Markov Decision Processes

Krishnendu Chatterjee and Thomas A. Henzinger
scribed by Mirco Giacobbe and Lenka Matejovikova

Summarizing the modeling formalisms we have seen up to now we have

• Graphs that model *choices* and

• Markov Chains model *probabilistic transitions*.

In this lecture notes we will introduce Markov Decision Processes (MDPs) that combine *nondeterministic* transitions (choices) with *probabilistic* transitions.

**Example 1.** For instance consider the game represented in the following table:
The player starts at cell st. From each cell he can make a decision and go up (u), left (l) or right (r). The player looses the game with the probability indicated in the cell, while he wins if he reaches the smiley cell. This is modeled by the following MDP:

\[
\begin{array}{c|c}
\text{st} & 0.8 & \text{Smiley} \\
\end{array}
\]

\[
\begin{array}{c|c}
st & 0.2 \\
\end{array}
\]

**Definition 1 (Markov Decision Process).** A Markov Decision Process (MDP) is a tuple \((S, E, \delta, S_1, S_p)\) such that

• \(S\) is a set of locations partitioned in \(S_1 \subseteq S\) nondeterministic locations and \(S_p \subseteq S\) probabilistic locations \((S_1 \cap S_p = \emptyset)\),

• \(E \subseteq S \times S\) is a set of edges and

• \(\delta: S_p \to \mathcal{D}(S)\) gives a probability distribution over successors of probabilistic locations, s.t. \(s \in S_p\) and \((s, t) \in E\) iff \(\delta(s)(t) > 0\).
MDPs generalize both graphs and Markov chains, namely if \( S_1 = \emptyset \) then we have Markov chain, if \( S_p = \emptyset \) then we have a graph.

Choices in MPDs are made according to some strategy. A strategy is a recipe that looks at the history of the current MPD execution (finite path) and specifies how to extend it.

**Definition 2** (Strategy). A strategy is a function \( \sigma : S^* \times S_1 \to S \) s.t. for all \( \omega \in S^* \), \( s \in S_1 \) holds \( (s, \sigma(\omega, s)) \in E \).

In addition, we define the concept of memoryless strategy, which is a strategy that makes a choice only according to the current nondeterministic state.

**Definition 3** (Memoryless strategy). A memoryless strategy is a function \( \sigma : S_1 \to S \) s.t. for all \( s \in S_1 \) holds \( (s, \sigma(\omega, s)) \in E \).

MDPs define a probability measure \( P_\sigma(\cdot) \) with respect to a strategy \( \sigma \). Let \( \varphi \) be some property, then we define the outcomes of

- quantitative analysis as \( \sup_\sigma P_\sigma(\varphi) \),
- qualitative analysis as

\[
\text{Almost}(\varphi) = \{ s \in S \mid \exists \sigma. P_\sigma(\varphi) = 1 \}
\]

(with \( \sigma \) memoryless).

In the following we will discuss the analysis of MDPs in terms of the usual notions of

- reachability \( \Diamond T \)
- liveness \( 
\square \Diamond T \)

**Example 2.** Consider a spouse that have two lovers. When at home (H) the spouse can either decide to go by lover 1 (L1) or to go by lover 2 (L2). The lover must ensure to visit both infinitely often. The reader can easily see that \( L_2 \) one cannot construct a memoryless strategy. Memory is needed to guarantee interleaving among lovers.

**Observation 1.** Strategies with memory are more powerful.

**Example 3.** The figure below models a message sender with 2 options: wait (w) or send (s). In case of sending failure the sender can retry (R). In case of two consecutive failures, it stops running. A proper sender visits \( \infty \) infinitely often almost surely \( (s_0 \in \text{Almost}(\square \Diamond \infty)) \). The strategy that never retries satisfies the property.
Definition 4 (Random Attractors). Let the set $Z \subseteq S$ be a subset of states. We inductively define the construction of the random attractors as

$$
Z_0 := Z \\
Z_{i+1} := Z_i \cup \{s \in S_p \mid \exists (s, t) \in E \text{ and } t \in Z_i\} \\
\quad \cup \{s \in S_1 \mid \forall (s, t) \in E, t \in Z_i\},
$$

where the random attractors are defined as the fixpoint (see the illustration below)

$$
Attra(Z) = \bigcup_{i \geq 0} Z_i.
$$

Lemma 1 (Random attractor lemma). If $Z$ is a set of states, then from $Attra(Z)$ the set $Z$ is reached with positive probability independently of Player 1 strategy.

Proof. by induction

How to use this lemma? We use it to compute $Almost(\lozenge T)$ as follows:

- make the set $T$ absorbing;
- compute backward reachability $B(\lozenge T)$ as the MDP was a graph,
- take the nonreaching states $Z = S \setminus B(\lozenge T)$,
- compute the random attractors on the nonreaching states $Attra(Z)$,
Figure 1: (i) $Z = S \backslash B(\Diamond T)$  (ii) remove all states which cannot reach $T$ avoiding $Attra(Z)$  (iii) repeat the steps (i) and (ii) until fixpoint to obtain $Z^*$ (all removed states) and $A^*$.

- repeat on the MDP induced by the state space $A = B(\Diamond T) \backslash Attra(Z)$ until fixpoint.

What is the complexity? With $n = |S|$, it is $O(n \cdot Random\ Attractor)$

Proof of correctness. Let $A^*$ and $Z^*$ be the sets $A$ and $Z$ at fixpoint. We know by their construction that $A^* \subseteq S \backslash Z^*$ and that $A^* \subseteq B(\Diamond T)$. We can observe two properties:

(i) any probabilistic state in $A^*$ has no edge to $Z^*$,

(ii) any state in $A^*$ can reach $T$ by going through states in $A^*$.

By definition of qualitative analysis, if $A = Almost(\Diamond T)$ then for all $s \in A$ there exist a memoryless strategy $\sigma$ that leads to $T$ almost surely, i.e. with $P_s^T(\Diamond T) = 1$. Let $N(s)$ be the shortest path length from $s$ to any state $t \in T$, i.e. for all $s \in A^* \backslash T$ there exists $(s,t) \in E$ with $t \in A^*$ s.t. $N(t) < N(s)$. We define a memoryless strategy $\sigma: S_1 \cup A^* \rightarrow A^*$ as follows:

$$\sigma(s) := t \text{ s.t. } N(t) < N(s) .$$

By the properties observed above we know that the strategy is well defined.

A fixed memoryless strategy over an MDP induces a Markov chain (nondeterministic choices are made deterministic, then only the probabilistic component remains). We want to prove that $A^* = Almost(\Diamond T)$ by proving that $\sigma$ induces a Markov chain that reaches $T$ almost surely.

Assume toward contradiction that $\sigma$ does not induce a Markov chain which reaches $T$ almost surely. Then there must exist a bottom SCC $X \subseteq A^*$ s.t. $X \cap T = \emptyset$. Let $x^*$ be the minimum distance from $X$ to $T$, i.e. $x^* = \min_{s \in X} N(s)$. Let $s \in X$ be any state at such distance, i.e. $N(s) = x^*$.

(i) If $s \in S_p$ then all its outgoing edges lead to $X$, because $X$ is a bottom SCC. Moreover there must exist $t \in X. N(t) < N(s)$. But $N(s) = x^*$. Contradiction.

(ii) If $s \in S_1$ then $\sigma(s) = t \text{ s.t. } N(t) < N(s)$ and $t \in X$ because $X$ is a bottom SCC. But $N(s) = x^*$. Contradiction.
Then $X \cap T \neq \emptyset$, then the induced Markov chain reaches $T$ almost surely (because of some lemmas). □