There is also a bisimilarity metric on MDPs [9, 10, 11]. This metric compares the rewards when taking the same action at each time step in both MDPs. It takes the actions that maximize the bisimilarity metric. This gives rise to a notion of bisimilarity uncertainty on MDPs.

Definition 7.4 (Bisimilarity uncertainty)**.** The *bisimilarity uncertainty* of an uncertain MDP $\mathcal{M}$ is defined as

$$U_B(\mathcal{M}) = \sup_{M,N \in [\![\mathcal{M}]\!]} d_B(M,N)$$

In the proof of Lemma 4.7, we have seen that the bisimilarity metric on Markov chains is sensitive to the timing of a split between two possible paths. Similarly, the bisimilarity metric on MDPs is very sensitive to the timing of certain choices in MDPs. Take the MDPs in Figure 10. The only difference between these two MDPs is whether the choice between the top and bottom path is in the first time step or in the second. As a result, their maximum rewards will differ by a factor of γ: $\frac{1}{1-\gamma}$ for M and $\frac{\gamma}{1-\gamma}$ for N. However, the bisimilarity metric between these two MDPs is 1, the maximal possible value. This is because taking action a in the first step and b in the second gives a reward of $\frac{1}{1-\gamma}$ for M, and a reward of 0 for N.

7.4 Action uncertainty

Another way of looking at uncertainty is to look at the uncertainty in terms of which actions could be optimal instead of in terms of value. The motivation for this is as follows: Imagine we would like to formulate a policy which optimizes for reward for an MDP, but we do not have direct access to its transition probabilities. Instead, we can obtain data about these probabilities from data collected by interacting with the system. This data is then used to construct an uncertain MDP, from which we can use the robust policy. In this context, we can use uncertainty functions to determine whether we need to collect more data.

For this purpose, the previous uncertainty functions have a weakness. It might be the case that the optimal policies are identical for each MDP in the uncertainty set, but that the value of those policies is still uncertain. To address this, an alternative is to measure the number of actions or policies which could be optimal. In other words, instead of quantifying the uncertainty of the value of the uncertain MDP, we can quantify the uncertainty of which actions or policies are optimal.

This could be approximated by taking all actions whose value in the optimistic case is greater than the value of the robust policy in that state. These are actions that, if taken, could provide more reward than the robust policy. To quantify how much uncertainty such states involve, we can use the difference between the optimistic value of an action, and the value of the robust policy in that state. This ensures that a small change in the possible transition probabilities only changes the function by a small amount. If there is only one action that can be optimal in a given state, there is no uncertainty in that state, so we subtract the difference between the robust and optimistic values of the state from the total. This gives us the following definition.

Definition 7.5. The action uncertainty of an uncertain MDP $\mathcal{M}$ is:

$$U_A(\mathcal{M}) = \sum_{s \in S} \left(\left(\sum_{a \in A} \mathit{potential}(s, a) \right) - \left(\overline{V}(s) - \underline{V}(s) \right) \right)$$

where the potential of an action is $\mathit{potential}(s, a) = \max(\overline{Q}^*(s, a) - \underline{V}(s), 0)$

There are several drawbacks to this method. Firstly, in this function all states are weighed equally. However, the probability of reaching certain states is much higher than the probability of reaching other states. In the worst case, this could mean that the uncertainty is primarily in states that are never reached by any optimal policy of any MDP in the uncertainty set.

Secondly, the values of this uncertainty function are heavily dependent the number of states and actions. This means that comparing the uncertainty of uncertain MDPs of different sizes in not meaningful.

Thirdly, while the goal of action uncertainty is to indicate the uncertainty of which actions are optimal, it is still possible for action uncertainty to be positive, while the optimal actions are identical in all MDPs in the uncertainty set. In this case, only the value of these actions would still be uncertain.

8 Related Work

8.1 Other metrics on Markov chains and MDPs

In this thesis, we defined uncertainty functions based on various distances on Markov chains: the bisimilarity metric, the trace metric, the value distribution metric and the expected value metric. In the last section, we also sketched how this could be done for MDPs. Now we will look at some other metrics on Markov chains and MDPs from the literature.

An alternative distance on MDPs that is more scalable is MICo (Matching under Independent Couplings) [7]. This distance function is similar to the bisimilarity metric, but uses an independent coupling instead of optimizing over all couplings in the Kantorovich metric. This is much more efficient, but it does mean that the distance from a state to itself can be positive. As a result, using the maximum MICo distance to estimate the uncertainty could result in a positive uncertainty for a Markov chain or MDP without any uncertainty. This makes it unsuitable for the purpose of estimating the uncertainty in a Markov chain or MDP.

In [6], Castro introduces several methods to approximate the bisimilarity metric on MDPs. One of these is the notion of on-policy bisimulation. This could be used to estimate uncertainty on an MDP using an uncertainty function on Markov chains.

In this thesis, we defined a trace metric on Markov chains. This is a different metric from the stutter and strong trace distances that are used in, for example, [1]. That trace distance is a pseudometric for labelled Markov chains. The strong trace distance is based on the largest difference in probability for a single trace. The stutter trace distance is similar, but treats traces with multiple states with the same label as identical to a trace with only one state with that label. There are two reasons why analogues of these distance functions for Markov chains with reward would be impractical for the purposes of constructing uncertainty functions. First, while all labels in a labelled Markov chain can be treated as fundamentally different, reward values exist on a spectrum, and a distance function should not be too sensitive to small differences. Second, in this thesis we work with expected discounted reward. This also means that the differences in value over time should also be discounted in distance functions in some way.

8.2 Bisimulation metrics on labelled Markov chains

In this thesis, we looked at Markov chains with reward, i.e. Markov reward models, as a step towards MDPs. In contrast, existing work on bisimulation metrics on Markov chains is about labelled Markov chains. For labelled Markov chains, the distance between states with a different label is defined to be 1, and the distances between states with the same label are based on that. This is different from Markov reward models, where the distance is based on the difference in reward at each step. As a result, the algorithms for labelled Markov chains in for example [8, 2, 30] are not directly applicable here, even though the ideas might be transferable.

In the dissertation of Tang [30], the bisimulation pseudometric for labelled Markov chains is shown to be a unique fixed point in Theorem 2.1.32. The function of which it is a pseudometric is similar to the one used in this thesis, but immediately assigns values 0 and 1 to pairs of states that are bisimilar or have diffing labels respectively. The proof uses the Knaster-Tarski theorem to establish the existence of least and greatest fixed points, and then shows that they must be equal by showing the points with the greatest distance form a bisimulation relation. This makes it different from the proof in this thesis that bisimulation metrics for Markov reward models are unique fixed points, which uses the Banach fixed point theorem.

8.3 Uncertainty quantification

Uncertainty quantification is a field that tries to quantify uncertainty in a wide variety of contexts [28]. The problems studied in uncertainty quantification can be split into two categories. The first is forward propagation, where the uncertainty of a model's outputs are estimated based on various sources of uncertainty. The second is the inverse problem, where based on the outputs of a model, the bias of the model and values of unknown parameters are estimated. The problem in this thesis is an example of forward propagation, because we try to estimate the uncertainty of an uncertain Markov chain as a whole based on the uncertainty of the transition probabilities.

Most methods from uncertainty quantification focus on probabilistic uncertainty, while the uncertainty sets in uncertain Markov chains do not have a probability distribution associated with them. This means that those methods are not applicable to our problem.

9 Conclusion

In this thesis we examined ways to quantify uncertainty in uncertain Markov chains. This is a stepping stone to quantifying uncertainty in uMDPs. In particular, we used the supremum of the distance between any two possible Markov chains to measure the uncertainty.

We first introduced the notion of uncertainty functions on uncertain Markov chains. We also showed that one of its properties, monotonicity under bisimilarity, is equivalent to the combination of stability under bisimilarity and monotonicity.

To construct such uncertainty functions, we can use pseudometrics on Markov chains. The supremum of the distance between any two Markov chains in the uncertainty set is a measure of the uncertainty in the uncertain Markov chain. If the distance between bisimilar Markov chains is zero, this forms the uncertainty function corresponding to that pseudometric.

We looked at four pseudometrics on Markov chains: the bisimilarity metric, the trace metric, the expected value metric and the value distribution metric. Of these, the bisimilarity metric was an existing metric, while the others are new. They each consider different aspects of the Markov chains.

We proved that these four pseudometrics can be strictly ordered, with the expected value metric being the least, then the value distribution metric, then the trace metric and finally the bisimilarity metric. Because the difference between bisimilar Markov chains is zero under the bisimilarity metric, the same must be true for the other three metrics. As a result, each of them can be used to construct an uncertainty function.

Then, we discussed the computation of the metrics and their induced uncertainty functions. All the metrics can be approximated. However, the algorithm to approximate the trace metric is infeasible in practice.

Of the induced uncertainty functions, we only know how to compute the one induced by the expected value metric. This makes this uncertainty function the only usable one in practical applications. The other uncertainty functions are more granular, so they distinguish more types of differences. In particular, compared to the expected value metric, the value distribution metric also distinguishes Markov chains with identical expected value, but with different distributions. The trace and bisimilarity metrics add to this a distinction between the timing of reward. If efficient algorithms can be found for these uncertainty functions, they might be useful in situations where information is required about more than just the expected value.

In future work, the definition of uncertainty functions as well as specific uncertainty functions could be extended to MDPs. A sketch for this is made in Section 7. Specifically, the definition of uncertainty functions can be straightforwardly extended to MDPs. This could be used to extend the metrics and induced uncertainty functions in this thesis to MDPs in a variety of ways. Other future work in this area includes defining and calculating other uncertainty functions on MDPs.

One particular approach highlighted in Section 7.4 is to define uncertainty not in terms of the variety of possible values or traces, but in terms of possible optimal policies. It is not clear whether it is possible to do this in such a way that it forms an uncertainty function.

When it comes to uncertainty on Markov chains, possible future work includes calculating the proposed induced uncertainty functions as well as finding and implementing algorithms for calculating the trace metric and the value distribution metric.

References

[1] Giorgio Bacci et al. "Converging from Branching to Linear Metrics on Markov Chains". In: *Mathematical Structures in Computer Science* 29.1 (Jan. 2019), pp. 3–37. DOI: 10.1017/S0960129517000160.

[2] Giorgio Bacci et al. "On-the-Fly Exact Computation of Bisimilarity Distances". In: *Tools and Algorithms for the Construction and Analysis of Systems*. TACAS 2013 (Rome, Italy, Mar. 16–24, 2013). Ed. by Nir Piterman and Scott A. Smolka. Vol. 7795. Lecture Notes in Computer Science. Berlin, Germany: Springer, 2013, pp. 1–15. DOI: 10.1007/978-3-642-36742-7_1.

[3] Christel Baier and Joost-Pieter Katoen. *Principles of Model Checking*. Cambridge, Massachusetts, USA: MIT press, 2008. ISBN: 978-0-262-02649-9.

[4] Jaco de Bakker and Erik de Vink. *Control Flow Semantics*. Cambridge, Massachusetts, USA: MIT press, 1996. ISBN: 9780262041546.

[5] Stefan Banach. "Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales". In: *Fundamenta Mathematicae* 3 (1922), pp. 133–181. DOI: 10.4064/fm-3-1-133-181.

[6] Pablo Samuel Castro. "Scalable Methods for Computing State Similarity in Deterministic Markov Decision Processes". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 34.6. AAAI-20 (New York, New York, USA, Feb. 7–12, 2020). Palo Alto, California, USA: AAAI Press, 2020, pp. 10069–10076. DOI: 10.1609/aaai.v34i06.6564.

[7] Pablo Samuel Castro et al. "MICo: Improved Representations via Sampling-Based State Similarity for Markov Decision Processes". In: *Advances in Neural Information Processing Systems 35*. NeurIPS 2021 (online, Dec. 6–14, 2021). Ed. by M. Ranzato et al. Red Hook, New York, USA: Curran Associates, Inc., 2021, pp. 30113–30126. URL: https://openreview.net/forum?id=wFp6kmQELgu.

[8] Di Chen, Franck Van Breugel, and James Worrell. "On the Complexity of Computing Probabilistic Bisimilarity". In: *Proceedings of the 15th international conference on Foundations of Software Science and Computational Structures*. FoSSaCS 2012 (Tallinn, Estonia, Mar. 24–Apr. 1, 2012). Ed. by Lars Birkedal. Vol. 7213. Lecture Notes in Computer Science. Berlin, Germany: Springer, 2012, pp. 437–451. DOI: 10.1007/978-3-642-28729-9_29.

[9] Norm Ferns. "Metrics for Markov Decision Processes". MA thesis. McGill University, Dec. 2003.

[10] Norm Ferns, Prakash Panangaden, and Doina Precup. "Metrics for Finite Markov Decision Processes". In: *Proceedings of the Twentieth Conference on Uncertainty in Artificial Intelligence*. UAI 2004 (Banff, Canada, July 7–11, 2004). Arlington, Virginia, USA: AUAI Press, 2004, pp. 162–169.

[11] Norm Ferns, Prakash Panangaden, and Doina Precup. *Metrics for Finite Markov Decision Processes*. July 11, 2012. arXiv: 1207.4114 [cs.AI]. URL: https://arxiv.org/abs/1207.4114.

[12] Norm Ferns and Doina Precup. "Bisimulation Metrics are Optimal Value Functions". In: *Proceedings of the Thirtieth Conference on Uncertainty in Artificial Intelligence*. UAI 2014 (Quebec City, Canada, July 23–27, 2014). Ed. by Nevin L. Zhang and Jin Tian. Arlington, Virginia, USA: AUAI Press, 2014, pp. 210–219.

[13] Norm Ferns et al. “Methods for Computing State Similarity in Markov Decision Processes”. In: *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence.* UAI 2006 (Cambridge, Massachusetts, USA, July 13–16, 2006). Ed. by Rina Dechter and Thomas Richardson. Arlington, Virginia, USA: AUAI Press, 2006, pp. 174–181.

[14] David H. Fremlin. *Measure Theory.* Vol. 2: *Broad Foundations.* Colchester, England: Torres Fremlin, 2003. ISBN: 978-0-9538129-2-9.

[15] Javier García, Álvaro Visús, and Fernando Fernández. “A Taxonomy for Similarity Metrics between Markov Decision Processes”. In: *Machine Learning* 111 (Nov. 2022), pp. 4217–4247. DOI: 10.1007/s10994-022-06242-4.

[16] Robert Givan, Thomas Dean, and Matthew Greig. “Equivalence Notions and Model Minimization in Markov Decision Processes”. In: *Artificial intelligence* 147.1–2 (2003), pp. 163–223. DOI: 10.1016/S0004-3702(02)00376-4.

[17] Vineet Goyal and Julien Grand-Clément. “Robust Markov Decision Process: Beyond Rectangularity”. In: *Mathematics of Operations Research* 48.1 (Feb. 2023), pp. 203–226. DOI: 10.1287/moor.2022.1259.

[18] Garud N. Iyengar. “Robust Dynamic Programming”. In: *Mathematics of Operations Research* 30.2 (May 2005), pp. 257–280. DOI: 10.1287/moor.1040.0129.

[19] C. J. Jagtenberg, S. Bhulai, and R. D. van der Mei. “Optimal Ambulance Dispatching”. In: *Markov Decision Processes in Practice.* Ed. by Richard J. Boucherie and Nico M. van Dijk. Vol. 248. International Series in Operations Research & Management Science. Cham, Germany: Springer International Publishing, 2017, pp. 269–291. DOI: 10.1007/978-3-319-47766-4_9.

[20] Leonid Vitalyevich Kantorovich. “Об одном эффективном методе решения некоторых классов экстремальных проблем”. [English transl.: A new method of solving some classes of extremal problems]. In: Доклады Академии Наук СССР *[English transl.: Proceedings of the USSR Academy of Sciences]* 28.3 (1940), pp. 212–215.

[21] Pia L. Kempker et al. “Smart Charging of Electric Vehicles”. In: *Markov Decision Processes in Practice.* Ed. by Richard J. Boucherie and Nico M. van Dijk. Vol. 248. International Series in Operations Research & Management Science. Cham, Germany: Springer International Publishing, 2017, pp. 387–404. DOI: 10.1007/978-3-319-47766-4_14.

[22] Vidyadhar G. Kulkarni. *Modeling and Analysis of Stochastic Systems.* Boca Raton, Florida, USA: CRC Press, Taylor & Francis Group, 2017. ISBN: 978-1-4987-5661-7.

[23] Kim G. Larsen and Arne Skou. “Bisimulation through Probabilistic Testing”. In: *Information and Computation* 94.1 (Sept. 1991), pp. 1–28. DOI: 10.1016/0890-5401(91)90030-6.

[24] Arnab Nilim and Laurent El Ghaoui. “Robust Control of Markov Decision Processes with Uncertain Transition Matrices”. In: *Operations Research* 53.5 (Oct. 2005), pp. 780–798. DOI: 10.1287/opre.1050.0216.

[25] James Orlin. “A Faster Strongly Polynomial Minimum Cost Flow Algorithm”. In: *Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing.* STOC ’88 (Chicago, Illinois, USA, May 2–4, 1988). Chicago, Illinois, USA: ACM Press, 1988, pp. 377–387. DOI: 10.1145/62212.62249.

[26] Yossi Rubner, Carlo Tomasi, and Leonidas J. Guibas. “The Earth Mover’s Distance as a Metric for Image Retrieval”. In: *International Journal of Computer Vision* 40.2 (Nov. 2000), pp. 99–121. DOI: `10.1023/A:1026543900054`.

[27] Jay K. Satia and Roy E. Lave Jr. “Markovian Decision Processes with Uncertain Transition Probabilities”. In: *Operations Research* 21.3 (June 1973), pp. 728–740. DOI: `10.1287/opre.21.3.728`.

[28] Ralph C. Smith. *Uncertainty Quantification. Theory, Implementation, and Applications.* Philadelphia, Pennsylvania, USA: Society for Industrial and Applied Mathematics, 2014. ISBN: 978-1-61197-321-1.

[29] Marnix Suilen et al. “Robust Anytime Learning of Markov Decision Processes”. In: *Advances in Neural Information Processing Systems 35.* NeurIPS 2022 (New Orleans, Louisiana, USA, Nov. 28–Dec. 9, 2022). Ed. by S. Koyejo et al. Red Hook, New York, USA: Curran Associates, Inc., 2022, pp. 28790–28802.

[30] Qiyi Tang. “Computing Probabilistic Bisimilarity Distances”. PhD thesis. York University, Nov. 2018.

[31] Cédric Villani. *Topics in Optimal Transportation.* Graduate Studies in Mathematics 58. Providence, Rhode Island, USA: American Mathematical Society, 2003. ISBN: 978-0-8218-7232-1.

[32] Douglas J. White. “Real Applications of Markov Decision Processes”. In: *Interfaces* 15.6 (Dec. 1985), pp. 73–83. DOI: `10.1287/inte.15.6.73`.

[33] Wolfram Wiesemann, Daniel Kuhn, and Berç Rustem. “Robust Markov Decision Processes”. In: *Mathematics of Operations Research* 38.1 (Feb. 2013), pp. 153–183. DOI: `10.1287/moor.1120.0566`.

A Appendix

In this example, we will compute the induced uncertainty function of the bisimilarity metric on the uMC $\mathcal{M}$ in Figure 8. To determine the bisimilarity uncertainty of $\mathcal{M}$, we will compute the bisimilarity metric on two Markov chains $M, N \in [\![\mathcal{M}]\!]$. Let p and q be the probabilities of the transition from s_0 to s_1 in M and N respectively. For s_1 and s_2, we will omit the superscript, because s_1^M is bisimilar to s_1^N, and s_2^M to s_2^N. To find $d_B(M,N)$, we will compute ρ^* using Lemma 4.3.

We will start by looking at the distances between s_1 and s_2. All of these are trivial, since they are in the deterministic part of the Markov chain.

$$\begin{aligned}\rho^*(s_1,s_1) &= 0\\ \rho^*(s_2,s_2) &= 0\\ \rho^*(s_1,s_2) &= 1\end{aligned}$$

The next case that we will look at the distances $\rho^*(s_0^M, s_1)$ and $\rho^*(s_0^M, s_2)$. First we will write out the definitions for $\rho^*(s_0^M, s_1)$.

$$\begin{aligned}\rho^*(s_0^M, s_1) &= \theta_\gamma(s_0^M, s_1) + \gamma\mathcal{K}(\rho^*)(P_{s_0^M}, P_{s_1})\\ &= (1-\gamma)0.5 + \gamma\left(p\rho^*(s_1,s_2) + (1-p)\rho^*(s_0^M, s_2)\right)\\ &= (1-\gamma)0.5 + \gamma p\cdot 1 + \gamma(1-p)\rho^*(s_0^M, s_2)\end{aligned}$$

Now we can do the same for $\rho^*(s_0^M, s_2)$.

$$\begin{aligned}\rho^*(s_0^M, s_2) &= \theta_\gamma(s_0^M, s_2) + \gamma\mathcal{K}(\rho^*)(P_{s_0^M}, P_{s_2})\\ &= (1-\gamma)0.5 + \gamma\left(p\rho^*(s_1,s_1) + (1-p)\rho^*(s_0^M, s_1)\right)\\ &= (1-\gamma)0.5 + \gamma(1-p)\rho^*(s_0^M, s_1)\end{aligned}$$

Substituting this formula for $\rho^*(s_0^M, s_2)$ into the previous formula, we get

$$\rho^*(s_0^M, s_1) = (1-\gamma)0.5 + \gamma p + \gamma(1-p)((1-\gamma)0.5 + \gamma(1-p)\rho^*(s_0^M, s_1))$$

Solving for $\rho^*(s_0^M, s_1)$ results in

$$\rho^*(s_0^M, s_1) = \frac{(1-\gamma)0.5 + \gamma p + \gamma(1-p)(1-\gamma)0.5}{1-\gamma^2(1-p)^2} = \frac{\gamma p}{1-\gamma^2(1-p)^2} + \frac{(1-\gamma)0.5}{1-\gamma(1-p)}.$$

The same equality holds for N, with q instead of p. Now we can calculate $\rho^*(s_0^M, s_0^N)$. To do so, we need to find the coupling λ between $P_{s_0^M}$ and $P_{s_0^N}$ that minimizes the sum $\sum_{s,t}\rho^*(s,t)\lambda(s,t)$. For simplicity, we will assume that $p \leq q$. Note that the distance $\rho^*(s_1,s_1) = 0$. By Lemma 2.7, it follows that there exists a minimizing coupling where $\lambda(s_1,s_1) = \min(p,q) = p$. Then the remaining distances can be derived by the properties of couplings:

$$\begin{aligned}\sum_s \lambda(s,s_1) = P_{s_0^N}(s_1) = q &\implies \lambda(s_0^M, s_1) = q-p\\ \sum_s \lambda(s_1,s) = P_{s_0^M}(s_1) = p &\implies \lambda(s_1, s_0^N) = 0\\ \sum_{s,t}\lambda(s,t) = 1 &\implies \lambda(s_0^M, s_0^N) = 1-(q-p)-p = 1-q\end{aligned}$$

With this, we can compute the distance $\rho^*(s_0^M, s_0^N)$.

$$\begin{aligned}\rho^*(s_0^M, s_0^N) &= \theta_\gamma(s_0^M, s_0^N) + \gamma\mathcal{K}(\rho^*)(P_{s_0^M}, P_{s_0^N})\\ &= \gamma(p\rho^*(s_1, s_1) + (q-p)\rho^*(s_0^M, s_1) + (1-q)\rho^*(s_0^M, s_0^N))\\ &= \gamma(q-p)\rho^*(s_0^M, s_1) + \gamma(1-q)\rho^*(s_0^M, s_0^N)\end{aligned}$$

Moving the term with $\rho^*(s_0^M, s_0^N)$ to the left-hand side, and dividing by $\gamma(1-q)$ will give us:

$$\begin{aligned}\rho^*(s_0^M, s_0^N) &= \frac{\gamma(q-p)\rho^*(s_0^M, s_1)}{1-\gamma(1-q)}\\ &= \frac{\gamma(q-p)}{1-\gamma(1-q)}\left(\frac{\gamma p}{1-\gamma^2(1-p)^2} + \frac{(1-\gamma)0.5}{1-\gamma(1-p)}\right)\\ &= \frac{\gamma^2(q-p)p}{(1-\gamma^2(1-p)^2)(1-\gamma(1-q))} + \frac{(1-\gamma)0.5\gamma(q-p)}{(1-\gamma(1-p))(1-\gamma(1-q))}\end{aligned}$$

To determine the maximum of this expression, we take the partial derivatives with regard to p and q.

$$\begin{aligned}&\frac{\partial}{\partial p}\rho^*(s_0^M, s_0^N)\\ &= \frac{\gamma^2\left(-2p(1-\gamma^2(1-p)^2) - (q-p)p\cdot 2\gamma^2(1-p)\right)}{(1-\gamma^2(1-p)^2)^2(1-\gamma(1-q))} + \frac{(1-\gamma)0.5\gamma(-p(1-\gamma(1-p)) - (q-p)\gamma)}{(1-\gamma(1-q))(1-\gamma(1-p))^2}\\ &= -\frac{\gamma^2\left(2p(1-\gamma^2(1-p)^2) + (q-p)p\cdot 2\gamma^2(1-p)\right)}{(1-\gamma^2(1-p)^2)^2(1-\gamma(1-q))} - \frac{(1-\gamma)0.5\gamma(p(1-\gamma(1-p)) + (q-p)\gamma)}{(1-\gamma(1-q))(1-\gamma(1-p))^2}\end{aligned}$$

This term is negative for all $0 \leq p < q \leq 1$ and $\gamma \in (0,1)$. As a result, for the maximal distance, we have $p = 0$. If we put this into the formula for $\rho^*(s_0^M, s_0^N)$, the first term falls away, and we are left with

$$\rho^*(s_0^M, s_0^N) = \frac{(1-\gamma)0.5\gamma q}{(1-\gamma)(1-\gamma(1-q))} = \frac{0.5\gamma q}{1-\gamma(1-q)}.$$

Computing the partial derivative with respect to q on this, we get:

$$\frac{\partial}{\partial q}\rho^*(s_0^M, s_0^N) = \frac{0.5\gamma((1-\gamma(1-q)) + q\gamma)}{(1-\gamma(1-q))^2}$$

This partial derivative is always positive, given that $0 < q \leq 1$ and $\gamma \in (0,1)$. So the value of q that maximizes the distance $\rho^*(s_0^M, s_0^N)$ is $q = 1$. If we fill in this value as well, we get

$$U_{d_B}(\mathcal{M}) = \frac{0.5\gamma\cdot 1}{(1-\gamma(1-1))} = 0.5\gamma.$$