Computing Behavioral Distances, Compositionally*

Giorgio Bacci, Giovanni Bacci, Kim G. Larsen, and Radu Mardare

Department of Computer Science, Aalborg University, Denmark
{grbacci,giovbacci,kgl,mardare}@cs.aau.dk

Abstract. We propose a general definition of composition operator on Markov Decision Processes with rewards (MDPs) and identify a well behaved class of operators, called *safe*, that are guaranteed to be *non-extensive* w.r.t. the bisimilarity pseudometrics of Ferns et al. [10], which measure behavioral similarities between MDPs. For MDPs built using safe/non-extensive operators, we present the first method that exploits the structure of the system for (exactly) computing the bisimilarity distance on MDPs. Experimental results show significant improvements upon the non-compositional technique.

1 Introduction

Probabilistic bisimulation of Larsen and Skou [13] is the standard equivalence for analyzing the behaviour of Markov chains. In [12], this notion has been extended to Markov Decision Processes with rewards (MDPs) with the intent of reducing the size of large systems to help the computation of optimal policies.

However, when the numerical values of probabilities are based on statistical sampling or subject to error estimates, any behavioral analysis based on a notion of equivalence is too fragile, as it only relates processes with identical behaviors. This is a common issue in applications such as systems biology [15], games [4], or planning [7]. Such problems motivated the study of *behavioral distances* (pseudometrics) for probabilistic systems, firstly developed for Markov chains [9,17,16] and later extended to MDPs [10]. These distances support approximate reasoning on probabilistic systems, providing a way to measure the behavioral similarity between states. They allow one to analyze models obtained as approximations of others, more accurate but less manageable, still ensuring that the obtained solution is close to the real one. For instance, in [2,3] the pseudometric of [10] is used to compute (approximated) optimal polices for MDPs in applications for artificial intelligence. These arguments motivate the development of methods to efficiently compute behavioral distances for MDPs.

Realistic models are usually specified compositionally by means of operators that describe the interactions between the subcomponents. These specifications may thus suffer from an exponential growth of the state space, e.g. the parallel

* Work supported by the VKR Center of Excellence MT-LAB and by the Sino-Danish Basic Research Center IDEA4CPS.

composition of n subsystems with m states may cause the main system to have m^n states. To cope with this problem, algorithms like [10,7,5] that need to investigate the entire state space of the system and even more recent proposals [1], that avoid the entire state space exploration using on-the-fly techniques, are not sufficient: one needs to reason compositionally.

Classically, the exact behavior of systems can be analyzed compositionally if the considered behavioral equivalence (e.g. bisimilarity) is a congruence w.r.t. the composition operators. When the behavior of processes is approximated by means of behavioral distances, congruence is generalized by the notion of *non-extensiveness* of the composition operators, that describes the relation between the distances of the subcomponents to that of the composite system [9].

In this paper we study to which extent compositionality on MDPs can be exploited in the computation of the behavioral pseudometrics of [10], hence how the compositional structure of processes can be used in an approximated analysis of behaviors. To this end we introduce a general notion of composition operator on MDPs and characterize a class of operators, called *safe*, that are guaranteed to be non-extensive. This class is shown to cover a wide range of known operators (e.g. synchronous and asynchronous parallel composition), moreover its defining property provides an easy systematic way to check non-extensiveness.

We provide an algorithm to compute the bisimilarity pseudometric by exploiting both the on-the-fly state space exploration in the spirit of [1], and the compositional structure of MDPs built over safe operators. Experimental results show that the compositional optimization yields a significant additional improvement on top of that obtained by the on-the-fly method. In the best cases, the exploitation of compositionality achieves a reduction of computation time by a factor of 10, and for least significant cases the reduction is that of a factor of 2.

2 Markov Decision Processes and Behavioral Metrics

In this section we recall the definitions of *finite discrete-time Markov Decision Process with rewards* (MDP), and of *bisimulation relation* on MDPs [12]. Then we recall the definition of *bisimilarity pseudometric* introduced in [10], which measures behavioral similarities between states.

We start recalling a few facts related to probability distributions that are essential in what follows. A *probability distribution* over a finite set S is a function $\mu\colon S \to [0,1]$ such that $\sum_{s\in S}\mu(s) = 1$. We denote by $\Delta(S)$ the set of probability distributions over S. Given $\mu,\nu \in \Delta(S)$, a distribution $\omega \in \Delta(S\times S)$ is a *matching* for (μ,ν) if for all $u,v \in S$, $\sum_{s\in S}\omega(u,s) = \mu(u)$ and $\sum_{s\in S}\omega(s,v) = \nu(v)$; we denote by $\Pi(\mu,\nu)$ the set of matchings for (μ,ν). For a (pseudo)metric $d\colon S\times S \to [0,\infty)$ over a finite set S, the *Kantorovich (pseudo)metric* is defined by $\mathcal{T}_d(\mu,\nu) = \min_{\omega\in\Pi(\mu,\nu)}\sum_{u,v\in S}\omega(u,v)d(u,v)$, for arbitrary $\mu,\nu \in \Delta(S)$[1].

Definition 1 (Markov Decision Process). *A Markov Decision Process is a tuple $\mathcal{M} = (S, A, \tau, \rho)$ consisting of a finite nonempty set S of* states*, a finite*

[1] Since S is finite, $\Pi(\mu,\nu)$ describes a *bounded* transportation polytope [8], hence the minimum in the definition of $\mathcal{T}_d(\mu,\nu)$ exists and can be achieved at some vertex.

nonempty set A of actions*, a* transition function $\tau\colon S \times A \to \Delta(S)$*, and a* reward function $\rho\colon S \times A \to \mathbb{R}$*.*

The operational behavior of an MDP $\mathcal{M} = (S, A, \tau, \rho)$ is as follows: the process in the state $s_0 \in S$ chooses nondeterministically an action $a \in A$ and it changes the state to $s_1 \in S$, with probability $\tau(s_0, a)(s_1)$. The choice of a in s_0 is rewarded by $\rho(s_0, a)$. The *executions* are transition sequences $w = (s_0, a_0)(s_1, a_1)\ldots$; the challenge is to find *strategies* for choosing the actions in order to maximize the reward $R_\lambda(w) = \lim_{n\to\infty} \sum_{i=0}^{n} \lambda^i \rho(s_i, a_i)$, where $\lambda \in (0, 1)$ is a *discount factor*. A *strategy* is given by a function $\pi\colon S \to \Delta(A)$, called *policy*, where $\pi(s_0)(a)$ is the probability of choosing the action a at state s_0. Each policy π induces a probability distribution over executions defined, for an arbitrary $w = (s_0, a_0)(s_1, a_1)\ldots$, by $P^\pi(w) = \lim_{n\to\infty} \prod_{i=0}^{n} \pi(s_i)(a_i) \cdot \tau(s_i, a_i)(s_{i+1})$. The *value of $s \in S$ according to π*, written $V_\lambda^\pi(s)$, is the expected value of R_λ w.r.t. P^π on the measurable cylinder set of the executions starting from s. The mapping $V_\lambda^\pi\colon S \to \mathbb{R}$ is the *value function according to π*. The value functions induce a preorder on policies defined by $\pi \preceq \pi'$ iff $V_\lambda^\pi(s) \leq V_\lambda^{\pi'}(s)$, for all $s \in S$. A policy π^* is *optimal* for an MDP $\mathcal{M}$ if it is maximal w.r.t. $\preceq$ among all policies for $\mathcal{M}$. Given $\mathcal{M}$, there always exists an optimal policy π^*, but it might not be unique; it has a unique value function $V_\lambda^{\pi^*}$ satisfying the following system of equations known as the *Bellman optimality equations*: $V_\lambda^{\pi^*}(s) = \max_{a\in A} \left(\rho(s, a) + \lambda \sum_{t\in S} \tau(s, a)(t) \cdot V_\lambda^{\pi^*}(t)\right)$, for all $s \in S$. As reference on MDPs we recommend to consult [14].

Definition 2 (Stochastic Bisimulation). *Let $\mathcal{M} = (S, A, \tau, \rho)$ be an MDP. An equivalence relation $R \subseteq S \times S$ is a* stochastic bisimulation *if whenever $(s, t) \in R$ then, for all $a \in A$, $\rho(s, a) = \rho(t, a)$ and, for all R-equivalence classes C, $\tau(s, a)(C) = \tau(t, a)(C)$. Two states $s, t \in S$ are* stochastic bisimilar*, written $s \sim_{\mathcal{M}} t$, if they are related by some stochastic bisimulation on $\mathcal{M}$.*

To cope with the problem of measuring how similar two MDPs are, Ferns et al. [10] defined a bisimilarity pseudometrics that measure the behavioural similarity of two non-bisimilar MDPs. This is defined as the least fixed point of a transformation operator on functions in $[0, \infty)^{S\times S}$.

Let $\mathcal{M} = (S, A, \tau, \rho)$ be an MDP and $\lambda \in (0, 1)$ be a discount factor. The set $[0, \infty)^{S\times S}$ of $[0, \infty)$-valued maps on $S \times S$ equipped with the point-wise partial order defined by $d \sqsubseteq d'$ iff $d(s, t) \leq d'(s, t)$, for all $s, t \in S$, forms an ω-complete partial order with bottom the constant zero-function $\mathbf{0}$, and greatest lower bound given by $(\bigsqcap_{i\in\mathbb{N}} d_i)(s, t) = \inf_{i\in\mathbb{N}} d_i(s, t)$, for all $s, t \in S$. We define a fixed point operator $F_\lambda^{\mathcal{M}}$ on $[0, \infty)^{S\times S}$, for $d\colon S \times S \to [0, \infty)$ and $s, t \in S$, as follows:

$$F_\lambda^{\mathcal{M}}(d)(s, t) = \max_{a\in A} \left(|\rho(s, a) - \rho(t, a)| + \lambda \cdot \mathcal{T}_d(\tau(s, a), \tau(t, a))\right).$$

$F_\lambda^{\mathcal{M}}$ is monotonic [10], thus, by Tarski's fixed point theorem, it admits a least fixed point. This fixed point is the *bisimilarity pseudometric*.

Definition 3 (Bisimilarity pseudometric). *Let $\mathcal{M}$ be an MDP and $\lambda \in (0, 1)$ be a discount factor, then the λ-discounted* bisimilarity pseudometric *for $\mathcal{M}$, written $\delta_\lambda^{\mathcal{M}}$, is the least fixed point of $F_\lambda^{\mathcal{M}}$.*

The pseudometric $\delta_\lambda^{\mathcal{M}}$ enjoys the property that two states are at zero distance if and only if they are bisimilar. Moreover, in [6] it has been proved, using Banach's fixed point theorem, that for $\lambda \in (0,1)$, $F_\lambda^{\mathcal{M}}$ has a *unique* fixed point.

3 Non-Extensiveness and Compositional Reasoning

In this section we give a general definition of composition operator on MDPs that subsumes most of the known composition operators such as the synchronous, asynchronous, and CCS-like parallel compositions. We introduce the notion of *safeness* for an operator and prove that it implies non-extensiveness. Recall that, non-extensiveness corresponds to the quantitative analogue of congruence when one aims to reason with behavioral distances, as advocated e.g. in [11,9].

Definition 4 (Composition Operator). *Let $\mathcal{M}_i = (S_i, A_i, \tau_i, \rho_i)$, $i = 1..n$, be MDPs. A* composition operator *on $\mathcal{M}_1, \ldots, \mathcal{M}_n$ is a tuple $op = (A, op_\tau, op_\rho)$ consisting of a nonempty set A of actions and the following operations*

- ***on transitions functions:*** $op_\tau \colon \prod_{i=1}^n \Delta(S_i)^{S_i \times A_i} \to \Delta(S)^{S \times A}$,
- ***on reward functions:*** $op_\rho \colon \prod_{i=1}^n \mathbb{R}^{S_i \times A_i} \to \mathbb{R}^{S \times A}$.

where, $S = \prod_{i=1}^n S_i$ denotes the cartesian product of S_i, $i = 1..n$. We denote by $op(\mathcal{M}_i, \ldots, \mathcal{M}_n)$ the composite MDP $(S, A, op_\tau(\tau_1, \ldots, \tau_n), op_\rho(\rho_1, \ldots, \rho_n))$.

Below we present examples, for two fixed MDPs $\mathcal{M}_X = (X, A_X, \tau_X, \rho_X)$ and $\mathcal{M}_Y = (Y, A_Y, \tau_Y, \rho_Y)$, of some of the known parallel composition operators.

Example 5. **Synchronous Parallel Composition** can be given as a binary composition operator $| = (A_X \cap A_Y, |_\tau, |_\rho)$, where

$$(\tau_X \mid_\tau \tau_Y)((x,y),a)(u,v) = \tau_X(x,a)(u) \cdot \tau_Y(y,a)(v)\,,$$
$$(\rho_X \mid_\rho \rho_Y)((x,y),a) = \rho_X(x,a) + \rho_Y(y,a)\,.$$

The process $\mathcal{M}_X \mid \mathcal{M}_Y$ reacts iff $\mathcal{M}_X$ and $\mathcal{M}_Y$ can react synchronously. Actions are rewarded by summing up the rewards of the components. ∎

Example 6. **CCS-like Parallel Composition** can be defined by the composition operator $\| = (A_X \cup A_Y, \|_\tau, \|_\rho)$, where

$$(\tau_X \parallel_\tau \tau_Y)((x,y),a)(u,v) = \begin{cases} \tau_X(x,a)(u) & \text{if } a \notin A_Y \text{ and } v = y \\ \tau_Y(y,a)(v) & \text{if } a \notin A_X \text{ and } u = x \\ \tau_X(x,a)(u) \cdot \tau_Y(y,a)(v) & \text{if } a \in A_X \cap A_Y \\ 0 & \text{otherwise} \end{cases}$$

$$(\rho_X \parallel_\rho \rho_Y)((x,y),a) = \begin{cases} \rho_X(x,a) & \text{if } a \notin A_Y \\ \rho_Y(y,a) & \text{if } a \notin A_X \\ \rho_X(x,a) + \rho_Y(y,a) & \text{if } a \in A_X \cap A_Y \end{cases}$$

In the process $\mathcal{M}_X \parallel \mathcal{M}_Y$, the components synchronize on the same action, otherwise they proceed asynchronously. Asynchronous parallel composition can be defined as above, requiring that the MDPs have disjoint set of actions. ∎

Before introducing the concept of non-extensiveness for a composition operator, we provide some preliminary notations. Consider the sets X_i, the functions $d_i \colon X_i \times X_i \to [0, \infty)$, for $i = 1..n$, and $p \in [1, \infty]$. We define the *p-norm* function $\|d_1, \ldots, d_n\|_p \colon \prod_{i=1}^n X_i \times \prod_{i=1}^n X_i \to [0, \infty)$ as follows:

$$\|d_1, \ldots, d_n\|_p((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = (\textstyle\sum_{i=1}^n d_i(x_i, y_i)^p)^{\frac{1}{p}} \qquad \text{if } p < \infty\,,$$
$$\|d_1, \ldots, d_n\|_\infty((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = \max_{1 \le i \le n} d_i(x_i, y_i)\,.$$

Note that, if (X_i, d_i) are (pseudo)metric spaces, $\|d_1, \ldots, d_n\|_p$ is a (pseudo)metric on $\prod_{i=1}^n X_i$, known in the literature as the *p-product (pseudo)metric*.

Definition 7. *Let $p \in [1, \infty]$. A composition operator op on MDPs $\mathcal{M}_1, \ldots, \mathcal{M}_n$ is* p-non-extensive *if $\delta_\lambda^{op(\mathcal{M}_1, \ldots, \mathcal{M}_n)} \sqsubseteq \|\delta_\lambda^{\mathcal{M}_1}, \ldots, \delta_\lambda^{\mathcal{M}_n}\|_p$. A composition operator is* non-extensive *if it is p-non-extensive for some p.*

Non-extensiveness for a composition operator ensures that bisimilarity is a congruence with respect to it —direct consequence of Theorem 4.5 in [10].

Lemma 8. *Let $\mathcal{M}_i = (S_i, A_i, \tau_i, \rho_i)$ be an MDP and $s_i, t_i \in S_i$, for $i = 1..n$, and op be a p-non-extensive composition operator on $\mathcal{M}_1, \ldots, \mathcal{M}_n$. Then,*

i) if $p < \infty$, $\delta_\lambda^{op(\mathcal{M}_1, \ldots, \mathcal{M}_n)}((s_1, \ldots, s_n), (t_1, \ldots, t_n)) \le (\sum_{i=1}^n \delta_\lambda^{\mathcal{M}_i}(s_i, t_i)^p)^{\frac{1}{p}}$
ii) if $p = \infty$, $\delta_\lambda^{op(\mathcal{M}_1, \ldots, \mathcal{M}_n)}((s_1, \ldots, s_n), (t_1, \ldots, t_n)) \le \max_{i=1}^n \delta_\lambda^{\mathcal{M}_i}(s_i, t_i)$.

Corollary 9. *Let $\mathcal{M}_i = (S_i, A_i, \tau_i, \rho_i)$ be an MDP, $s_i, t_i \in S_i$, for $i = 1..n$, and op be a non-extensive composition operator on $\mathcal{M}_1, \ldots, \mathcal{M}_n$. If $s_i \sim_{\mathcal{M}_i} t_i$ for all $i = 1..n$, then $(s_1, \ldots, s_n) \sim_{op(\mathcal{M}_1, \ldots, \mathcal{M}_n)} (t_1, \ldots, t_n)$.*

In general, proving non-extensiveness for a composition operator on MDPs is not a simple task, since one needs to consider the pseudometrics $\delta_\lambda^{\mathcal{M}_i}$ which are defined as the least fixed point of $F_\lambda^{\mathcal{M}_i}$. A simpler sufficient condition that ensures non-extensiveness is the following:

Definition 10. *Let $\mathcal{M}_i = (S_i, A_i, \tau_i, \rho_i)$, for $i = 1..n$, be MDPs and $p \in [1, \infty]$. A composition operator op on $\mathcal{M}_1, \ldots, \mathcal{M}_n$ is* p-safe *if, for any d_i pseudometric on S_i, such that $d_i \sqsubseteq F_\lambda^{\mathcal{M}_i}(d_i)$, it holds*

$$F_\lambda^{op(\mathcal{M}_1, \ldots, \mathcal{M}_n)}(\|d_1, \ldots, d_n\|_p) \sqsubseteq \|F_\lambda^{\mathcal{M}_1}(d_1), \ldots, F_\lambda^{\mathcal{M}_n}(d_n)\|_p\,.$$

A composition operator on MDPs is safe *if it is p-safe for some $p \in [1, \infty]$.*

Theorem 11. *Any safe composition operator on MDPs is non-extensive.*

The examples of compositional operators that we have presented in this section are all 1-safe, hence non-extensive.

Proposition 12. *The composition operators of Examples 5–6 are 1-safe.*

4 Alternative Characterization of the Pseudometric

In this section we give an alternative characterization of $\delta_\lambda^{\mathcal{M}}$ based on the notion of *coupling* that allows us to transfer the results previously proven for Markov chains in [1,5] to MDPs. Then, we show how to relate this characterization to the concept of non-extensiveness for compositional operators on MDPs.

Definition 13 (Coupling). *Let $\mathcal{M} = (S, A, \tau, \rho)$ be an MDP. A* coupling *for $\mathcal{M}$ is a pair $\mathcal{C} = (\rho, \omega)$, where $\omega\colon (S \times S) \times A \to \Delta(S \times S)$ is such that, for any $s, t \in S$ and $a \in A$, $\omega((s,t),a) \in \Pi(\tau(s,a), \tau(t,a))$.*

Given a coupling $\mathcal{C} = (\rho, \omega)$ for $\mathcal{M}$ and a discount factor $\lambda \in (0,1)$, we define the operator $\Gamma_\lambda^{\mathcal{C}}\colon [0,\infty)^{S\times S} \to [0,\infty)^{S\times S}$, for $d \in [0,\infty)^{S\times S}$ and $s, t \in S$, by

$$\Gamma_\lambda^{\mathcal{C}}(d)(s,t) = \max_{a\in A}\left(|\rho(s,a) - \rho(t,a)| + \lambda \textstyle\sum_{u,v\in S} d(u,v)\cdot\omega((s,t),a)(u,v)\right).$$

Note that, any coupling $\mathcal{C} = (\rho, \omega)$ for $\mathcal{M}$ induces an MDP $\mathcal{C}^* = (S\times S, A, \omega, \rho^*)$, defined for any $s, t \in S$ and $a \in A$ by $\rho^*((s,t),a) = |\rho(s,a) - \rho(t,a)|$, and $\Gamma_\lambda^{\mathcal{C}}$ corresponds to the *Bellman optimality operator* on $\mathcal{C}^*$. This operator is monotonic and has a unique fixed point, hereafter denoted by $\gamma_\lambda^{\mathcal{C}}$, corresponding to the value function for $\mathcal{C}^*$ (see [14, §6.2]).

Next we see that the bisimilarity pseudometric $\delta_\lambda^{\mathcal{M}}$ can be characterized as the minimum $\gamma_\lambda^{\mathcal{C}}$ among all the couplings $\mathcal{C}$ for $\mathcal{M}$.

Theorem 14. *Let $\mathcal{M}$ be an MDP. Then, $\delta_\lambda^{\mathcal{M}} = \min\left\{\gamma_\lambda^{\mathcal{C}} \mid \mathcal{C} \text{ coupling for } \mathcal{M}\right\}$.*

Theorem 14 allows us to transfer the compositional reasoning on couplings. To this end, we introduce the notion of composition operator on couplings.

Definition 15. *Let $\mathcal{M}_i = (S_i, A_i, \tau_i, \rho_i)$ be MDPs, for $i = 1..n$. A* coupling composition operator *for $\mathcal{M}_1, \ldots, \mathcal{M}_n$ is a tuple $op^* = (A, op^*_\rho, op^*_\omega)$ consisting of a nonempty set A, and the following operations, where $S = \prod_{i=1}^n S_i$.*

- $op^*_\rho\colon \prod_{i=1}^n \mathbb{R}^{S_i\times A_i} \to \mathbb{R}^{S\times A}$,
- $op^*_\omega\colon \prod_{i=1}^n \Delta(S_i \times S_i)^{S_i\times S_i\times A_i} \to \Delta(S\times S)^{S\times S\times A}$,

Let $\mathcal{C}_i = (\rho_i, \omega_i)$ be a coupling for $\mathcal{M}_i$, for $i = 1..n$, we denote by $op^*(\mathcal{C}_1, \ldots, \mathcal{C}_n)$ the composite coupling $(op^*_\rho(\rho_1, \ldots, \rho_n), op^*_\omega(\omega_1, \ldots, \omega_n))$. Moreover, op^* is called lifting *of a composition operator op on $\mathcal{M}_1, \ldots, \mathcal{M}_n$ if, for all $i = 1..n$ and $\mathcal{C}_i$ coupling for $\mathcal{M}_i$, $op^*(\mathcal{C}_1, \ldots, \mathcal{C}_n)$ is a coupling for $op(\mathcal{M}_1, \ldots, \mathcal{M}_n)$.*

It is not always possible to find coupling composition operators that lift a composition operator on MDPs. Nevertheless, the composite operators presented in Examples 5–6 can be lifted on couplings. We show in the next example how this can be done for the CCS-like parallel composition. For the other example the construction is similar.

Example 16. The composition operator of Example 6 can be lifted on couplings by the operator $\|^* = (A_X \cup A_Y, \|_\rho, \|_\omega)$

$$(\omega_X \|_\omega \omega_Y)(((x,y),(x',y')),a)((u,v),(u',v')) = \\ = \begin{cases} \omega_X((x,x'),a)(u,u') & \text{if } a \notin A_Y,\ (v,v') = (y,y') \\ \omega_Y((y,y'),a)(v,v') & \text{if } a \notin A_X,\ (u,u') = (x,x') \\ \omega_X((x,x'),a)(u,u') \cdot \omega_Y((y,y'),a)(v,v') & \text{if } a \in A_X \cap A_Y \\ 0 & \text{otherwise} \end{cases}$$

Note how the definition above mimics the one in Example 6. ■

Next we adapt the concept of safeness to coupling composition operators.

Definition 17. *Let $\mathcal{M}_i = (S_i, A_i, \tau_i, \rho_i)$ be MDPs, $i = 1..n$ and $p \in [1,\infty]$. A coupling composition operator op^* on $\mathcal{M}_1, \ldots, \mathcal{M}_n$ is* p-safe *if, for all $i = 1..n$, $\mathcal{C}_i$ coupling for $\mathcal{M}_i$ and $d_i \colon S_i \times S_i \to [0,\infty)$ such that $d_i \sqsubseteq \Gamma_\lambda^{\mathcal{C}_i}(d_i)$, it holds*

$$\Gamma_\lambda^{op^*(\mathcal{C}_1,\ldots,\mathcal{C}_n)}(\|d_1,\ldots,d_n\|_p) \sqsubseteq \|\Gamma_\lambda^{\mathcal{C}_1}(d_1),\ldots,\Gamma_\lambda^{\mathcal{C}_n}(d_n)\|_p\,.$$

A coupling composition operator is safe *if it is p-safe for some $p \in [1,\infty]$.*

As done for Proposition 12, the lifting in Example 16 can be shown to be 1-safe.

Non-extensiveness for an operator is ensured if it admits a lifting composition operator on couplings that is safe, as proven by the following theorem.

Theorem 18. *Let op^* be a coupling composition operator that lifts a composition operator op on $\mathcal{M}_1, \ldots, \mathcal{M}_n$. If op^* is safe, then op is non-extensive.*

5 Exact Computation of Bisimilarity Distance

Inspired by the characterization given in Theorem 14, in this section we propose a procedure to exactly compute the bisimilarity pseudometric. This extends to MDPs a method that has been proposed in [1] for Markov chains. We also show how this strategy can be optimized to cope well with composite MDPs.

For a discount factor $\lambda \in (0,1)$, the set of couplings for $\mathcal{M}$ can be endowed with the preorder $\trianglelefteq_\lambda$, defined by $\mathcal{C} \trianglelefteq_\lambda \mathcal{D}$ iff $\gamma_\lambda^{\mathcal{C}} \sqsubseteq \gamma_\lambda^{\mathcal{D}}$. Theorem 14 suggests to look for a coupling for $\mathcal{M}$ which is minimal w.r.t. $\trianglelefteq_\lambda$. The enumeration of all the couplings is clearly unfeasible, therefore it is crucial to provide an efficient search strategy which prevents us to do that.

A Greedy Search Strategy: We provide a greedy strategy that explores the set of couplings until an optimal one is eventually reached.

Let $\mathcal{M} = (S, A, \tau, \rho)$ and $\mathcal{C} = (\rho, \omega)$ be a coupling for $\mathcal{M}$. Given $s, t \in S$, $a \in A$, and $\mu \in \Pi(\tau(s,a), \tau(t,a))$, we denote by $\mathcal{C}[(s,t), a/\mu]$ the coupling (ρ, ω') for $\mathcal{M}$, where ω' is such that $\omega'((s,t),a) = \mu$ and $\omega'((s',t'),a') = \omega((s',t'),a')$ for all $s', t' \in S$ and $a' \in A$ with $((s',t'),a') \neq ((s,t),a)$.

Lemma 19. *Let $\mathcal{M} = (S, A, \tau, \rho)$ be an MDP, $\mathcal{C}$ be a coupling for $\mathcal{M}$, $s, t \in S$, $a \in A$, $\mu \in \Pi(\tau(s,a), \tau(t,a))$, and $\mathcal{D} = \mathcal{C}[(s,t), a/\mu]$. If $\Gamma^{\mathcal{D}}_{\lambda}(\gamma^{\mathcal{C}}_{\lambda})(s,t) < \gamma^{\mathcal{C}}_{\lambda}(s,t)$, then $\gamma^{\mathcal{D}}_{\lambda} \sqsubset \gamma^{\mathcal{C}}_{\lambda}$.*

The lemma above states that $\mathcal{C}$ can be improved w.r.t. $\trianglelefteq_{\lambda}$ by locally updating it as $\mathcal{C}[(s,t), a/\mu]$, with a matching $\mu \in \Pi(\tau(s,a), \tau(t,a))$ such that

$$\textstyle\sum_{u,v \in S} \gamma^{\mathcal{C}}_{\lambda}(u,v) \cdot \mu(u,v) < \sum_{u,v \in S} \gamma^{\mathcal{C}}_{\lambda}(u,v) \cdot \omega((s,t),a)(u,v)\,,$$

where $a \in A$ is the action that maximizes $\Gamma^{\mathcal{C}}(\gamma^{\mathcal{C}}_{\lambda})_{\lambda}(s,t)$. A matching μ satisfying the condition above can be obtained as a solution of a Transportation Problem [8] with cost matrix $(\gamma^{\mathcal{C}}_{\lambda}(u,v))_{u,v \in S}$ and marginals $\tau(s,a)$ and $\tau(t,a)$, hereafter denoted by $TP(\gamma^{\mathcal{C}}_{\lambda}, \tau(s,a), \tau(t,a))$. This gives us a strategy for moving toward $\delta^{\mathcal{M}}_{\lambda}$ by successive improvements on the couplings.

Now we give a necessary and sufficient condition for termination.

Lemma 20. *Let $\mathcal{M} = (S, A, \tau, \rho)$ be an MDP and $\mathcal{C}$ be a coupling for $\mathcal{M}$. If $\gamma^{\mathcal{C}}_{\lambda} \neq \delta^{\mathcal{M}}_{\lambda}$, then there exist $s, t \in S$, $a \in A$, and $\mu \in \Pi(\tau(s,a), \tau(t,a))$ such that $\Gamma^{\mathcal{D}}_{\lambda}(\gamma^{\mathcal{C}}_{\lambda})(s,t) < \gamma^{\mathcal{C}}_{\lambda}(s,t)$, where $\mathcal{D} = \mathcal{C}[(s,t), a/\mu]$.*

The above result ensures that, unless $\mathcal{C}$ is optimal w.r.t $\trianglelefteq_{\lambda}$, the hypotheses of Lemma 19 are satisfied, so that, we can further improve $\mathcal{C}$ following the same strategy. The next statement proves that this search strategy is correct.

Theorem 21. *$\delta^{\mathcal{M}}_{\lambda} = \gamma^{\mathcal{C}}_{\lambda}$ iff there exists no coupling $\mathcal{D}$ for $\mathcal{M}$ s.t. $\Gamma^{\mathcal{D}}_{\lambda}(\gamma^{\mathcal{C}}_{\lambda}) \sqsubset \gamma^{\mathcal{C}}_{\lambda}$.*

Remark 22. In general, there could be an infinite number of couplings (ρ, ω). However, for each fixed $d \in [0, \infty)^{S \times S}$, the linear function mapping $\omega((s,t),a)$ to $\sum_{u,v \in S} d(u,v) \cdot \omega((s,t)a)(u,v)$ achieves its minimum at some vertex of the transportation polytope $P = \Pi(\tau(s,a), \tau(t,a))$. Since the number of such vertices is finite, using the optimal transportation schedule (which is a vertex in P) for the update ensures that the search strategy is always terminating. ∎

Compositional Heuristic: Assume we want to compute the bisimilarity distance for a composite MDP $\mathcal{M} = op(\mathcal{M}_1, \ldots, \mathcal{M}_n)$. The greedy strategy described above moves toward an optimal coupling for $\mathcal{M}$ starting from an arbitrary one. Clearly, the better is the initial coupling the fewer are the steps to the optimal one. The following result gives a heuristic for choosing such a coupling when *op* admits a safe lifting coupling composition operator.

Proposition 23. *Let op be a composition operator on $\mathcal{M}_1, \ldots, \mathcal{M}_n$, and op^* be a p-safe coupling composition operator that lifts op. Then,*

(i) $\gamma^{op^*(\mathcal{C}_1,\ldots,\mathcal{C}_n)}_{\lambda} \sqsubseteq \|\gamma^{\mathcal{C}_1}_{\lambda}, \ldots, \gamma^{\mathcal{C}_n}_{\lambda}\|_p$, *for any $\mathcal{C}_i$ coupling for $\mathcal{M}_i$;*

(ii) $\delta^{op(\mathcal{M}_1,\ldots,\mathcal{M}_n)}_{\lambda} \sqsubseteq \gamma^{op^*(\mathcal{D}_1,\ldots,\mathcal{D}_n)}_{\lambda} \sqsubseteq \|\delta^{\mathcal{M}_1}_{\lambda}, \ldots, \delta^{\mathcal{M}_n}_{\lambda}\|_p$, *where $\mathcal{D}_i$ is a coupling for $\mathcal{M}_i$ which is minimal w.r.t. $\trianglelefteq_{\lambda}$.*

Proposition 23(ii) suggests to start from the coupling $op^*(\mathcal{D}_1, \ldots, \mathcal{D}_n)$, i.e., the one given as the composite of the optimal couplings $\mathcal{D}_i$ for the subcomponents $\mathcal{M}_i$. This ensures that the first over-approximation of $\delta^{\mathcal{M}}_{\lambda}$, that is $\gamma^{op^*(\mathcal{D}_1,\ldots,\mathcal{D}_n)}_{\lambda}$, is at least as good as the upper bound given by non-extensiveness of *op*.

Algorithm 1 On-the-Fly Bisimilarity Pseudometric

Input: MDP $\mathcal{M} = (S, A, \tau, \rho)$; discount factor $\lambda \in (0, 1)$; query $Q \subseteq S \times S$.

1. $\mathcal{C} \leftarrow (\rho, \mathsf{empty})$; $d \leftarrow \mathsf{empty}$; $\mathsf{visited} \leftarrow \emptyset$; $\mathsf{exact} \leftarrow \emptyset$; $\mathsf{toComp} \leftarrow Q$; // Initialize
2. **while** $\exists (s, t) \in \mathsf{toComp}$ **do**
3. **for all** $a \in A$ **do** guess $\mu \in \Pi(\tau(s, a), \tau(t, a))$; $\mathit{UpdateC}(\mathcal{M}, (s, t), a, \mu)$
4. $d \leftarrow \mathit{BellmanOpt}(\lambda, \mathcal{C}, d)$ // update the current estimate
5. **while** $\mathcal{C}[(u, v), a]$ is not optimal for $\mathit{TP}(d, \tau(u, a), \tau(v, a))$ **do**
6. $\mu \leftarrow$ optimal schedule for $\mathit{TP}(d, \tau(u, a), \tau(v, a))$
7. $\mathit{UpdateC}(\mathcal{M}, (u, v), a, \mu)$ // improve the current coupling
8. $d \leftarrow \mathit{BellmanOpt}(\lambda, \mathcal{C}, d)$ // update the current estimate
9. **end while**
10. $\mathsf{exact} \leftarrow \mathsf{exact} \cup \mathsf{visited}$ // add new exact distances
11. $\mathsf{toComp} \leftarrow \mathsf{toComp} \setminus \mathsf{exact}$ // remove exactly computed pairs
12. **end while**
13. **return** $d\restriction_Q$ // return the distance restricted to the pairs in Q

6 A Compositional On-the-Fly Algorithm

In this section we provide an on-the-fly algorithm for computing the bisimilarity distance making full use of the greedy strategy presented in Section 5. Then, we describe how to optimize the computation on composite MDPs.

Let $\mathcal{M} = (S, A, \tau, \rho)$ be an MDP, $Q \subseteq S \times S$, and assume we want to compute $\delta_\lambda^{\mathcal{M}}$ restricted to Q, written $\delta_\lambda^{\mathcal{M}}\restriction_Q$. Our strategy has the following features:

- when a coupling $\mathcal{C}$ is considered, $\gamma_\lambda^{\mathcal{C}}$ can be computed solving the Bellman optimality equation system associated with it;
- the current coupling $\mathcal{C}$ can be improved by a *local* update $\mathcal{C}[(u, v), a/\mu]$ that satisfies the hypotheses of Lemma 19.

Note that, $\gamma_\lambda^{\mathcal{C}}\restriction_Q$ can be computed considering only the smallest independent subsystem containing the variables associated with the pairs in Q. Therefore, we do not need to store the entire coupling, but we can construct it on-the-fly.

The computation of $\delta_\lambda^{\mathcal{M}}\restriction_Q$ is implemented by Algorithm 1. We assume the following global variables to store: $\mathcal{C}$, the current partial coupling; d, the current partial over-approximation of $\delta_\lambda^{\mathcal{M}}$; toComp, the pairs of states for which the distance has to be computed; exact, the pairs of states (s, t) such that $d(s, t) = \delta_\lambda^{\mathcal{M}}(s, t)$; $\mathsf{visited}$, the pair of states considered so far.

At the beginning $\mathcal{C}$ and d are empty, there are no visited states and no exact distances. While there are pairs (s, t) left to be computed we update $\mathcal{C}$ calling the subroutine *UpdateC* on a matching $\mu \in \Pi(\tau(s, a), \tau(t, a))$, for each $a \in A$. Then, d is updated on all visited pairs with the over-approximation $\gamma_\lambda^{\mathcal{C}}$ by calling *BellmanOpt*. According to the greedy strategy, $\mathcal{C}$ is successively improved and d is consequently updated, until no further improvements are possible. Each improvement is demanded by the existence of a better transportation schedule. When line 10 is reached, $d(u, v) = \delta_\lambda^{\mathcal{M}}(u, v)$ for all $(u, v) \in \mathsf{visited}$, therefore $\mathsf{visited}$ is added to exact and removed from toComp. If no more pairs have to be considered, the exact distance on Q is returned.

Algorithm 2 *UpdateC*$(\mathcal{M}, (s,t), a, \mu)$

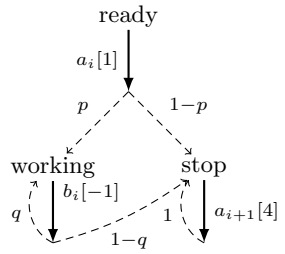
Input: MDP $\mathcal{M} = (S, A, \tau, \rho)$; $s, t \in S$; $a \in A$, $\mu \in \Pi(\tau(s,a), \tau(t,a))$

1. $\mathcal{C} \leftarrow \mathcal{C}[(s,t), a/\mu]$ // update the coupling
2. $\mathsf{visited} \leftarrow \mathsf{visited} \cup \{(s,t)\}$ // set (s,t) as visited
3. **for all** $(u,v) \in \{(u',v') \mid \mu(u',v') > 0\} \setminus \mathsf{visited}$ **do** // for all demanded pairs
4. $\quad \mathsf{visited} \leftarrow \mathsf{visited} \cup \{(u,v)\}$
5. $\quad$ // propagate the construction
6. $\quad$ **for all** $a \in A$ **do** guess $\mu' \in \Pi(\tau(u,a), \tau(v,a))$; *UpdateC*$(\mathcal{M}, (u,v), a, \mu')$
7. **end for**

The subroutine *UpdateC* (Algorithm 2) updates the coupling $\mathcal{C}$ and recursively populates it on all demanded pairs. *BellmanOpt*$(\lambda, \mathcal{C}, d)$ solves the smallest independent subsystem of the Bellman optimality equation system on the MDP induced by $\mathcal{C}$, that contains all the visited pairs. Notice that, the equation system can be further reduced by Gaussian elimination, substituting the variables associated with pairs $(u,v) \in \mathsf{exact}$ with $d(u,v)$.

Compositional Optimizations: Algorithm 1 can be modified to handle composite MDPs efficiently. Assume $\mathcal{M} = op(\mathcal{M}_1, \ldots, \mathcal{M}_n)$ and to have a safe coupling composition operator op^* that lifts op. The compositional heuristic described in Section 5 suggests to start from the coupling $op^*(\mathcal{D}_1, \ldots, \mathcal{D}_n)$ obtained by composing the optimal couplings $\mathcal{D}_i$ for each $\mathcal{M}_i$. This is done running Algorithm 1 in two modalities: master/slave. For each $\mathcal{M}_i$, the master shares the data structures $\mathcal{C}_i$, d_i, $\mathsf{visited}_i$, toComp_i and exact_i with the corresponding slave to keep track of the computation of $\delta_\lambda^{\mathcal{M}_i}$. When a new pair $((s_i, \ldots, s_n), (t_1, \ldots, t_n))$ is considered, the master runs (possibly in parallel) n slave threads of Algorithm 1 on the query $\{(s_i, t_i)\}$. At the end of these subcomputations, the couplings $\mathcal{C}_i$ are optimal, and they are composed to obtain a coupling for $\mathcal{M}$. Note that, the master can reuse the values stored by the slaves in their previous computations.

Experimental Results: For Markov chains, in [1] it has already been shown that an on-the-fly strategy yields, on average, significant improvements with respect to the corresponding iterative algorithms.



Here we focus on how the compositional optimization affects the performances. To this end we consider a simple yet meaningful set of experiments performed on a collection of MDPs, parametric in the probabilities, modeling a pipeline. The figure aside specifies an element $E_i(p,q)$ of the pipeline with actions $A_i = \{a_i, a_{i+1}, b_i\}$. Pipelines are modeled as the parallel composition of different processing elements, that are connected in series by means of synchronization on shared actions. Table 1 reports the computation times of the tests[2] we have run both

[2] The tests have been made using a prototype implementation coded in Mathematica® (available at `http://people.cs.aau.dk/~giovbacci/tools.html`) running on an Intel Core-i5 2.4 GHz processor with 4GB of RAM.

Query	Instance	OTF	COTF	# States
All pairs	$E_0 \parallel E_1$	0.654791	0.97248	9
	$E_1 \parallel E_2$	0.702105	0.801121	9
	$E_0 \parallel E_0 \parallel E_1$	48.5982	13.5731	27
	$E_0 \parallel E_1 \parallel E_2$	23.1984	19.9137	27
	$E_0 \parallel E_1 \parallel E_1$	126.335	13.6483	27
	$E_0 \parallel E_0 \parallel E_0$	49.1167	14.1075	27
Single pair	$E_0 \parallel E_0 \parallel E_0 \parallel E_1 \parallel E_1$	16.7027	11.6919	243
	$E_0 \parallel E_1 \parallel E_0 \parallel E_1 \parallel E_1$	20.2666	16.6274	243
	$E_2 \parallel E_1 \parallel E_0 \parallel E_1 \parallel E_1$	22.8357	10.4844	243
	$E_1 \parallel E_2 \parallel E_0 \parallel E_0 \parallel E_2$	11.7968	6.76188	243
	$E_1 \parallel E_2 \parallel E_0 \parallel E_0 \parallel E_2 \parallel E_2$	Time-out	79.902	729

Table 1. Comparison between the on-the-fly algorithm (OTF) and its compositional optimization (COTF); $E_0 = E_0(0.7, 0.2)$, $E_1 = E_1(0.6, 0.2)$, and $E_2 = E_2(0.5, 0.3)$.

on all-pairs queries and single-pair queries for several pipeline instances; timings are expressed in seconds and, as for the single-pair case, they represent the average of 20 randomly chosen queries. Table 1 shows that the required overhead for maintaining the additional data structure for the subcomponents, affects the performances only on very small systems. In all other cases the compositional optimization yields a significant reduction of the computation time that varies from a factor of 2 up to a factor of 10. Notably, on single-pair queries the compositional version can manage (relatively) large systems whereas the non-compositional one exceeds a time-bound of 3 minutes. Interestingly, we observe better reductions on all-pairs queries than in single-pairs; this may be due to fact that the exact distances collected during the computation are used to further reduce the size of the equation systems that are successively encountered.

7 Conclusions and Future Work

We have proposed a general notion of composition operator on MDPs and identified safeness as a sufficient condition for ensuring non-extensiveness. We showed that the class of safe operators is general enough to cover a wide range of known composition operators. Moreover, we presented an algorithm for computing bisimilarity distances on MDPs, which is able to exploit the compositional structure of the system and applies on MDPs built over any safe operators. This is the first proposal for a compositional algorithm for computing bisimilarity distances; before our contribution, the known tools were based on iterative methods that, by their nature, cannot take advantage of the structure of the systems.

Our work can be extended in several directions. For instance, the notion of safeness can be easily adapted to other contexts where bisimilarity pseudometrics have a fixed point characterization. In the same spirit, one may obtain a sufficient condition that ensures continuity of operators, which is the natural generalization of non-extensiveness.

References

1. G. Bacci, G. Bacci, K. G. Larsen, and R. Mardare. On-the-Fly Exact Computation of Bisimilarity Distances. In N. Piterman and S. A. Smolka, editors, *Tools and Algorithms for the Construction and Analysis of Systems*, volume 7795 of *Lecture Notes in Computer Science*, pages 1–15. Springer Berlin Heidelberg, 2013.
2. P. S. Castro and D. Precup. Using bisimulation for policy transfer in MDPs. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*, volume 1 of *AAMAS '10*, pages 1399–1400, Richland, SC, 2010. International Foundation for Autonomous Agents and Multiagent Systems.
3. P. S. Castro and D. Precup. Automatic Construction of Temporally Extended Actions for MDPs Using Bisimulation Metrics. In S. Sanner and M. Hutter, editors, *Recent Advances in Reinforcement Learning*, volume 7188 of *Lecture Notes in Computer Science*, pages 140–152. Springer Berlin Heidelberg, 2012.
4. K. Chatterjee, L. de Alfaro, R. Majumdar, and V. Raman. Algorithms for Game Metrics. *Logical Methods in Computer Science*, 6(3), 2010.
5. D. Chen, F. van Breugel, and J. Worrell. On the Complexity of Computing Probabilistic Bisimilarity. In L. Birkedal, editor, *FoSSaCS*, volume 7213 of *Lecture Notes in Computer Science*, pages 437–451. Springer, 2012.
6. G. Comanici, P. Panangaden, and D. Precup. On-the-Fly Algorithms for Bisimulation Metrics. *International Conference on Quantitative Evaluation of Systems*, pages 94–103, 2012.
7. G. Comanici and D. Precup. Basis function discovery using spectral clustering and bisimulation metrics. In *AAMAS '11*, volume 3, pages 1079–1080, Richland, SC, 2011. International Foundation for Autonomous Agents and Multiagent Systems.
8. G. B. Dantzig. Application of the Simplex method to a transportation problem. In T. Koopmans, editor, *Activity analysis of production and allocation*, pages 359–373. J. Wiley, New York, 1951.
9. J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for labelled Markov processes. *Theoretical Computer Science*, 318(3):323–354, 2004.
10. N. Ferns, P. Panangaden, and D. Precup. Metrics for finite Markov Decision Processes. In *Proceedings of the 20th conference on Uncertainty in Artificial Intelligence*, UAI, pages 162–169. AUAI Press, 2004.
11. A. Giacalone, C. Jou, and S. A. Smolka. Algebraic reasoning for probabilistic concurrent systems. In *Proc. IFIP TC2 Working Conference on Programming Concepts and Methods*, pages 443–458. North-Holland, 1990.
12. R. Givan, T. Dean, and M. Greig. Equivalence notions and model minimization in Markov decision processes. *Artificial Intelligence*, 147(1-2):163–223, 2003.
13. K. G. Larsen and A. Skou. Bisimulation through probabilistic testing. *Information and Computation*, 94(1):1–28, 1991.
14. M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., New York, NY, USA, 1st edition, 1994.
15. D. Thorsley and E. Klavins. Approximating stochastic biochemical processes with Wasserstein pseudometrics. *IET Systems Biology*, 4(3):193–211, 2010.
16. F. van Breugel, B. Sharma, and J. Worrell. Approximating a Behavioural Pseudometric without Discount for Probabilistic Systems. *Logical Methods in Computer Science*, 4(2):1–23, 2008.
17. F. van Breugel and J. Worrell. Approximating and computing behavioural distances in probabilistic transition systems. *Theoretical Computer Science*, 360(1-3):373–385, 2006.

A Transportation Problem

In 1941 Hitchcock and, independently, in 1947 Koopmans considered the problem which is usually referred to as the (homogeneous) *transportation problem*. This problem can be intuitively described as: a homogeneous product is to be shipped in the amounts $a_1, \ldots, a_m$ respectively, from each of m shipping *origins* and received in amounts $b_1, \ldots, b_n$ respectively, by each of n shipping *destinations*. The cost of shipping a unit amount from the i-th origin to the j-th destination is $c_{i,j}$ and is known for all combinations (i, j). The problem is to determine an optimal *shipping schedule*, i.e. the amount $x_{i,j}$ to be shipped over all routes (i, j), which minimizes the total cost of transportation.

It can be easily formalized as a linear programming problem

$$\begin{array}{lll} \textsf{minimize} & \sum_{i=1}^{m} \sum_{j=1}^{n} c_{i,j} \cdot x_{i,j} & \\ \textsf{such that} & \sum_{j=1}^{n} x_{i,j} = a_i & (i = 1, \ldots, m) \\ & \sum_{i=1}^{m} x_{i,j} = b_j & (j = 1, \ldots, n) \\ & x_{i,j} \geq 0 & (i = 1, \ldots, m \text{ and } j = 1, \ldots, n) \end{array}$$

The set of schedules feasible for a transportation problem, which is formalized as a conjunction of linear constraints, describes a (bounded) convex polytope in $\mathbb{R}^2$, often called transportation polytope.

There are several algorithms in literature which efficiently solve (not necessarily homogeneous) transportation problems. Among these we recall [8].

B Technical proofs

In this appendix we provide the proofs for all the results mentioned so far.

Proof (of Lemma 8). By $s_i \sim_{\mathcal{M}_i} t_i$ and Theorem 4.5 in [10], we have that $\delta_\lambda^{\mathcal{M}_i}(s_i, t_i) = 0$. The thesis follows by p-non-extensiveness of *op* and the definition of p-product metric. □

Proof (of Corollary 9). It follows by Lemma 8 and Theorem 4.5 in [10]. □

Lemma 24. *Let $\mathcal{C}$ be a coupling for $\mathcal{M}$. Then, $\delta_\lambda^{\mathcal{M}} \sqsubseteq \gamma_\lambda^{\mathcal{C}}$.*

Proof. Assume $\mathcal{M} = (S, A, \tau, \rho)$ and $\mathcal{C} = (\rho, \omega)$. In order to prove $\delta_\lambda^{\mathcal{M}} \sqsubseteq \gamma_\lambda^{\mathcal{C}}$, by Tarski's fixed point theorem, it suffices to show that $F_\lambda^{\mathcal{M}}(\gamma_\lambda^{\mathcal{C}}) \sqsubseteq \gamma_\lambda^{\mathcal{C}}$. Let $s, t \in S$. Since $\omega((s,t), a) \in \Pi(\tau(s,a), \tau(t,a))$ we have

$$\begin{aligned} F_\lambda^{\mathcal{M}}(\gamma_\lambda^{\mathcal{C}})(s,t) &= \max_{a \in A} \left(|\rho(s,a) - \rho(t,a)| + \lambda \mathcal{T}_{\gamma_\lambda^{\mathcal{C}}}(\tau(s,a), \tau(t,a)) \right) \\ &\leq \max_{a \in A} \left(|\rho(s,a) - \rho(t,a)| + \lambda \sum_{u,v \in S} \gamma_\lambda^{\mathcal{C}}(u,v)\, \omega((s,t),a)(u,v) \right) \\ &= \Gamma_\lambda^{\mathcal{C}}(\gamma_\lambda^{\mathcal{C}})(s,t)\,. \end{aligned}$$

Since $\gamma_\lambda^{\mathcal{C}} = \Gamma_\lambda^{\mathcal{C}}(\gamma_\lambda^{\mathcal{C}})$, we have $F_\lambda^{\mathcal{M}}(\gamma_\lambda^{\mathcal{C}}) \sqsubseteq \gamma_\lambda^{\mathcal{C}}$. □

Proof (of Theorem 11). Assume *op* is a safe composition operator on the MDPs $\mathcal{M}_1, \ldots, \mathcal{M}_n$, and let $\mathcal{M} = op(\mathcal{M}_1, \ldots, \mathcal{M}_n)$. Then, for some $p \in [1, \infty]$ it holds

$$F_\lambda^{\mathcal{M}}(\|\delta_\lambda^{\mathcal{M}_1}, \ldots, \delta_\lambda^{\mathcal{M}_n}\|_p) \sqsubseteq \|F_\lambda^{\mathcal{M}_1}(\delta_\lambda^{\mathcal{M}_1}), \ldots, F_\lambda^{\mathcal{M}_n}(\delta_\lambda^{\mathcal{M}_n})\|_p \,.$$

By definition $F_\lambda^{\mathcal{M}_1}(\delta_\lambda^{\mathcal{M}_1}) = \delta_\lambda^{\mathcal{M}_1}$, therefore $\|\delta_\lambda^{\mathcal{M}_1}, \ldots, \delta_\lambda^{\mathcal{M}_n}\|_p$ is a prefix point of $F_\lambda^{\mathcal{M}}$. By Tarski's fixed point theorem it follows that $\delta_\lambda^{\mathcal{M}} \sqsubseteq \|\delta_\lambda^{\mathcal{M}_1}, \ldots, \delta_\lambda^{\mathcal{M}_n}\|_p$. □

Proof (of Proposition 12). Let $\mathcal{M}_X = (X, A_X, \tau_X, \rho_X)$, $\mathcal{M}_Y = (Y, A_Y, \tau_Y, \rho_Y)$ be MDPs, and $d_1 \sqsubseteq F_\lambda^{\mathcal{M}_X}(d_1)$, $d_2 \sqsubseteq F_\lambda^{\mathcal{M}_Y}(d_2)$.

1. Assume $\mathcal{M} = \mathcal{M}_X \mid \mathcal{M}_Y$, $\tau = \tau_X \mid_\tau \tau_Y$, and $\rho = \rho_X \mid_\rho \rho_Y$. We have to show that $F_\lambda^{\mathcal{M}}(\|d_1, d_2\|_1) \sqsubseteq \|F_\lambda^{\mathcal{M}_X}(d_1), F_\lambda^{\mathcal{M}_Y}(d_2)\|_1$. Note that, for $d \in A_X \cap A_Y$, $\omega_X \in \Pi(\tau_X(x,d), \tau_X(x',d))$ and $\omega_Y \in \Pi(\tau_Y(y,d), \tau_Y(y',d))$, the function $\omega\colon (X \times Y)^2 \to [0,1]$, defined by $\omega((u,v),(u',v')) = \omega_X(u,u') \cdot \omega_Y(v,v')$, is a probability distribution on $(X \times Y)^2$. Moreover, by definition of $|$, it holds:
   i. $\omega \in \Pi(\tau((x,y),d), \tau((x',y'),d))$;
   ii. $\rho((x,y),d) = \rho_X(x,d) + \rho_Y(y,d)$ and $\rho((x',y'),d) = \rho_X(x',d) + \rho_Y(y',d)$.

   Let $x, x' \in X$, $y, y' \in Y$, $c \in A_X \cap A_Y$ and $\tilde{\omega} \in \Pi(\tau((x,y),c), \tau((x',y'),c))$ be such that

$$F_\lambda^{\mathcal{M}}(\|d_1, d_2\|_1)((x,y),(x',y')) = |\rho((x,y),c) - \rho((x',y'),c)| + \lambda \sum_{\substack{u,u' \in X \\ v,v' \in Y}} (d_1(u,u') + d_2(v,v')) \cdot \tilde{\omega}((u,v),(u',v'))\,,$$

   i.e., $\tilde{\omega}$ is the matching that minimizes the last summand, and c is the action that maximize the whole expression (see the definition of $F_\lambda^{\mathcal{M}}$). To ease the notation, let P and Q denote the following expressions:

$$P = |\rho_X(x,c) - \rho_X(x',c)| + \lambda \sum_{\substack{u,u' \in X \\ v,v' \in Y}} d_1(u,u') \cdot \tilde{\omega}((u,v),(u',v'))\,,$$

$$Q = |\rho_Y(y,c) - \rho_Y(y',c)| + \lambda \sum_{\substack{u,u' \in X \\ v,v' \in Y}} d_2(v,v') \cdot \tilde{\omega}((u,v),(u',v'))\,.$$

   By (ii) and $|(A-B)+(C-D)| \leq |A-B| + |C-D|$, it is easy to see that $F_\lambda^{\mathcal{M}}(\|d_1, d_2\|_1)((x,y),(x',y')) \leq P + Q$. To prove the thesis it suffices to show that $P \leq F_\lambda^{\mathcal{M}_X}(d_1)(x,x')$ and $Q \leq F_\lambda^{\mathcal{M}_Y}(d_2)(y,y')$. Indeed

$$F_\lambda^{\mathcal{M}_X}(d_1)(x,x') + F_\lambda^{\mathcal{M}_Y}(d_2)(y,y') = \|F_\lambda^{\mathcal{M}_X}(d_1), F_\lambda^{\mathcal{M}_Y}(d_2)\|_1((x,y),(x',y'))\,.$$

   Here we show only $P \leq F_\lambda^{\mathcal{M}_X}(d_1)(x,x')$; the other inequality can be proven similarly. Let $a \in A_X$ and $\tilde{\omega}_X \in \Pi(\tau_X(x,a), \tau_X(x',a))$ be such that

$$F_\lambda^{\mathcal{M}_X}(d_1)(x,x') = |\rho_X(x,a) - \rho_X(x',a)| + \lambda \sum_{u,u' \in X} d_1(u,u') \tilde{\omega}_X(u,u')\,,$$

   Case $c = a$**:** By (i) we have that for any $\omega_Y \in \Pi(\tau_Y(y,a), \tau_Y(y',a))$

$$\begin{aligned} P &\leq |\rho_X(x,a) - \rho_X(x',a)| + \lambda \sum_{\substack{u,u' \in X \\ v,v' \in Y}} d_1(u,u')\, \tilde{\omega}_X(u,u')\, \omega_Y(v,v') \\ &= |\rho_X(x,a) - \rho_X(x',a)| + \lambda \sum_{u,u' \in X} d_1(u,u')\, \tilde{\omega}_X(u,u') \\ &= F_\lambda^{\mathcal{M}_X}(d_1)(x,x')\,. \end{aligned}$$

Case $c \neq a$**:** By (i) we have that for any $\omega_X \in \Pi(\tau_X(x,c), \tau_Y(x',c))$ and $\omega_Y \in \Pi(\tau_Y(y,c), \tau_Y(y',c))$

$$\begin{aligned} P &\leq |\rho_X(x,c) - \rho_X(x',c)| + \lambda \sum_{\substack{u,u' \in X \\ v,v' \in Y}} d_1(u,u')\, \omega_X(u,u')\, \omega_Y(v,v') \\ &= |\rho_X(x,c) - \rho_X(x',c)| + \lambda \sum_{u,u' \in X} d_1(u,u')\, \omega_X(u,u') \\ &\leq F_\lambda^{\mathcal{M}_X}(d_1)(x,x')\,. \end{aligned}$$

2. Assume $\mathcal{M} = \mathcal{M}_X \parallel \mathcal{M}_Y$, $\tau = \tau_X \parallel_\tau \tau_Y$, and $\rho = \rho_X \parallel_\rho \rho_Y$. We want to prove $F_\lambda^{\mathcal{M}}(\|d_1, d_2\|_1) \sqsubseteq \|F_\lambda^{\mathcal{M}_X}(d_1), F_\lambda^{\mathcal{M}_Y}(d_2)\|_1$. Let $x, x' \in X$, $y, y' \in Y$, $c \in A_X \cup A_Y$ and $\tilde{\omega} \in \Pi(\tau((x,y),c), \tau((x',y'),c))$ be such that

$$\begin{aligned} F_\lambda^{\mathcal{M}}(\|d_1,d_2\|_1)((x,y),(x',y')) = |\rho((x,y),c) - \rho((x',y'),c)| + \\ \lambda \sum_{\substack{u,u' \in X \\ v,v' \in Y}} (d_1(u,u') + d_2(v,v')) \cdot \tilde{\omega}((u,v),(u',v'))\,. \end{aligned} \quad (1)$$

Case $c \notin A_Y$**:** Notice first that, by definition of $\|$, for $v, v' \in Y$ and arbitrary $\omega \in \Pi(\tau((x,y),c), \tau((x',y'),c))$ it holds that

i. $v \neq y$ or $v' \neq y' \implies \omega((u,v),(u',v')) = 0$;
ii. $\omega((\cdot,y),(\cdot,y')) \in \Pi(\tau_X(x,c), \tau_X(x',c))$;
iii. $\rho((x,y),c) = \rho_X(x,c)$, and $\rho((x',y'),c) = \rho_X(x',c)$.

Given (i)–(iii), it is easy to see that (1) can be rewritten as follows:

$$\begin{aligned} F_\lambda^{\mathcal{M}}(\|d_1,d_2\|_1)((x,y),(x',y')) = |\rho_X(x,c) - \rho_X(x',c)| + \\ \lambda \sum_{u,u' \in X} d_1(u,u') \cdot \tilde{\omega}((u,y),(u',y')) + \lambda d_2(y,y')\,. \end{aligned}$$

By (i) and (ii), it also holds that

$$\begin{aligned} F_\lambda^{\mathcal{M}_X}(d_1)(x,x') \geq |\rho_X(x,c) - \rho_X(x',c)| + \\ \lambda \sum_{u,u' \in X} d_1(u,u') \cdot \tilde{\omega}((u,y),(u',y'))\,. \end{aligned}$$

From the above considerations, we have

$$\begin{aligned} F_\lambda^{\mathcal{M}}(\|d_1,d_2\|_1)((x,y),(x',y')) &\leq F_\lambda^{\mathcal{M}_X}(d_1)(x,x') + \lambda d_2(y,y') \\ &\leq F_\lambda^{\mathcal{M}_X}(d_1)(x,x') + \lambda F_\lambda^{\mathcal{M}_Y}(d_2)(y,y') \\ &\leq F_\lambda^{\mathcal{M}_X}(d_1)(x,x') + F_\lambda^{\mathcal{M}_Y}(d_2)(y,y') \\ &= \|F_\lambda^{\mathcal{M}_X}(d_1), F_\lambda^{\mathcal{M}_Y}(d_2)\|_1((x,y),(x',y'))\,. \end{aligned}$$

Case $c \notin A_X$**:** Similar to the previous case.
Case $c \in A_X \cap A_Y$**:** Similar to 1. □

Proof (of Theorem 14). Let $\mathcal{M} = (S, A, \tau, \rho)$. For any fixed $d \in [0,\infty)^{S \times S}$ there exists a coupling $\mathcal{C} = (\rho, \omega)$ for $\mathcal{M}$ such that $\Gamma_\lambda^{\mathcal{C}}(d) = F_\lambda^{\mathcal{M}}(d)$. This can be defined by taking ω such that, for all $s, t \in S$, $\mathcal{T}_d(\tau(s,a), \tau(t,a)) = \sum_{u,v \in S} d(u,v) \cdot \omega((s,t),a)(u,v)$.

Let $\mathcal{D}$ be a coupling for $\mathcal{M}$ such that $\Gamma_\lambda^{\mathcal{D}}(\delta_\lambda^{\mathcal{M}}) = F_\lambda^{\mathcal{M}}(\delta_\lambda^{\mathcal{M}})$. By definition, $F_\lambda^{\mathcal{M}}(\delta_\lambda^{\mathcal{M}}) = \delta_\lambda^{\mathcal{M}}$, therefore $\delta_\lambda^{\mathcal{M}}$ is a fixed point for $\Gamma_\lambda^{\mathcal{D}}$. By uniqueness of the fixed point of $\Gamma_\lambda^{\mathcal{D}}$ we have $\delta_\lambda^{\mathcal{M}} = \gamma_\lambda^{\mathcal{D}}$. Let $D = \{\gamma_\lambda^{\mathcal{C}} \mid \mathcal{C} \text{ coupling for } \mathcal{M}\}$, we have $\delta_\lambda^{\mathcal{M}} \in D$ and, by Lemma 24, $\delta_\lambda^{\mathcal{M}}$ is a lower bound for D. Hence, $\delta_\lambda^{\mathcal{M}} = \min D$. □

Proof (of Theorem 18). By Theorem 14, for all $1 \leq i \leq n$, there exist a coupling $\mathcal{C}_i$ for $\mathcal{M}_i$, such that $\delta_\lambda^{\mathcal{M}_i} = \gamma_\lambda^{\mathcal{C}_i}$. By hypothesis there exists $1 \leq p \leq \infty$, such that

$$\Gamma_\lambda^{op^*(\mathcal{C}_1,\ldots,\mathcal{C}_n)}(\|\gamma_\lambda^{\mathcal{C}_1},\ldots,\gamma_\lambda^{\mathcal{C}_n}\|_p) \sqsubseteq \|\Gamma_\lambda^{\mathcal{C}_1}(\gamma_\lambda^{\mathcal{C}_1}),\ldots,\Gamma_\lambda^{\mathcal{C}_n}(\gamma_\lambda^{\mathcal{C}_n})\|_p\,.$$

Since, for $1 \leq i \leq n$ it holds $\Gamma_\lambda^{\mathcal{C}_i}(\gamma_\lambda^{\mathcal{C}_i}) = \gamma_\lambda^{\mathcal{C}_i} = \delta_\lambda^{\mathcal{M}_i}$, we have that $\|\delta_\lambda^{\mathcal{M}_1},\ldots,\delta_\lambda^{\mathcal{M}_n}\|_p$ is a pre-fixed point of $\Gamma_\lambda^{op^*(\mathcal{C}_1,\ldots,\mathcal{C}_n)}$. Therefore, by Tarski's fixed point theorem, it holds $\gamma_\lambda^{op^*(\mathcal{C}_1,\ldots,\mathcal{C}_n)} \sqsubseteq \|\delta_\lambda^{\mathcal{M}_1},\ldots,\delta_\lambda^{\mathcal{M}_n}\|_p$. By hypothesis op^* is a lifting of op, that is $op^*(\mathcal{C}_1,\ldots,\mathcal{C}_n)$ is a coupling for $op(\mathcal{M}_1,\ldots,\mathcal{M}_n)$ and, by Lemma 24, we have $\delta_\lambda^{op(\mathcal{M}_1,\ldots,\mathcal{M}_n)} \sqsubseteq \gamma_\lambda^{op^*(\mathcal{C}_1,\ldots,\mathcal{C}_n)}$. Therefore $\delta_\lambda^{op(\mathcal{M}_1,\ldots,\mathcal{M}_n)} \sqsubseteq \|\delta_\lambda^{\mathcal{M}_1},\ldots,\delta_\lambda^{\mathcal{M}_n}\|_p$. □

Proof (of Lemma 19). By Tarski's fixed point theorem, it suffices to show that $\Gamma^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}}) \sqsubset \gamma_\lambda^{\mathcal{C}}$. Assume $\mathcal{C} = (\rho,\omega)$, $\mathcal{D} = (\rho,\bar{\omega})$, and let $s',t' \in S$. If $(s',t') \neq (s,t)$, by definition of $\mathcal{D}$, we have that for all $a' \in A$, $\bar{\omega}((s',t'),a') = \omega((s',t'),a')$, hence $\Gamma_\lambda^{\mathcal{C}}(\gamma_\lambda^{\mathcal{C}})(s',t') = \Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}})(s',t')$. Since, by hypothesis, $\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}})(s,t) < \gamma_\lambda^{\mathcal{C}}(s,t)$, we have $\Gamma^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}}) \sqsubset \gamma_\lambda^{\mathcal{C}}$. □

Proof (of Lemma 20). If for all $s,t \in S$, $a \in A$, and $\mu \in \Pi(\tau(s,a),\tau(t,a))$, $\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}})(s,t) \geq \gamma_\lambda^{\mathcal{C}}(s,t)$, then $\gamma_\lambda^{\mathcal{C}} = F_\lambda^{\mathcal{M}}(\gamma_\lambda^{\mathcal{C}})$. Since $F_\lambda^{\mathcal{M}}$ has a unique fixed point, $\gamma_\lambda^{\mathcal{C}} = \delta_\lambda^{\mathcal{M}}$. □

Proof (of Theorem 21). We prove that $\delta_\lambda \neq \gamma_\lambda^{\mathcal{C}}$ if and only if there exists a coupling $\mathcal{D}$ for $\mathcal{M}$ such that $\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}}) \sqsubset \gamma_\lambda^{\mathcal{C}}$.

($\Rightarrow$) Assume $\delta_\lambda \neq \gamma_\lambda^{\mathcal{C}}$. From Lemma 20, there exist $s,t \in S$, $a \in A$ and $\mu \in \Pi(\tau(s,a),\tau(s,a))$ s.t. $\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}})(s,t) < \gamma_\lambda^{\mathcal{C}}(s,t)$, where $\mathcal{D} = \mathcal{C}[(s,t),a/\mu]$. As in the proof of Lemma 19, we have $\Gamma^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}}) \sqsubset \gamma_\lambda^{\mathcal{C}}$.

($\Leftarrow$) Let $\mathcal{D}$ be such that $\Gamma_\lambda^{\mathcal{D}}(\gamma_\lambda^{\mathcal{C}}) \sqsubset \gamma_\lambda^{\mathcal{C}}$. By Tarski's fixed point theorem $\gamma_\lambda^{\mathcal{D}} \sqsubset \gamma_\lambda^{\mathcal{C}}$. By Lemma 24, $\delta_\lambda \sqsubseteq \gamma_\lambda^{\mathcal{D}} \sqsubset \gamma_\lambda^{\mathcal{C}}$. □

Proof (of Proposition 23). We prove the two items separately.

(i) By hypothesis op^* is a p-safe, hence for all $\mathcal{C}_i$ coupling for $\mathcal{M}_i$, it holds $\Gamma_\lambda^{op^*(\mathcal{C}_1,\ldots,\mathcal{C}_n)}(\|\gamma_\lambda^{\mathcal{C}_1},\ldots,\gamma_\lambda^{\mathcal{C}_n}\|_p) \sqsubseteq \|\gamma_\lambda^{\mathcal{C}_1},\ldots,\gamma_\lambda^{\mathcal{C}_n}\|_p$. By Tarski's fixed point theorem, $\gamma_\lambda^{op^*(\mathcal{C}_1,\ldots,\mathcal{C}_n)} \sqsubseteq \|\gamma_\lambda^{\mathcal{C}_1},\ldots,\gamma_\lambda^{\mathcal{C}_n}\|_p$.

(ii) Since $\mathcal{D}_i$ are minimal w.r.t. $\trianglelefteq_\lambda$, by Theorem 14, we have $\delta_\lambda^{\mathcal{M}_i} = \gamma_\lambda^{\mathcal{D}_i}$. By (i), it holds $\gamma_\lambda^{op^*(\mathcal{D}_1,\ldots,\mathcal{D}_n)} \sqsubseteq \|\delta_\lambda^{\mathcal{M}_1},\ldots,\delta_\lambda^{\mathcal{M}_n}\|_p$. Since op^* is a lifting of op, by Lemma 24, $\delta_\lambda^{op(\mathcal{M}_1,\ldots,\mathcal{M}_n)} \sqsubseteq \gamma_\lambda^{op^*(\mathcal{D}_1,\ldots,\mathcal{D}_n)}$. □