Trace Refinement in Labelled Markov Decision Processes

Nathanaël Fijalkow, Stefan Kiefer, and Mahsa Shirmohammadi

University of Oxford, United Kingdom

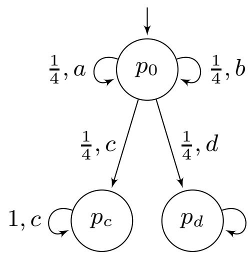
Abstract. Given two labelled Markov decision processes (MDPs), the trace-refinement problem asks whether for all strategies of the first MDP there exists a strategy of the second MDP such that the induced labelled Markov chains are trace-equivalent. We show that this problem is decidable in polynomial time if the second MDP is a Markov chain. The algorithm is based on new results on a particular notion of bisimulation between distributions over the states. However, we show that the general trace-refinement problem is undecidable, even if the first MDP is a Markov chain. Decidability of those problems was stated as open in 2008. We further study the decidability and complexity of the trace-refinement problem provided that the strategies are restricted to be memoryless.

1 Introduction

We consider labelled Markov chains (MCs) whose transitions are labelled with symbols from an alphabet L. Upon taking a transition, the MC emits the associated label. In this way, an MC defines a *trace-probability* function $Tr : \mathsf{L}^* \to [0,1]$ which assigns to each finite trace $w \in \mathsf{L}^*$ the probability that the MC emits w during its first $|w|$ transitions. Consider the MC depicted in Figure 1 with initial state p_0. For example, see that if in state p_0, with probability $\frac{1}{4}$, a transition to state p_c is taken and c is emitted. We have, e.g., $Tr(abc) = \frac{1}{4} \cdot \frac{1}{4} \cdot \frac{1}{4}$. Two MCs over the same alphabet L are called *equivalent* if their trace-probability functions are equal.

The study of such MCs and their equivalence has a long history, going back to Schützenberger [18] and Paz [15]. Schützenberger and Paz studied *weighted* and *probabilistic* automata, respectively. Those models generalize labelled MCs, but the respective equivalence problems are essentially the same. It can be extracted from [18] that equivalence is decidable in polynomial time, using a technique based on linear algebra. Variants of this technique were developed, see e.g. [19,7]. Tzeng [20] considered the path-equivalence problem for nondeterministic automata which asks, given nondeterministic automata A and B, whether each word has the same number of accepting paths in A as in B. He gives an NC algorithm[1] for deciding path equivalence which can be straightforwardly adapted to yield an NC algorithm for equivalence of MCs.

[1] The complexity class NC is the subclass of P containing those problems that can be solved in polylogarithmic parallel time (see e.g. [9]).



trace w	$Tr(w)$
$w \in (a+b)^n$	$\frac{1}{4^n}$
$w \in (a+b)^n c^+$	$\frac{1}{4^{n+1}}$
$w \in (a+b)^n d^+$	$\frac{1}{4^{n+1}}$
otherwise	0

Figure 1. An MC with its trace-probability function. This MC, denoted by $\mathcal{C}(\mathcal{A})$, is also used in the reduction from universality of probabilistic automata to $\mathsf{MC} \sqsubseteq \mathsf{MDP}$.

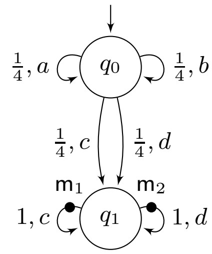


Figure 2. An MDP where the choice of controller is relevant only in q_1. Two available moves $\mathsf{m}_1, \mathsf{m}_2$ are shown with small black circles.

More recently, the efficient decidability of the equivalence problem was exploited, both theoretically and practically, for the verification of probabilistic systems, see e.g. [11,12,16,14,13]. In those works, equivalence naturally expresses properties such as obliviousness and anonymity, which are difficult to formalize in temporal logic. The *inclusion problem* for two probabilistic automata asks whether for each word the acceptance probability in the first automaton is less than or equal to the acceptance probability in the second automaton. Despite its semblance to the equivalence problem, the inclusion problem is undecidable [5], even for automata of fixed dimension [2]. This is unfortunate, especially because deciding language inclusion is often at the heart of verification algorithms.

We study another "inclusion-like" generalization of the equivalence problem: trace refinement in labelled Markov decision processes (MDPs). MDPs extend MCs by nondeterminism; in each state, a controller chooses, possibly randomly and possibly depending on the history, one out of finitely many *moves*[2]. A move determines a probability distribution over the emitted label and the successor state. In this way, an MDP and a strategy of the controller induce an MC.

The *trace-refinement problem* asks, given two MDPs $\mathcal{D}$ and $\mathcal{E}$, whether for all strategies for $\mathcal{D}$ there is a strategy for $\mathcal{E}$ such that the induced MCs are equivalent. Consider the MDP depicted in Figure 2 where in state q_1 there are two available moves; one move generates the label c with probability 1, the other move generates d with probability 1. A strategy of the controller that chooses the last generated label in the state q_1, either c or d, with probability 1, induces the same trace-probability function as the MC shown in Figure 1; the MDP thus refines that MC. The described strategy needs one bit of memory to keep track of the last generated label. It was shown in [7] that the strategy for $\mathcal{E}$ may require *infinite memory*, even if $\mathcal{D}$ is an MC. The decidability of trace refinement was posed as an open problem, both in the introduction and in the conclusion

[2] As in [7] we speak of moves rather than of actions, to avoid possible confusion with the label alphabet L.

of [7]. The authors of [7] also ask about the decidability of subcases, where $\mathcal{D}$ or $\mathcal{E}$ are restricted to be MCs. In this paper we answer all those questions. We show that trace refinement is undecidable, even if $\mathcal{D}$ is an MC. In contrast, we show that trace refinement is decidable efficiently (in NC, hence in P), if $\mathcal{E}$ is an MC. Moreover, we prove that the trace-refinement problem becomes decidable if one imposes suitable *restrictions on the strategies* for $\mathcal{D}$ and $\mathcal{E}$, respectively. More specifically, we consider *memoryless* (i.e., no dependence on the history) and *pure memoryless* (i.e., no randomization and no dependence on the history) strategies, establishing various complexity results between NP and PSPACE.

To obtain the aforementioned NC result, we demonstrate a link between trace refinement and a particular notion of *bisimulation* between two MDPs that was studied in [10]. This variant of bisimulation is not defined between two states as in the usual notion, but between two *distributions* on states. An exponential-time algorithm that decides (this notion of) bisimulation was provided in [10]. We sharpen this result by exhibiting a coNP algorithm that decides bisimulation between two MDPs, and an NC algorithm for the case where one of the MDPs is an MC. For that we refine the arguments devised in [10]. The model considered in [10] is more general than ours in that they also consider continuous state spaces, but more restricted than ours in that the label is determined by the move.

2 Preliminaries

A trace over a finite set L of labels is a finite sequence $w = a_1 \cdots a_n$ of labels where the length of the trace is $|w| = n$. The empty trace ϵ has length zero. For $n \geq 0$, let L^n be the set of all traces with length n; we denote by L^* the set of all (finite) traces over L.

For a function $d : S \to [0,1]$ over a finite set S, define the *norm* $\|d\| := \sum_{s \in S} d(s)$. The *support* of d is the set $\mathsf{Supp}(d) = \{s \in S \mid d(s) > 0\}$. The function d is a *probability subdistribution* over S if $\|d\| \leq 1$; it is a *probability distribution* if $\|d\| = 1$. We denote by $\mathsf{subDist}(S)$ (resp. $\mathsf{Dist}(S)$) the set of all probability subdistributions (resp. distributions) over S. Given $s \in S$, the *Dirac distribution* on s assigns probability 1 to s; we denote it by d_s. For a non-empty subset $T \subseteq S$, the *uniform distribution* over T assigns probability $\frac{1}{|T|}$ to every element in T.

2.1 Labelled Markov Decision Processes

A *labelled Markov decision process* (MDP) $\mathcal{D} = \langle Q, \mu_0, \mathsf{L}, \delta \rangle$ consists of a finite set Q of states, an initial distribution $\mu_0 \in \mathsf{Dist}(Q)$, a finite set L of labels, and a finite probabilistic transition relation $\delta \subseteq Q \times \mathsf{Dist}(\mathsf{L} \times Q)$ where states are in relation with distributions over pairs of labels and successors. We assume that for each state $q \in Q$ there exists some distribution $d \in \mathsf{Dist}(\mathsf{L} \times Q)$ where $\langle q, d \rangle \in \delta$. The set of *moves* in q is $\mathsf{moves}(q) = \{d \in \mathsf{Dist}(\mathsf{L} \times Q) \mid \langle q, d \rangle \in \delta\}$; denote by $\mathsf{moves} = \bigcup_{q \in Q} \mathsf{moves}(q)$ the set of all moves.

For the complexity results, we assume that probabilities of transitions are rational and given as fractions of integers represented in binary.

We describe the behaviour of an MDP as a trace generator running in steps. The MDP starts in the first step in state q with probability $\mu_0(q)$. In each step, if the MDP is in state q the controller chooses $\mathsf{m} \in \mathsf{moves}(q)$; then, with probability $\mathsf{m}(a, q')$, the label a is generated and the next step starts in the successor state q'.

Given $q \in Q$, denote by $\mathsf{post}(q)$ the set $\{(a, q') \in \mathsf{Supp}(\mathsf{m}) \mid \mathsf{m} \in \mathsf{moves}(q)\}$. A *path* in $\mathcal{D}$ is a sequence $\rho = q_0 a_1 q_1 \ldots a_n q_n$ such that $(a_{i+1}, q_{i+1}) \in \mathsf{post}(q_i)$ for all $0 \leq i < n$. The path ρ has the last state $\mathsf{last}(\rho) = q_n$; and the generated trace after ρ is $a_1 a_2 \cdots a_n$, denoted by $\mathsf{trace}(\rho)$. We denote by $\mathsf{Paths}(\mathcal{D})$ the set of all paths in $\mathcal{D}$, and by $\mathsf{Paths}(w) = \{\rho \in \mathsf{Paths}(\mathcal{D}) \mid \mathsf{trace}(\rho) = w\}$ the set of all path generating w.

Strategies. A *randomized strategy* (or simply a strategy) for an MDP $\mathcal{D}$ is a function $\alpha : \mathsf{Paths}(\mathcal{D}) \to \mathsf{Dist}(\mathsf{moves})$ that, given a finite path ρ, returns a probability distribution $\alpha(\rho) \in \mathsf{Dist}(\mathsf{moves}(\mathsf{last}(\rho)))$ over the set of moves in $\mathsf{last}(\rho)$, used to generate a label a and select a successor state q' with probability $\sum_{\mathsf{m} \in \mathsf{moves}(q)} \alpha(\rho)(\mathsf{m}) \cdot \mathsf{m}(a, q')$ where $q = \mathsf{last}(\rho)$.

A strategy α is *pure* if for all $\rho \in \mathsf{Paths}(\mathcal{D})$, we have $\alpha(\rho)(\mathsf{m}) = 1$ for some $\mathsf{m} \in \mathsf{moves}$; we thus view pure strategies as functions $\alpha : \mathsf{Paths}(\mathcal{D}) \to \mathsf{moves}$. A strategy α is *memoryless* if $\alpha(\rho) = \alpha(\rho')$ for all paths ρ, ρ' with $\mathsf{last}(\rho) = \mathsf{last}(\rho')$; we thus view memoryless strategies as functions $\alpha : Q \to \mathsf{Dist}(\mathsf{moves})$. A strategy α is *trace-based* if $\alpha(\rho) = \alpha(\rho')$ for all ρ, ρ' where $\mathsf{trace}(\rho) = \mathsf{trace}(\rho')$ and $\mathsf{last}(\rho) = \mathsf{last}(\rho')$; we view trace-based strategies as functions $\alpha : \mathsf{L}^* \times Q \to \mathsf{Dist}(\mathsf{moves})$.

Trace-probability function. For an MDP $\mathcal{D}$ and a strategy α, the probability of a single path is inductively defined by $\mathsf{Pr}_{\mathcal{D},\alpha}(q) = \mu_0(q)$ and

$$\mathsf{Pr}_{\mathcal{D},\alpha}(\rho a q) = \mathsf{Pr}_{\mathcal{D},\alpha}(\rho) \cdot \sum_{\mathsf{m} \in \mathsf{moves}(\mathsf{last}(\rho))} \alpha(\rho)(\mathsf{m}) \cdot \mathsf{m}(a, q).$$

The *trace-probability* function $Tr_{\mathcal{D},\alpha} : \mathsf{L}^* \to [0, 1]$ is, given a trace w, defined by

$$Tr_{\mathcal{D},\alpha}(w) = \sum_{\rho \in \mathsf{Paths}(w)} \mathsf{Pr}_{\mathcal{D},\alpha}(\rho).$$

We may drop the subscript $\mathcal{D}$ or α from $Tr_{\mathcal{D},\alpha}$ if it is understood. We denote by $subDis_{\mathcal{D},\alpha}(w) \in \mathsf{subDist}(Q)$, the subdistribution after generating a traces w, that is

$$subDis_{\mathcal{D},\alpha}(w)(q) = \sum_{\rho \in \mathsf{Paths}(w) : \mathsf{last}(\rho) = q} \mathsf{Pr}_{\mathcal{D},\alpha}(\rho).$$

We have:

$$Tr_{\mathcal{D},\alpha}(w) = \| subDis_{\mathcal{D},\alpha}(w) \| \tag{1}$$

A version of the following lemma was proved in [7, Lemma 1]:

Lemma 1. *Let $\mathcal{D}$ be an MDP and α be a strategy. There exists a trace-based strategy β such that $Tr_\alpha = Tr_\beta$.*

Here, by $Tr_\alpha = Tr_\beta$ we mean $Tr_\alpha(w) = Tr_\beta(w)$ for all traces $w \in \mathsf{L}^*$.

Labeled Markov Chains. A finite-state labeled Markov chain (MC for short) is an MDP where only a single move is available in each state, and thus controller's choice plays no role. An MC $\mathcal{C} = \langle Q, \mu_0, \mathsf{L}, \delta \rangle$ is an MDP where $\delta : Q \to \mathsf{Dist}(\mathsf{L} \times Q)$ is a probabilistic transition function. Since MCs are MDPs, we analogously define paths, and the probability of a single path inductively as follows: $\mathsf{Pr}_{\mathcal{C}}(q) = \mu_0(q)$ and $\mathsf{Pr}_{\mathcal{C}}(\rho a q) = \mathsf{Pr}_{\mathcal{C}}(\rho) \cdot \delta(q')(a, q)$ where $q' = \mathsf{last}(\rho)$. The notations $subDis_{\mathcal{C}}(w)$ and $Tr_{\mathcal{C}}$ are defined analogously.

2.2 Trace Refinement

Given two MDPs $\mathcal{D}$ and $\mathcal{E}$ with the same set L of labels, we say that $\mathcal{E}$ *refines* $\mathcal{D}$, denoted by $\mathcal{D} \sqsubseteq \mathcal{E}$, if for all strategies α for $\mathcal{D}$ there exists some strategy β for $\mathcal{E}$ such that $Tr_{\mathcal{D}} = Tr_{\mathcal{E}}$. We are interested in the problem $\mathsf{MDP} \sqsubseteq \mathsf{MDP}$, which asks, for two given MDPs $\mathcal{D}$ and $\mathcal{E}$, whether $\mathcal{D} \sqsubseteq \mathcal{E}$. The decidability of this problem was posed as an open question in [7]. We show in Theorem 2 that the problem $\mathsf{MDP} \sqsubseteq \mathsf{MDP}$ is undecidable.

We consider various subproblems of $\mathsf{MDP} \sqsubseteq \mathsf{MDP}$, which asks whether $\mathcal{D} \sqsubseteq \mathcal{E}$ holds. Specifically, we speak of the problem

- $\mathsf{MDP} \sqsubseteq \mathsf{MC}$ when $\mathcal{E}$ is restricted to be an MC;
- $\mathsf{MC} \sqsubseteq \mathsf{MDP}$ when $\mathcal{D}$ is restricted to be an MC;
- $\mathsf{MC} \sqsubseteq \mathsf{MC}$ when both $\mathcal{D}$ and $\mathcal{E}$ are restricted to be MCs.

We show in Theorem 2 that even the problem $\mathsf{MC} \sqsubseteq \mathsf{MDP}$ is undecidable. Hence we consider further subproblems. Specifically, we denote by $\mathsf{MC} \sqsubseteq \mathsf{MDP_m}$ the problem where the MDP is restricted to use only memoryless strategies, and by $\mathsf{MC} \sqsubseteq \mathsf{MDP_{pm}}$ the problem where the MDP is restricted to use only pure memoryless strategies. When both MDPs $\mathcal{D}$ and $\mathcal{E}$ are restricted to use only pure memoryless strategies, the trace-refinement problem is denoted by $\mathsf{MDP_{pm}} \sqsubseteq \mathsf{MDP_{pm}}$. The problem $\mathsf{MC} \sqsubseteq \mathsf{MC}$ equals the *trace-equivalence problem* for MCs: given two MCs $\mathcal{C}_1, \mathcal{C}_2$ we have $\mathcal{C}_1 \sqsubseteq C_2$ if and only if $Tr_{\mathcal{C}_1} = Tr_{\mathcal{C}_2}$ if and only if $\mathcal{C}_2 \sqsubseteq \mathcal{C}_1$. This problem is known to be in NC [20], hence in P.

3 Undecidability Results

In this section we show:

Theorem 2. *The problem* $\mathsf{MC} \sqsubseteq \mathsf{MDP}$ *is undecidable. Hence a fortiori,* $\mathsf{MDP} \sqsubseteq \mathsf{MDP}$ *is undecidable.*

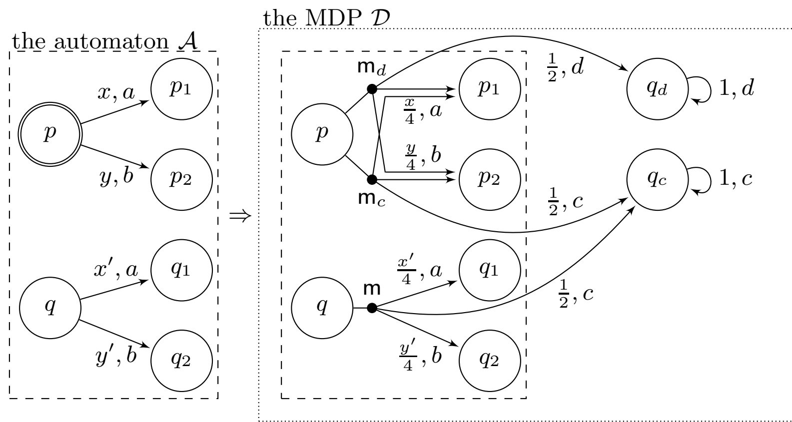


Figure 3. Sketch of the construction of the MDP $\mathcal{D}$ from the probabilistic automaton $\mathcal{A}$, for the undecidability result of $\mathsf{MC} \sqsubseteq \mathsf{MDP}$. Here, p is an accepting state whereas q is not. To read the picture, note that in p there is a transition to the state p_1 with probability x and label a: $\delta(p, a)(p_1) = x$.

Proof. To show that the problem $\mathsf{MC} \sqsubseteq \mathsf{MDP}$ is undecidable, we establish a reduction from the universality problem for probabilistic automata. A *probabilistic automaton* is a tuple $\mathcal{A} = \langle Q, \mu_0, \mathsf{L}, \delta, F\rangle$ consisting of a finite set Q of states, an initial distribution $\mu_0 \in \mathsf{Dist}(Q)$, a finite set L of letters, a transition function $\delta : Q \times \mathsf{L} \to \mathsf{Dist}(Q)$ assigning to every state and letter a distribution over states, and a set F of final states. For a word $w \in \mathsf{L}^*$ we write $dis_{\mathcal{A}}(w) \in \mathsf{Dist}(Q)$ for the distribution such that, for all $q \in Q$, we have that $dis_{\mathcal{A}}(w)(q)$ is the probability that, after inputting w, the automaton $\mathcal{A}$ is in state q. We write $\mathsf{Pr}_{\mathcal{A}}(w) = \sum_{q \in F} dis_{\mathcal{A}}(w)(q)$ to denote the probability that $\mathcal{A}$ accepts w. The *universality problem* asks, given a probabilistic automaton $\mathcal{A}$, whether $\mathsf{Pr}_{\mathcal{A}}(w) \geq \frac{1}{2}$ holds for all words w. This problem is known to be undecidable [15].

Let $\mathcal{A} = \langle Q, \mu_0, \mathsf{L}, \delta, F\rangle$ be a probabilistic automaton; without loss of generality we assume that $\mathsf{L} = \{a, b\}$. We construct an MDP $\mathcal{D}$ such that $\mathcal{A}$ is universal if and only if $\mathcal{C} \sqsubseteq \mathcal{D}$ where $\mathcal{C}$ is the MC shown in Figure 1. The MDP $\mathcal{D}$ is constructed from $\mathcal{A}$ as follows; see Figure 3.

Its set of states is $Q \cup \{q_c, q_d\}$, and its initial distribution is μ_0. (Here and in the following we identify subdistributions $\mu \in \mathsf{subDist}(Q)$ and $\mu \in \mathsf{subDist}(Q \cup \{q_c, q_d\})$ if $\mu(q_c) = \mu(q_d) = 0$.) We describe the transitions of $\mathcal{D}$ using the transition function δ of $\mathcal{A}$. Consider a state $q \in Q$:

- If $q \in F$, there are two available moves $\mathsf{m}_c, \mathsf{m}_d$; both emit a with probability $\frac{1}{4}$ and simulate the probabilistic automaton $\mathcal{A}$ reading the letter a, or emit b with probability $\frac{1}{4}$ and simulate the probabilistic automaton $\mathcal{A}$ reading the letter b. With the remaining probability of $\frac{1}{2}$, m_c emits c and leads to q_c

and m_d emits d and leads to q_d. Formally, $\mathsf{m}_c(c, q_c) = \frac{1}{2}$, $\mathsf{m}_d(d, q_d) = \frac{1}{2}$ and $\mathsf{m}_c(e, q') = \mathsf{m}_d(e, q') = \frac{1}{4}\delta(q, e)(q')$ where $q' \in Q$ and $e \in \{a, b\}$.
- If $q \notin F$, there is a single available move m such that $\mathsf{m}(d, q_d) = \frac{1}{2}$ and $\mathsf{m}(e, q') = \frac{1}{4}\delta(q, e)(q')$ where $q' \in Q$ and $e \in \{a, b\}$.
- The only move from q_c is the Dirac distribution on (c, q_c); likewise the only move from q_d is the Dirac distribution on (d, q_d).

This MDP $\mathcal{D}$ "is almost" an MC, in the sense that a strategy α does not influence its behaviour until eventually a transition to q_c or q_d is taken. Indeed, for all α and for all $w \in \{a, b\}^*$ we have $subDis_{\mathcal{D},\alpha}(w) = \frac{1}{4^{|w|}} dis_{\mathcal{A}}(w)$. In particular, it follows $Tr_{\mathcal{D},\alpha}(w) = \|subDis_{\mathcal{D},\alpha}(w)\| = \frac{1}{4^{|w|}}\|dis_{\mathcal{A}}(w)\| = \frac{1}{4^{|w|}}$. Further, if α is trace-based we have:

$$\begin{aligned}
Tr_{\mathcal{D},\alpha}(wc) &= \|subDis_{\mathcal{D},\alpha}(wc)\| && \text{by (1)} \\
&= subDis_{\mathcal{D},\alpha}(wc)(q_c) && \text{structure of } \mathcal{D} \\
&= \sum_{q \in F} subDis_{\mathcal{D},\alpha}(w)(q) \cdot \alpha(w, q)(\mathsf{m}_c) \cdot \frac{1}{2} && \text{structure of } \mathcal{D} \\
&= \frac{1}{4^{|w|}} \sum_{q \in F} dis_{\mathcal{A}}(w)(q) \cdot \alpha(w, q)(\mathsf{m}_c) \cdot \frac{1}{2} && \text{as argued above}
\end{aligned} \tag{2}$$

We show that $\mathcal{A}$ is universal if and only if $\mathcal{C} \sqsubseteq \mathcal{D}$. Let $\mathcal{A}$ be universal. Define a trace-based strategy α with $\alpha(w, q)(\mathsf{m}_c) = \frac{1}{2\mathsf{Pr}_{\mathcal{A}}(w)}$ for all $w \in \{a, b\}^*$ and $q \in F$. Note that $\alpha(w, q)(\mathsf{m}_c)$ is a probability as $\mathsf{Pr}_{\mathcal{A}}(w) \geq \frac{1}{2}$. Let $w \in \{a, b\}^*$. We have:

$$\begin{aligned}
Tr_{\mathcal{D},\alpha}(w) &= \frac{1}{4^{|w|}} && \text{as argued above} \\
&= Tr_{\mathcal{C}}(w) && \text{Figure 1}
\end{aligned}$$

Further we have:

$$\begin{aligned}
Tr_{\mathcal{D},\alpha}(wc) &= \frac{1}{4^{|w|}} \sum_{q \in F} dis_{\mathcal{A}}(w)(q) \cdot \alpha(w, q)(\mathsf{m}_c) \cdot \frac{1}{2} && \text{by (2)} \\
&= \frac{1}{4^{|w|}} \sum_{q \in F} dis_{\mathcal{A}}(w)(q) \cdot \frac{1}{\mathsf{Pr}_{\mathcal{A}}(w)} \cdot \frac{1}{4} && \text{definition of } \alpha \\
&= \frac{1}{4^{|w|+1}} && \mathsf{Pr}_{\mathcal{A}}(w) = \sum_{q \in F} dis_{\mathcal{A}}(w)(q) \\
&= Tr_{\mathcal{C}}(wc) && \text{Figure 1}
\end{aligned}$$

It follows from the definitions of $\mathcal{D}$ and $\mathcal{C}$ that for all $k \geq 1$, we have $Tr_{\mathcal{D},\alpha}(wc^k) = Tr_{\mathcal{D},\alpha}(wc) = Tr_{\mathcal{C}}(wc) = Tr_{\mathcal{C}}(wc^k)$. We have $\sum_{e \in \{a,b,c,d\}} Tr_{\mathcal{D},\alpha}(we) = Tr_{\mathcal{D},\alpha}(w) = Tr_{\mathcal{C}}(w) = \sum_{e \in \{a,b,c,d\}} Tr_{\mathcal{C}}(we)$. Since for $e \in \{a, b, c\}$ we also proved that $Tr_{\mathcal{D},\alpha}(we) = Tr_{\mathcal{C}}(we)$ it follows that $Tr_{\mathcal{D},\alpha}(wd) = Tr_{\mathcal{C}}(wd)$. Hence, as above, $Tr_{\mathcal{D},\alpha}(wd^k) = Tr_{\mathcal{C}}(wd^k)$ for all $k \geq 1$. Finally, if $w \notin (a + b)^* \cdot (c^* + d^*)$ then $Tr_{\mathcal{D},\alpha}(w) = 0 = Tr_{\mathcal{C}}(w)$.

For the converse, assume that $\mathcal{A}$ is not universal. Then there is $w \in \{a,b\}^*$ with $\mathsf{Pr}_{\mathcal{A}}(w) < \frac{1}{2}$. Let α be a trace-based strategy. Then we have:

$$
\begin{aligned}
Tr_{\mathcal{D},\alpha}(wc) &= \frac{1}{4^{|w|}} \sum_{q \in F} dis_{\mathcal{A}}(w)(q) \cdot \alpha(w,q)(\mathsf{m}_c) \cdot \frac{1}{2} && \text{by (2)} \\
&\leq \frac{1}{4^{|w|}} \cdot \frac{1}{2} \cdot \sum_{q \in F} dis_{\mathcal{A}}(w)(q) && \alpha(w,q)(\mathsf{m}_c) \leq 1 \\
&= \frac{1}{4^{|w|}} \cdot \frac{1}{2} \cdot \mathsf{Pr}_{\mathcal{A}}(w) && \mathsf{Pr}_{\mathcal{A}}(w) = \sum_{q \in F} dis_{\mathcal{A}}(w)(q) \\
&< \frac{1}{4^{|w|}} \cdot \frac{1}{2} \cdot \frac{1}{2} && \text{definition of } w \\
&= Tr_{\mathcal{C}}(wc) && \text{Figure 1}
\end{aligned}
$$

We conclude that there is no trace-based strategy α with $Tr_{\mathcal{D},\alpha} = Tr_{\mathcal{C}}$. By Lemma 1 there is *no* strategy α with $Tr_{\mathcal{D},\alpha} = Tr_{\mathcal{C}}$. Hence $\mathcal{C} \not\sqsubseteq \mathcal{D}$. $\square$

A straightforward reduction from $\mathsf{MDP} \sqsubseteq \mathsf{MDP}$ now establishes:

Theorem 3. *The problem that, given two MDPs $\mathcal{D}$ and $\mathcal{E}$, asks whether $\mathcal{D} \sqsubseteq \mathcal{E}$ and $\mathcal{E} \sqsubseteq \mathcal{D}$ is undecidable.*

4 Decidability for Memoryless Strategies

Given two MCs $\mathcal{C}_1$ and $\mathcal{C}_2$, the (symmetric) trace-equivalence relation $\mathcal{C}_1 \sqsubseteq \mathcal{C}_2$ is polynomial-time decidable [20]. An MDP $\mathcal{D}$ under a memoryless strategy α induces a finite MC $\mathcal{D}(\alpha)$, and thus once a memoryless strategy is fixed for the MDP, its relation to another given MC in the trace-equivalence relation $\sqsubseteq$ can be decided in P. Theorems 4 and 5 provide tight complexity bounds of the trace-refinement problems for MDPs that are restricted to use only pure memoryless strategies. In Theorems 6 and 7 we establish bounds on the complexity of the problem when randomization is allowed for memoryless strategies.

4.1 Pure Memoryless Strategies

In this subsection, we show that the two problems $\mathsf{MC} \sqsubseteq \mathsf{MDP}_{\mathsf{pm}}$ and $\mathsf{MDP}_{\mathsf{pm}} \sqsubseteq \mathsf{MDP}_{\mathsf{pm}}$ are NP-complete and Π_2^p-complete.

Membership of $\mathsf{MC} \sqsubseteq \mathsf{MDP}_{\mathsf{pm}}$ in NP and $\mathsf{MDP}_{\mathsf{pm}} \sqsubseteq \mathsf{MDP}_{\mathsf{pm}}$ in Π_2^p are obtained as follows. Given an MC $\mathcal{C}$ and an MDP $\mathcal{D}$, the polynomial witness of $\mathcal{C} \sqsubseteq \mathcal{D}$ is a pure memoryless strategy α for $\mathcal{D}$. Once α is fixed, then $\mathcal{C} \sqsubseteq \mathcal{D}(\alpha)$ can be decided in P. Given another MDP $\mathcal{E}$, for all pure memoryless strategies β of $\mathcal{E}$ whether there exists a polynomial witness α for $\mathcal{E} \sqsubseteq \mathcal{D}$ such that $\mathcal{E}(\beta) \sqsubseteq \mathcal{D}(\alpha)$ can be decided in P.

The hardness results are by reductions from the *subset-sum problem* and a variant of the *quantified subset-sum* problem. Given a set $\{s_1, s_2, \cdots, s_n\}$ of

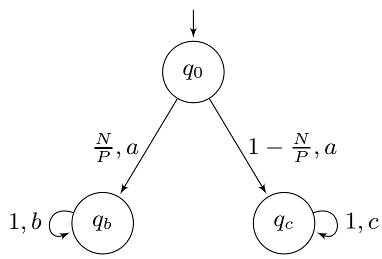


Figure 4. The MC $\mathcal{C}$ in the reduction for NP-hardness of $\mathsf{MC} \sqsubseteq \mathsf{MDP_m}$.

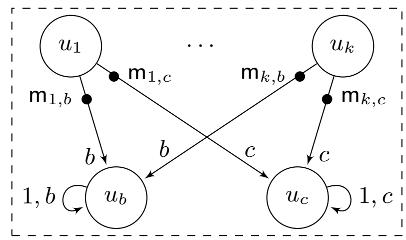


Figure 5. The gadget G_u for the set $\{u_1, \cdots, u_k\}$.

natural numbers and $N \in \mathbb{N}$, the subset-sum problem asks whether there exists a subset $S \subseteq \{s_1, \cdots, s_n\}$ such that $\sum_{s \in S} s = N$. The subset-sum problem is known to be NP-complete [6]. The quantified version of subset sum is a game between a *universal player* and an *existential player*. Given $k, N \in \mathbb{N}$, the game is played turn-based for k rounds. For each round $1 \leq i \leq k$, two sets $\{s_1, s_2, \cdots, s_n\}$ and $\{t_1, t_2, \cdots, t_m\}$ of natural numbers are given. In each round i, the universal player first chooses $S_i \subseteq \{s_1, \cdots, s_n\}$ and then the existential player chooses $T_i \subseteq \{t_1, \cdots, t_m\}$. The existential player wins if and only if

$$\sum_{s \in S_1} s + \sum_{t \in T_1} t + \cdots + \sum_{s \in S_k} s + \sum_{t \in T_k} t = N.$$

The quantified subset sum is known to be PSPACE-complete [8]. The proof therein implies that the variant of the problem with a fixed number k of rounds is Π_{2k}^p-complete.

To establish the NP-hardness of $\mathsf{MC} \sqsubseteq \mathsf{MDP_{pm}}$, consider an instance of subset sum, i.e., a set $\{s_1, \cdots, s_n\}$ and $N \in \mathbb{N}$. We construct an MC $\mathcal{C}$ and an MDP $\mathcal{D}$ such that there exists $S \subseteq \{s_1, \cdots, s_n\}$ with $\sum_{s \in S} s = N$ if and only if $\mathcal{C} \sqsubseteq \mathcal{D}$ when $\mathcal{D}$ uses only pure memoryless strategies.

The MC $\mathcal{C}$ is shown in Figure 4. it generates traces in ab^+ with probability $\frac{N}{P}$ and traces in ac^+ with probability $1 - \frac{N}{P}$ where $P = s_1 + \cdots + s_n$.

For a set $\{u_1, \cdots, u_k\}$, we define a gadget G_u that is an MDP with $k+2$ states: $u_1, \cdots, u_k$ and u_b, u_c; see Figure 5. For all states u_i, two moves $\mathsf{m}_{i,b}$ and $\mathsf{m}_{i,c}$ are available, the Dirac distributions on (a, u_b) and (a, u_c). The states u_b, u_c emit only the single labels b and c. The MDP $\mathcal{D}$ is exactly the gadget G_s for $\{s_1, \cdots, s_n\}$ equipped with the initial distribution μ_0 where $\mu_0(s_i) = \frac{s_i}{P}$ for all $1 \leq i \leq n$. Choosing b in s_i simulates the membership of s_i in S by adding $\frac{s_i}{P}$ to the probability of generating ab^+.

Theorem 4. *The problem* $\mathsf{MC} \sqsubseteq \mathsf{MDP_{pm}}$ *is* NP*-complete.*

To establish the Π_2^p-hardness of $\mathsf{MDP_{pm}} \sqsubseteq \mathsf{MDP_{pm}}$, consider an instance of quantified subset sum, i.e., $N \in \mathbb{N}$ and two sets $\{s_1, \cdots, s_n\}$ and $\{t_1, \cdots, t_m\}$.

We construct MDPs $\mathcal{E}_{univ}, \mathcal{E}_{exist}$ such that the existential player wins in one round if and only if $\mathcal{E}_{univ} \sqsubseteq \mathcal{E}_{exist}$ restricted to use pure memoryless strategies.

Let $P = s_1 + \cdots + s_n$ and $R = t_1 + \cdots + t_m$. Pick a small real number $0 < x < 1$ so that $0 < xP, xR, xN < 1$. Pick real numbers $0 \leq y_1, y_2 \leq 1$ such that $y_1 + xN = y_2 + xR$.

The MDPs $\mathcal{E}_{univ}$ and $\mathcal{E}_{exist}$ have symmetric constructions. To simulate the choice of the universal player, the MDP $\mathcal{E}_{univ}$ is the gadget G_s for the set $\{s_1, \cdots, s_n\}$ where two new states s_r, s_y are added. The transitions in s_r and s_y are the Dirac distributions on (a, s_b) and (a, s_c), respectively. The initial distribution μ_0 for $\mathcal{E}_{univ}$ is such that $\mu_0(s_y) = \frac{1}{2}y_1$ and $\mu_0(s_r) = 1 - \frac{1}{2}(xP + y_1)$, and $\mu_0(s_i) = \frac{1}{2}xs_i$ for all $1 \leq i \leq n$. Similarly, to simulate the counter-attack of the existential player, the MDP $\mathcal{E}_{exist}$ is the gadget G_t for $\{t_1, \cdots, t_m\}$ with two new states t_r, t_y. The transitions in t_r and t_y are the Dirac distributions on (a, t_b) and (a, t_c), respectively. The initial distribution μ'_0 is $\mu'_0(p_y) = \frac{1}{2}y_2$ and $\mu'_0(p_r) = 1 - \frac{1}{2}(xT + y_2)$, and $\mu'_0(t_j) = \frac{1}{2}xt_j$ for all $1 \leq j \leq m$. Choosing b in a set of states s_i by the universal player must be defended by choosing c in a right set of states t_j by existential player such that the probabilities of emitting ab^+ in MDPs are equal.

Theorem 5. *The problem* $\mathsf{MDP_{pm}} \sqsubseteq \mathsf{MDP_{pm}}$ *is* Π_2^p*-complete.*

4.2 Memoryless Strategies

In this subsection, we provide upper and lower complexity bounds for the problem $\mathsf{MC} \sqsubseteq \mathsf{MDP_m}$: a reduction to the existential theory of the reals and a reduction from nonnegative factorization of matrices.

A formula of the *existential theory of the reals* is of the form $\exists x_1 \ldots \exists x_m \, R(x_1, \ldots, x_n)$, where $R(x_1, \ldots, x_n)$ is a boolean combination of comparisons of the form $p(x_1, \ldots, x_n) \sim 0$, where $p(x_1, \ldots, x_n)$ is a multivariate polynomial and $\sim \in \{<, >, \leq, \geq, =, \neq\}$. The validity of closed formulas (i.e., when $m = n$) is decidable in PSPACE [3,17], and is not known to be PSPACE-hard.

Theorem 6. *The problem* $\mathsf{MC} \sqsubseteq \mathsf{MDP_m}$ *is polynomial-time reducible to the existential theory of the reals, hence in* PSPACE.

Given a nonnegative matrix $M \in \mathbb{R}^{n \times m}$, a *nonnegative factorization* of M is any representation of the form $M = A \cdot W$ where $A \in \mathbb{R}^{n \times r}$ and $W \in \mathbb{R}^{r \times m}$ are nonnegative matrices (see [4,21,1] for more details). The *NMF problem* asks, given a nonnegative matrix $M \in \mathbb{R}^{n \times m}$ and a number $r \in \mathbb{N}$, whether there exists a factorization $M = A \cdot W$ with nonnegative matrices $A \in \mathbb{R}^{n \times r}$ and $W \in \mathbb{R}^{r \times m}$. The NMF problem is known to be NP-hard, but membership in NP is open [21].

Below, we present a reduction from the NMF problem to $\mathsf{MC} \sqsubseteq \mathsf{MDP_m}$. To establish the reduction, consider an instance of the NMF problem, i.e., a nonnegative matrix $M \in \mathbb{R}^{n \times m}$ and a number $r \in \mathbb{N}$. We construct an MC $\mathcal{C}$ and an MDP $\mathcal{D}$ such that the NMF instance is a yes-instance if and only if $\mathcal{C} \sqsubseteq \mathcal{D}$ where $\mathcal{D}$ is restricted to use only memoryless strategies.

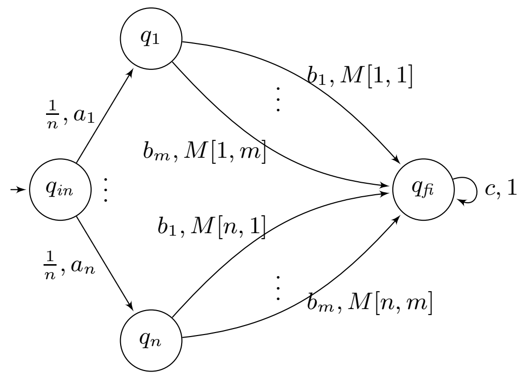


Figure 6. The MC $\mathcal{C}$ of the reduction from NMF to $\mathsf{MC} \sqsubseteq \mathsf{MDP_m}$.

We assume, without loss of generality, that M is a stochastic matrix, that is $\sum_{j=1}^{m} M[i,j] = 1$ for all rows $1 \leq i \leq n$. We know, by [1, Section 5], that there exists a nonnegative factorization of M with rank r if and only if there exist two stochastic matrices $A \in \mathbb{R}^{n \times r}$ and $W \in \mathbb{R}^{r \times m}$ such that $M = A \cdot W$.

The transition probabilities in the MC $\mathcal{C}$ encode the entries of matrix M. The initial distribution of the MC is the Dirac distribution on q_{in}; see Figure 6. There are $n + m + 1$ labels $a_1, \cdots, a_n, b_1, \cdots, b_m, c$. The transition in q_{in} is the uniform distribution over $\{(a_i, q_i) \mid 1 \leq i \leq n\}$. In each state q_i, each label b_j is emitted with probability $M[i,j]$, and a transition to q_{fi} is taken. In state q_{fi} only c is emitted. Observe that for all $1 \leq i \leq n$ and $1 \leq j \leq m$ we have $Tr_{\mathcal{C}}(a_i) = \frac{1}{n}$ and $Tr_{\mathcal{C}}(a_i \cdot b_j \cdot c^*) = \frac{1}{n} M[i,j]$.

The initial distribution of the MDP $\mathcal{D}$ is the uniform distribution over $\{p_1, \cdots, p_n\}$; see Figure 7. In each p_i (where $1 \leq i \leq n$), there are r moves $\mathsf{m}_{i,1}, \mathsf{m}_{i,2}, \cdots, \mathsf{m}_{i,r}$ where $\mathsf{m}_{i,k}(a_i, \ell_k) = 1$ and $1 \leq k \leq r$. In each ℓ_k, there are m moves $\mathsf{m}'_{k,1}, \mathsf{m}'_{k,2}, \cdots, \mathsf{m}'_{k,m}$ where $\mathsf{m}'_{k,j}(b_j, p_{fi}) = 1$ where $1 \leq j \leq m$. In state p_{fi}, only c is emitted. The probabilities of choosing the move $m_{i,k}$ in p_i and choosing $m'_{k,j}$ in ℓ_k simulate the entries of $A[i,k]$ and $W[k,j]$.

Theorem 7. *The NMF problem is polynomial-time reducible to* $\mathsf{MC} \sqsubseteq \mathsf{MDP_m}$*, hence* $\mathsf{MC} \sqsubseteq \mathsf{MDP_m}$ *is* NP*-hard.*

Recall that it is open whether the NMF problem is in NP and whether the existential theory of the reals is PSPACE-hard. So Theorems 6 and 7 show that proving NP-completeness or PSPACE-completeness of $\mathsf{MC} \sqsubseteq \mathsf{MDP_m}$ requires a breakthrough in those areas.

5 Bisimulation

In this section we show:

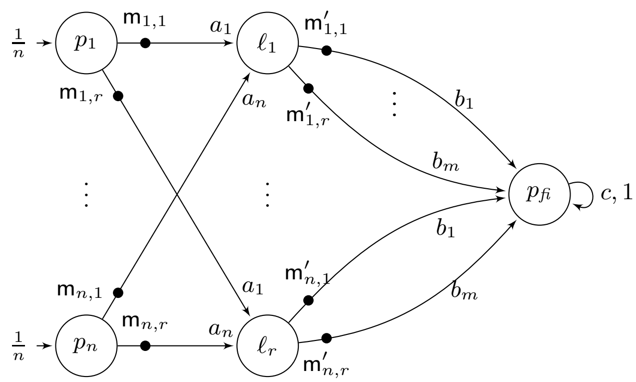


Figure 7. The MDP $\mathcal{D}$ of the reduction from NMF to $\mathsf{MC} \sqsubseteq \mathsf{MDP_m}$.

Theorem 8. *The problem* $\mathsf{MDP} \sqsubseteq \mathsf{MC}$ *is in* NC*, hence in* P.

We prove Theorem 8 in two steps: First, in Proposition 9 below, we establish a link between trace refinement and a notion of bisimulation between distributions that was studied in [10]. Second, we show that this notion of bisimulation can be decided efficiently (in NC, hence in P) if one of the MDPs is an MC. Proposition 9 then implies Theorem 8. Along the way, we prove that bisimulation between two MDPs can be decided in coNP, improving the exponential-time result from [10]. We rebuild a detailed proof from scratch, not referring to [10], as the authors were unable to verify some of the technical claims made in [10].

A *local strategy* for an MDP $\mathcal{D} = \langle Q, \mu_0, \mathsf{L}, \delta \rangle$ is a function $\alpha : Q \to \mathsf{Dist}(\mathsf{moves})$ that maps each state q to a distribution $\alpha(q) \in \mathsf{Dist}(\mathsf{moves}(q))$ over moves in q. We call α *pure* if for all states q there is a move m such that $\alpha(q)(\mathsf{m}) = 1$. For a subdistribution $\mu \in \mathsf{subDist}(Q)$, a local strategy α, and a label $a \in \mathsf{L}$, define the *successor* subdistribution $\mathsf{Succ}(\mu, \alpha, a)$ with

$$\mathsf{Succ}(\mu, \alpha, a)(q') = \sum_{q \in Q} \mu(q) \cdot \sum_{\mathsf{m} \in \mathsf{moves}(q)} \alpha(q)(\mathsf{m}) \cdot \mathsf{m}(a, q')$$

for all $q' \in Q$. Let $\mathcal{D} = \langle Q_{\mathcal{D}}, \mu_0^{\mathcal{D}}, \mathsf{L}, \delta_{\mathcal{D}} \rangle$ and $\mathcal{E} = \langle Q_{\mathcal{E}}, \mu_0^{\mathcal{E}}, \mathsf{L}, \delta_{\mathcal{E}} \rangle$ be two MDPs over the same set L of labels. A *bisimulation* is a relation $\mathcal{R} \subseteq \mathsf{subDist}(Q_{\mathcal{D}}) \times \mathsf{subDist}(Q_{\mathcal{E}})$ such that whenever $\mu_{\mathcal{D}} \mathrel{\mathcal{R}} \mu_{\mathcal{E}}$ then

- $\|\mu_{\mathcal{D}}\| = \|\mu_{\mathcal{E}}\|$;
- for all local strategies $\alpha_{\mathcal{D}}$ there exists a local strategy $\alpha_{\mathcal{E}}$ such that for all $a \in \mathsf{L}$ we have $\mathsf{Succ}(\mu_{\mathcal{D}}, \alpha_{\mathcal{D}}, a) \mathrel{\mathcal{R}} \mathsf{Succ}(\mu_{\mathcal{E}}, \alpha_{\mathcal{E}}, a)$;
- for all local strategies $\alpha_{\mathcal{E}}$ there exists a local strategy $\alpha_{\mathcal{D}}$ such that for all $a \in \mathsf{L}$ we have $\mathsf{Succ}(\mu_{\mathcal{D}}, \alpha_{\mathcal{D}}, a) \mathrel{\mathcal{R}} \mathsf{Succ}(\mu_{\mathcal{E}}, \alpha_{\mathcal{E}}, a)$.

As usual, a union of bisimulations is a bisimulation. Denote by $\sim$ the union of all bisimulations, i.e., $\sim$ is the largest bisimulation. We write $\mathcal{D} \sim \mathcal{E}$ if $\mu_0^{\mathcal{D}} \sim \mu_0^{\mathcal{E}}$.

In general, the set $\sim$ is uncountably infinite, so methods for computing state-based bisimulation (e.g., partition refinement) are not applicable. The following proposition establishes a link between trace refinement and bisimulation.

Proposition 9. *Let $\mathcal{D}$ be an MDP and $\mathcal{C}$ be an MC. Then $\mathcal{D} \sim \mathcal{C}$ if and only if $\mathcal{D} \sqsubseteq \mathcal{C}$.*

We often view a subdistribution $d \in \mathsf{subDist}(Q)$ as a row vector $d \in [0,1]^Q$. For a local strategy α and a label a, define the *transition matrix* $\Delta_\alpha(a) \in [0,1]^{Q \times Q}$ with $\Delta_\alpha(a)[q,q'] = \sum_{\mathsf{m} \in \mathsf{moves}(q)} \alpha(q)(\mathsf{m}) \cdot \mathsf{m}(a,q')$. Viewing subdistributions μ as row vectors, we have:

$$\mathsf{Succ}(\mu, \alpha, a) = \mu \cdot \Delta_\alpha(a) \tag{3}$$

In the following we consider MDPs $\mathcal{D} = \langle Q, \mu_0^{\mathcal{D}}, \mathsf{L}, \delta \rangle$ and $\mathcal{E} = \langle Q, \mu_0^{\mathcal{E}}, \mathsf{L}, \delta \rangle$ over the same state space. This is without loss of generality, since we might take the disjoint union of the state spaces. Since $\mathcal{D}$ and $\mathcal{E}$ differ only in the initial distribution, we will focus on $\mathcal{D}$.

Let $B \in \mathbb{R}^{Q \times k}$ with $k \geq 1$. Assume the label set is $\mathsf{L} = \{a_1, \ldots, a_{|\mathsf{L}|}\}$. For $\mu \in \mathsf{subDist}(Q)$ and a local strategy α we define a point $p(\mu, \alpha) \in \mathbb{R}^{|\mathsf{L}| \cdot k}$ such that

$$p(\mu, \alpha) \; = \; \big(\mu \Delta_\alpha(a_1) B \quad \mu \Delta_\alpha(a_2) B \quad \cdots \quad \mu \Delta_\alpha(a_{|\mathsf{L}|}) B\big).$$

For the reader's intuition, we remark that we will choose matrices $B \in \mathbb{R}^{Q \times k}$ so that if two subdistributions $\mu_{\mathcal{D}}, \mu_{\mathcal{E}}$ are bisimilar then $\mu_{\mathcal{D}} B = \mu_{\mathcal{E}} B$. (In fact, one can compute B so that the converse holds as well, i.e., $\mu_{\mathcal{D}} \sim \mu_{\mathcal{E}}$ if and only if $\mu_{\mathcal{D}} B = \mu_{\mathcal{E}} B$.) It follows that, for subdistributions $\mu_{\mathcal{D}}, \mu_{\mathcal{E}}$ and local strategies $\alpha_{\mathcal{D}}, \alpha_{\mathcal{E}}$, if $\mathsf{Succ}(\mu_{\mathcal{D}}, \alpha_{\mathcal{D}}, a) \sim \mathsf{Succ}(\mu_{\mathcal{E}}, \alpha_{\mathcal{E}}, a)$ holds for all $a \in \mathsf{L}$ then $p(\mu_{\mathcal{D}}, \alpha_{\mathcal{D}}) = p(\mu_{\mathcal{E}}, \alpha_{\mathcal{E}})$. Let us also remark that for fixed $\mu \in \mathsf{subDist}(Q)$, the set $P_\mu = \{p(\mu, \alpha) \mid \alpha \text{ is a local strategy}\} \subseteq \mathbb{R}^{|\mathsf{L}| \cdot k}$ is a (bounded and convex) polytope. As a consequence, if $\mu_{\mathcal{D}} \sim \mu_{\mathcal{E}}$ then the polytopes $P_{\mu_{\mathcal{D}}}$ and $P_{\mu_{\mathcal{E}}}$ must be equal. In the next paragraph we define *"extremal"* strategies $\widehat{\alpha}$, which intuitively are local strategies such that $p(\mu, \widehat{\alpha})$ is a vertex of the polytope P_μ.

Let $\boldsymbol{v} \in \mathbb{R}^{|\mathsf{L}| \cdot k}$ be a *column* vector; we denote column vectors in boldface. We view $\boldsymbol{v}$ as a "direction". Recall that d_q is the Dirac distribution on the state q. A pure local strategy $\widehat{\alpha}$ is *extremal in direction $\boldsymbol{v}$ with respect to B* if

$$p(d_q, \alpha)\boldsymbol{v} \leq p(d_q, \widehat{\alpha})\boldsymbol{v} \tag{4}$$

$$p(d_q, \alpha)\boldsymbol{v} = p(d_q, \widehat{\alpha})\boldsymbol{v} \quad \text{implies} \quad p(d_q, \alpha) = p(d_q, \widehat{\alpha}) \tag{5}$$

for all states $q \in Q$ and all pure local strategies α.

By linearity, if (4) and (5) hold for all pure local strategies α then (4) and (5) hold for all local strategies α. We say a local strategy $\widehat{\alpha}$ is *extremal with respect to B* if there is a direction $\boldsymbol{v}$ such that $\widehat{\alpha}$ is extremal in direction $\boldsymbol{v}$ with respect to B.

Proposition 10. *Let $\mathcal{D} = \langle Q, \mu_0, \mathsf{L}, \delta \rangle$ be an MDP. Let $B_1 \in \mathbb{R}^{Q \times k_1}$ and $B_2 \in \mathbb{R}^{Q \times k_2}$ for $k_1, k_2 \geq 1$. Denote by $\mathcal{V}_1, \mathcal{V}_2 \subseteq \mathbb{R}^Q$ the subspaces spanned by the* *columns of B_1, B_2, respectively. Assume $\mathcal{V}_1 \subseteq \mathcal{V}_2$. If a local strategy $\widehat{\alpha}$ is extremal with respect to B_1 then $\widehat{\alpha}$ is extremal with respect to B_2.*

In light of this fact, we may define that $\widehat{\alpha}$ be *extremal with respect to a column-vector space* $\mathcal{V}$ if $\widehat{\alpha}$ is extremal with respect to a matrix B whose column space equals $\mathcal{V}$.

The following proposition describes a vector space $\mathcal{V}$ so that two subdistributions are bisimilar if and only if their difference (viewed as a row vector) is orthogonal to $\mathcal{V}$.

Proposition 11. *Let $\mathcal{D} = \langle Q, \mu_0, \mathsf{L}, \delta\rangle$ be an MDP. Let $\mathcal{V} \subseteq \mathbb{R}^Q$ be the smallest column-vector space such that*

- *$\mathbf{1} = (1\ 1 \cdots 1)^T \in \mathcal{V}$ (where T denotes transpose) and*
- *$\Delta_{\widehat{\alpha}}(a)\boldsymbol{u} \in \mathcal{V}$ for all $\boldsymbol{u} \in \mathcal{V}$, all labels $a \in \mathsf{L}$ and local strategies $\widehat{\alpha}$ that are extremal with respect to $\mathcal{V}$.*

Then for all $\mu_\mathcal{D}, \mu_\mathcal{E} \in \mathsf{subDist}(Q)$, we have $\mu_\mathcal{D} \sim \mu_\mathcal{E}$ if and only if $\mu_\mathcal{D}\boldsymbol{u} = \mu_\mathcal{E}\boldsymbol{u}$ for all $\boldsymbol{u} \in \mathcal{V}$.

Proposition 11 allows us to prove the following theorem:

Theorem 12. *The problem that, given two MDPs $\mathcal{D}$ and $\mathcal{E}$, asks whether $\mathcal{D} \sim \mathcal{E}$ is in* coNP.

In the following, without loss of generality, we consider an MDP $\mathcal{D} = \langle Q, \mu_0^\mathcal{D}, \mathsf{L}, \delta\rangle$ and an MC $\mathcal{C} = \langle Q_\mathcal{C}, \mu_0^\mathcal{C}, \mathsf{L}, \delta\rangle$ with $Q_\mathcal{C} \subseteq Q$. We may view subdistributions $\mu_\mathcal{C} \in \mathsf{subDist}(Q_\mathcal{C})$ as $\mu_\mathcal{C} \in \mathsf{subDist}(Q)$ in the natural way. The following proposition is analogous to Proposition 11.

Proposition 13. *Let $\mathcal{D} = \langle Q, \mu_0^\mathcal{D}, \mathsf{L}, \delta\rangle$ be an MDP and $\mathcal{C} = \langle Q_\mathcal{C}, \mu_0^\mathcal{C}, \mathsf{L}, \delta\rangle$ be an MC with $Q_\mathcal{C} \subseteq Q$. Let $\mathcal{V} \subseteq \mathbb{R}^Q$ be the smallest column-vector space such that*

- *$\mathbf{1} = (1\ 1 \cdots 1)^T \in \mathcal{V}$ (where T denotes transpose) and*
- *$\Delta_\alpha(a)\boldsymbol{u} \in \mathcal{V}$ for all $\boldsymbol{u} \in \mathcal{V}$, all labels $a \in \mathsf{L}$ and all local strategies α.*

Then for all $\mu_\mathcal{D} \in \mathsf{subDist}(Q)$ and all $\mu_\mathcal{C} \in \mathsf{subDist}(Q_\mathcal{C})$, we have $\mu_\mathcal{D} \sim \mu_\mathcal{C}$ if and only if $\mu_\mathcal{D}\boldsymbol{u} = \mu_\mathcal{C}\boldsymbol{u}$ for all $\boldsymbol{u} \in \mathcal{V}$.

Notice the differences to Proposition 11: there we considered all extremal local strategies (potentially exponentially many), here we consider all local strategies (in general infinitely many). However, we show that one can efficiently find few local strategies that span all local strategies. This allows us to reduce (in logarithmic space) the bisimulation problem between an MDP and an MC to the bisimulation problem between two MCs, which is equivalent to the trace-equivalence problem in MCs (by Proposition 9). The latter problem is known to be in NC [20]. Theorem 8 then follows with Proposition 9.

References

1. S. Arora, R. Ge, R. Kannan, and A. Moitra. Computing a nonnegative matrix factorization - provably. In *STOC*, pages 145–162. ACM, 2012.
2. V. D. Blondel and V. Canterini. Undecidable problems for probabilistic automata of fixed dimension. *Theoretical Computer Science*, 36 (3):231–245, 2003.
3. J. Canny. Some algebraic and geometric computations in PSPACE. In *STOC*, pages 460–467, 1988.
4. J. Cohen and U. Rothblum. Nonnegative ranks, decompositions, and factorizations of nonnegative matrices. *Linear Algebra and its Applications*, 190:149–168, 1993.
5. A. Condon and R. Lipton. On the complexity of space bounded interactive proofs (extended abstract). In *FOCS*, pages 462–467, 1989.
6. T. Cormen, C. Stein, R. Rivest, and C. E. Leiserson. *Introduction to algorithms*. McGraw-Hill Higher Education, 2nd edition, 2001.
7. L. Doyen, T.A. Henzinger, and J.-F. Raskin. Equivalence of labeled Markov chains. *International Journal on Foundations of Computer Science*, 19(3):549–563, 2008.
8. J. Fearnley and M. Jurdzinski. Reachability in two-clock timed automata is pspace-complete. *Information and Computation*, 243:26–36, 2015.
9. R. Greenlaw, H.J. Hoover, and W.L. Ruzzo. *Limits to parallel computation: P-completeness theory*. Oxford University Press, 1995.
10. H. Hermanns, J. Krčál, and J. Křetínský. Probabilistic bisimulation: Naturally on distributions. In *CONCUR*, volume 8704 of *LNCS*, pages 249–265. Springer, 2014. Technical report at `http://arxiv.org/abs/1404.5084`.
11. S. Kiefer, A.S. Murawski, J. Ouaknine, B. Wachter, and J. Worrell. Language equivalence for probabilistic automata. In *CAV*, volume 6806 of *LNCS*, pages 526–540. Springer, 2011.
12. S. Kiefer, A.S. Murawski, J. Ouaknine, B. Wachter, and J. Worrell. APEX: An analyzer for open probabilistic programs. In *CAV*, volume 7358 of *LNCS*, pages 693–698. Springer, 2012.
13. L. Li and Y. Feng. Quantum Markov chains: Description of hybrid systems, decidability of equivalence, and model checking linear-time properties. *Information and Computation*, 244:229–244, 2015.
14. T.M. Ngo, M. Stoelinga, and M. Huisman. Confidentiality for probabilistic multi-threaded programs and its verification. In *Engineering Secure Software and Systems*, volume 7781 of *LNCS*, pages 107–122. Springer, 2013.
15. A. Paz. *Introduction to Probabilistic Automata*. Academic Press, 1971.
16. S. Peyronnet, M. de Rougemont, and Y. Strozecki. Approximate verification and enumeration problems. In *ICTAC*, volume 7521 of *LNCS*, pages 228–242. Springer, 2012.
17. J. Renegar. On the computational complexity and geometry of the first-order theory of the reals. Parts I–III. *Journal of Symbolic Computation*, 13(3):255–352, 1992.
18. M.-P. Schützenberger. On the definition of a family of automata. *Information and Control*, 4:245–270, 1961.
19. W. Tzeng. A polynomial-time algorithm for the equivalence of probabilistic automata. *SIAM Journal on Computing*, 21(2):216–227, 1992.
20. W. Tzeng. On path equivalence of nondeterministic finite automata. *Information Processing Letters*, 58(1):43–46, 1996.
21. S. Vavasis. On the complexity of nonnegative matrix factorization. *SIAM Journal on Optimization*, 20(3):1364–1377, 2009.

A Proofs of Section 2

In this section we prove Lemma 1 from the main text:

Lemma 1. *Let $\mathcal{D}$ be an MDP and α be a strategy. There exists a trace-based strategy β such that $Tr_\alpha = Tr_\beta$.*

Proof. Let α be a strategy of the MDP $\mathcal{D} = \langle Q, \mu_0, \mathsf{L}, \delta \rangle$, and let $\mathsf{Paths}(w, q) = \{\rho \in \mathsf{Paths}(\mathcal{D}) \mid \mathsf{trace}(\rho) = w \text{ and } \mathsf{last}(\rho) = q\}$. We define a trace-based strategy $\beta : \mathsf{L}^* \times Q \to \mathsf{Dist}(\mathsf{moves})$, given a pair of trace w and state q, such that

$$\beta(w, q) = \sum_{\rho \in \mathsf{Paths}(w,q)} \frac{\mathsf{Pr}_{\mathcal{D},\alpha}(\rho)}{subDis_{\mathcal{D},\alpha}(w)(q)} \cdot \alpha(\rho)$$

if $q \in S_{w,\alpha}$; and the uniform distribution on $\mathsf{moves}(q)$ otherwise.

We prove that $subDis_{\mathcal{D},\alpha}(w) = subDis_{\mathcal{D},\beta}(w)$ for all traces $w \in \mathsf{L}^*$. The proof is by induction on the length of w. The induction base holds since for the empty trace ϵ, we have $subDis_{\mathcal{D},\alpha}(\epsilon) = subDis_{\mathcal{D},\beta}(\epsilon) = \mu_0$. For the induction step, we assume that two subdistributions $subDis_{\mathcal{D},\alpha}(v) = subDis_{\mathcal{D},\beta}(v)$ are equal for all traces v with length $|v| \leq k$. Let $w = v \cdot a$ be a trace of length $k+1$. By definition, $subDis_{\mathcal{D},\alpha}(w)(q) = \sum_{\rho \in \mathsf{Paths}(w,q)} \mathsf{Pr}_{\mathcal{D},\alpha}(\rho)$ for all states q. Since each path ρ with the trace w is a continuation of a path ρ' such that $\mathsf{trace}(\rho') = v$, and ρ' ends in some state $q' \in S_{v,\alpha}$ where the label a is generated under the strategy α. We thus have

$$\begin{aligned}
subDis_{\mathcal{D},\alpha}(w)(q) &= \sum_{q' \in S_{v,\alpha}} \sum_{\rho' \in \mathsf{Paths}(v,q')} \mathsf{Pr}_{\mathcal{D},\alpha}(\rho') \sum_{m \in \mathsf{moves}(q')} \alpha(\rho')(\mathsf{m}) \cdot \mathsf{m}(a, q) \\
&= \sum_{q' \in S_{v,\alpha}} \sum_{m \in \mathsf{moves}(q')} \sum_{\rho' \in \mathsf{Paths}(v,q')} \mathsf{Pr}_{\mathcal{D},\alpha}(\rho') \cdot \alpha(\rho')(\mathsf{m}) \cdot \mathsf{m}(a, q) \\
&= \sum_{q' \in S_{v,\alpha}} \sum_{m \in \mathsf{moves}(q')} subDis_{\mathcal{D},\alpha}(v)(q') \cdot \\
&\qquad \left(\sum_{\rho' \in \mathsf{Paths}(v,q')} \frac{\mathsf{Pr}_{\mathcal{D},\alpha}(\rho')}{subDis_{\mathcal{D},\alpha}(v)(q')} \cdot \alpha(\rho')(\mathsf{m}) \cdot \mathsf{m}(a, q) \right)
\end{aligned}$$

and based on the definition of β and the induction hypothesis, we have

$$= \sum_{q' \in S_{v,\beta}} \sum_{m \in \mathsf{moves}(q')} subDis_{\mathcal{D},\beta}(v)(q') \cdot \beta(v, q')(\mathsf{m}) \cdot \mathsf{m}(a, q)$$

and as β is traced-based, then

$$\begin{aligned}
&= \sum_{q' \in S_{v,\beta}} subDis_{\mathcal{D},\beta}(v)(q') \sum_{m \in \mathsf{moves}(q')} \beta(v, q')(\mathsf{m}) \cdot \mathsf{m}(a, q) \\
&= subDis_{\mathcal{D},\beta}(w)(q).
\end{aligned}$$

The proof of the inductive step is complete; we have proved that $subDis_{\mathcal{D},\alpha}(w) = subDis_{\mathcal{D},\beta}(w)$ for all traces $w \in \mathsf{L}^*$.

This implies that $Tr_\alpha = Tr_\beta$, and completes the proof. $\square$

B Proofs of Section 3

Theorem 3. *The problem that, given two MDPs $\mathcal{D}$ and $\mathcal{E}$, asks whether $\mathcal{D} \sqsubseteq \mathcal{E}$ and $\mathcal{E} \sqsubseteq \mathcal{D}$ is undecidable.*

Proof. We reduce the problem $\mathsf{MDP} \sqsubseteq \mathsf{MDP}$ to the problem that asks, given two MDPs $\mathcal{D}_1$ and $\mathcal{D}_2$, whether both $\mathcal{D}_1 \sqsubseteq \mathcal{D}_2$ and $\mathcal{D}_2 \sqsubseteq \mathcal{D}_1$ holds.

The reduction is as follows; given two MDPs $\mathcal{D} = \langle Q, \mu_0, \mathsf{L}, \delta\rangle$ and $\mathcal{E} = \langle Q', \mu_0', \mathsf{L}, \delta'\rangle$, construct the MDP $\mathcal{D}+\mathcal{E}$ by simply having a copy of each MDP, adding a new label $\#$ and a new state p_0. The initial distribution of $\mathcal{D}+\mathcal{E}$ is the Dirac distribution on p_0, where there are two available moves $\mathsf{m}_{\mathcal{D}}$ and $\mathsf{m}_{\mathcal{E}}$. Let $\mathsf{m}_{\mathcal{D}}(\#, q) = \mu_0(q)$ for all $q \in Q$ and $\mathsf{m}_{\mathcal{D}}(\#, q) = 0$ otherwise; and let $\mathsf{m}_{\mathcal{E}}(\#, q) = \mu_0'(q)$ for all $q \in Q'$ and $\mathsf{m}_{\mathcal{D}}(\#, q) = 0$ otherwise. We see that $\mathcal{D}+\mathcal{E}$ always starts by generating label $\#$ with probability 1, and based on the strategic choice can simulate $\mathcal{D}$ or $\mathcal{E}$.

We construct $\mathcal{E}_2$ from $\mathcal{E}$ as follows. We extend the set of labels with the new label $\#$ and set Q' of states with a new state q_0. The new initial distribution of $\mathcal{D}$ is the Dirac distribution on q_0, where there is only one available move m such that $\mathsf{m}(\#, q) = \mu_0'(q)$. We see that $\mathcal{E}_2$ always starts by generating label $\#$ with probability 1, and then simply behaves as $\mathcal{E}$.

We argue that $\mathcal{D} \sqsubseteq \mathcal{E}$ if, and only if, $\mathcal{E}_2 \sqsubseteq \mathcal{D}+\mathcal{E}$ and $\mathcal{D}+\mathcal{E} \sqsubseteq \mathcal{E}_2$. This follows from three simple observations:

- the relation $\mathcal{E}_2 \sqsubseteq \mathcal{D}+\mathcal{E}$ always holds. Strategies of $\mathcal{D}+\mathcal{E}$ can choose to simulate $\mathcal{E}_2$ by playing $\mathsf{m}_{\mathcal{E}}$ with probability 1.
- if $\mathcal{D} \sqsubseteq \mathcal{E}$ then $\mathcal{D}+\mathcal{E} \sqsubseteq \mathcal{E}_2$: a strategy γ of $\mathcal{D}+\mathcal{E}$, at first step, would play $\mathsf{m}_{\mathcal{D}}$ and $\mathsf{m}_{\mathcal{E}}$ with probabilities $\gamma(p_0)(\mathsf{m}_{\mathcal{D}})$ and $\gamma(p_0)(\mathsf{m}_{\mathcal{E}})$. Next, it follows a strategy α for the copy of $\mathcal{D}$ and a strategy β for the copy of $\mathcal{E}$. Since $\mathcal{D} \sqsubseteq \mathcal{E}$, then there exists some strategy β' of $\mathcal{E}$ such that $Tr_{\mathcal{D},\alpha} = Tr_{\mathcal{E},\beta'}$. Then a strategy of $\mathcal{E}_2$ that, at all steps, plays β' with probability $\gamma(p_0)(\mathsf{m}_{\mathcal{D}})$ and plays β with probability $\gamma(p_0)(\mathsf{m}_{\mathcal{E}})$ would result in the same trace-probability function as γ of $\mathcal{D}+\mathcal{E}$ does. It gives $\mathcal{D}+\mathcal{E} \sqsubseteq \mathcal{E}_2$.
- if $\mathcal{D}+\mathcal{E} \sqsubseteq \mathcal{E}_2$ then $\mathcal{D} \sqsubseteq \mathcal{E}$: consider a strategy α of $\mathcal{D}$. Construct strategy α' of $\mathcal{D}+\mathcal{E}$ such that $\alpha'(p_0)(\mathsf{m}_{\mathcal{D}}) = 1$ and $\alpha'(p_0\#\rho) = \alpha(\rho)$ for all paths ρ. Since $\mathcal{D}+\mathcal{E} \sqsubseteq \mathcal{E}_2$, there must be some strategy β' such that $Tr_{\mathcal{D}+\mathcal{E},\alpha'} = Tr_{\mathcal{E}_2,\beta'}$. For the strategy β where $\beta(\rho) = \beta'(q_0\#\rho)$ for all paths ρ, we have $Tr_{\mathcal{D},\alpha} = Tr_{\mathcal{E},\beta}$, and thus $\mathcal{D} \sqsubseteq \mathcal{E}$.

C Proofs of Section 4

C.1 Proofs of Theorems 4 and 5

Theorem 4. *The problem* $\mathsf{MC} \sqsubseteq \mathsf{MDP}_{\mathsf{pm}}$ *is* NP*-complete.*

Proof. Given an instance of the subset-sum problem, a set $\{s_1, s_2, \cdots, s_n\}$ and $N \in \mathbb{N}$, we construct the MC $\mathcal{C}$ and an MDP $\mathcal{D}$ as described in subsection 4.1. We prove that there exists a subset $S \subseteq \{s_1, s_2, \cdots, s_n\}$ where $\sum_{s \in S} s = N$ if and only if $\mathcal{C} \sqsubseteq \mathcal{D}$ when $\mathcal{D}$ uses only pure memoryless strategies.

Observe that, in one hand, $Tr_{\mathcal{C}}(ab^+) = \frac{N}{P}$ and $Tr_{\mathcal{C}}(ac^+) = 1 - \frac{N}{P}$ holds for the MC $\mathcal{C}$. On the other hand, in the MDP $\mathcal{D}$, strategies would purely choose between moves $\mathsf{m}_{i,b}$ and $\mathsf{m}_{i,c}$ in states s_i. For a pure strategy α for $\mathcal{D}$, let S_α be the set of states s_i where $\alpha(s_i) = m_{i,b}$. Then, $Tr_D(ab^+) = \sum_{s \in S_\alpha} \frac{s}{P}$ and $Tr_D(ac^+) = 1 - Tr_D(ab^+)$. The above arguments provide that $\mathcal{C} \sqsubseteq \mathcal{D}$ if and only if there exists a strategy α for $\mathcal{D}$ such that $\sum_{s \in S_\alpha} \frac{s}{P} = \frac{N}{P}$. It implies that the instance of subset problem is positive, meaning that there exists a subset $S \subseteq \{s_1, s_2, \cdots, s_n\}$ such that $\sum_{s \in S} s = N$, if and only if $\mathcal{C} \sqsubseteq \mathcal{D}$ when $\mathcal{D}$ uses only pure memoryless strategies. The NP-hardness results follows.

Theorem 5. *The problem* $\mathsf{MDP_{pm}} \sqsubseteq \mathsf{MDP_{pm}}$ *is* Π_2^p*-complete.*

Proof. Given an instance of the quantified subset-sum problem, two sets $\{s_1, s_2, \cdots, s_n\}$ and $\{t_1, t_2, \cdots, t_m\}$ and $N \in \mathbb{N}$, we construct the MDPs $\mathcal{E}_{univ}$ and $\mathcal{E}_{exist}$ as described in subsection 4.1. We prove that the existential player wins in one round if and only if $\mathcal{E}_{univ} \sqsubseteq \mathcal{E}_{exist}$ where the MDPs use only pure memoryless strategies.

We see that the strategic choices in the MDPs are relevant only in states s_i and t_j. For a pure strategy α of $\mathcal{E}_{univ}$, let S_α be the set of states s_i where $\alpha(s_i) = m_{i,b}$. We therefore have $Tr_{\mathcal{E}_{univ}}(ab^+) = \frac{1}{2}y_1 + \frac{1}{2}\sum_{s \in S_\alpha} \frac{xs}{P}$. For a pure strategy β of $\mathcal{E}_{exist}$, let T_β be the set of states t_j where $\beta(t_j) = m_{i,c}$. Then, $Tr_{\mathcal{E}_{exist}}(ab^+) = \frac{1}{2}y_2 + \frac{1}{2}R(1 - \sum_{t \in T_\beta} \frac{xt}{R})$. Since $y_1 + xN = y_2 + xR$, to achieve $Tr_{\mathcal{E}_{univ}} = Tr_{\mathcal{E}_{exist}}$ the equality $\sum_{s \in S_\alpha} s = N - \sum_{t \in T_\beta} t$ must be guaranteed. It shows that the existential player wins in one round, meaning that for all subsets $S \subseteq \{s_1, s_2, \cdots, s_n\}$ there exists a subset $T \subseteq \{t_1, t_2, \cdots, t_n\}$ such that $\sum_{s \in S} s + \sum_{t \in T} t = N$, if and only if for all strategies α of $\mathcal{D}$ there exists some strategy β for $\mathcal{E}$ such that $\sum_{s \in S_\alpha} s = N - \sum_{t \in T_\beta} t$ implying $Tr_{\mathcal{E}_{univ}} = Tr_{\mathcal{E}_{exist}}$. Note that α and β are chosen pure and memoryless. The Π_2^p-hardness result follows.

C.2 Proof of Theorem 6

In this section we prove Theorem 6 from the main text:

Theorem 6. *The problem* $\mathsf{MC} \sqsubseteq \mathsf{MDP_m}$ *is polynomial-time reducible to the existential theory of the reals, hence in* PSPACE.

Given an MC $\mathcal{C} = \langle Q, \mu_0, \mathsf{L}, \delta \rangle$, to each label $a \in \mathsf{L}$ we associate a *transition matrix* $\Delta(a) \in [0,1]^{Q \times Q}$ with $\Delta(a)[q, q'] = \delta(q)(a, q')$. We view subdistributions μ_0 over states as row vectors $\mu_0 \in [0,1]^Q$. We denote column vectors in boldface; in particular, $\mathbf{1} \in \{1\}^Q$ and $\mathbf{0} \in \{0\}^Q$ are column vectors all whose entries are 1 and 0, respectively. We build on [**?**, Proposition 10] which reads—translated to our framework—as follows:

Proposition 14. *Let $\mathcal{C}_1 = \langle Q_1, \mu_0, \mathsf{L}, \delta\rangle$ and $\mathcal{C}_2 = \langle Q_2, \mu_0', \mathsf{L}, \delta'\rangle$ be MCs with Q as the disjoint union of Q_1, Q_2. Then $Tr_{\mathcal{C}_1} = Tr_{\mathcal{C}_2}$ if and only if there exists a matrix $F \in \mathbb{R}^{Q\times Q}$ such that*

- *the first row of F equals $(\mu_0, -\mu_0')$,*
- *$F\mathbf{1} = \mathbf{0}$,*

and moreover, for all labels $a \in \mathsf{L}$ there exist matrices $M(a) \in \mathbb{R}^{Q\times Q}$ such that

$$F\begin{pmatrix} \Delta(a) & 0 \\ 0 & \Delta'(a)\end{pmatrix} = M(a)F$$

where $\Delta(a), \Delta(a)'$ are the transition matrices of C_1 and $\mathcal{C}_2$ for the label a.

With this at hand we prove Theorem 6:

Proof (of Theorem 6).
Let $\mathcal{C} = \langle Q_1, \mu_0, \mathsf{L}, \delta\rangle$ be an MC and $\mathcal{D} = \langle Q_2, \mu_0', \mathsf{L}, \delta'\rangle$ be an MDP with Q as the disjoint union of Q_1, Q_2. A memoryless strategy α of $\mathcal{D}$ can be characterized by numbers $x_{q,\mathsf{m}} \in [0,1]$ where $q \in Q_2$ and $\mathsf{m} \in \mathsf{moves}(q)$, such that $x_{q,\mathsf{m}} = \alpha(q)(\mathsf{m})$. We have $\sum_{\mathsf{m}\in\mathsf{moves}(q)} x_{q,\mathsf{m}} = 1$ for all states q. We write $\overline{x}$ for the collection $(x_{q,\mathsf{m}})_{q\in Q_2,\ \mathsf{m}\in\mathsf{moves}(q)}$, and $\alpha(\overline{x})$ for the memoryless strategy characterized by $\overline{x}$. We have:

$\mathcal{C} \sqsubseteq \mathcal{D}$ for $\mathcal{D}$ restricted to memoryless strategies	
$\iff \exists$ memoryless strategy $\alpha : Tr_{\mathcal{C}} = Tr_{\mathcal{D},\alpha}$	definition
$\iff$ *Cond*	Proposition 14,

where *Cond* is the following condition:

There exist
- $x_{q,\mathsf{m}} \in [0,1]$ for all $q \in Q_2$ and all $\mathsf{m} \in \mathsf{moves}(q)$
- matrices $M(a) \in \mathbb{R}^{Q\times Q}$ for all labels $a \in \mathsf{L}$,
- a matrix $F \in \mathbb{R}^{Q\times Q}$

such that
- $\sum_{\mathsf{m}\in\mathsf{moves}(q)} x_{q,\mathsf{m}} = 1$ for all $q \in Q_2$,
- the first row of F equals $(\mu_0, -\mu_0')$,
- $F\mathbf{1} = \mathbf{0}$,
- for all labels $a \in \mathsf{L}$,

$$F\begin{pmatrix} \Delta(a) & 0 \\ 0 & \Delta'(a)\end{pmatrix} = M(a)F$$

where $\Delta(a), \Delta'(a)$ are the transition matrices of C and the finite MC $\mathcal{D}(\alpha(\overline{x}))$ induced by $\mathcal{D}$ under the strategy $\alpha(\overline{x})$.

This condition *Cond* is a closed formula in the existential theory of the reals. □

C.3 Proof of Theorem 7

Theorem 7. *The NMF problem is polynomial-time reducible to* $\mathsf{MC} \sqsubseteq \mathsf{MDP_m}$*, hence* $\mathsf{MC} \sqsubseteq \mathsf{MDP_m}$ *is* NP*-hard.*

Proof. Given a nonnegative matrix $M \in \mathbb{R}^{n\times m}$ and rank r, we construct the MC $\mathcal{C}$ and the MDP $\mathcal{D}$ as described in subsection 4.2. We prove that there is a nonnegative factorization for $M = A \cdot W$ such that $A \in \mathbb{R}^{n\times r}$ and $W \in \mathbb{R}^{r\times m}$ if and only if $\mathcal{C} \sqsubseteq \mathcal{D}$ where $\mathcal{D}$ is restricted to use only memoryless strategies.

We establish the correctness of the reduction as follows. First, assume that there is a nonnegative factorization for M. Thus, there are stochastic matrices $A \in \mathbb{R}^{n\times r}$ and $W \in \mathbb{R}^{r\times m}$ such that $M = A \cdot W$. To prove that $\mathcal{C} \sqsubseteq \mathcal{D}$, we construct a memoryless strategy α such that $Tr_{\mathcal{C}} = Tr_{\mathcal{D},\alpha}$. For all states q of $\mathcal{D}$, strategy α is defined by

$$\alpha(q) = \begin{cases} d \in \mathsf{Dist}(\mathsf{moves}(p_i)) & \text{if } q = p_i \text{ and } 1 \le i \le n, \\ \text{where } d(\mathsf{m}_{i,k}) = A[i,k] \text{ for all } 1 \le k \le r & \\ & \\ d \in \mathsf{Dist}(\mathsf{moves}(\ell_k)) & \text{if } q = p_i \text{ and } 1 \le k \le r, \\ \text{where } d(\mathsf{m}'_{k,j}) = W[k,j] \text{ for all } 1 \le j \le m & \\ & \\ \text{the Dirac distribution on } (c, p_{\mathit{fi}}) & \text{if } q = p_{\mathit{fi}}. \end{cases}$$

The trace-probability function for $\mathcal{D}$ and α is such that for all $1 \le i \le n$ and all $1 \le j \le m$, we have $Tr_{\mathcal{D},\alpha}(a_i) = \frac{1}{n}$, and

$$\begin{aligned} Tr_{\mathcal{D},\alpha}(a_i \cdot b_j \cdot c^*) &= \frac{1}{n} \sum_{k=1}^{r} \alpha(p_i)(\mathsf{m}_{i,k}) \cdot \alpha(\ell_k)(\mathsf{m}'_{k,j}) \\ &= \frac{1}{n} \sum_{k=1}^{r} A[i,k] \cdot W[k,j] = \frac{1}{n} M[i,j]. \end{aligned}$$

This gives $Tr_{\mathcal{D},\alpha} = Tr_{\mathcal{C}}$, and thus $\mathcal{C} \sqsubseteq \mathcal{D}$ where $\mathcal{D}$ uses a memoryless strategy.

Second, assume that there exists a memoryless strategy β for the MDP $\mathcal{D}$ such that $Tr_{\mathcal{C}} = Tr_{\mathcal{D},\beta}$. We present a factorization $M = A \cdot W$ where $A \in \mathbb{R}^{n\times r}$ and $W \in \mathbb{R}^{r\times m}$. For all $1 \le i \le n$, $1 \le k \le r$ and $1 \le j \le m$, let

$$A[i,k] = \beta(p_i)(\mathsf{m}_{i,k}) \qquad \text{and} \qquad W[k,j] = \beta(\ell_k)(\mathsf{m}'_{k,j}).$$

Since $\mathcal{D}$ under the strategy β refines $\mathcal{C}$, then for all $1 \le i \le n$ and all $1 \le j \le m$

$$Tr_{\mathcal{D},\beta}(a_i \cdot b_j \cdot c^*) = Tr_{\mathcal{C}}(a_i \cdot b_j \cdot c^*) = \frac{1}{n} M[i,j].$$

Since the probability of generating $a_i \cdot b_j \cdot c^*$ is $\frac{1}{n} \sum_{k=1}^{r} \beta(p_i)(\mathsf{m}_{i,k}) \cdot \beta(\ell_k)(\mathsf{m}'_{k,j})$ then we have $\sum_{k=1}^{r} A[i,k] \cdot W[k,j] = M[i,j]$. This completes the proof. ◻

D Proofs of Section 5

D.1 Proof of Proposition 9

For a trace-based strategy $\alpha : \mathsf{L}^* \times Q \to \mathsf{Dist}(\mathsf{moves})$ and a trace $w \in \mathsf{L}^*$, define the local strategy $\alpha[w] : Q \to \mathsf{Dist}(\mathsf{moves})$ with $\alpha[w](q) = \alpha(w, q)$ for all $q \in Q$. We have the following lemma.

Lemma 15. *Let $\mathcal{D} = \langle Q, \mu_0, \mathsf{L}, \delta \rangle$ be an MDP. Let $\alpha : \mathsf{L}^* \times Q \to \mathsf{Dist}(\mathsf{moves})$ be a trace-based strategy. Let $w \in \mathsf{L}^*$ and $a \in \mathsf{L}$. Then:*

$$subDis_{\mathcal{D},\alpha}(wa) = subDis_{\mathcal{D},\alpha}(w) \cdot \mathit{\Delta}_{\alpha[w]}(a)$$

Proof. Let $q' \in Q$. We have:

$$\begin{aligned}
& subDis_{\mathcal{D},\alpha}(wa)(q') \\
&= \sum_{\rho \in \mathsf{Paths}(w)} \mathsf{Pr}_{\mathcal{D},\alpha}(\rho a q') && \text{definition of } subDis \\
&= \sum_{q \in Q} \sum_{\substack{\rho \in \mathsf{Paths}(w) \\ \mathsf{last}(\rho) = q}} \mathsf{Pr}_{\mathcal{D},\alpha}(\rho) \cdot \sum_{\mathsf{m} \in \mathsf{moves}(q)} \alpha(\rho)(\mathsf{m}) \cdot \mathsf{m}(a, q') && \text{definition of } \mathsf{Pr} \\
&= \sum_{q \in Q} \sum_{\substack{\rho \in \mathsf{Paths}(w) \\ \mathsf{last}(\rho) = q}} \mathsf{Pr}_{\mathcal{D},\alpha}(\rho) \cdot \sum_{\mathsf{m} \in \mathsf{moves}(q)} \alpha(w, q)(\mathsf{m}) \cdot \mathsf{m}(a, q') && \alpha \text{ is trace-based} \\
&= \sum_{q \in Q} subDis_{\mathcal{D},\alpha}(w)(q) \cdot \sum_{\mathsf{m} \in \mathsf{moves}(q)} \alpha(w, q)(\mathsf{m}) \cdot \mathsf{m}(a, q') && \text{definition of } subDis \\
&= \sum_{q \in Q} subDis_{\mathcal{D},\alpha}(w)(q) \cdot \sum_{\mathsf{m} \in \mathsf{moves}(q)} \alpha[w](q)(\mathsf{m}) \cdot \mathsf{m}(a, q') && \text{definition of } \alpha[w] \\
&= \sum_{q \in Q} subDis_{\mathcal{D},\alpha}(w)(q) \cdot \mathit{\Delta}_{\alpha[w]}(a)[q, q'] && \text{definition of } \mathit{\Delta}_{\alpha[w]}(a) \\
&= \left(subDis_{\mathcal{D},\alpha}(w) \cdot \mathit{\Delta}_{\alpha[w]}(a) \right) (q')
\end{aligned}$$

□

The following lemma allows us to view strategies as a composition of local strategies, and conversely.

Lemma 16. *Let $\mathcal{D} = \langle Q, \mu_0, \mathsf{L}, \delta \rangle$ be an MDP. Let $w = a_1 a_2 \cdots a_n \in \mathsf{L}^*$. Let $\mu_1, \mu_2, \ldots, \mu_n$ be subdistributions over Q. Then there is a strategy $\alpha : \mathsf{Paths}(\mathcal{D}) \to \mathsf{Dist}(\mathsf{moves})$ with*

$$\mu_i = subDis_{\mathcal{D},\alpha}(a_1 a_2 \cdots a_i) \qquad \textit{for all } i \in \{0, 1, \ldots, n\}$$

if and only if there are local strategies $\alpha_0, \alpha_1, \ldots, \alpha_{n-1} : Q \to \mathsf{Dist}(\mathsf{moves})$ with

$$\mu_{i+1} = \mathsf{Succ}(\mu_i, \alpha_i, a_{i+1}) \qquad \textit{for all } i \in \{0, 1, \ldots, n-1\}.$$

Proof. We prove the two implications from the lemma in turn.

"$\Longrightarrow$": Let α be a strategy with $\mu_i = \mathit{subDis}_{\mathcal{D},\alpha}(a_1 a_2 \cdots a_i)$ for all $i \in \{0,1,\ldots,n\}$. By Lemma 1 we can assume that α is trace-based. For all $i \in \{0,1,\ldots,n-1\}$ define a local strategy α_i with $\alpha_i = \alpha[a_1 a_2 \cdots a_i]$. Then we have for all $i \in \{0,1,\ldots,n-1\}$:

$$\begin{aligned}
\mu_{i+1} &= \mathit{subDis}_{\mathcal{D},\alpha}(a_1 a_2 \cdots a_{i+1}) && \text{definition of } \alpha \\
&= \mathit{subDis}_{\mathcal{D},\alpha}(a_1 a_2 \cdots a_i) \cdot \mathit{\Delta}_{\alpha[a_1 a_2 \cdots a_i]}(a_{i+1}) && \text{Lemma 15} \\
&= \mu_i \cdot \mathit{\Delta}_{\alpha_i}(a_{i+1}) && \text{definitions of } \alpha, \alpha_i \\
&= \mathsf{Succ}(\mu_i, \alpha_i, a_{i+1}) && \text{by (3)}
\end{aligned}$$

"$\Longleftarrow$": Let $\alpha_0, \alpha_1, \ldots, \alpha_{n-1}$ be local strategies with $\mu_{i+1} = \mathsf{Succ}(\mu_i, \alpha_i, a_{i+1})$ for all $i \in \{0,1,\ldots,n-1\}$. Define a trace-based strategy α such that $\alpha[a_1 a_2 \cdots a_i] = \alpha_i$ for all $i \in \{0,1,\ldots,n-1\}$. (This condition need not completely determine α.) We prove by induction on i that $\mu_i = \mathit{subDis}_{\mathcal{D},\alpha}(a_1 a_2 \cdots a_i)$ for all $i \in \{0,1,\ldots,n\}$. For $i = 0$ this is trivial. For the step, we have:

$$\begin{aligned}
\mu_{i+1} &= \mathsf{Succ}(\mu_i, \alpha_i, a_{i+1}) && \text{definition of } \alpha_i \\
&= \mu_i \cdot \mathit{\Delta}_{\alpha_i}(a_{i+1}) && \text{by (3)} \\
&= \mathit{subDis}_{\mathcal{D},\alpha}(a_1 a_2 \cdots a_i) \cdot \mathit{\Delta}_{\alpha_i}(a_{i+1}) && \text{induction hypothesis} \\
&= \mathit{subDis}_{\mathcal{D},\alpha}(a_1 a_2 \cdots a_i) \cdot \mathit{\Delta}_{\alpha[a_1 a_2 \cdots a_i]}(a_{i+1}) && \text{definition of } \alpha \\
&= \mathit{subDis}_{\mathcal{D},\alpha}(a_1 a_2 \cdots a_{i+1}) && \text{Lemma 15}
\end{aligned}$$

$\square$

Now we can prove Proposition 9 from the main text:

Proposition 9. *Let $\mathcal{D}$ be an MDP and $\mathcal{C}$ be an MC. Then $\mathcal{D} \sim \mathcal{C}$ if and only if $\mathcal{D} \sqsubseteq \mathcal{C}$.*

Proof. Let $\mathcal{D} = \langle Q_{\mathcal{D}}, \mu_0^{\mathcal{D}}, \mathsf{L}, \delta^{\mathcal{D}} \rangle$ and $\mathcal{C} = \langle Q_{\mathcal{C}}, \mu_0^{\mathcal{C}}, \mathsf{L}, \delta^{\mathcal{C}} \rangle$.

"$\Longrightarrow$": Let $\mathcal{D} \sim \mathcal{C}$. Hence $\mu_0^{\mathcal{D}} \sim \mu_0^{\mathcal{C}}$. We show that $\mathcal{D} \sqsubseteq \mathcal{C}$. Let $\alpha^{\mathcal{D}}$ be a strategy for $\mathcal{D}$. Let $w = a_1 a_2 \cdots a_n \in \mathsf{L}^*$. Let $\mu_0^{\mathcal{D}}, \mu_1^{\mathcal{D}}, \ldots, \mu_n^{\mathcal{D}}$ be the subdistributions with $\mu_i^{\mathcal{D}} = \mathit{subDis}_{\mathcal{D},\alpha^{\mathcal{D}}}(a_1 a_2 \cdots a_i)$ for all i. By Lemma 16 there exist local strategies $\alpha_0^{\mathcal{D}}, \alpha_1^{\mathcal{D}}, \ldots, \alpha_{n-1}^{\mathcal{D}}$ with $\mu_{i+1}^{\mathcal{D}} = \mathsf{Succ}(\mu_i^{\mathcal{D}}, \alpha_i^{\mathcal{D}}, a_{i+1})$ for all i. Since $\mu_0^{\mathcal{D}} \sim \mu_0^{\mathcal{C}}$, there exist local strategies $\alpha_0^{\mathcal{C}}, \alpha_1^{\mathcal{C}}, \ldots, \alpha_{n-1}^{\mathcal{C}}$ for $\mathcal{C}$ and subdistributions $\mu_1^{\mathcal{C}}, \mu_2^{\mathcal{C}}, \ldots, \mu_n^{\mathcal{C}}$ with $\mu_{i+1}^{\mathcal{C}} = \mathsf{Succ}(\mu_i^{\mathcal{C}}, \alpha_i^{\mathcal{C}}, a_{i+1})$ for all i and $\mu_i^{\mathcal{D}} \sim \mu_i^{\mathcal{C}}$ for all i. Since $\mathcal{C}$ is an MC, the local strategies $\alpha_i^{\mathcal{C}}$ are, in fact, irrelevant. By Lemma 16 we have $\mu_i^{\mathcal{C}} = \mathit{subDis}_{\mathcal{C}}(a_1 a_2 \cdots a_i)$ for all i. So we have:

$$\begin{aligned}
\mathit{Tr}_{\mathcal{D},\alpha^{\mathcal{D}}}(w) &= \|\mathit{subDis}_{\mathcal{D},\alpha^{\mathcal{D}}}(w)\| && \text{by (1)} \\
&= \|\mu_n^{\mathcal{D}}\| && \mu_n^{\mathcal{D}} = \mathit{subDis}_{\mathcal{D},\alpha^{\mathcal{D}}}(w) \\
&= \|\mu_n^{\mathcal{C}}\| && \mu_n^{\mathcal{D}} \sim \mu_n^{\mathcal{C}} \\
&= \|\mathit{subDis}_{\mathcal{C}}(w)\| && \mu_n^{\mathcal{C}} = \mathit{subDis}_{\mathcal{C}}(w) \\
&= \mathit{Tr}_{\mathcal{C}}(w) && \text{by (1)}
\end{aligned}$$

Since $\alpha^{\mathcal{D}}$ and w were chosen arbitrarily, we conclude that $\mathcal{D} \sqsubseteq \mathcal{C}$.

"$\Longleftarrow$": Let $\mathcal{D} \sqsubseteq \mathcal{C}$. We show $\mu_0^{\mathcal{D}} \sim \mu_0^{\mathcal{C}}$. Define a relation $\mathcal{R} \subseteq \mathsf{subDist}(Q_{\mathcal{D}}) \times \mathsf{subDist}(Q_{\mathcal{C}})$ such that $\mu^{\mathcal{D}} \ \mathcal{R} \ \mu^{\mathcal{C}}$ if and only if there exist a strategy $\alpha^{\mathcal{D}}$ for $\mathcal{D}$ and a trace w with $\mu^{\mathcal{D}} = \mathit{subDis}_{\mathcal{D},\alpha^{\mathcal{D}}}(w)$ and $\mu^{\mathcal{C}} = \mathit{subDis}_{\mathcal{C}}(w)$. We claim that $\mathcal{R}$ is a bisimulation. To prove the claim, consider any $\mu^{\mathcal{D}}, \mu^{\mathcal{C}}$ with $\mu^{\mathcal{D}} \ \mathcal{R} \ \mu^{\mathcal{C}}$. Then there exist a strategy $\alpha^{\mathcal{D}}$ for $\mathcal{D}$ and a trace w with $\mu^{\mathcal{D}} = \mathit{subDis}_{\mathcal{D},\alpha^{\mathcal{D}}}(w)$ and $\mu^{C} = \mathit{subDis}_{\mathcal{C}}(w)$. Since $\mathcal{D} \sqsubseteq \mathcal{C}$, we have $\mathit{Tr}_{\mathcal{D},\alpha^{\mathcal{D}}}(w) = \mathit{Tr}_{\mathcal{C}}(w)$. So we have:

$$
\begin{aligned}
\|\mu^{\mathcal{D}}\| &= \|\mathit{subDis}_{\mathcal{D},\alpha^{\mathcal{D}}}(w)\| && \mu^{\mathcal{D}} = \mathit{subDis}_{\mathcal{D},\alpha^{\mathcal{D}}}(w) \\
&= \mathit{Tr}_{\mathcal{D},\alpha^{\mathcal{D}}}(w) && \text{by (1)} \\
&= \mathit{Tr}_{\mathcal{C}}(w) && \text{as argued above} \\
&= \|\mathit{subDis}_{\mathcal{C}}(w)\| && \text{by (1)} \\
&= \|\mu^{\mathcal{C}}\| && \mu^{\mathcal{C}} = \mathit{subDis}_{\mathcal{C}}(w)
\end{aligned}
$$

This proves the first condition for $\mathcal{R}$ being a bisimulation.

For the rest of the proof assume $w = a_1 a_2 \cdots a_n$. Write $\mu_n^{\mathcal{D}} = \mu^{\mathcal{D}}$ and $\mu_n^{\mathcal{C}} = \mu^{\mathcal{C}}$. Let $\alpha_n^{\mathcal{D}}$ be a local strategy for $\mathcal{D}$. Let $a_{n+1} \in \mathsf{L}$. Define $\mu_{n+1}^{\mathcal{D}} = \mathsf{Succ}(\mu_n^{\mathcal{D}}, \alpha_n^{\mathcal{D}}, a_{n+1})$, and $\mu_{n+1}^{\mathcal{C}} = \mathsf{Succ}(\mu_n^{\mathcal{C}}, \alpha_n^{\mathcal{C}}, a_{n+1})$ for an arbitrary (and unimportant) local strategy $\alpha_n^{\mathcal{C}}$ for $\mathcal{C}$. For the second and the third condition of $\mathcal{R}$ being a bisimulation we need to prove $\mu_{n+1}^{\mathcal{D}} \ \mathcal{R} \ \mu_{n+1}^{\mathcal{C}}$. Define $\mu_1^{\mathcal{D}}, \mu_2^{\mathcal{D}}, \ldots, \mu_{n-1}^{\mathcal{D}}$ such that $\mu_i^{\mathcal{D}} = \mathit{subDis}_{\mathcal{D},\alpha^{\mathcal{D}}}(a_1 a_2 \cdots a_i)$ for all $i \in \{0, 1, \ldots, n\}$. By Lemma 16 there are local strategies $\alpha_0^{\mathcal{D}}, \alpha_1^{\mathcal{D}}, \ldots, \alpha_{n-1}^{\mathcal{D}}$ such that $\mu_{i+1}^{\mathcal{D}} = \mathsf{Succ}(\mu_i^{\mathcal{D}}, \alpha_i^{\mathcal{D}}, a_{i+1})$ for all $i \in \{0, 1, \ldots, n-1\}$. We also have $\mu_{n+1}^{\mathcal{D}} = \mathsf{Succ}(\mu_n^{\mathcal{D}}, \alpha_n^{\mathcal{D}}, a_{n+1})$, so again by Lemma 16 there is a strategy $\beta^{\mathcal{D}}$ with $\mu_i^{\mathcal{D}} = \mathit{subDis}_{\mathcal{D},\beta^{\mathcal{D}}}(a_1 a_2 \cdots a_i)$ for all $i \in \{0, 1, \ldots, n+1\}$. In particular, $\mu_{n+1}^{\mathcal{D}} = \mathit{subDis}_{\mathcal{D},\beta^{\mathcal{D}}}(w a_{n+1})$. Similarly, we have $\mu_{n+1}^{\mathcal{C}} = \mathit{subDis}_{\mathcal{C}}(w a_{n+1})$. Thus, $\mu_{n+1}^{\mathcal{D}} \ \mathcal{R} \ \mu_{n+1}^{\mathcal{C}}$. Hence we have proved that $\mathcal{R}$ is a bisimulation.

Considering the empty trace, we see that $\mu_0^{\mathcal{D}} \ \mathcal{R} \ \mu_0^{\mathcal{C}}$. Since $\mathcal{R} \subseteq \sim$, we also have $\mu_0^{\mathcal{D}} \sim \mu_0^{\mathcal{C}}$, as desired.

□

D.2 On the Notion of *Extremal with Respect to a Vector Space*

We prove here:

Proposition 10. *Let $\mathcal{D} = \langle Q, \mu_0, \mathsf{L}, \delta \rangle$ be an MDP. Let $B_1 \in \mathbb{R}^{Q \times k_1}$ and $B_2 \in \mathbb{R}^{Q \times k_2}$ for $k_1, k_2 \geq 1$. Denote by $\mathcal{V}_1, \mathcal{V}_2 \subseteq \mathbb{R}^Q$ the subspaces spanned by the columns of B_1, B_2, respectively. Assume $\mathcal{V}_1 \subseteq \mathcal{V}_2$. If a local strategy $\widehat{\alpha}$ is extremal with respect to B_1 then $\widehat{\alpha}$ is extremal with respect to B_2.*

Proof. Let $\boldsymbol{v}_1 \in \mathbb{R}^{|\mathsf{L}| \cdot k_1}$ be a direction in which $\widehat{\alpha}$ is extremal with respect to B_1. Since $\mathcal{V}_1 \subseteq \mathcal{V}_2$, there exists a matrix $T \in \mathbb{R}^{k_2 \times k_1}$ with $B_1 = B_2 T$. Define columns

$\boldsymbol{v}_{1,1}, \boldsymbol{v}_{1,2}, \ldots, \boldsymbol{v}_{1,|\mathsf{L}|} \in \mathbb{R}^{k_1}$ such that:

$$\boldsymbol{v}_1 = \begin{pmatrix} \boldsymbol{v}_{1,1} \\ \boldsymbol{v}_{1,2} \\ \vdots \\ \boldsymbol{v}_{1,|\mathsf{L}|} \end{pmatrix}$$

Define $\boldsymbol{v}_2 \in \mathbb{R}^{|\mathsf{L}| \cdot k_2}$ by:

$$\boldsymbol{v}_2 = \begin{pmatrix} T\boldsymbol{v}_{1,1} \\ T\boldsymbol{v}_{1,2} \\ \vdots \\ T\boldsymbol{v}_{1,|\mathsf{L}|} \end{pmatrix}$$

For $\mu \in \mathsf{subDist}(Q)$ and $\alpha : Q \to \mathsf{Dist}(\mathsf{moves})$, let us write $p_1(\mu, \alpha) \in \mathbb{R}^{|\mathsf{L}| \cdot k_1}$ (resp., $p_2(\mu, \alpha) \in \mathbb{R}^{|\mathsf{L}| \cdot k_2}$) for the point $p(\mu, \alpha)$ defined in terms of B_1 (resp., B_2). We have:

$$\begin{aligned} p_1(\mu, \alpha)\boldsymbol{v}_1 &= \sum_{i=1}^{|\mathsf{L}|} \mu \Delta_\alpha(a_i) B_1 \boldsymbol{v}_{1,i} && \text{definitions of } p_1 \text{ and } \boldsymbol{v}_{1,i} \\ &= \sum_{i=1}^{|\mathsf{L}|} \mu \Delta_\alpha(a_i) B_2 T \boldsymbol{v}_{1,i} && B_1 = B_2 T \\ &= p_2(\mu, \alpha)\boldsymbol{v}_2 && \text{definitions of } p_2 \text{ and } \boldsymbol{v}_2 \end{aligned}$$

It follows that $\widehat{\alpha}$ is extremal in direction $\boldsymbol{v}_2$ with respect to B_2. □

D.3 Further Geometrical Facts about Extremal Strategies

In this section we prove facts about extremal local strategies that will be needed later.

Lemma 17. *Let $\mathcal{D} = \langle Q, \mu_0, \mathsf{L}, \delta \rangle$ be an MDP. Let $B \in \mathbb{R}^{Q \times k}$ with $k \geq 1$. Let $\mu \in \mathsf{subDist}(Q)$. Let $\alpha, \widehat{\alpha}$ be local strategies. Suppose $\boldsymbol{v} \in \mathbb{R}^{|\mathsf{L}| \cdot k}$ is a direction in which $\widehat{\alpha}$ is extremal and $p(\mu, \alpha)\boldsymbol{v} = p(\mu, \widehat{\alpha})\boldsymbol{v}$. Then $p(\mu, \alpha) = p(\mu, \widehat{\alpha})$.*

Proof. We have:

$$\begin{aligned} \sum_{q \in Q} \mu(q) \cdot p(d_q, \alpha)\boldsymbol{v} &= p(\mu, \alpha)\boldsymbol{v} && \text{definition of } p \\ &= p(\mu, \widehat{\alpha})\boldsymbol{v} && \text{assumption on } \widehat{\alpha} \\ &= \sum_{q \in Q} \mu(q) \cdot p(d_q, \widehat{\alpha})\boldsymbol{v} && \text{definition of } p \end{aligned}$$

With (4) it follows that for all $q \in \mathsf{Supp}(\mu)$ we have $p(d_q, \alpha)\boldsymbol{v} = p(d_q, \widehat{\alpha})\boldsymbol{v}$. Hence by (5) we obtain $p(d_q, \alpha) = p(d_q, \widehat{\alpha})$ for all $q \in \mathsf{Supp}(\mu)$. Thus:

$$p(\mu, \alpha) = \sum_{q \in Q} \mu(q) \cdot p(d_q, \alpha) = \sum_{q \in Q} \mu(q) \cdot p(d_q, \widehat{\alpha}) = p(\mu, \widehat{\alpha})$$

$\square$

For a subdistribution μ define the bounded, convex polytope $P_\mu \subseteq \mathbb{R}^{|\mathsf{L}| \cdot k}$ with

$$P_\mu = \{p(\mu, \alpha) \mid \alpha : Q \to \mathsf{Dist}(\mathsf{moves})\}.$$

Comparing two polytopes $P_{\mu_{\mathcal{D}}}$ and $P_{\mu_{\mathcal{E}}}$ for subdistributions $\mu_{\mathcal{D}}, \mu_{\mathcal{E}}$ will play a key role for deciding bisimulation. First we prove the following lemma, which states that any vertex of the polytope P_μ can be obtained by applying an extremal local strategy. Although this is intuitive, the proof is not very easy.

Lemma 18. *Let $\mathcal{D} = \langle Q, \mu_0, \mathsf{L}, \delta \rangle$ be an MDP. Let $B \in \mathbb{R}^{Q \times k}$ with $k \geq 1$. Let $\mu \in \mathsf{subDist}(Q)$. If $x \in P_\mu$ is a vertex of P_μ then there is an extremal local strategy $\widehat{\alpha}$ with $x = p(\mu, \widehat{\alpha})$.*

Proof. Let $x \in P_\mu$ be a vertex of P_μ. Let $\alpha_1 : Q \to \mathsf{Dist}(\mathsf{moves})$ be a local strategy so that $x = p(\mu, \alpha_1)$. Since x is a vertex, we can assume that α_1 is pure. Since x is a vertex of P_μ, there is a hyperplane $H \subseteq \mathbb{R}^{|\mathsf{L}| \cdot k}$ such that $\{x\} = P_\mu \cap H$. Let $\boldsymbol{v}_1 \in \mathbb{R}^{|\mathsf{L}| \cdot k}$ be a normal vector of H. Since $\{x\} = P_\mu \cap H$, we have $x\boldsymbol{v}_1 = \max_{y \in P_\mu} y\boldsymbol{v}_1$ or $x\boldsymbol{v}_1 = \min_{y \in P_\mu} y\boldsymbol{v}_1$; without loss of generality, say $x\boldsymbol{v}_1 = \max_{y \in P_\mu} y\boldsymbol{v}_1$. Since $\{x\} = P_\mu \cap H$, we have for all $q \in \mathsf{Supp}(\mu)$ and all α:

$$p(d_q, \alpha)\boldsymbol{v}_1 = p(d_q, \alpha_1)\boldsymbol{v}_1 \quad \text{implies} \quad p(d_q, \alpha) = p(d_q, \alpha_1). \tag{6}$$

For all $q \in Q \backslash \mathsf{Supp}(\mu)$, redefine the pure local strategy $\alpha_1(q)$ so that all $q \in Q$ and all local strategies α satisfy $p(d_q, \alpha)\boldsymbol{v}_1 \leq p(d_q, \alpha_1)\boldsymbol{v}_1$. Since Q and moves are finite, there is $\varepsilon > 0$ such that all $q \in Q$ and all *pure* local strategies α either satisfy $p(d_q, \alpha)\boldsymbol{v}_1 = p(d_q, \alpha_1)\boldsymbol{v}_1$ or $p(d_q, \alpha)\boldsymbol{v}_1 \leq p(d_q, \alpha_1)\boldsymbol{v}_1 - \varepsilon$.

Define

$$\Sigma = \{\alpha : Q \to \mathsf{Dist}(\mathsf{moves}) \mid \forall\, q \in Q : p(d_q, \alpha)\boldsymbol{v}_1 = p(d_q, \alpha_1)\boldsymbol{v}_1\}.$$

Consider the bounded, convex polytope $P_2 \subseteq \mathbb{R}^{|\mathsf{L}| \cdot k}$ defined by

$$P_2 = \left\{ \sum_{q \in Q} p(d_q, \alpha) \;\middle|\; \alpha \in \Sigma \right\}.$$

By an argument similar to the one above, there are a pure local strategy $\widehat{\alpha} \in \Sigma$, a vertex $x_2 = \sum_{q \in Q} p(d_q, \widehat{\alpha})$ of P_2, and a vector $\boldsymbol{v}_2 \in \mathbb{R}^{|\mathsf{L}| \cdot k}$ such that for all $q \in Q$ and all $\alpha \in \Sigma$, we have $p(d_q, \alpha)\boldsymbol{v}_2 \leq p(d_q, \widehat{\alpha})\boldsymbol{v}_2$, and if $p(d_q, \alpha)\boldsymbol{v}_2 = p(d_q, \widehat{\alpha})\boldsymbol{v}_2$ then $p(d_q, \alpha) = p(d_q, \widehat{\alpha})$. By scaling down $\boldsymbol{v}_2$ by a small positive scalar, we can assume that all $q \in Q$ and all local strategies α satisfy

$$|p(d_q, \alpha)\boldsymbol{v}_2| \leq \frac{\varepsilon}{3}. \tag{7}$$

Since $\widehat{\alpha} \in \Sigma$, all $q \in Q$ satisfy $p(d_q, \widehat{\alpha})\boldsymbol{v}_1 = p(d_q, \alpha_1)\boldsymbol{v}_1$. By (6) all $q \in \mathsf{Supp}(\mu_{\mathcal{D}})$ satisfy $p(d_q, \widehat{\alpha}) = p(d_q, \alpha_1)$. Hence:

$$p(\mu_{\mathcal{D}}, \widehat{\alpha}) = \sum_{q \in \mathsf{Supp}(\mu_{\mathcal{D}})} \mu_{\mathcal{D}}(q) p(d_q, \widehat{\alpha}) = \sum_{q \in \mathsf{Supp}(\mu_{\mathcal{D}})} \mu_{\mathcal{D}}(q) p(d_q, \alpha_1) = p(\mu_{\mathcal{D}}, \alpha_1) = x$$

It remains to show that there is a direction $\boldsymbol{v}$ in which $\widehat{\alpha}$ is extremal. Take $\boldsymbol{v} = \boldsymbol{v}_1 + \boldsymbol{v}_2$. Let $q \in Q$ and let α be a pure local strategy. We consider two cases:

- Assume $p(d_q, \alpha)\boldsymbol{v}_1 = p(d_q, \alpha_1)\boldsymbol{v}_1$. Then there is $\beta \in \Sigma$ with $\alpha(q) = \beta(q)$, hence $p(d_q, \alpha) = p(d_q, \beta)$. We have:

$$\begin{aligned}
p(d_q, \alpha)\boldsymbol{v} &= p(d_q, \beta)\boldsymbol{v} && p(d_q, \alpha) = p(d_q, \beta) \\
&= p(d_q, \beta)\boldsymbol{v}_1 + p(d_q, \beta)\boldsymbol{v}_2 && \text{definition of } \boldsymbol{v} \\
&= p(d_q, \alpha_1)\boldsymbol{v}_1 + p(d_q, \beta)\boldsymbol{v}_2 && \beta \in \Sigma \\
&= p(d_q, \widehat{\alpha})\boldsymbol{v}_1 + p(d_q, \beta)\boldsymbol{v}_2 && \widehat{\alpha} \in \Sigma \\
&\leq p(d_q, \widehat{\alpha})\boldsymbol{v}_1 + p(d_q, \widehat{\alpha})\boldsymbol{v}_2 && \text{definition of } \widehat{\alpha} \\
&= p(d_q, \widehat{\alpha})\boldsymbol{v} && \text{definition of } \boldsymbol{v}
\end{aligned}$$

  Hence (4) holds for $\widehat{\alpha}$. To show (5), assume $p(d_q, \alpha)\boldsymbol{v} = p(d_q, \widehat{\alpha})\boldsymbol{v}$. Then all terms in the computation above are equal, and $p(d_q, \beta)\boldsymbol{v}_2 = p(d_q, \widehat{\alpha})\boldsymbol{v}_2$. By the definition of $\widehat{\alpha}$, this implies $p(d_q, \beta) = p(d_q, \widehat{\alpha})$. Hence $p(d_q, \alpha) = p(d_q, \beta) = p(d_q, \widehat{\alpha})$. Hence (5) holds for $\widehat{\alpha}$.
- Assume $p(d_q, \alpha)\boldsymbol{v}_1 \neq p(d_q, \alpha_1)\boldsymbol{v}_1$. By the definition of ε it follows $p(d_q, \alpha)\boldsymbol{v}_1 \leq p(d_q, \alpha_1)\boldsymbol{v}_1 - \varepsilon$. We have:

$$\begin{aligned}
p(d_q, \alpha)\boldsymbol{v} &= p(d_q, \alpha)\boldsymbol{v}_1 + p(d_q, \alpha)\boldsymbol{v}_2 && \text{definition of } \boldsymbol{v} \\
&\leq p(d_q, \alpha_1)\boldsymbol{v}_1 - \varepsilon + p(d_q, \alpha)\boldsymbol{v}_2 && \text{as argued above} \\
&= p(d_q, \widehat{\alpha})\boldsymbol{v}_1 - \varepsilon + p(d_q, \alpha)\boldsymbol{v}_2 && \widehat{\alpha} \in \Sigma \\
&\leq p(d_q, \widehat{\alpha})\boldsymbol{v}_1 - \varepsilon + \frac{\varepsilon}{3} && \text{by (7)} \\
&\leq p(d_q, \widehat{\alpha})\boldsymbol{v}_1 + p(d_q, \widehat{\alpha})\boldsymbol{v}_2 - \varepsilon + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} && \text{by (7)} \\
&< p(d_q, \widehat{\alpha})\boldsymbol{v}_1 + p(d_q, \widehat{\alpha})\boldsymbol{v}_2 && \varepsilon > 0 \\
&= p(d_q, \widehat{\alpha})\boldsymbol{v} && \text{definition of } \boldsymbol{v}
\end{aligned}$$

  This implies (4) and (5) for $\widehat{\alpha}$.

Hence, $\widehat{\alpha}$ is extremal in direction $\boldsymbol{v}$. □

The following lemma states the intuitive fact that in order to compare the polytopes $P_{\mu_{\mathcal{D}}}$ and $P_{\mu_{\mathcal{E}}}$, it suffices to compare the vertices obtained by applying extremal local strategies:

Lemma 19. *Let $\mathcal{D} = \langle Q, \mu_0, \mathsf{L}, \delta \rangle$ be an MDP. Let $B \in \mathbb{R}^{Q \times k}$ with $k \geq 1$. Then for all $\mu_{\mathcal{D}}, \mu_{\mathcal{E}} \in \mathsf{subDist}(Q)$ we have $P_{\mu_{\mathcal{D}}} = P_{\mu_{\mathcal{E}}}$ if and only if for all extremal local strategies $\widehat{\alpha}$ we have $p(\mu_{\mathcal{D}}, \widehat{\alpha}) = p(\mu_{\mathcal{E}}, \widehat{\alpha})$.*

Proof. We prove the two implications from the lemma in turn.

"$\Longrightarrow$": Suppose $P_{\mu_\mathcal{D}} = P_{\mu_\mathcal{E}}$. Let $\widehat{\alpha}$ be a local strategy that is extremal in direction $\boldsymbol{v}$. Since $P_{\mu_\mathcal{D}} = P_{\mu_\mathcal{E}}$, there are $\alpha_\mathcal{E}$ and $\alpha_\mathcal{D}$ such that $p(\mu_\mathcal{D}, \widehat{\alpha}) = p(\mu_\mathcal{E}, \alpha_\mathcal{E})$ and $p(\mu_\mathcal{E}, \widehat{\alpha}) = p(\mu_\mathcal{D}, \alpha_\mathcal{D})$. We have:

$$\begin{aligned}
p(\mu_\mathcal{D}, \widehat{\alpha})\boldsymbol{v} &= p(\mu_\mathcal{E}, \alpha_\mathcal{E})\boldsymbol{v} && p(\mu_\mathcal{D}, \widehat{\alpha}) = p(\mu_\mathcal{E}, \alpha_\mathcal{E}) \\
&\leq p(\mu_\mathcal{E}, \widehat{\alpha})\boldsymbol{v} && \widehat{\alpha} \text{ is extremal in direction } \boldsymbol{v} \\
&= p(\mu_\mathcal{D}, \alpha_\mathcal{D})\boldsymbol{v} && p(\mu_\mathcal{E}, \widehat{\alpha}) = p(\mu_\mathcal{D}, \alpha_\mathcal{D}) \\
&\leq p(\mu_\mathcal{D}, \widehat{\alpha})\boldsymbol{v} && \widehat{\alpha} \text{ is extremal in direction } \boldsymbol{v}
\end{aligned}$$

So all inequalities are in fact equalities. In particular, we have $p(\mu_\mathcal{D}, \widehat{\alpha})\boldsymbol{v} = p(\mu_\mathcal{D}, \alpha_\mathcal{D})\boldsymbol{v}$. It follows:

$$\begin{aligned}
p(\mu_\mathcal{D}, \widehat{\alpha}) &= p(\mu_\mathcal{D}, \alpha_\mathcal{D}) && \text{Lemma 17} \\
&= p(\mu_\mathcal{E}, \widehat{\alpha}) && \text{definition of } \alpha_\mathcal{D}
\end{aligned}$$

"$\Longleftarrow$": Let x be a vertex of $P_{\mu_\mathcal{D}}$. By Lemma 18 there exists an extremal local strategy $\widehat{\alpha}$ with $x = p(\mu_\mathcal{D}, \widehat{\alpha})$. By the assumption we have $p(\mu_\mathcal{D}, \widehat{\alpha}) = p(\mu_\mathcal{E}, \widehat{\alpha})$. Hence $x = p(\mu_\mathcal{D}, \widehat{\alpha}) = p(\mu_\mathcal{E}, \widehat{\alpha}) \in P_{\mu_\mathcal{E}}$. Since x is an arbitrary vertex of $P_{\mu_\mathcal{D}}$, and $P_{\mu_\mathcal{D}}, P_{\mu_\mathcal{E}}$ are bounded, convex polytopes, it follows $P_{\mu_\mathcal{D}} \subseteq P_{\mu_\mathcal{E}}$. The reverse inclusion is shown similarly.

$\square$

D.4 Proof of Proposition 11

For $n \geq 0$, define a relation $\sim_n \subseteq \mathsf{subDist}(Q) \times \mathsf{subDist}(Q)$ as follows. Let $\mu_\mathcal{D}, \mu_\mathcal{E} \in \mathsf{subDist}(Q)$. Define $\sim_0$ such that $\mu_\mathcal{D} \sim_0 \mu_\mathcal{E}$ if and only if $\|\mu_\mathcal{D}\| = \|\mu_\mathcal{E}\|$. For $n \geq 0$, define $\sim_{n+1}$ such that $\mu_\mathcal{D} \sim_{n+1} \mu_\mathcal{E}$ if and only if

- $\|\mu_\mathcal{D}\| = \|\mu_\mathcal{E}\|$;
- for all local strategies $\alpha_\mathcal{D}$ there exists a local strategy $\alpha_\mathcal{E}$ such that for all $a \in \mathsf{L}$ we have $\mathsf{Succ}(\mu_\mathcal{D}, \alpha_\mathcal{D}, a) \sim_n \mathsf{Succ}(\mu_\mathcal{E}, \alpha_\mathcal{E}, a)$;
- for all local strategies $\alpha_\mathcal{E}$ there exists a local strategy $\alpha_\mathcal{D}$ such that for all $a \in \mathsf{L}$ we have $\mathsf{Succ}(\mu_\mathcal{D}, \alpha_\mathcal{D}, a) \sim_n \mathsf{Succ}(\mu_\mathcal{E}, \alpha_\mathcal{E}, a)$.

Lemma 20. *We have:*

1. $\sim_n \supseteq \sim_{n+1} \supseteq \sim$ *for all* $n \geq 0$.
2. *If* $\sim_n = \sim_{n+1}$ *then* $\sim_n = \sim$.

Proof. Item 1. follows from a straightforward induction.

For item 2., let $\sim_n = \sim_{n+1}$. By item 1. we have $\sim_n \supseteq \sim$, so it remains to prove $\sim_n \subseteq \sim$. It suffices to prove that $\sim_n$ is a bisimulation.

Suppose $\mu_\mathcal{D} \sim_n \mu_\mathcal{E}$. Since $\sim_n = \sim_{n+1}$, we have $\mu_\mathcal{D} \sim_{n+1} \mu_\mathcal{E}$. Thus:

- $\|\mu_\mathcal{D}\| = \|\mu_\mathcal{E}\|$;

– for all local strategies $\alpha_{\mathcal{D}}$ there exists a local strategy $\alpha_{\mathcal{E}}$ such that for all $a \in \mathsf{L}$ we have $\mathsf{Succ}(\mu_{\mathcal{D}}, \alpha_{\mathcal{D}}, a) \sim_n \mathsf{Succ}(\mu_{\mathcal{E}}, \alpha_{\mathcal{E}}, a)$;
– for all local strategies $\alpha_{\mathcal{E}}$ there exists a local strategy $\alpha_{\mathcal{D}}$ such that for all $a \in \mathsf{L}$ we have $\mathsf{Succ}(\mu_{\mathcal{D}}, \alpha_{\mathcal{D}}, a) \sim_n \mathsf{Succ}(\mu_{\mathcal{E}}, \alpha_{\mathcal{E}}, a)$.

Hence we have shown that $\sim_n$ is a bisimulation. □

We will show later that we have $\sim_n = \sim$ for $n = |Q| - 1$.

The following lemma reduces the membership problem of $\sim_{n+1}$ to the membership problem of $\sim_n$ and a polytope-comparison problem.

Lemma 21. *Let $\mathcal{D} = \langle Q, \mu_0, \mathsf{L}, \delta \rangle$ be an MDP. Let $n \geq 0$ and $k \geq 1$. Suppose that a matrix $B \in \mathbb{R}^{Q \times k}$ is such that for all $\mu_{\mathcal{D}}, \mu_{\mathcal{E}} \in \mathsf{subDist}(Q)$ we have $\mu_{\mathcal{D}} \sim_n \mu_{\mathcal{E}}$ if and only if $\mu_{\mathcal{D}} B = \mu_{\mathcal{E}} B$. Then for all $\mu_{\mathcal{D}}, \mu_{\mathcal{E}} \in \mathsf{subDist}(Q)$ we have $\mu_{\mathcal{D}} \sim_{n+1} \mu_{\mathcal{E}}$ if and only if $\mu_{\mathcal{D}} B = \mu_{\mathcal{E}} B$ and $P_{\mu_{\mathcal{D}}} = P_{\mu_{\mathcal{E}}}$.*

Proof. Let $\mu_{\mathcal{D}}, \mu_{\mathcal{E}} \in \mathsf{subDist}(Q)$. For any local strategies $\alpha_{\mathcal{D}}, \alpha_{\mathcal{E}}$ we have:

$$\begin{array}{lll} & \forall\, a \in \mathsf{L} : \mathsf{Succ}(\mu_{\mathcal{D}}, \alpha_{\mathcal{D}}, a) \sim_n \mathsf{Succ}(\mu_{\mathcal{E}}, \alpha_{\mathcal{E}}, a) & \\ \iff & \forall\, a \in \mathsf{L} : \mu_{\mathcal{D}} \Delta_{\alpha_{\mathcal{D}}}(a) \sim_n \mu_{\mathcal{E}} \Delta_{\alpha_{\mathcal{E}}}(a) & \text{by (3)} \\ \iff & \forall\, a \in \mathsf{L} : \mu_{\mathcal{D}} \Delta_{\alpha_{\mathcal{D}}}(a) B = \mu_{\mathcal{E}} \Delta_{\alpha_{\mathcal{E}}}(a) B & \text{assumption on } B \\ \iff & p(\mu_{\mathcal{D}}, \alpha_{\mathcal{D}}) = p(\mu_{\mathcal{E}}, \alpha_{\mathcal{E}}) & \text{definition of } p \end{array} \tag{8}$$

We prove the two implications from the lemma in turn.

"$\Longrightarrow$": Let $\mu_{\mathcal{D}} \sim_{n+1} \mu_{\mathcal{E}}$. Since $\sim_{n+1} \subseteq \sim_n$, we have $\mu_{\mathcal{D}} \sim_n \mu_{\mathcal{E}}$. By the assumption on B it follows $\mu_{\mathcal{D}} B = \mu_{\mathcal{E}} B$.
To show $P_{\mu_{\mathcal{D}}} = P_{\mu_{\mathcal{E}}}$, choose an arbitrary local strategy $\alpha_{\mathcal{D}}$. Since $\mu_{\mathcal{D}} \sim_{n+1} \mu_{\mathcal{E}}$, there is $\alpha_{\mathcal{E}}$ such that for all $a \in \mathsf{L}$ we have $\mathsf{Succ}(\mu_{\mathcal{D}}, \alpha_{\mathcal{D}}, a) \sim_n \mathsf{Succ}(\mu_{\mathcal{E}}, \alpha_{\mathcal{E}}, a)$. By (8) we have $p(\mu_{\mathcal{D}}, \alpha_{\mathcal{D}}) = p(\mu_{\mathcal{E}}, \alpha_{\mathcal{E}})$. Since $\alpha_{\mathcal{D}}$ was chosen arbitrarily, we have shown $P_{\mu_{\mathcal{D}}} \subseteq P_{\mu_{\mathcal{E}}}$. The reverse inclusion is shown similarly.

"$\Longleftarrow$": Suppose $\mu_{\mathcal{D}} B = \mu_{\mathcal{E}} B$ and $P_{\mu_{\mathcal{D}}} = P_{\mu_{\mathcal{E}}}$. By the assumption on B it follows $\mu_{\mathcal{D}} \sim_n \mu_{\mathcal{E}}$, hence $\|\mu_{\mathcal{D}}\| = \|\mu_{\mathcal{E}}\|$. It remains to show:
- for all local strategies $\alpha_{\mathcal{D}}$ there exists a local strategy $\alpha_{\mathcal{E}}$ such that for all $a \in \mathsf{L}$ we have $\mathsf{Succ}(\mu_{\mathcal{D}}, \alpha_{\mathcal{D}}, a) \sim_n \mathsf{Succ}(\mu_{\mathcal{E}}, \alpha_{\mathcal{E}}, a)$;
- for all local strategies $\alpha_{\mathcal{E}}$ there exists a local strategy $\alpha_{\mathcal{D}}$ such that for all $a \in \mathsf{L}$ we have $\mathsf{Succ}(\mu_{\mathcal{D}}, \alpha_{\mathcal{D}}, a) \sim_n \mathsf{Succ}(\mu_{\mathcal{E}}, \alpha_{\mathcal{E}}, a)$.

Let $\alpha_{\mathcal{D}}$ be a local strategy. Since $P_{\mu_{\mathcal{D}}} = P_{\mu_{\mathcal{E}}}$, there is a local strategy $\alpha_{\mathcal{E}}$ such that $p(\mu_{\mathcal{D}}, \alpha_{\mathcal{D}}) = p(\mu_{\mathcal{E}}, \alpha_{\mathcal{E}})$. By (8) we obtain that for all $a \in \mathsf{L}$ we have $\mathsf{Succ}(\mu_{\mathcal{D}}, \alpha_{\mathcal{D}}, a) \sim_n \mathsf{Succ}(\mu_{\mathcal{E}}, \alpha_{\mathcal{E}}, a)$. We have shown the first condition. The second condition is shown similarly.

□

The characterization of $\sim_{n+1}$ provided by Lemma 21 depends strongly on comparing two polytopes $P_{\mu_{\mathcal{D}}}$ and $P_{\mu_{\mathcal{E}}}$. Using Lemma 19, we can instead formulate a characterization in terms of matrices and extremal local strategies:

Lemma 22. *Let $\mathcal{D} = \langle Q, \mu_0, \mathsf{L}, \delta \rangle$ be an MDP. Let $n \geq 0$ and $k \geq 1$. Suppose that a matrix $B \in \mathbb{R}^{Q \times k}$ is such that for all $\mu_\mathcal{D}, \mu_\mathcal{E} \in \mathsf{subDist}(Q)$ we have $\mu_\mathcal{D} \sim_n \mu_\mathcal{E}$ if and only if $\mu_\mathcal{D} B = \mu_\mathcal{E} B$. Then for all $\mu_\mathcal{D}, \mu_\mathcal{E} \in \mathsf{subDist}(Q)$ we have $\mu_\mathcal{D} \sim_{n+1} \mu_\mathcal{E}$ if and only if $\mu_\mathcal{D} B = \mu_\mathcal{E} B$ and $\mu_\mathcal{D} \Delta_{\widehat{\alpha}}(a) B = \mu_\mathcal{E} \Delta_{\widehat{\alpha}}(a) B$ holds for all $a \in \mathsf{L}$ and for all extremal local strategies $\widehat{\alpha}$.*

Proof. By combining Lemmas 21 and 19, and the definitions of $p(\mu_\mathcal{D}, \widehat{\alpha})$ and $p(\mu_\mathcal{E}, \widehat{\alpha})$. □

Now we can prove Proposition 11 from the main text:

Proposition 11. *Let $\mathcal{D} = \langle Q, \mu_0, \mathsf{L}, \delta \rangle$ be an MDP. Let $\mathcal{V} \subseteq \mathbb{R}^Q$ be the smallest column-vector space such that*

- *$\mathbf{1} = (1\ 1 \cdots 1)^T \in \mathcal{V}$ (where T denotes transpose) and*
- *$\Delta_{\widehat{\alpha}}(a) \boldsymbol{u} \in \mathcal{V}$ for all $\boldsymbol{u} \in \mathcal{V}$, all labels $a \in \mathsf{L}$ and local strategies $\widehat{\alpha}$ that are extremal with respect to $\mathcal{V}$.*

Then for all $\mu_\mathcal{D}, \mu_\mathcal{E} \in \mathsf{subDist}(Q)$, we have $\mu_\mathcal{D} \sim \mu_\mathcal{E}$ if and only if $\mu_\mathcal{D} \boldsymbol{u} = \mu_\mathcal{E} \boldsymbol{u}$ for all $\boldsymbol{u} \in \mathcal{V}$.

Proof. Define $\mathcal{V}_0 = \{r\mathbf{1} \mid r \in \mathbb{R}\} \subseteq \mathbb{R}^Q$. For all $n \geq 0$ define $\mathcal{V}_{n+1}$ to be the smallest vector space such that

- $\mathcal{V}_n \subseteq \mathcal{V}_{n+1}$
- $\Delta_{\widehat{\alpha}}(a) \boldsymbol{u} \in \mathcal{V}_{n+1}$, for all $\boldsymbol{u} \in \mathcal{V}_n$ and all $a \in \mathsf{L}$ and all local strategies $\widehat{\alpha}$ that are extremal with respect to $\mathcal{V}_n$.

We have $\mathcal{V}_n \subseteq \mathcal{V}$ for all $n \geq 0$. We claim for all $\mu_\mathcal{D}, \mu_\mathcal{E} \in \mathsf{subDist}(Q)$:

$$\mu_\mathcal{D} \sim_n \mu_\mathcal{E} \quad \Longleftrightarrow \quad \mu_\mathcal{D} \boldsymbol{u} = \mu_\mathcal{E} \boldsymbol{u} \quad \text{for all } \boldsymbol{u} \in \mathcal{V}_n \tag{9}$$

We proceed by induction on n. For $n = 0$, we have:

$$\begin{aligned} \mu_\mathcal{D} \sim_0 \mu_\mathcal{E} &\Longleftrightarrow \|\mu_\mathcal{D}\| = \|\mu_\mathcal{E}\| && \text{definition of } \sim_0 \\ &\Longleftrightarrow \mu_\mathcal{D} \mathbf{1} = \mu_\mathcal{E} \mathbf{1} && \text{definition of } \|\cdot\| \\ &\Longleftrightarrow \mu_\mathcal{D} \boldsymbol{u} = \mu_\mathcal{E} \boldsymbol{u} \quad \forall\, \boldsymbol{u} \in \mathcal{V}_0 && \text{definition of } \mathcal{V}_0 \end{aligned}$$

For the induction step, let $n \geq 0$. Define B as a matrix whose columns span $\mathcal{V}_n$. We have:

$$\begin{aligned} \mu_\mathcal{D} \sim_n \mu_\mathcal{E} &\Longleftrightarrow \mu_\mathcal{D} \boldsymbol{u} = \mu_\mathcal{E} \boldsymbol{u} \quad \forall\, \boldsymbol{u} \in \mathcal{V}_n && \text{induction hypothesis} \\ &\Longleftrightarrow \mu_\mathcal{D} B = \mu_\mathcal{E} B && \text{definition of } B \end{aligned} \tag{10}$$

In the following, let $\widehat{\alpha}$ range over all local strategies that are extremal with respect to B:

$$
\begin{aligned}
&\mu_{\mathcal{D}} \sim_{n+1} \mu_{\mathcal{E}} \\
&\iff \mu_{\mathcal{D}} B = \mu_{\mathcal{E}} B \ \text{ and } \ \mu_{\mathcal{D}} \Delta_{\widehat{\alpha}}(a) B = \mu_{\mathcal{E}} \Delta_{\widehat{\alpha}}(a) B \qquad \forall\, a \in \mathsf{L} \quad \forall\, \widehat{\alpha} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(10), Lemma 22} \\
&\iff \mu_{\mathcal{D}} \boldsymbol{u} = \mu_{\mathcal{E}} \boldsymbol{u} \ \text{ and } \ \mu_{\mathcal{D}} \Delta_{\widehat{\alpha}}(a) \boldsymbol{u} = \mu_{\mathcal{E}} \Delta_{\widehat{\alpha}}(a) \boldsymbol{u} \qquad \forall\, \boldsymbol{u} \in \mathcal{V}_n \quad \forall\, a \in \mathsf{L} \quad \forall\, \widehat{\alpha} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{definition of } B \\
&\iff \mu_{\mathcal{D}} \boldsymbol{u} = \mu_{\mathcal{E}} \boldsymbol{u} \qquad \forall\, \boldsymbol{u} \in \mathcal{V}_{n+1} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{definition of } \mathcal{V}_{n+1}
\end{aligned}
$$

Hence (9) is shown.

Let s be the smallest number with $\mathcal{V}_s = \mathcal{V}_{s+1}$. We have $s \le |Q| - 1$, since $\mathcal{V}_0, \mathcal{V}_1, \ldots$ are increasing subspaces of $\mathbb{R}^Q$. By the definition, the subspace $\mathcal{V}_{n+1}$ depends only on $\mathcal{V}_n$, so we have $\mathcal{V}_s = \mathcal{V}_t$ for all $t \ge s$. The vector space $\mathcal{V}_s$ has the closure properties required by the definition of $\mathcal{V}$, hence $\mathcal{V}_s \supseteq \mathcal{V}$. Since $\mathcal{V}_n \subseteq \mathcal{V}$ for all n, it follows that $\mathcal{V}_s = \mathcal{V}$.

Let $\mu_{\mathcal{D}}, \mu_{\mathcal{E}} \in \mathsf{subDist}(Q)$. We have:

$$
\begin{aligned}
\mu_{\mathcal{D}} \sim_s \mu_{\mathcal{E}} &\iff \mu_{\mathcal{D}} \boldsymbol{u} = \mu_{\mathcal{E}} \boldsymbol{u} \quad \text{for all } \boldsymbol{u} \in \mathcal{V}_s && \text{by (9)} \\
&\iff \mu_{\mathcal{D}} \boldsymbol{u} = \mu_{\mathcal{E}} \boldsymbol{u} \quad \text{for all } \boldsymbol{u} \in \mathcal{V}_{s+1} && \mathcal{V}_s = \mathcal{V}_{s+1} \\
&\iff \mu_{\mathcal{D}} \sim_{s+1} \mu_{\mathcal{E}} && \text{by (9)}
\end{aligned}
$$

By Lemma 20.2. this implies $\sim_s = \sim$. We have:

$$
\begin{aligned}
\mu_{\mathcal{D}} \sim \mu_{\mathcal{E}} &\iff \mu_{\mathcal{D}} \sim_s \mu_{\mathcal{E}} && \sim = \sim_s \\
&\iff \mu_{\mathcal{D}} \boldsymbol{u} = \mu_{\mathcal{E}} \boldsymbol{u} \quad \text{for all } \boldsymbol{u} \in \mathcal{V}_s && \text{by (9)} \\
&\iff \mu_{\mathcal{D}} \boldsymbol{u} = \mu_{\mathcal{E}} \boldsymbol{u} \quad \text{for all } \boldsymbol{u} \in \mathcal{V} && \mathcal{V}_s = \mathcal{V}
\end{aligned}
$$

□

D.5 Proof of Theorem 12

We prove Theorem 12 from the main text. The proof is mainly based on Proposition 11.

Theorem 12. *The problem that, given two MDPs $\mathcal{D}$ and $\mathcal{E}$, asks whether $\mathcal{D} \sim \mathcal{E}$ is in* coNP.

Proof. Without loss of generality we assume $\mathcal{D} = \langle Q, \mu_0^{\mathcal{D}}, \mathsf{L}, \delta \rangle$ and $\mathcal{E} = \langle Q, \mu_0^{\mathcal{E}}, \mathsf{L}, \delta \rangle$. Hence we wish to decide in NP whether $\mu_0^{\mathcal{D}} \not\sim \mu_0^{\mathcal{E}}$.

Let $\mathcal{V} \subseteq \mathbb{R}^Q$ be the vector space defined in Proposition 11. We have:

$$
\begin{aligned}
\mu_0^{\mathcal{D}} \not\sim \mu_0^{\mathcal{E}} \quad &\iff \quad \exists\, \boldsymbol{u} \in \mathcal{V} \text{ with } \mu_0^{\mathcal{D}} \boldsymbol{u} \neq \mu_0^{\mathcal{E}} \boldsymbol{u} && \text{Proposition 11} \\
&\iff \quad \textit{Cond} && \text{from the definition of } \mathcal{V},
\end{aligned}
$$

where *Cond* is the following condition:

There are $k \in \{1, 2, \ldots, |Q|\}$ and $\boldsymbol{b}_0 = \mathbf{1}, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_{k-1} \in \mathbb{R}^Q$ and $i_0, i_1, \ldots, i_{k-1} \in \{0, 1, \ldots, k-2\}$ and $a_1, a_2, \ldots, a_{k-1} \in \mathsf{L}$ and pure local strategies $\widehat{\alpha}_1, \widehat{\alpha}_2, \ldots, \widehat{\alpha}_{k-1}$ such that for all $j \in \{1, 2, \ldots, k-1\}$
- $\widehat{\alpha}_j$ is extremal with respect to the vector space spanned by $\boldsymbol{b}_0, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_{j-1}$ and
- $i_j < j$ and
- $\boldsymbol{b}_j = \Delta_{\widehat{\alpha}_j}(a_j)\boldsymbol{b}_i$,

and $\mu_0^{\mathcal{D}}\boldsymbol{b}_{k-1} \neq \mu_0^{\mathcal{E}}\boldsymbol{b}_{k-1}$.

It remains to argue that *Cond* can be checked in NP. We can nondeterministically guess $k \leq |Q|$ and $i_0, i_1, \ldots, i_{k-1} \leq k-2$ and $a_1, a_2, \ldots, a_{k-1} \in \mathsf{L}$ and pure local strategies $\widehat{\alpha}_1, \widehat{\alpha}_2, \ldots, \widehat{\alpha}_{k-1}$. This determines $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_{k-1}$. All conditions in *Cond* are straightforward to check in polynomial time, except the condition that for all $j \in \{1, 2, \ldots, k-1\}$ we have that $\widehat{\alpha}_j$ is extremal with respect to the vector space spanned by $\boldsymbol{b}_0, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_{j-1}$. In the remainder of the proof, we argue that this can also be checked in polynomial time.

Let $j \in \{1, 2, \ldots, k-1\}$. Let $B \in \mathbb{R}^{Q \times j}$ be the matrix with columns $\boldsymbol{b}_0, \boldsymbol{b}_1, \ldots, \boldsymbol{b}_{j-1}$. We want to check that $\widehat{\alpha}_j$ is extremal with respect to B. For all $q \in Q$, compute in polynomial time the set $\mathsf{eqmoves}(q) \subseteq \mathsf{moves}(q)$ defined by

$$\mathsf{eqmoves}(q) = \{\mathsf{m} \in \mathsf{moves}(q) \mid p(d_q, \alpha_{q,\mathsf{m}}) = p(d_q, \widehat{\alpha}_j)\},$$

where $\alpha_{q,\mathsf{m}}$ is a pure local strategy with $\alpha_{q,\mathsf{m}}(q)(\mathsf{m}) = 1$ (it does not matter how $\alpha_{q,\mathsf{m}}(q')$ is defined for $q' \neq q$). We want to verify that (4) and (5) holds for $\widehat{\alpha}_j$. By linearity, it suffices to check (4) and (5) for all *pure* local strategies α. Hence we need to find $\boldsymbol{v} \in \mathbb{R}^{|\mathsf{L}| \cdot j}$ so that for all $q \in Q$ and all $\mathsf{m} \in \mathsf{moves}(q) \setminus \mathsf{eqmoves}(q)$ we have $p(d_q, \alpha_{q,\mathsf{m}})\boldsymbol{v} < p(d_q, \widehat{\alpha}_j)\boldsymbol{v}$. If such a vector $\boldsymbol{v}$ exists, it can be scaled up by a large positive scalar so that we have:

$$p(d_q, \alpha_{q,\mathsf{m}})\boldsymbol{v} + 1 \leq p(d_q, \widehat{\alpha}_j)\boldsymbol{v} \quad \forall\, q \in Q \quad \forall\, \mathsf{m} \in \mathsf{moves}(q) \setminus \mathsf{eqmoves}(q) \qquad (11)$$

Hence it suffices to check if there exists a vector $\boldsymbol{v}$ that satisfies (11). This amounts to a feasibility check of a linear program of polynomial size. Such a check can be carried out in polynomial time. $\square$

D.6 Proof of Proposition 13

The following Lemma is analogous to Lemma 22.

Lemma 23. *Let $\mathcal{D} = \langle Q, \mu_0^{\mathcal{D}}, \mathsf{L}, \delta\rangle$ be an MDP and $\mathcal{C} = \langle Q_{\mathcal{C}}, \mu_0^{\mathcal{C}}, \mathsf{L}, \delta\rangle$ be an MC with $Q_{\mathcal{C}} \subseteq Q$. Let $n \geq 0$ and $k \geq 1$. Suppose that a matrix $B \in \mathbb{R}^{Q \times k}$ is such that for all $\mu_{\mathcal{D}} \in \mathsf{subDist}(Q)$ and all $\mu_{\mathcal{C}} \in \mathsf{subDist}(Q_{\mathcal{C}})$ we have $\mu_{\mathcal{D}} \sim_n \mu_{\mathcal{C}}$ if and only if $\mu_{\mathcal{D}}B = \mu_{\mathcal{C}}B$. Then for all $\mu_{\mathcal{D}} \in \mathsf{subDist}(Q)$ and all $\mu_{\mathcal{C}} \in \mathsf{subDist}(Q_{\mathcal{C}})$ we have $\mu_{\mathcal{D}} \sim_{n+1} \mu_{\mathcal{C}}$ if and only if $\mu_{\mathcal{D}}B = \mu_{\mathcal{C}}B$ and $\mu_{\mathcal{D}}\Delta_\alpha(a)B = \mu_{\mathcal{C}}\Delta_\alpha(a)B$ holds for all $a \in \mathsf{L}$ and for all local strategies α.*

Proof. By combining Lemma 21 and the definition of $p(\mu, \alpha)$. $\square$

Now we can prove Proposition 13 from the main text:

Proposition 13. *Let $\mathcal{D} = \langle Q, \mu_0^{\mathcal{D}}, \mathsf{L}, \delta\rangle$ be an MDP and $\mathcal{C} = \langle Q_{\mathcal{C}}, \mu_0^{\mathcal{C}}, \mathsf{L}, \delta\rangle$ be an MC with $Q_{\mathcal{C}} \subseteq Q$. Let $\mathcal{V} \subseteq \mathbb{R}^Q$ be the smallest column-vector space such that*

- *$\mathbf{1} = (1\ 1 \cdots 1)^T \in \mathcal{V}$ (where T denotes transpose) and*
- *$\Delta_\alpha(a)\boldsymbol{u} \in \mathcal{V}$ for all $\boldsymbol{u} \in \mathcal{V}$, all labels $a \in \mathsf{L}$ and all local strategies α.*

Then for all $\mu_{\mathcal{D}} \in \mathsf{subDist}(Q)$ and all $\mu_{\mathcal{C}} \in \mathsf{subDist}(Q_{\mathcal{C}})$, we have $\mu_{\mathcal{D}} \sim \mu_{\mathcal{C}}$ if and only if $\mu_{\mathcal{D}}\boldsymbol{u} = \mu_{\mathcal{C}}\boldsymbol{u}$ for all $\boldsymbol{u} \in \mathcal{V}$.

Proof. The proof is very similar to the one of Proposition 11. We give it explicitly for completeness.

Define $\mathcal{V}_0 = \{r\mathbf{1} \mid r \in \mathbb{R}\} \subseteq \mathbb{R}^Q$. For all $n \geq 0$ define $\mathcal{V}_{n+1}$ to be the smallest vector space such that

- $\mathcal{V}_n \subseteq \mathcal{V}_{n+1}$
- $\Delta_\alpha(a)\boldsymbol{u} \in \mathcal{V}_{n+1}$, for all $\boldsymbol{u} \in \mathcal{V}_n$ and all $a \in \mathsf{L}$ and all local strategies α.

We have $\mathcal{V}_n \subseteq \mathcal{V}$ for all $n \geq 0$. We claim for all $\mu_{\mathcal{D}} \in \mathsf{subDist}(Q)$ and all $\mu_{\mathcal{C}} \in \mathsf{subDist}(Q_{\mathcal{C}})$:

$$\mu_{\mathcal{D}} \sim_n \mu_{\mathcal{C}} \quad\Longleftrightarrow\quad \mu_{\mathcal{D}}\boldsymbol{u} = \mu_{\mathcal{C}}\boldsymbol{u} \quad \text{for all } \boldsymbol{u} \in \mathcal{V}_n \tag{12}$$

We proceed by induction on n. For $n = 0$, we have:

$$\begin{aligned}
\mu_{\mathcal{D}} \sim_0 \mu_{\mathcal{C}} &\Longleftrightarrow \|\mu_{\mathcal{D}}\| = \|\mu_{\mathcal{C}}\| && \text{definition of } \sim_0\\
&\Longleftrightarrow \mu_{\mathcal{D}}\mathbf{1} = \mu_{\mathcal{C}}\mathbf{1} && \text{definition of } \|\cdot\|\\
&\Longleftrightarrow \mu_{\mathcal{D}}\boldsymbol{u} = \mu_{\mathcal{C}}\boldsymbol{u} \quad \forall\, \boldsymbol{u} \in \mathcal{V}_0 && \text{definition of } \mathcal{V}_0
\end{aligned}$$

For the induction step, let $n \geq 0$. Define B as a matrix whose columns span $\mathcal{V}_n$. We have:

$$\begin{aligned}
\mu_{\mathcal{D}} \sim_n \mu_{\mathcal{C}} &\Longleftrightarrow \mu_{\mathcal{D}}\boldsymbol{u} = \mu_{\mathcal{C}}\boldsymbol{u} \quad \forall\, \boldsymbol{u} \in \mathcal{V}_n && \text{induction hypothesis}\\
&\Longleftrightarrow \mu_{\mathcal{D}}B = \mu_{\mathcal{C}}B && \text{definition of } B
\end{aligned} \tag{13}$$

In the following, let α range over all local strategies:

$$\begin{aligned}
&\mu_{\mathcal{D}} \sim_{n+1} \mu_{\mathcal{C}}\\
&\Longleftrightarrow \mu_{\mathcal{D}}B = \mu_{\mathcal{C}}B \text{ and } \mu_{\mathcal{D}}\Delta_\alpha(a)B = \mu_{\mathcal{C}}\Delta_\alpha(a)B \qquad \forall\, a \in \mathsf{L} \quad \forall\, \alpha\\
&\hspace{20em} \text{(13), Lemma 23}\\
&\Longleftrightarrow \mu_{\mathcal{D}}\boldsymbol{u} = \mu_{\mathcal{C}}\boldsymbol{u} \text{ and } \mu_{\mathcal{D}}\Delta_\alpha(a)\boldsymbol{u} = \mu_{\mathcal{C}}\Delta_\alpha(a)\boldsymbol{u} \qquad \forall\, \boldsymbol{u} \in \mathcal{V}_n \quad \forall\, a \in \mathsf{L} \quad \forall\, \alpha\\
&\hspace{20em} \text{definition of } B\\
&\Longleftrightarrow \mu_{\mathcal{D}}\boldsymbol{u} = \mu_{\mathcal{C}}\boldsymbol{u} \qquad \forall\, \boldsymbol{u} \in \mathcal{V}_{n+1}\\
&\hspace{20em} \text{definition of } \mathcal{V}_{n+1}
\end{aligned}$$

Hence (12) is shown.

Let s be the smallest number with $\mathcal{V}_s = \mathcal{V}_{s+1}$. We have $s \le |Q| - 1$, since $\mathcal{V}_0, \mathcal{V}_1, \ldots$ are increasing subspaces of $\mathbb{R}^Q$. By the definition, the subspace $\mathcal{V}_{n+1}$ depends only on $\mathcal{V}_n$, so we have $\mathcal{V}_s = \mathcal{V}_t$ for all $t \ge s$. The vector space $\mathcal{V}_s$ has the closure properties required by the definition of $\mathcal{V}$, hence $\mathcal{V}_s \supseteq \mathcal{V}$. Since $\mathcal{V}_n \subseteq \mathcal{V}$ for all n, it follows that $\mathcal{V}_s = \mathcal{V}$.

Let $\mu_{\mathcal{D}} \in \mathsf{subDist}(Q)$ and $\mu_{\mathcal{C}} \in \mathsf{subDist}(Q_{\mathcal{C}})$. We have:

$$\begin{aligned} \mu_{\mathcal{D}} \sim_s \mu_{\mathcal{C}} &\iff \mu_{\mathcal{D}} \boldsymbol{u} = \mu_{\mathcal{C}} \boldsymbol{u} \quad \text{for all } \boldsymbol{u} \in \mathcal{V}_s && \text{by (12)} \\ &\iff \mu_{\mathcal{D}} \boldsymbol{u} = \mu_{\mathcal{C}} \boldsymbol{u} \quad \text{for all } \boldsymbol{u} \in \mathcal{V}_{s+1} && \mathcal{V}_s = \mathcal{V}_{s+1} \\ &\iff \mu_{\mathcal{D}} \sim_{s+1} \mu_{\mathcal{C}} && \text{by (12)} \end{aligned}$$

By Lemma 20.2. this implies $\sim_s = \sim$. We have:

$$\begin{aligned} \mu_{\mathcal{D}} \sim \mu_{\mathcal{C}} &\iff \mu_{\mathcal{D}} \sim_s \mu_{\mathcal{C}} && \sim = \sim_s \\ &\iff \mu_{\mathcal{D}} \boldsymbol{u} = \mu_{\mathcal{C}} \boldsymbol{u} \quad \text{for all } \boldsymbol{u} \in \mathcal{V}_s && \text{by (12)} \\ &\iff \mu_{\mathcal{D}} \boldsymbol{u} = \mu_{\mathcal{C}} \boldsymbol{u} \quad \text{for all } \boldsymbol{u} \in \mathcal{V} && \mathcal{V}_s = \mathcal{V} \end{aligned}$$

□

D.7 Proof of Theorem 8

We prove Theorem 8 from the main text:

Theorem 8. *The problem* $\mathsf{MDP} \sqsubseteq \mathsf{MC}$ *is in* NC*, hence in* P.

Proof. Let $\mathcal{D} = \langle Q, \mu_0^{\mathcal{D}}, \mathsf{L}, \delta \rangle$ be an MDP and $\mathcal{C} = \langle Q_{\mathcal{C}}, \mu_0^{\mathcal{C}}, \mathsf{L}, \delta \rangle$ be an MC with $Q_{\mathcal{C}} \subseteq Q$.

Let α_0 denote an arbitrary pure local strategy. For each $q \in Q$ and each $\mathsf{m} \in \mathsf{moves}(q)$ denote by $\alpha_{q,\mathsf{m}}$ the pure local strategy such that $\alpha_{q,\mathsf{m}}(q)(\mathsf{m}) = 1$ and $\alpha_{q,\mathsf{m}}(q') = \alpha_0(q')$ for all $q' \in Q \setminus \{q\}$. Define

$$\begin{aligned} \Sigma &= \{\alpha_0\} \cup \{\alpha_{q,\mathsf{m}} \mid q \in Q,\ \mathsf{m} \in \mathsf{moves}(q)\} && \text{and} \\ \mathcal{M} &= \{\Delta_\alpha(a) \in \mathbb{R}^{Q \times Q} \mid \alpha \in \Sigma,\ a \in \mathsf{L}\} && \text{and} \\ \mathcal{M}_\infty &= \left\{\Delta_\alpha(a) \in \mathbb{R}^{Q \times Q} \;\middle|\; \alpha \text{ is a local strategy, } a \in \mathsf{L}\right\}. \end{aligned}$$

By the definition of $\mathcal{M}_\infty$ the vector space $\mathcal{V} \subseteq \mathbb{R}^Q$ from Proposition 13 is the smallest column-vector space such that

- $\mathbf{1} = (1\ 1 \cdots 1)^T \in \mathcal{V}$ and
- $M\boldsymbol{u} \in \mathcal{V}$, for all $\boldsymbol{u} \in \mathcal{V}$ and all $M \in \mathcal{M}_\infty$.

We have $\mathcal{M} \subseteq \mathcal{M}_\infty$, where $|\mathcal{M}|$ is finite and $|\mathcal{M}_\infty|$ is infinite. Every matrix in $\mathcal{M}_\infty$ can be expressed as a linear combination of matrices from $\mathcal{M}$: Indeed, let α be a local strategy. Then for all $a \in \mathsf{L}$ we have:

$$\Delta_\alpha(a) = \Delta_{\alpha_0}(a) + \sum_{q \in Q} \left(-\Delta_{\alpha_0}(a) + \sum_{\mathsf{m} \in \mathsf{moves}(q)} \alpha(q)(\mathsf{m}) \cdot \Delta_{\alpha_{q,\mathsf{m}(q)}}(a) \right)$$

So by linearity, the vector space $\mathcal{V}$ is the smallest column-vector space such that

– $\mathbf{1} = (1\ 1 \cdots 1)^T \in \mathcal{V}$ and
– $M\boldsymbol{u} \in \mathcal{V}$, for all $\boldsymbol{u} \in \mathcal{V}$ and all $M \in \mathcal{M}$.

Define a finite set of labels $\mathsf{L}' = \{b_{\alpha,a} \mid \alpha \in \Sigma,\ a \in \mathsf{L}\}$, and for each $\alpha \in \Sigma$ and each $a \in \mathsf{L}$ a matrix

$$\Delta'(b_{\alpha,a}) = \frac{1}{|\Sigma|}\Delta_\alpha(a).$$

The matrix $\sum_{b\in\mathsf{L}'} \Delta'(b)$ is stochastic. Define the MCs $\mathcal{D}' = \langle Q, \mu_0^{\mathcal{D}}, \mathsf{L}', \delta'\rangle$ and $\mathcal{C}' = \langle Q, \mu_0^{\mathcal{C}}, \mathsf{L}', \delta'\rangle$ such that δ' induces the transition matrices $\Delta'(b)$ for all $b \in \mathsf{L}'$. The MCs $\mathcal{D}'$ and $\mathcal{C}'$ are computable in logarithmic space. Let $\mathcal{V}' \subseteq \mathbb{R}^Q$ be the smallest column-vector space such that

– $\mathbf{1} = (1\ 1 \cdots 1)^T \in \mathcal{V}$ and
– $\Delta'(b)\boldsymbol{u} \in \mathcal{V}$, for all $\boldsymbol{u} \in \mathcal{V}$ and all $b \in \mathsf{L}'$.

Since the matrices in $\mathcal{M}$ and the matrices $\Delta'(b)$ are scalar multiples of each other, we have $\mathcal{V} = \mathcal{V}'$. It holds:

$$\begin{aligned}
\mathcal{D} \sqsubseteq \mathcal{C} &\iff \mathcal{D} \sim \mathcal{C} \text{ in } \mathcal{D} && \text{Proposition 9}\\
&\iff \mu_0^{\mathcal{D}} \sim \mu_0^{\mathcal{C}} \text{ in } \mathcal{D} && \text{definition}\\
&\iff \forall\, \boldsymbol{u} \in \mathcal{V} : \mu_0^{\mathcal{D}}\boldsymbol{u} = \mu_0^{\mathcal{C}}\boldsymbol{u} && \text{Proposition 13}\\
&\iff \forall\, \boldsymbol{u} \in \mathcal{V}' : \mu_0^{\mathcal{D}}\boldsymbol{u} = \mu_0^{\mathcal{C}}\boldsymbol{u} && \mathcal{V} = \mathcal{V}'\\
&\iff \mu_0^{\mathcal{D}} \sim \mu_0^{\mathcal{C}} \text{ in } \mathcal{D}' && \text{Proposition 13}\\
&\iff \mathcal{D}' \sim \mathcal{C}' \text{ in } \mathcal{D}' && \text{definition}\\
&\iff \mathcal{D}' \sqsubseteq \mathcal{C}' && \text{Proposition 9}
\end{aligned}$$

As mentioned in Section 2.2, deciding whether $\mathcal{D}' \sqsubseteq \mathcal{C}'$ holds amounts to the trace-equivalence problem for MCs. It follows from Tzeng [20] that the latter is decidable in NC, hence in P. □