Synthesis of H-safe strategies for MDPs[*]

Soham Joshi
sohamjoshi@cse.iitb.ac.in
IIT Bombay

December 2, 2023

Abstract

Markov decision processes can be viewed as transformers of probability distributions. Although basic reachability and safety problems are known to be computationally intractable (i.e. Skolem hard) to solve in such models, we find that there are special cases of MDPs of practical importance that are easy to solve for such problems. We provide an algorithm to synthesize safe strategies for 2-state MDPs in polynomial time and further show that memoryless strategies are enough for safety of 2-state MDPs.

[*] This was a credited research project under the supervision of Prof. S. Akshay at IIT Bombay.

1 Introduction

Markov Decision Processes (MDPs) are a classical model for probabilistic decision making systems. They can be viewed as modelling transition systems, and hence, can be viewed as transformers of probability distributions. Hence, restricting the set of attainable distributions to a “safe" set of distributions is an important problem for MDPs.

The *initialised safety* problem asks the question given an initial distribution over the set of states, if there exists a strategy which yields a probability distribution in the safe set for all steps. As it turns out, this question is Skolem-Hard even for markov chains, and hence skolem-hard for MDPs [AAOW15]. The approach taken in [ACMŽ23] in order to tackle this problem is an over-approximation route where the algorithm synthesizes affine inductive invariants along with the strategy, instead of merely synthesizing the strategy. This gave a PSPACE algorithm for synthesis of general strategies, which is sound.

The *existential, universal* safety problems have also been addressed in 2018 by [AGV18], where deciding the existence of an initial distribution and a corresponding safe strategy was shown to be PTIME-complete. Moreover, deciding if there exists a safe strategy for all initial distributions was shown to be co-NP complete.

This raises the question, do there exist simple instances of the initialised safety problem? We show an algorithm which decides the initialised safety of a 2-state MDP with affine safety set in `PTIME`. Moreover, we also show that memoryless strategies suffice for the intialised safety of 2-state MDPs with affine safety sets, and synthesize such strategies in `PTIME`.

2 Preliminaries

In this section, we use the same notation as defined in [AGV18], in order to define MDPs, associated strategies and safety.

2.1 Notation

Let $S = s_1, ..., s_n$ be a set of states, Σ a finite alphabet of actions. For all $1 \leq i \leq n$, we use $\boldsymbol{s_i}$ to denote the n-dimensional vector, which has 1 in position i and 0 elsewhere. We use Δ_1, Δ_2 etc. to denote arbitrary (probability) distributions over S ,i.e., n dimensional vectors $\Delta \in [0, 1]^n$ such that $\sum_{i=1}^{n} \Delta(i) = \sum_{i=1}^{n} \boldsymbol{s_i}.\Delta = 1$. We also use M, M' etc. to denote n-dimensional row stochastic matrices . Any such matrix can be seen as defining the transition matrix of a Markov chain over the set of states S.

Definition 2.1 (Markov Decision Process)**.** A markov decision process is a tuple $\mathcal{A} = (S, \Sigma, (M_\alpha)_{\alpha \in \Sigma})$ where S is a set of states , Σ is the alphabet of actions, and $(M_\alpha)_{\alpha \in \Sigma}$ is a set of stochastic matrices, which will define how the probability mass in a state $s_i \in S$ is transformed while playing any action $\alpha \in \Sigma$.

An example of a 2-state MDP is shown in fig. 1

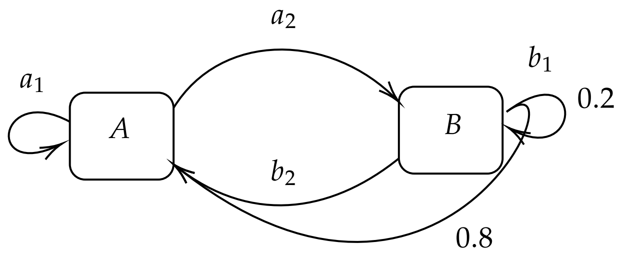

Figure 1: Example of a 2-state MDP with actions a_1, a_2 associated with state A, and actions b_1, b_2 associated with state B

In order to define the safety problem for MDPs, first, we need to define the notion of a one-step strategies for the same.

Definition 2.2 (One step strategy)**.** A one step strategy of an MDP over S, Σ is a function $\tau : \Sigma \times S \longrightarrow [0,1]$ such that for all $s \in S$, $\sum_{\alpha \in \Sigma} \tau(\alpha, s) = 1$. A one step strategy τ is associated with the stochastic matrix,

$$M_\tau = \sum_{\alpha \in \Sigma, i \leq n} \tau(\alpha, s_i) M_{(\alpha, i)}$$

So, given this definition of a one-step strategy, we can define a general strategy of an MDP (σ) as a sequence one strategies $\tau_1, \tau_2, \cdots$. Finally, we define the notion of probability distribution Δ_m^σ obtained using a strategy σ at a step m,

$$\Delta_m^\sigma = \Delta . M_{\tau_1} \cdots M_{\tau_m}$$

where Δ is the initial probability distribution over states. Note that the strategies defined here are randomized. That is, in one step, a strategy is a probability distribution over the outgoing actions from a state.
We now define the notion of a memoryless strategy,

Definition 2.3 (Memoryless strategies)**.** A memoryless strategy corresponds to a fixed choice in each state, independent of history, i. e., a mapping $\pi : S \longrightarrow \Delta(Act)$. Fixing such a strategy induces a finite state markov chain.

where $\Delta(X)$ denotes the space of probability distributions over X. For example, in fig. 1, a possible memoryless strategy is $\pi(A) = \{0.5, 0.5\}, \pi(B) = \{0.3, 0.7\}$ which translates to taking

actions a_1, a_2 with probability 0.5 each and taking action b_1, b_2 with probabilities 0.3, 0.7 respectively, for each time step.
The objective we will focus on in this report is distributional safety. That is, we mark some probability distributions as “safe" and we don’t want the probability distribution on the states of the MDP to lie outside this set in any time step. We will define the notion of a safe strategy,

Definition 2.4 (H-safe strategies)**.** Let $\mathcal{A}$ be an MDP over n states and let H be an affine polytope in $\mathbb{R}^n$ (i.e. H is a finite intersection of half-spaces in $\mathbb{R}^n$, where each half space is a linear inequality). Then, a strategy $\sigma = \tau_{1 \ldots}$ is said to be H-safe from $\Delta_0 \in H$ iff for all $m \in \mathbb{N}$, $\Delta_m^\sigma \in H$

That is, σ is a strategy of $\mathcal{A}$ that allows us to stay forever in H when starting from Δ_1.
Now that we have these definitions, we can define the initialised safety problem for MDPs.

Definition 2.5 (Initialised safety problem for MDPs)**.** Given a MDP $\mathcal{A}$ over n states and an affine polytope $H \in \mathbb{R}^n$, with an initial distribution $\Delta_0 \in \mathbb{R}^n$ the *initialised safety problem* asks if there exists a H-safe strategy of $\mathcal{A}$ from Δ_0.

2.2 State of the Art

Similar to the notion of initialised safety defined above, we can also define the notion of existential, universal safety.

Definition 2.6 (Existential, Universal safety for MDPs)**.** Given a MDP $\mathcal{A}$ over n states and an affine polytope $H \in \mathbb{R}^n$,

1. the *existential safety problem* asks if there exists a dsitribution $\Delta_0 \in H$, H-safe strategy (σ) of $\mathcal{A}$ such that σ is H-safe from Δ_0.

2. the *universal safety problem* asks if for all distributions $\Delta_0 \in H$ there exists an H-safe strategy of $\mathcal{A}$ from Δ

In 2018, it was showed by [AGV18] that existential safety of MDPs is PTIME-complete and universal safety of MDPs is co-NP complete.
Moreover, the safety problem in the initialised setting for affine safety sets was addressed in [ACMŽ23], where inductive invariants were used to give a PSPACE algorithm for synthesizing memoryless strategies and a sound PSPACE algorithm to synthesize template based strategies for MDPs, but the algorithm is not complete. More specifically, in [ACMŽ23], the algorithms shown lie in the existential theory of reals ($\exists\mathbb{R}$), a complexity class between NP and PSPACE.
The initialised safety problem, if H is allowed to be a closed convex polytope is actually known to be Skolem-hard. Moreover, if H is an intersection of finite half-spaces the problem is again skolem hard [AAOW15].

3 Motivation for the problem

In [ACMŽ23] it was shown that the distributional safety problem does not admit a memoryless strategy for the MDP given in fig. 2, where the initial distribution is given as $\mu_0 = \{\frac{3}{4}, \frac{1}{4}, 0\}$.

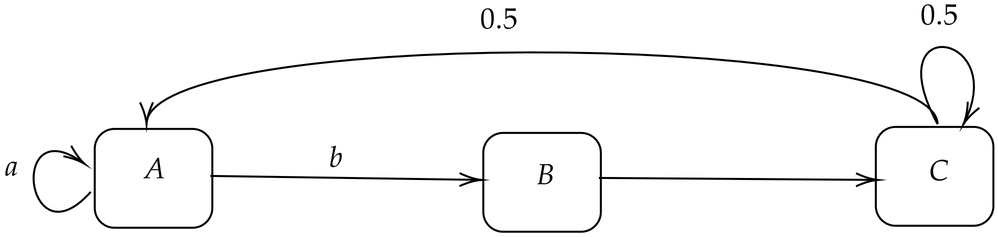

Figure 2: Example of a 3-state MDP which does not admit a memoryless strategy for H given as $\{\mu : \mu_B = \frac{1}{4}\}$, where $\mu_0 = \{\frac{3}{4}, \frac{1}{4}, 0\}$

We can claim that the probbility distribution at step i is given by

$$\mu = \{\frac{1}{4} + \frac{1}{2^{i+1}}, \frac{1}{4}, \frac{1}{2} - \frac{1}{2^{i+1}}\}$$

We can show this via induction. Clearly, at $i = 0$, the probability distribution is $\{\frac{3}{4}, \frac{1}{4}, 0\}$. Now, assuming that the probability distribution at step i is given by

$$\mu = \{\frac{1}{4} + \frac{1}{2^{i+1}}, \frac{1}{4}, \frac{1}{2} - \frac{1}{2^{i+1}}\}$$

We know that at the step $i + 1$, the probability at state B must remain $\frac{1}{4}$. So probability of state C is given by $0.5\mu_C + \mu_B$,

$$\begin{aligned} 0.5\mu_C + \mu_B &= 0.5(\frac{1}{2} - \frac{1}{2^{i+1}}) + \frac{1}{4} \\ &= \frac{1}{2} - \frac{1}{2^{i+2}} \end{aligned}$$

Hence, showing the inductive hypothesis. Let the probability of taking action b in step i be given as β_i. Then, $\mu_A\beta_i = \frac{1}{4}$, since the probability of state B must remain $\frac{1}{4}$. Hence, at step i,

$$\begin{aligned} \mu_A\beta_i &= \frac{1}{4} \\ (\frac{1}{4} + \frac{1}{2^{i+1}})\beta_i &= \frac{1}{4} \\ \beta_i &= \frac{\frac{1}{4}}{\frac{1}{4} + \frac{1}{2^{i+1}}} \end{aligned}$$

Hence, taking action b with probability β_i is the only safe strategy, which is clearly not memoryless. Hence, in 3-state MDPs (and for n-state MDPs with $n \geq 3$), memoryless strategies do not suffice for safety.

This motivates the question, what can be said about 2-state MDPs? Precisely, does there exist an example of a 2-state MDP with an affine safety set where memoryless strategies do not suffice. In addition to this, [ACMŽ23], showed a sound and relatively complete algorithm for synthesis of memoryless safe strategies, which lies in the existential theory of reals, which is contained in PSPACE. This also motivates the question, what is the complexity of synthesis of safe memoryless strategies for 2-state MDPs? In the sections which follow we show,

1. For any 2-state MDP, if there exists a safe strategy, then there exists a safe memoryless strategy. In other words, memoryless strategies suffice for safety.
2. Synthesis of memoryless safe strategies for 2-state MDPs has a polynomial time algorithm. That is, the algorithm takes input as an affine safe set H, and a 2-state MDP $\mathcal{A}$, with an initial distribution $\Delta_0 \in H$ and outputs the safe memoryless strategy if it exists, otherwise outputs $\perp$

4 Results

In this section, we show that any 2-state MDP with affine safety set under distributions admits a memoryless strategy for safety. In order to get started, we show a lemma.

4.1 Reduction of 2-state MDPs

In order to proceed with the main result, we need a reduction from 2-state MDPs with many outgoing actions to 2-state MDPs with atmost 2 outgoing actions per state.

Lemma 4.1 (Reduction of 2-state MDPs)**.** *Given a 2-state MDP $\mathcal{A}$, there exists another 2-state MDP $\mathcal{A}'$ with atmost 2 actions outgoing from each state such that for any $\Delta \in H$, τ is a one-step strategy of $\mathcal{A}$ iff there exists another one-step strategy τ' for $\mathcal{A}'$ such that $M_\tau(\Delta) = M_{\tau'}(\Delta)$*

Proof. Let us call the states of the MDP $\mathcal{A}$ as s_1, s_2. Let the set of actions outgoing from state s_1 parametrised by the probability that a self loop is taken be given by $\alpha_1, \alpha_2, \cdots, \alpha_m$, and those outgoing from s_2 be parametrized by the probability of reaching state s_1, given by $\beta_1, \beta_2, \cdots, \beta_n$ (where $n, m \geq 2$). Now, the distribution over states given by $\Delta = (\mu, 1-\mu)$ and let $\Delta_1 = \mu$. Now, let the distribution of actions outgoing from states s_1, s_2 be given as p_1, p_2 respectively. Hence, in a single step, $\Delta_1 = \mu$ is transformed to,

$$M_\tau(\Delta)_1 = \sum_{i=1}^{m} p_i \alpha_i \mu + \sum_{j=1}^{m} q_j \beta_j (1-\mu)$$

But, $M_\tau(\Delta)_1$ is maximised if, $p_i = 1, q_j = 1$ if $i = \arg\max_i \alpha_i, j = \arg\max_j \beta_j$, and 0 otherwise. Correspondingly, $M_\tau(\Delta)_1$ is minimised for $p_i = 1, q_j = 1$ if $i = \arg\min_i \alpha_i, j = \arg\min_j \beta_j$. Without loss of generality let the variables be ordered as, $\alpha_1 < \cdots < \alpha_m$ and $\beta_1 < \cdots < \beta_n$. Then we have, $\alpha_1\mu + \beta_1(1-\mu) \leq M_\tau(\Delta)_1 \leq \alpha_m\mu + \beta_n(1-\mu)$, and that any value that $M_\tau(\Delta)_1$ can be expressed as a linear combination of $\alpha_1\mu + \beta_1(1-\mu)$ and $\alpha_m\mu + \beta_n(1-\mu)$. Moreover, any such linear combination can be attained by substitution of appropriate p', q', i.e. some strategy τ'. Hence, the MDP can be reduced to one with actions corresponding to $\alpha_1, \alpha_m, \beta_1, \beta_n$, denoted by $\mathcal{A}'$.
Now, if $m = 1$ or $n = 1$, i.e. one of the states has exactly one outgoing action, the procedure above gives an MDP $\mathcal{A}'$ with actions $\alpha_1, \beta_1, \beta_m$, which satisfies the constraint of having atmost 2 actions per state. □

Now that we have this reduction, we can tackle the question of memoryless strategies for 2-state MDPs. In this result we use the idea that the affine safety set is just a mixed interval for μ and any transformation must eventually guarantee either a fixed point (i.e. application of strategy does not change distribution) or that repeated action of an action is enough for safety.

4.2 Memoryless strategy for 2-state MDPs

Now that we have the reduction as described in the previous section, we show that any 2-state MDP with affine safety set under distributions admits a memoryless strategy for safety.

Theorem 4.2 (Memoryless strategy for 2-state MDP)**.** *Given a 2-state MDP $\mathcal{A}$ with any affine safety objective, if there exists a safe strategy then there exists a safe memoryless strategy*

Proof. Given the MDP $\mathcal{A}$, there exists an MDP $\mathcal{A}'$ given by 4.1 parametrized by actions $\alpha_1, \alpha_2, \beta_1, \beta_2$, with $\alpha_1 < \alpha_2, \beta_1 < \beta_2$ and a safety set H.
This proof will proceed via cases. Let us define memoryless strategies σ_1 which takes actions corresponding α_1 for s_1, β_1 for s_2. Similarly, define σ_2. Hence, given Δ such that $\Delta = \mu$, let's denote $M_{\sigma_1}(\mu) = \alpha_1\mu + \beta_1(1-\mu)$, and $M_{\sigma_2}(\mu) = \alpha_2\mu + \beta_2(1-\mu)$ (there is an abuse of notation here, since $M_\sigma(\Delta)$ denotes a probability distribution in the conventional notation, but here we modify the notation to $M_\sigma(\mu)$ to denote the first component of the transformed probability distribution). Define $g(\mu) = \mu$. Now, $M_{\sigma_1}(\mu), M_{\sigma_2}(\mu) : [0,1] \to [0,1]$ are univariate linear functions in μ.

Let the line g intersect M_{σ_1} at μ_1, and M_{σ_2} at μ_2, where it follows that $\mu_1 < \mu_2$ since $M_{\sigma_1} < M_{\sigma_2}$ for all μ. Now, $M_{\sigma_1}(\mu) < M_{\sigma_2}(\mu)$ for all $\mu \in [0,1]$. Moreover, for $\mu > \mu_1$, $g(\mu) > M_{\sigma_1}(\mu)$, and for $\mu < \mu_1$, $g(\mu) < M_{\sigma_1}(\mu)$. Similarly, the claim follows for M_{σ_2} and μ_2.
Now, any value between $M_{\sigma_1}(\mu)$ and $M_{\sigma_2}(\mu)$ can be attained by a one-step transformation of the 2-state MDP $\mathcal{A}'$ by Lemma 4.1.

Case 1 : $H \subseteq [\mu_1, \mu_2]$.
In this case, for any $\mu \in H$, we have $M_{\sigma_1}(\mu) \leq g(\mu) \leq M_{\sigma_2}(\mu)$. Hence, the MDP can attain probability $g(\mu)$ for state s_1, where $g(\mu) = \mu \in H$ resulting in a fixed point, and correspondingly a 1-step strategy to attain a fixed point. (as shown in fig. 3)

Case 2 : $H \subseteq [0, \mu_1)$
In this case, for any $\mu \in H$, we have $g(\mu) < M_{\sigma_1}(\mu)$. If μ has a safe strategy, then $\lambda M_{\sigma_1}(\mu) + (1-\lambda)M_{\sigma_2}(\mu) \in H$ for some λ. Now, transformation of μ can't exceed μ_1 since, $x < \mu_1 \forall x \in H$.
Moreover, since $\mu \in H$, $M_{\sigma_1}(\mu) > \mu = g(\mu)$, hence $M_{\sigma_1}(\mu)$ must belong to H for the next state to be safe. Hence, $M_{\sigma_1}(\mu) \in H$ if a safe strategy exists. It follows that $M_{\sigma_1}(\mu)$ is an increasing function since $M_{\sigma_1}(\mu) \in H < \mu_1 = M_{\sigma_1}(\mu_1)$. Let the initial distribution be μ_0, then the minimum value of the distribution at step k is $M^k_{\sigma_1}(\mu_0)$, and value of distribution is non-decreasing. Since H is contiguous and upper bounded, if a safe strategy exists, then always choosing α_1, β_1 must be safe. Hence, we get a memoryless strategy. (refer fig. 4)

Case 3 : $H \subseteq (\mu_2, 1]$

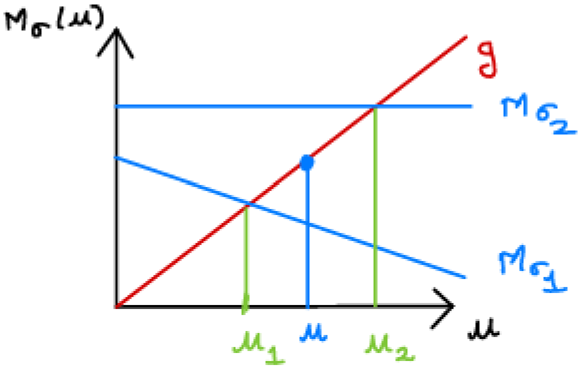

Figure 3: This graph shows the intuition behind the presence of a fixed point

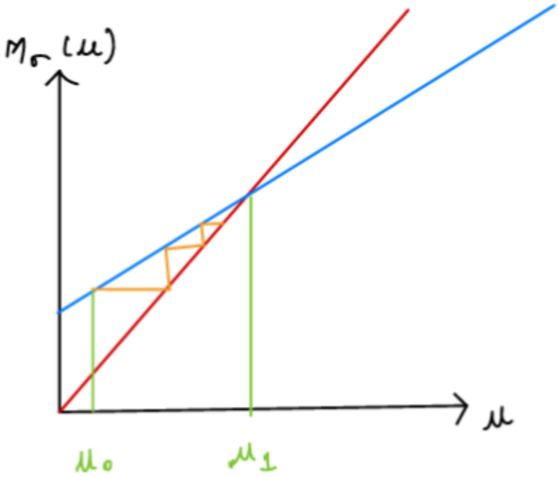

Figure 4: This figure shows the intuition behind the inductive hypothesis considered in case 2 of the proof.

Proof follows similar to case 2 by choosing only actions corresponding to α_2, β_2.

If the MDP had only 1 action available in some state, we could assign $\alpha_1 = \alpha_2$ or $\beta_1 = \beta_2$ and the same proof follows for each of the cases. Finally, for any other H, we have $H = H_1 \cup H_2 \cup H_3$, where a subset H_i corresponds to case i for $i \in \{1, 2, 3\}$. For any $\mu_0 \in H$, we get $\mu_0 \in H_i$ for some i.

Case 4: $\mu_0 \in H_2$
The proof follows immediately, with the presence of a fixed point.

Case 5: $\mu_0 \in H_1$ and $H_2 \neq \phi$
If the line M_{σ_1} has a positive (or zero) slope, we have $\mu_1 = \sup H_1$. Hence, playing actions corresponding to α_1, β_1, transforms μ to $M_{\sigma_1}(\mu)$ with $\mu_1 \geq M_{\sigma_1}(\mu) \geq \mu$. Hence the sequence

$M^i_{\sigma_1}(\mu_0)$ is increasing and bounded above. Hence, playing actions corresponding to α_1, β_1 gives us a memoryless strategy.
If not, the horizontal line $y = M_{\sigma_1}(\mu_0)$ intersects $y = x$ at a point (c, c) where $c \in H_2$. Now, consider the strategy of only playing action 1. We show that this strategy is safe. Let μ^i be the distribution attained at step i. Then, we claim that

$$\mu^i - \mu_1 \leq c - \mu_1$$
$$\mu_1 - \mu^i \leq \mu_1 - \mu_0$$

Base case : For $i = 0, 1$, clearly, $\mu^0 = \mu_0$, $\mu^1 = c$ the hypothesis holds.
Now, assuming hypothesis holds for all $i \leq k$,
If $\mu^i < \mu_1$, then $\mu^{i+1} = M_{\sigma_1}(\mu^i) > M_{\sigma_1}(\mu_1) = \mu_1$. If $\mu^i > \mu_1$, then $\mu^{i+1} = M_{\sigma_1}(\mu^i) < M_{\sigma_1}(\mu_1) = \mu_1$. Since M_{σ_1} is linear in μ, let it be denoted as $M_{\sigma_1}(\mu) = m\mu + c$, with $-1 \leq m < 0, c \geq 0$. Now, $\mu^{i+1} = M_{\sigma_1}(M_{\sigma_1}(\mu^{i-1})) = m(m\mu^{i-1} + c) + c$, where $m\mu_1 + c = \mu_1$
Case 5.1: Let $\mu^i < \mu_1$. Since M_{σ_1} is decreasing, and $\mu^i \geq \mu_0$ by inductive hypothesis, $\mu^{i+1} = M_{\sigma_1}(\mu^i) \leq M_{\sigma_1}(\mu_0) \leq c$. Now, $\mu^{i+1} \geq \mu_1, \mu_1 - \mu^{i+1} \leq \mu_1 - \mu_0$ and since $\mu^{i+1} \leq c$ implies $\mu^{i+1} - \mu_1 \leq c - \mu_1$. Hence, inductive hypothesis holds.
Case 5.2: Let $\mu^i > \mu_1$. Since M_{σ_1} is decreasing, and $\mu^i \leq c = \mu^1$ by inductive hypothesis, $\mu^{i+1} = M_{\sigma_1}(\mu^i) \geq M_{\sigma_1}(\mu^1)$. By strong induction, if $i > 2$, we have $M_{\sigma_1}(\mu^1) \geq \mu_0$ showing the claim by induction. So, we need to show this only for $i = 1$. That is, we need to show $M_{\sigma_1}(M_{\sigma_1}(\mu^0)) \geq \mu_0$. Using slope-intercept form, let $M_{\sigma_1}(\mu) = m\mu + c$

$$\begin{aligned} M_{\sigma_1}(M_{\sigma_1}(\mu^0)) &= M_{\sigma_1}(M_{\sigma_1}(\mu_0)) \\ &= m(m\mu + c) + c \\ &= m^2\mu + mc + c \end{aligned}$$

So, we need to solve $m^2\mu_0 + mc + c \geq \mu_0$. This follows iff $c(1+m) \geq \mu_0(1+m)(1-m)$ iff $c \geq \mu_0(1-m)$. Since $c > \mu_0$, the inductive hypothesis is true.
Now, since the inductive hypothesis is true, hence, the strategy is safe, since $\mu_0 \leq \mu^i \leq \mu_1$ and μ_0, μ_1 where both are safe distributions, hence for all i, μ^i is safe. (the intuition can be seen in fig. 5)

Case 6: $\mu_0 \in H_3$ and $H_2 \neq \phi$
Proof follows similar to case 5.
This completes all cases. Now, since we can find a safe memoryless strategy for $\mathcal{A}'$, it follows from lemma 4.1 that there exists a safe memoryless strategy for $\mathcal{A}$. □

4.3 Algorithm for safe strategy synthesis for 2-state MDPs

The proof of theorem 4.2 performs case work depending on the location of the initial distribution and the structure of the safe set to construct a safe strategy. Hence, we can rewrite the steps of the proof to obtain an algorithm, which synthesizes memoryless strategies.
Let us define `ALG2STATE` which takes input an MDP $\mathcal{A}$, H, and μ_0

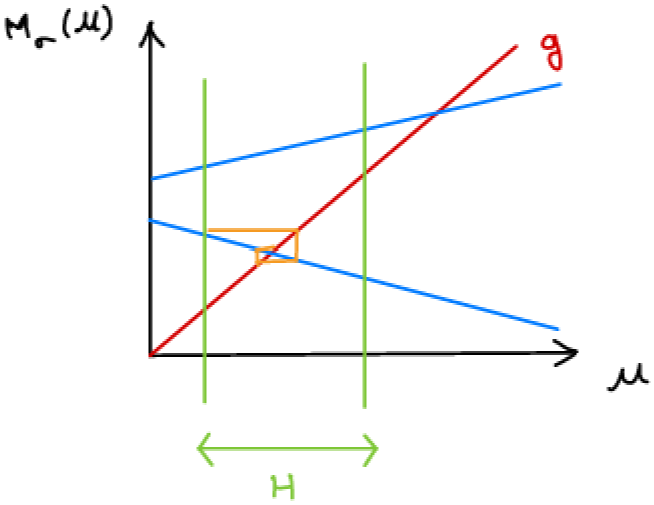

Figure 5: This figure shows the intuition behind the inductive hypothesis considered in case 5 of the proof.

1. Construct an MDP $\mathcal{A}'$ with actions $\alpha_1, \alpha_2, \beta_1, \beta_2$ as given in 4.1. Correspondingly obtain the points of intersection of M_{σ_1} with g (μ_1) and intersection of M_{σ_2} with g (μ_2).
2. Write $H = H_1 \cup H_2 \cup H_3$ where, $H_1 \subseteq [0, \mu_1)$, $H_2 \subseteq [\mu_1, \mu_2]$ and $H_3 \subseteq (\mu_2, 1]$
3. If $\mu_0 \in H_2$
    (a) Since $M_{\sigma_1}(\mu) \leq g(\mu) \leq M_{\sigma_2}(\mu)$, obtain $g(\mu)$ as a convex combination of $M_{\sigma_1}(\mu)$ and $M_{\sigma_2}(\mu)$ to get the memoryless strategy
4. If $\mu_0 \in H_1$ and slope of M_{σ_1} is positive
    (a) If H_1 is of the form $[a, \mu_1], [a, \mu_1), (b, \mu_1]$ or (b, μ_1) return action 1
    (b) Else, return $\perp$
5. If $\mu_0 \in H_1$ and slope of M_{σ_1} is non-negative
    (a) Return action 1 if $M_{\sigma_1}(\mu) \in H$
    (b) Otherwise, return $\perp$
6. If $\mu_0 \in H_3$ and slope of M_{σ_2} is positive
    (a) If H_3 is of the form $[\mu_2, a], [\mu_2, a), (\mu_2, b]$ or (μ_2, b) return action 2
    (b) Else, return $\perp$
7. If $\mu_0 \in H_3$ and slope of M_{σ_2} is non-negative
    (a) Return action 2 if $M_{\sigma_2}(\mu) \in H$

(b) Otherwise, return $\bot$

Lemma 4.3 (Synthesis of memoryless strategy for 2-state MDP). `ALG2STATE` *is a sound and complete algorithm for synthesis of memoryless strategies for 2-state MDPs. Moreover, the algorithm runs in* `PTIME`

Proof. It is enough to show that the cases taken in μ_0, H are correct.

1. If $\mu_0 \in H_2$, the algorithm always gives a memoryless strategy.
2. If $\mu_0 \in H_1$ and slope of M_{σ_1} is positive
   (a) If H_1 is of the form $[a, \mu_1], [a, \mu_1), (b, \mu_1]$ or (b, μ_1) and line M_{σ_1} has a positive (or zero) slope, we have $\mu_1 = \sup H_1$. Hence, playing actions corresponding to α_1, β_1, transforms μ to $M_{\sigma_1}(\mu)$ with $\mu_1 \geq M_{\sigma_1}(\mu) \geq \mu$. Hence the sequence $M^i_{\sigma_1}(\mu_0)$ is increasing and bounded above.
   (b) If not, showing that the sequence converges to μ_1 will show that only H of the form described above has an H-safe strategy. Now, we show that the sequence $M^i_{\sigma_1}(\mu_0)$ converges to μ_1. The sequence is monotonically increasing and bounded above by μ_1 (as shown in case 5 on 4.2), hence the sequence converges. Say the sequence converges to a μ such that $\mu < \mu_1$. Moreover, $M_{\sigma_1}(\mu) = m\mu + c$ is a linear function. Since $M_{\sigma_1}(\mu) > \mu$ and M_{σ_1} is continuous, there exists an $\varepsilon > 0$ such that $M_{\sigma_1}(\mu - \varepsilon) > \mu$. Now, consider the $\varepsilon - n_0$ definition of limits, taking ε obtained above, there exists a k such that $M^k_{\sigma_1}(\mu_0) \geq \mu - \varepsilon$. Hence. $M^{k+1}_{\sigma_1} \geq M_{\sigma_1}(\mu - \varepsilon) > \mu$, giving a contradiction. Since the sequence must converge to a value less than or equal to μ_1, hence it converges to μ_1.
3. If $\mu_0 \in H_1$ and slope of M_{σ_1} is non-negative
   (a) If $M_{\sigma_1}(\mu) \in H$, proof of case 5 in 4.2, shows that the strategy is safe
   (b) If $M_{\sigma_1}(\mu) \notin H$, any other strategy must transform μ_0 to a point outside H, since $M_\sigma \geq M_{\sigma_1}$
4. For $\mu_0 \in H_3$, similar proofs show the desired result.

Now that the correctness is established, let's analyse the running time of the algorithm. We have 7 cases in the algorithm, where each case requires checking some conditions on the safe set, evaluating a point μ on $M_{\sigma_1}, M_{\sigma_2}, g$, obtaining a point as a convex combination of two other points, intersection of linear functions, all of which are polynomial time operations. Hence the algorithm is `PTIME`. □

5 Discussion

There are many further interesting questions relating the work of [ACMŽ23] with the approach for getting strategies of 2-state MDPs described in this report. Some directions of future work are,

1. Is deciding the initialised safety of 3-state Markov Decision Processes NP-hard? If so, $n = 2$ would be the maximum value of n for which deciding initialised safety of n-state MDPs is easy.

2. [ACMŽ23] adopts a template based synthesis approach to give an algorithm which creates a strategy satisfies initialised safety. Since this algorithm uses existentially quantified reals in its template, showing hardness of such an approach using the `Co-NP` hardness of Quantified Linear Implication [AGV18] is a possibility.

3. In the template based approach described in [ACMŽ23], the probability of a state, action pair is of the form of a ratio of two linear functions in the current probability distribution. In order to show that the approach is not complete, one can try to show that there exists an MDP for which ratio of linear expressions in probability distributions don't suffice to generate a safe strategy. The same question can also be asked about strategies which are rational expressions in the distribution.

4. Does there exist an MDP with the expressions of H (the safe set), transition probabilites rational, but there does not exist a rational safe strategy. A rational safe strategy means that the probabilities of taking state, action pairs are rational for all steps.

5. Given an MDP with all transition matrices having real eigenvalues, does the initialised safety problem with affine safety sets become easier to solve?

References

[AAOW15] S. Akshay, Timos Antonopoulos, Joël Ouaknine, and James Worrell. Reachability problems for Markov chains. *Information Processing Letters*, 115(2):155–158, 2015.

[ACMŽ23] S. Akshay, Krishnendu Chatterjee, Tobias Meggendorfer, and Đorđe Žikelić. MDPs as Distribution Transformers: Affine Invariant Synthesis for Safety Objectives. In Constantin Enea and Akash Lal, editors, *Computer Aided Verification*, pages 86–112, Cham, 2023. Springer Nature Switzerland.

[AGV18] S. Akshay, Blaise Genest, and Nikhil Vyas. Distribution-Based Objectives for Markov Decision Processes. In *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science*, LICS '18, page 36–45, New York, NY, USA, 2018. Association for Computing Machinery.