Appendix

A SPECTRL syntax and semantics and the proof of Theorem 4

Syntax A specification in SPECTRL is defined in terms of *predicates* and *specification formulas*. An atomic predicate is a boolean function $a : \mathcal{X} \to \{\text{true}, \text{false}\}$ which for each system state specifies whether it satisfies the predicate, and a predicate is defined as a boolean combination of atomic predicates. Specification formulas in SPECTRL are then defined by the grammar

$$\phi := \text{achieve } b \mid \phi_1 \text{ ensuring } b \mid \phi_1; \phi_2 \mid \phi_1 \text{ or } \phi_2 \tag{1}$$

where b is a predicate and ϕ_1 and ϕ_2 are specification formulas. Intuitively, achieve b requires the agent to reach a state in which the predicate b is satisfied and ϕ_1 ensuring b requires the agent to satisfy the specification ϕ while only visiting states in which the predicate b is satisfied. The clause $\phi_1; \phi_2$ requires the agent to first satisfy specification ϕ_1 and then satisfy specification ϕ_2. Finally, ϕ_1 or ϕ_2 requires satisfaction of at least one of the specifications ϕ_1 or ϕ_2.

Semantics Given a trajectory $\rho = (\mathbf{x}_t, \mathbf{u}_t, \omega_t)_{t=0}^{\infty}$ and writing $\rho^K = (\mathbf{x}_t, \mathbf{u}_t, \omega_t)_{t=0}^{K}$ for its finite prefix of length K, the semantics of each SPECTRL clause are formally defined as follows:

$$\begin{array}{lll}
\rho & \models \phi & \exists K \in \mathbb{N}_0 \text{ s.t. } \rho^K \models \phi \\
\rho^K & \models \text{achieve } p & \exists\, t \leq K \text{ s.t. } p(\rho_t^K) = \text{true} \\
\rho^K & \models \phi_1 \text{ ensuring } p & \rho^K \models \phi_1 \wedge \forall\, t \leq K.\, p(\rho_t^K) = \text{true} \\
\rho^K & \models \phi_1; \phi_2 & \exists\, t \leq K \text{ s.t. } \rho^K_{[0:t]} \models \phi_1 \text{ and } \rho^K_{[t:K]} \models \phi_2 \\
\rho^K & \models \phi_1 \text{ or } \phi_2 & \rho^K \models \phi_1 \text{ or } \rho^K \models \phi_2
\end{array}$$

Here, ρ_t^K denotes the t-th state along ρ^K, $\rho^K_{[0:t]}$ denotes the prefix of ρ^K consisting of the first $t+1$ states along ρ^K and $\rho^K_{[t:K]}$ denotes the suffix of ρ^K that starts in the $(t+1)$-st state along ρ^K.

Theorem 4. *For each $\phi \in$ Finitary there exists $\phi' \in$ SPECTRL such that, for any word w, the word w is accepted by ϕ iff the word w is accepted by ϕ'.*

Proof. Let ϕ be a finitary specification defined over the set of atomic predicates AP. Since ϕ is finitary, there exist a finite time horizon H and a set L of words over AP of length H such that an infinite word over the alphabet AP is accepted by ϕ iff its prefix of length H is contained in L. Define a SPECTRL formula ϕ' via:

$$\phi' = \bigvee_{(w_1, \ldots, w_H) \in L} p(w_1)\,;\, p(w_2)\,;\, \ldots\,;\, p(w_H)$$

where each $p(w_i)$ is an atomic predicate associated to the i-th letter in the word $(w_1, \ldots, w_H)$ and ; denotes sequential composition of SPECTRL specifications. Then, an infinite word w is accepted by ϕ if and only if the prefix of w of length H is contained in L, which holds if and only if w is accepted by the SpectRL formula ϕ'. This completes our reduction. □

B Abstract Reachability Definition and Proof of Theorem 5

Given a trajectory $\rho = (\mathbf{x}_t, \mathbf{u}_t, \omega_t)_{t=0}^{\infty}$ of the system and an abstract graph $G = (V, E, \beta, s, t)$, we say that ρ satisfies *abstract reachability* for G (written $\rho \models G$) if it gives rise to a path in G that traverses G from s to t and satisfies reach-avoid specifications of all traversed edges. Formally, we require that there exists a sequence of time steps $0 = i_0 < i_1 < \cdots < i_k$ and a finite path $s = v_0, v_1, \ldots, v_k = t$ in G such that

1. $\mathbf{x}_{i_j} \in \beta(v_j)$ holds for each $0 \leq j \leq k$, and
2. $\mathbf{x}_t \in \beta(v_j, v_{j+1})$ holds for each $0 \leq j < k$ and $i_j \leq t \leq i_{j+1}$.

Intuitively, the first condition encodes that the trajectory satisfies reachability specifications of traversed vertices in G while the second condition encodes that it satisfies avoidance specifications of traversed edges in G. We then say that a policy π for the system satisfies *abstract reachability for G with probability $p \in [0, 1]$* at an initial state $\mathbf{x}_0 \in \mathcal{X}_0$, if we have that $\mathbb{P}_{\mathbf{x}_0}[\rho \in \Omega_{\mathbf{x}_0} \mid \rho \models G] \geq p$.

We now provide the proof of Theorem 5.

Theorem 5. *Consider a stochastic feedback loop system with an initial set of states $\mathcal{X}_0 \subseteq \mathcal{X}$ and let ϕ be a* SPECTRL *specification. Then there exists an abstract graph $G = (V, E, \beta, s, t)$ with $|V|$ in $\mathcal{O}(|\phi|)$ such that, for each trajectory ρ of the system, we have $\rho \models \phi$ if and only if $\rho \models G$. Hence, for each policy π and initial state $\mathbf{x}_0 \in \mathcal{X}_0$, we have $\mathbb{P}_{\mathbf{x}_0}[\rho \in \Omega_{\mathbf{x}_0} \mid \rho \models \phi] = \mathbb{P}_{\mathbf{x}_0}[\rho \in \Omega_{\mathbf{x}_0} \mid \rho \models G]$.*

Proof. Given a SPECTRL specification ϕ, one can construct an abstract graph G such that for each trajectory ρ of the system we have $\rho \models \phi$ iff $\rho \models G$ as follows. First, the specification ϕ is parsed according to the grammar of SPECTRL in eq. 1 in order to construct the parse tree of ϕ. We then start by constructing an abstract graph for each leaf formula in the parse tree, and traverse the parse tree bottom-up in order to construct abstract graphs of parent formulas. The abstract graph of the specification ϕ is then obtained by taking the abstract graph constructed for the root in the parse tree. The leaves of the parse tree are formulas of the form achieve p, for which we construct an abstract graph with two vertices s and t, a single edge $e = (s, t)$ and set $\beta(s) = \mathcal{X}_0$, $\beta(t) = \{\mathbf{x} \in \mathcal{X} \mid p(\mathbf{x}) = \text{true}\}$ and $\beta(e) = \mathcal{X}$. For a formula ϕ_1 ensuring p, we take an abstract graph (V, E, β, s, t) for the specification ϕ_1 which was already constructed for the child node and define the abstract graph $G = (V, E, \beta', s, t)$ by simply modifying the map β via $\beta'(e) = \beta(e) \cap \{\mathbf{x} \in \mathcal{X} \mid p(\mathbf{x}) = \text{true}\}$ for each $e \in E$. For a formula $\phi_1; \phi_2$, we take the abstract graph of the specifications ϕ_1 and ϕ_2 which were already constructed for the child nodes and merge them by identifying the target node of ϕ_1 with the source node of ϕ_2 and using the region associated to it by the abstract graph of ϕ_2. Finally, for a formula ϕ_1 or ϕ_2, we introduce a novel source node s with $\beta(s) = \mathcal{X}_0$, take the abstract graph of ϕ_1 and ϕ_2 and connect the novel source node s to them by an edge. Note that this construction yields a graph with V in $\mathcal{O}(|\phi|)$.

Since the above construction soundly encodes the semantics of each SPECTRL grammar element as a reach-avoid specification, it follows by induction on the depth of the parse tree that for each trajectory ρ of the system we have $\rho \models \phi$ iff $\rho \models G$. The claim of Theorem 5 follows. □

C Proof of Theorem 1

Theorem 1. *[Proof in Apendix C] The following two statements hold:*

1. *If a continuous function $V : \mathcal{X} \to \mathbb{R}$ is an (ϵ, λ)-additive RASM, then it is also a $(\frac{\lambda-\epsilon}{\lambda}, \min\{\epsilon, \lambda\}, \lambda)$-multiplicative RASM.*
2. *If a continuous function $V : \mathcal{X} \to \mathbb{R}$ is a $(\gamma, \delta, \lambda)$-multiplicative RASM, then it is also an $((1-\gamma)\cdot\delta, \lambda)$-additive RASM.*

Proof.

1. Let $\delta = \min\{\epsilon, \lambda\}$ and $\gamma = \frac{\lambda-\epsilon}{\lambda}$. To show that V is a $(\gamma, \delta, \lambda)$-multiplicative RASM, we need to show that the Strict positivity outside $\mathcal{X}_t$ and the Multiplicative expected decrease conditions hold. By the Additive expected decrease condition, for each $\mathbf{x} \in \mathcal{X}\backslash\mathcal{X}_t$ at which $V(\mathbf{x}) \leq \lambda$ we have $V(\mathbf{x}) \geq \epsilon$. So as $\delta = \min\{\epsilon, \lambda\}$, the Strict positivity outside $\mathcal{X}_t$ follows. On the other hand, observe that for every $\mathbf{x} \in \mathcal{X}\backslash\mathcal{X}_t$ at which $V(\mathbf{x}) \leq \lambda$, we have

$$\frac{\mathbb{E}_{\omega\sim d}[V(f(\mathbf{x}, \pi(\mathbf{x}), \omega))]}{V(\mathbf{x})} \leq \frac{V(\mathbf{x}) - \epsilon}{V(\mathbf{x})} \leq \frac{\lambda - \epsilon}{\lambda} = \gamma,$$

   where the first inequality follows by the Additive expected decrease condition and the second inequality follows since $\frac{z-\epsilon}{z}$ is monotonically increasing on the domain $z > \epsilon$. Hence, the Multiplicative expected decrease condition holds.

2. Let $\epsilon = (1-\gamma)\cdot\delta$. To show that V is an (ϵ, λ)-additive RASM, we need to show that the Additive expected decrease condition holds. We show this by observing that, for each $\mathbf{x} \in \mathcal{X}\backslash\mathcal{X}_t$ such that $V(\mathbf{x}) \leq \lambda$, we have

$$V(\mathbf{x}) - \mathbb{E}_{\omega\sim d}[V(f(\mathbf{x}, \pi(\mathbf{x}), \omega))] \geq V(\mathbf{x}) - \gamma \cdot V(\mathbf{x}) = (1-\gamma)\cdot V(\mathbf{x}) \geq (1-\gamma)\cdot\delta,$$

   where the first inequality holds by the Multiplicative expected decrease condition and the last inequality holds by the Strict positivity outside $\mathcal{X}_t$ condition. Hence, the Additive expected decrease condition is satisfied. □

D Proof of Theorem 2

We first provide an overview of definitions and results from martingale theory that we use in the proof. We then present the proof.

Probability theory A *probability space* is a triple $(\Omega, \mathcal{F}, \mathbb{P})$ of a state space Ω, a sigma-algebra $\mathcal{F}$ and a probability measure $\mathbb{P}$ that satisfies Kolmogorov axioms [58]. A *random variable* in $(\Omega, \mathcal{F}, \mathbb{P})$ is a function $X : \Omega \to \mathbb{R}$ that is $\mathcal{F}$-measurable, i.e. for each $a \in \mathbb{R}$ we have $\{\omega \in \Omega \mid X(\omega) \leq a\} \in \mathcal{F}$. A *(discrete-time) stochastic process* is a sequence $(X_i)_{i=0}^{\infty}$ of random variables in $(\Omega, \mathcal{F}, \mathbb{P})$.

Conditional expectation Let X be a random variable in $(\Omega, \mathcal{F}, \mathbb{P})$. Given a sub-sigma-algebra $\mathcal{F}' \subseteq \mathcal{F}$, a *conditional expectation* of X given $\mathcal{F}'$ is an $\mathcal{F}'$-measurable random variable Y such that, for each $A \in \mathcal{F}'$, we have

$$\mathbb{E}[X \cdot \mathbb{I}(A)] = \mathbb{E}[Y \cdot \mathbb{I}(A)].$$

Here, $\mathbb{I}(A) : \Omega \to \{0, 1\}$ is an *indicator function* of A, given by $\mathbb{I}(A)(\omega) = 1$ if $\omega \in A$, and $\mathbb{I}(A)(\omega) = 0$ if $\omega \notin A$. Intuitively, conditional expectation of X given $\mathcal{F}'$ is an $\mathcal{F}'$-measurable random variable that behaves like X upon evaluating its expected value on events in $\mathcal{F}'$. It is known that every nonnegative random variable admits a conditional expectation [58]. Moreover, the conditional expectation is almost-surely unique, meaning that for any two $\mathcal{F}'$-measurable random variables Y and Y' which are conditional expectations of X given $\mathcal{F}'$ we have $\mathbb{P}[Y = Y'] = 1$. Therefore, we pick any such random variable as a canonical conditional expectation and denote it by $\mathbb{E}[X \mid \mathcal{F}']$.

Supermartingales Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(\mathcal{F}_i)_{i=0}^{\infty}$ be an increasing sequence of sub-sigma-algebras in $\mathcal{F}$, i.e. $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \cdots \subseteq \mathcal{F}$. A nonnegative *supermartingale* with respect to $(\mathcal{F}_i)_{i=0}^{\infty}$ is a stochastic process $(X_i)_{i=0}^{\infty}$ such that each X_i is $\mathcal{F}_i$-measurable, and $X_i(\omega) \geq 0$ and $\mathbb{E}[X_{i+1} \mid \mathcal{F}_i](\omega) \leq X_i(\omega)$ hold for each $\omega \in \Omega$ and $i \geq 0$. Intuitively, the second condition is the expected decrease condition, and it is formally captured via conditional expectation.

We now present two results from martingale theory that will be used in the proof. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $(\mathcal{F}_i)_{i=0}^{\infty}$ be an increasing sequence of sub-σ-algebras in $\mathcal{F}$.

Theorem 6 (Supermartingale convergence theorem [58])**.** *Let $(X_i)_{i=0}^{\infty}$ be a nonnegative supermartingale with respect to $(\mathcal{F}_i)_{i=0}^{\infty}$. Then, there exists a random variable X_{∞} in $(\Omega, \mathcal{F}, \mathbb{P})$ to which the supermartingale converges to with probability* 1*, i.e. $\mathbb{P}[\lim_{i\to\infty} X_i = X_{\infty}] = 1$.*

Theorem 7 ([32])**.** *Let $(X_i)_{i=0}^{\infty}$ be a nonnegative supermartingale with respect to $(\mathcal{F}_i)_{i=0}^{\infty}$. Then, for every $\lambda > 0$, we have*

$$\mathbb{P}\Big[\sup_{i\geq 0} X_i \geq \lambda\Big] \leq \frac{\mathbb{E}[X_0]}{\lambda}.$$

Theorem 2. *[Proof in Appendix D] Let $\gamma \in (0, 1)$, $\delta > 0$ and $\lambda > 1$, and suppose that $V : \mathcal{X} \to \mathbb{R}$ is a $(\gamma, \delta, \lambda)$-multiplicative RASM with respect to $\mathcal{X}_t$ and $\mathcal{X}_u$. Suppose furthermore that V is Lipschitz continuous with a Lipschitz constant L_V, and that the system under policy π satisfies the* bounded step property*, i.e. that there exists $\Delta > 0$ such that $||\mathbf{x} - f(x, \pi(\mathbf{x}), \omega)||_1 \leq \Delta$ holds for each $\mathbf{x} \in \mathcal{X}$ and $\omega \in \mathcal{W}$. Let $N = \lfloor(\lambda - 1)/(L_V \cdot \Delta)\rfloor$. Then, for every $\mathbf{x}_0 \in \mathcal{X}_0$, we have that*

$$\mathbb{P}_{\mathbf{x}_0}\Big[\mathit{ReachAvoid}(\mathcal{X}_t, \mathcal{X}_u)\Big] \geq 1 - \frac{1}{\lambda} \cdot \gamma^N.$$

Proof. Fix an initial state $\mathbf{x}_0 \in \mathcal{X}_0$. We need to show that $\mathbb{P}_{\mathbf{x}_0}[\text{ReachAvoid}(\mathcal{X}_t, \mathcal{X}_u)] \geq 1 - \frac{1}{\lambda} \cdot \gamma^N$ with $N = \lfloor(\lambda - 1)/(L_V \cdot \Delta)\rfloor$.

Before proceeding with the proof, we define some notions. Consider the probability space $(\Omega_{\mathbf{x}_0}, \mathcal{F}_{\mathbf{x}_0}, \mathbb{P}_{\mathbf{x}_0})$ over the set of all system trajectories that start in $\mathbf{x}_0 \in \mathcal{X}_0$. For each time step $t \in \mathbb{N}_0$, define $\mathcal{F}_{\mathbf{x}_0,t} \subseteq \mathcal{F}_{\mathbf{x}_0}$ to be a sub-σ-algebra which contains events that are defined in terms of the first t states along a trajectory. Formally, for each $j \in \mathbb{N}_0$, let $C_j : \Omega_{\mathbf{x}_0} \to \mathcal{X}$ assign to each trajectory $\rho = (\mathbf{x}_t, \mathbf{u}_t, \omega_t)_{t\in\mathbb{N}_0} \in \Omega_{\mathbf{x}_0}$ the j-th state $\mathbf{x}_j$ along the trajectory. $\mathcal{F}_i$ is then defined as the smallest σ-algebra over $\Omega_{\mathbf{x}_0}$ with respect to which $C_0, C_1, \ldots, C_i$ are all measurable. The sequence $(\mathcal{F}_{\mathbf{x}_0,t})_{t=0}^{\infty}$ defines a filtration in $(\Omega_{\mathbf{x}_0}, \mathcal{F}_{\mathbf{x}_0}, \mathbb{P}_{\mathbf{x}_0})$.

Proceeding with the proof, we show that V induces a supermartingale in the probability space $(\Omega_{\mathbf{x}_0}, \mathcal{F}_{\mathbf{x}_0}, \mathbb{P}_{\mathbf{x}_0})$ over the set of all system trajectories that start in $\mathbf{x}_0 \in \mathcal{X}_0$. For each $t \in \mathbb{N}_0$, define a random variable X_t in $(\Omega_{\mathbf{x}_0}, \mathcal{F}_{\mathbf{x}_0}, \mathbb{P}_{\mathbf{x}_0})$ via

$$X_t(\rho) = \begin{cases} V(\mathbf{x}_t), & \text{if } \mathbf{x}_i \notin \mathcal{X}_t \text{ and } V(\mathbf{x}_i) < \lambda \text{ for all } 0 \leq i \leq t \\ 0, & \text{if } \mathbf{x}_i \in \mathcal{X}_t \text{ for some } 0 \leq i \leq t \text{ and } V(\mathbf{x}_j) < \lambda \text{ for all } 0 \leq j \leq i \\ \lambda, & \text{otherwise} \end{cases}$$

for each trajectory $\rho = (\mathbf{x}_t, \mathbf{u}_t, \omega_t)_{t \in \mathbb{N}_0} \in \Omega_{\mathbf{x}_0}$. Intuitively, X_t is equal to V at $\mathbf{x}_t$ until either the target set $\mathcal{X}_t$ is reached upon which X_t is set to 0, or some $V(\mathbf{x}_t) \geq \lambda$ is reached upon which X_t is set to λ. We claim that $(X_t)_{t=0}^{\infty}$ is a nonnegative supermartingale with respect to the filtration $(\mathcal{F}_{\mathbf{x}_0,t})_{t=0}^{\infty}$ in the probability space $(\Omega_{\mathbf{x}_0}, \mathcal{F}_{\mathbf{x}_0}, \mathbb{P}_{\mathbf{x}_0})$ and that it with probability 1 converges to either 0 or to a value that is $\geq \lambda$.

Clearly, each X_t is nonnegative. To prove that $(X_t)_{t=0}^{\infty}$ is a supermartingale, note first that each X_t is $\mathcal{F}_{\mathbf{x}_0,t}$ measurable as it is defined in terms of the ferst t states along a trajectory. To show that the expected decrease condition is satisfied, we show that $\mathbb{E}_{\mathbf{x}_0}[X_{t+1} \mid \mathcal{F}_{\mathbf{x}_0,t}](\rho) \leq X_t(\rho)$ holds for each $t \in \mathbb{N}_0$ and $\rho = (\mathbf{x}_t, \mathbf{u}_t, \omega_t)_{t \in \mathbb{N}_0}$. We consider 3 cases based on the definition of V:

1. If $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_t \notin \mathcal{X}_t$ and $V(\mathbf{x}_i) < \lambda$ for each $0 \leq i \leq t$, then

$$\begin{aligned}
&\mathbb{E}_{\mathbf{x}_0}[X_{t+1} \mid \mathcal{F}_{\mathbf{x}_0,t}](\rho) \\
&= \mathbb{E}_{\mathbf{x}_0}\Big[X_{t+1} \cdot \Big(\mathbb{I}(\mathbf{x}_{t+1} \notin \mathcal{X}_t \wedge V(\mathbf{x}_{t+1}) < \lambda) + \mathbb{I}(\mathbf{x}_{t+1} \in \mathcal{X}_t) + \mathbb{I}(V(\mathbf{x}_{t+1}) \geq \lambda)\Big) \mid \mathcal{F}_{\mathbf{x}_0,t}\Big](\rho) \\
&= \mathbb{E}_{\mathbf{x}_0}[X_{t+1} \cdot \mathbb{I}(\mathbf{x}_{t+1} \notin \mathcal{X}_t) \mid \mathcal{F}_{\mathbf{x}_0,t}](\rho) + 0 + \lambda \cdot \mathbb{E}[\mathbb{I}(V(\mathbf{x}_{t+1}) \geq \lambda) \mid \mathcal{F}_{\mathbf{x}_0,t}](\rho) \\
&\leq \mathbb{E}_{\omega \sim d}[V(f(\mathbf{x}_t, \mathbf{u}_t, \omega_t)) \cdot \mathbb{I}(\mathbf{x}_{t+1} \notin \mathcal{X}_t \wedge V(\mathbf{x}_{t+1}) < \lambda)] \\
&\quad + \mathbb{E}_{\omega \sim d}[V(f(\mathbf{x}_t, \mathbf{u}_t, \omega_t)) \cdot \mathbb{I}(\mathbf{x}_{t+1} \in \mathcal{X}_t)] \\
&\quad + \mathbb{E}_{\omega \sim d}[V(f(\mathbf{x}_t, \mathbf{u}_t, \omega_t)) \cdot \mathbb{I}(V(\mathbf{x}_{t+1}) \geq \lambda)] \\
&= \mathbb{E}_{\omega \sim d}[V(f(\mathbf{x}_t, \mathbf{u}_t, \omega_t))] \\
&\leq \gamma \cdot V(\mathbf{x}_t) \leq V(\mathbf{x}_t).
\end{aligned}$$

   The first equality follows by the law of total probability, the second equality follows by definition of X_t, the third inequality follows by observing that $V(\mathbf{x}_{t+1}) \geq X_{t+1}(\rho)$ if $\mathbf{x}_{t+1} \in \mathcal{X}_t$, the fourth equality is just the sum of expectations over disjoint sets, and finally the fifth inequality follows by the Multiplicative expected decrease condition of V and the assumption that $\mathbf{x}_t \notin \mathcal{X}_t$ and $V(\mathbf{x}_t) < \lambda$.

2. If $\mathbf{x}_i \in \mathcal{X}_t$ for some $0 \leq i \leq t$ and $V(\mathbf{x}_j) < \lambda$ for all $0 \leq j \leq i$, then we have $\mathbb{E}_{\mathbf{x}_0}[X_{t+1} \mid \mathcal{F}_{\mathbf{x}_0,t}](\rho) = \gamma \cdot X_{t+1} = X_{t+1}(\rho) = 0$.

3. Otherwise, we must have $V(\mathbf{x}_i) \geq \lambda$ and $\mathbf{x}_0, \ldots, \mathbf{x}_i \notin \mathcal{X}_t$ for some $0 \leq i \leq t$, thus $\mathbb{E}_{\mathbf{x}_0}[X_{t+1} \mid \mathcal{F}_{\mathbf{x}_0,t}](\rho) = X_{t+1}(\rho) = \lambda$.

Hence, we have proved that $(X_t)_{t=0}^{\infty}$ is a nonnegative supermartingale.

By Supermartingale Convergence Theorem, we then know that $(X_t)_{t=0}^{\infty}$ with probability 1 converges to some value. We claim furthermore that this value is either 0 or $\geq \lambda$ and that the value is attained. To see this, recall that in Theorem 1 we showed that V is also an (ϵ, λ)-additive RASM with $\epsilon = (1 - \gamma) \cdot \delta$. Then, the same sequence of inequalities as in the case one above shows that $\mathbb{E}_{\mathbf{x}_0}[X_{t+1} \mid \mathcal{F}_{\mathbf{x}_0,t}](\rho) \leq X_t(\rho) - \epsilon$ if $\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_t \notin \mathcal{X}_t$ and $V(\mathbf{x}_i) < \lambda$ for each $0 \leq i \leq t$. Thus, the limit to which $(X_t)_{t=0}^{\infty}$ converges cannot be in the open interval $(0, \lambda)$ and the claim follows.

To prove the theorem claim, we show that $(X_t)_{t=0}^{\infty}$ converges to a value with $\geq \lambda$ with probability at most $\frac{1}{\lambda} \cdot \gamma^N$. Then, since $(X_t)_{t=0}^{\infty}$ converging to 0 implies that system reaches a state in which $V < \delta$ while never reaching a state in which $V \geq \lambda$, which by the Safety and the Strict positivity outside $\mathcal{X}_t$ conditions implies that reach-avoidance is satisfied, this will imply the theorem claim. The proof until this point is analogous to the proof of [61, Theorem 1].

The technical novelty of our proof begins in the following step. We define another stochastic process $(Y_t)_{t=0}^{\infty}$ from $(X_t)_{t=0}^{\infty}$ by letting

$$Y_t = \begin{cases} X_t/\gamma^t, & \text{if } V(\mathbf{x}_i) < \lambda \text{ for all } 0 \leq i \leq t \\ Y_{t-1}, & \text{otherwise} \end{cases}$$

We claim that $(Y_t)_{t=0}^{\infty}$ is also a nonnegetive supermartingale with respect to the filtration $(\mathcal{F}_{\mathbf{x}_0,t})_{t=0}^{\infty}$. The nonnegativity part of the claim clearly holds since each X_t is nonnegative. To check the expected decrease condition of supermartingales, for each $t \in \mathbb{N}_0$ and for each $\rho \in \Omega_{\mathbf{x_0}}$ we have distinguish two cases:

1. If $V(\mathbf{x}_i) < \lambda$ for all $0 \leq i \leq t$, then

$$
\begin{aligned}
&\mathbb{E}_{\mathbf{x}_0}[Y_{t+1} \mid \mathcal{F}_{\mathbf{x}_0,t}](\rho) \\
&= \mathbb{E}_{\mathbf{x}_0}[Y_{t+1} \cdot \mathbb{I}(V(\mathbf{x}_{t+1}) < \lambda) \mid \mathcal{F}_{\mathbf{x}_0,t}](\rho) + \mathbb{E}_{\mathbf{x}_0}[Y_{t+1} \cdot \mathbb{I}(V(\mathbf{x}_{t+1}) \geq \lambda) \mid \mathcal{F}_{\mathbf{x}_0,t}](\rho) \\
&= \frac{1}{\gamma^{t+1}} \cdot \mathbb{E}_{\mathbf{x}_0}[X_{t+1} \cdot \mathbb{I}(V(\mathbf{x}_{t+1}) < \lambda) \mid \mathcal{F}_{\mathbf{x}_0,t}](\rho) + \mathbb{E}_{\mathbf{x}_0}[Y_t \cdot \mathbb{I}(V(\mathbf{x}_{t+1}) \geq \lambda) \mid \mathcal{F}_{\mathbf{x}_0,t}](\rho) \\
&= \frac{1}{\gamma^{t+1}} \cdot \mathbb{E}_{\mathbf{x}_0}[X_{t+1} \mid \mathcal{F}_{\mathbf{x}_0,t}](\rho) - \frac{1}{\gamma^{t+1}} \cdot \mathbb{E}_{\mathbf{x}_0}[X_{t+1} \cdot \mathbb{I}(V(\mathbf{x}_{t+1}) \geq \lambda) \mid \mathcal{F}_{\mathbf{x}_0,t}](\rho) \\
&+ \mathbb{E}_{\mathbf{x}_0}[Y_t \cdot \mathbb{I}(V(\mathbf{x}_{t+1}) \geq \lambda) \mid \mathcal{F}_{\mathbf{x}_0,t}](\rho) \\
&\leq \frac{1}{\gamma^{t+1}} \cdot \gamma \cdot X_t(\rho) - \mathbb{E}_{\mathbf{x}_0}[(\frac{1}{\gamma^{t+1}} \cdot X_{t+1} - Y_t) \cdot \mathbb{I}(V(\mathbf{x}_{t+1}) \geq \lambda) \mid \mathcal{F}_{\mathbf{x}_0,t}](\rho) \\
&= Y_t(\rho) - \mathbb{E}_{\mathbf{x}_0}[(\frac{1}{\gamma^{t+1}} \cdot X_{t+1} - Y_t) \cdot \mathbb{I}(V(\mathbf{x}_{t+1}) \geq \lambda) \mid \mathcal{F}_{\mathbf{x}_0,t}](\rho) \\
&\leq Y_t(\rho).
\end{aligned}
$$

   The first equality holds by the law of total probability. The second equality holds by the definition of Y_{t+1}. The third equality holds by the law of total probability. The fourth inequality holds since above we proved that $\mathbb{E}_{\mathbf{x}_0}[X_{t+1} \mid \mathcal{F}_{\mathbf{x}_0,t}](\rho) \leq \gamma \cdot X_t$ whenever $V(\mathbf{x}_i) < \lambda$ for all $0 \leq i \leq t$ (cases 1 and 2 above). The fifth equality holds by definition of Y_t and the assumption that $V(\mathbf{x}_i) < \lambda$ for all $0 \leq i \leq t$. Finally, the sixth inequality follows by observing that in the case when $V(\mathbf{x}_i) < \lambda$ for all $0 \leq i \leq t$ but $V(\mathbf{x}_{t+1}) \geq \lambda$, we have $X_{t+1}/\gamma^{t+1} \geq \lambda/\gamma^{t+1} \geq \lambda/\gamma^t \geq Y_t$.
2. If $V(\mathbf{x}_i) = \lambda$ for some $0 \leq i \leq t$, then $\mathbb{E}_{\mathbf{x}_0}[Y_{t+1} \mid \mathcal{F}_{\mathbf{x}_0,t}](\rho) = Y_t(\rho) = Y_i(\rho)$.

Hence, we have proved that $(Y_t)_{t=0}^{\infty}$ is a nonnegative supermartingale.

We conclude the theorem claim by observing that

$$
\begin{aligned}
&\mathbb{P}_{\mathbf{x}_0}\Big[\sup_{t\geq 0} X_t < \lambda\Big] = \mathbb{P}_{\mathbf{x}_0}\Big[\sup_{t\geq 0} \gamma^t \cdot Y_t < \lambda\Big] = \mathbb{P}_{\mathbf{x}_0}\Big[\sup_{t\geq N} \gamma^t \cdot Y_t < \lambda\Big] \\
&= \mathbb{P}_{\mathbf{x}_0}\Big[\gamma^N \cdot \sup_{t\geq N} \gamma^{t-N} \cdot Y_t < \lambda\Big] = \mathbb{P}_{\mathbf{x}_0}\Big[\sup_{t\geq N} \gamma^{t-N} \cdot Y_t < \frac{\lambda}{\gamma^N}\Big] \\
&\geq \mathbb{P}_{\mathbf{x}_0}\Big[\sup_{t\geq N} Y_t < \frac{\lambda}{\gamma^N}\Big] \geq \mathbb{P}_{\mathbf{x}_0}\Big[\sup_{t\geq 0} Y_t < \frac{\lambda}{\gamma^N}\Big] \geq 1 - \frac{1}{\lambda} \cdot \gamma^N.
\end{aligned}
$$

Three non-trivial steps are the first and the second equality and the last inequality. The first equality holds since, if $\sup_{t\geq 0} X_t < \lambda$, then we also have $Y_t = X_t/\gamma^t$ for each t by the definition of Y_t. The second equality holds since the system cannot reach a state in which $V \geq \lambda$ and so $X_t \geq \lambda$ in less than N time steps. On the other hand, for the last inequality we the inequality in Theorem 7. Applying the inequality to $(Y_t)_{t=0}^{\infty}$ and $\frac{\lambda}{\gamma^N}$ and observing that $\mathbb{E}[Y_0] = \mathbb{E}[X_0] \leq 1$ by the Initial condition in Definition 2 yields $\mathbb{P}_{\mathbf{x}_0}[\sup_{t\geq 0} Y_t \geq \lambda/\gamma^N] \leq \frac{\gamma^N}{\lambda}$ and thus the last inequality. □

E Learning Policies with Reach-avoid Supermartingales

We now present the POLICY+RASM subprocedure that we use for simultaneously learning a policy π_μ and an RASM V_θ, both of which are parametrized as neural networks with parameters μ and θ. The subprocedure POLICY+RASM is *identical* to the algorithm of [61], thus we keep this exposition brief and refer the reader to [61] for details. The reason why we can reuse this algorithm even though it learns additive RASMs is that additive and multiplicative RASMs are equivalent by Thereom 1. As we show below, the algorithm does not need to explicitly set an additive term ϵ or a multiplicative factor γ, thus it is applicable to learning both additive and multiplicative RASMs. We then show how to use V_θ to extract the bound in Theorem 2.

Analogously as in [61], the value $\lambda > 1$ in Definition 3 is an algorithm parameter and we initialize it to $\lambda = \frac{1}{1-p'}$ so that the Theorem **??** bound $1 - \frac{1}{\lambda} \cdot \gamma^N \geq 1 - \frac{1}{\lambda} = p'$ implies satisfaction of the desired probabilistic reach-avoid specification. If the algorithm succeeds in learning π_μ and V_θ with this value of λ, then the reach-avoid problem is solved. Otherwise, the algorithm gradually decreases the value of λ and tries to relearn π_μ and V_θ so that the resulting bound in Theorem 2 still exceeds p'. Thus, our new bound also yields an improvement in the algorithm.

The algorithm consists of two modules called *learner* and *verifier*, which are composed into a loop. In each loop iteration, the learner first learns a policy π_μ and an RASM candidate V_θ. These are then passed to the verifier which formally checks whether V_θ satisfies all conditions in Definition 2. If the verification is successful, the algorithm returns the policy. Otherwise, the verifier identifies *counterexample* states at which the additive RASM conditions are violated. These are then passed to the learner and are used to fine-tune the previously learned policy and RASM by refining the loss function using the computed counterexamples.

Learner A policy π_μ and an additive RASM candidate V_θ are learned by minimizing the loss function

$$\mathcal{L}(\theta, \nu) = \mathcal{L}_{\text{Init}}(\nu) + \mathcal{L}_{\text{Unsafe}}(\nu) + \mathcal{L}_{\text{Dec}}(\theta, \nu) + \mathcal{L}_{\text{Lipschitz}}(\theta, \nu).$$

The first three loss terms are constructed from the sets C_{init}, C_{unsafe} and C_{dec} which are initialized by computing finite discretizations of $\mathcal{X}_0$, $\mathcal{X}_u$ and $\mathcal{X} \backslash \mathcal{X}_t$ and are later extended by counterexamples computed by the verifier. The loss terms are used to guide the learner towards learning a true additive RASM which satisfies the Initial, Safety and Expected decrease conditions. Each loss term is designed to incur a loss at a counterexample whenever that counterexample violates the corresponding condition. In order for the Nonnegativity condition to be satisfied by default, the algorithm applies the softplus activation function to the output of V_θ. The loss term $\mathcal{L}_{\text{Lipschitz}}(\theta, \nu)$ is a regularization term that does not enforce any of the defining conditions of additive RASMs, however it helps in decreasing the Lipschitz constants of neural networks. Each loss term is defined as follows:

$$\mathcal{L}_{\text{Init}}(\nu) = \max_{\mathbf{x} \in C_{\text{init}}} \{V_\nu(\mathbf{x}) - 1, 0\}$$

$$\mathcal{L}_{\text{Unsafe}}(\nu) = \max_{\mathbf{x} \in C_{\text{unsafe}}} \{\frac{1}{1-p} - V_\nu(\mathbf{x}), 0\}$$

$$\mathcal{L}_{\text{Decrease}}(\theta, \nu) = \frac{1}{|C_{\text{dec}}|} \cdot$$

$$\sum_{\mathbf{x} \in C_{\text{decrease}}} \Big(\max \Big\{ \sum_{\omega_1, \ldots, \omega_N \sim \mathcal{N}} \frac{V_\nu\big(f(\mathbf{x}, \pi_\theta(\mathbf{x}), \omega_i)\big)}{N} - V_\theta(\mathbf{x}) + \tau \cdot K, 0 \Big\} \Big)$$

The last loss term $\mathcal{L}_{\text{Lipschitz}}(\theta, \nu) = t \cdot (\mathcal{L}_{\text{Lipschitz}}(\theta) + \mathcal{L}_{\text{Lipschitz}}(\nu))$ is the regularization term used to guide the learner towards learning neural networks whose Lipschitz constants are below a tolerable threshold ρ, where $t > 0$ is a regularization constant. By preferring networks with small Lipschitz constants, we allow the verifier to use a wider mesh and thus make verification condition easier to satisfy. We have The regularization term for π_θ (and analogously for V_ν) is defined via

$$\mathcal{L}_{\text{Lipschitz}}(\theta) = \max \Big\{ \prod_{W, b \in \theta} \max_j \sum_i |W_{i,j}| - \rho, 0 \Big\},$$

where W and b weight matrices and bias vectors for each layer in π_θ.

Verifier The verifier checks whether V_θ satisfies the defining properties of additive RASMs in Definition 2. Recall, the Nonnegativity condition is satisfied by default due to the softplus activation function applied to the output layer of V_θ. Hence, the verifier only needs to check the Initial, Safety and Expected decrease conditions.

Since f, π_μ and V_θ are continuous functions defined over a compact domain $\mathcal{X}$ thus also Lipschitz continuous, the verifier may check the (both Multiplicative and Additive) expected decrease condition by checking a slightly stricter condition at finitely many discretization points. A *discretization* of the state space $\mathcal{X}$ with *mesh* $\tau > 0$ is a finite set $\tilde{\mathcal{X}} \subseteq \mathcal{X}$ such that, for every $\mathbf{x} \in \mathcal{X}$, there exists $\tilde{\mathbf{x}} \in \tilde{\mathcal{X}}$ such that $||\mathbf{x} - \tilde{\mathbf{x}}||_1 < \tau$. The discretization is computed by taking a grid of mesh τ. Then, to check the expected decrease condition, it was showed in [61] that it suffices to check for each $\tilde{\mathbf{x}} \in \tilde{\mathcal{X}}$ whose

adjacent discretization grid cells contain a non-target state and over which V attains a value that is less than or equal to λ that

$$\mathbb{E}_{\omega\sim d}\Big[V_\theta\Big(f(\tilde{\mathbf{x}},\pi(\tilde{\mathbf{x}}),\omega)\Big)\Big] < V_\theta(\tilde{\mathbf{x}}) - \tau\cdot K,$$

where $K = L_V\cdot(L_f\cdot(L_\pi+1)+1)$ and L_f, L_π and L_V are Lipschitz constants of f, π_μ and V_θ. It is assumed that L_f is provided and L_π and L_V are computed by the method of [52]. To verify the Initial condition, the verifier collects the set $\text{Cells}_{\mathcal{X}_0}$ of all cells of the discretization grid that intersect the initial set $\mathcal{X}_0$. Then, for each cell $\in \text{Cells}_{\mathcal{X}_0}$, it checks if $\sup_{\mathbf{x}\in\text{cell}} V_\theta(\mathbf{x}) \leq 1$, where the supremum is bounded from above by using interval arithmetic abstract interpretation (IA-AI) [20, 24] to propagate across neural network layers the extreme values that V_θ can attain over a cell. Similarly, to verify the Unsafe condition, the verifier collects the set $\text{Cells}_{\mathcal{X}_u}$ of all cells of the discretization grid that intersect the initial set $\mathcal{X}_u$. Then, for each cell $\in \text{Cells}_{\mathcal{X}_u}$, it uses IA-AI to check if $\inf_{\mathbf{x}\in\text{cell}} V_\theta(\mathbf{x}) \geq \lambda$.

If the verifier shows that V_θ satisfies the above checks, it concludes that V_θ is an additive (and therefore multiplicative) RASM for the system under the policy π_μ and returns the policy together with the lower bound on the probability of satisfying the reach-avoid specification as in Theorem **??**. The fact that the verifier is correct was proved in [61, Theorem 2]. Otherwise, if a counterexample $\tilde{\mathbf{x}}$ to any of the checks is found, it is added to one of the three counterexample sets C_{init}, C_{unsafe} and C_{dec} that are then used by the learner to fine-tune V_θ and π_μ.

Soundness and computation of γ. The following theorem establishes that the above is a sound verification procedure and provides a closed-form expression for the values of $\delta > 0$ and $\gamma \in (0,1)$ for which V_θ is a (γ,δ,λ)-multiplicative RASM. Hence, to compute the lower bound on the probability of satisfying reach-avoidance in Theorem 2, one may use the value of γ implied by the theorem together with λ which is fixed by the algorithm, L_V which is computed by the algorithm and the maximal step size Δ which we assume is provided by the user.

Theorem 8. *If the verifier returns neural networks π_μ and V_θ, then V_θ is a (γ,δ,λ)-multiplicative RASM with*

$$\delta = \min\Big\{\min_{\tilde{\mathbf{x}}\in\tilde{\mathcal{X}}}\Big(V_\theta(\tilde{\mathbf{x}}) - \tau\cdot K - \mathbb{E}_{\omega\sim d}[V_\theta(f(\tilde{\mathbf{x}},\pi(\tilde{\mathbf{x}}),\omega))]\Big),\ \lambda\Big\},\quad \gamma = 1-\frac{\delta}{\lambda}.$$

Proof. Since it was shown in [61] that the verifier provides a sound verification procedure for checking that V_θ is an additive RASM, by the equivalence in Theorem 1 it follows that it is also sound for checking that V_θ is a (γ,δ,λ)-multiplicative RASM for some values of γ and δ. Hence, it remains to show that the values of γ and δ in the theorem statement are correct.

First, we show that the Strict positivity outside $\mathcal{X}_t$ condition is satisfied with the above value of δ. To see this, let $\mathbf{x}\in\mathcal{X}\backslash\mathcal{X}_t$. Since $\delta\leq\lambda$, suppose without loss of generality that $V_\theta(\mathbf{x})\leq\lambda$. Then, $\mathbf{x}$ is contained in a discretization grid cell which contains a non-target state and over which V attains a value that is $\leq\lambda$. Hence, the verifier has shown that $\mathbb{E}_{\omega\sim d}[V_\theta(f(\tilde{\mathbf{x}},\pi(\tilde{\mathbf{x}}),\omega))] < V_\theta(\tilde{\mathbf{x}}) - \tau\cdot K$ holds for each $\tilde{\mathbf{x}}\in\tilde{\mathcal{X}}$ which is a vertex of this cell. Taking a vertex $\tilde{\mathbf{x}}\in\tilde{\mathcal{X}}$ of this cell for which $||\mathbf{x}-\tilde{\mathbf{x}}||_1\leq\tau$, by the definition of Lipscthiz constants we have

$$\begin{aligned}
&\mathbb{E}_{\omega\sim d}\Big[V\Big(f(\mathbf{x},\pi(\mathbf{x}),\omega)\Big)\Big]\\
&\leq \mathbb{E}_{\omega\sim d}\Big[V\Big(f(\tilde{\mathbf{x}},\pi(\tilde{\mathbf{x}}),\omega)\Big)\Big] + ||f(\tilde{\mathbf{x}},\pi(\tilde{\mathbf{x}}),\omega) - f(\mathbf{x},\pi(\mathbf{x}),\omega)||_1\cdot L_V\\
&\leq \mathbb{E}_{\omega\sim d}\Big[V\Big(f(\tilde{\mathbf{x}},\pi(\tilde{\mathbf{x}}),\omega)\Big)\Big] + ||(\tilde{\mathbf{x}},\pi(\tilde{\mathbf{x}}),\omega) - (\mathbf{x},\pi(\mathbf{x}),\omega)||_1\cdot L_V\cdot L_f\\
&\leq \mathbb{E}_{\omega\sim d}\Big[V\Big(f(\tilde{\mathbf{x}},\pi(\tilde{\mathbf{x}}),\omega)\Big)\Big] + ||\tilde{\mathbf{x}}-\mathbf{x}||_1\cdot L_V\cdot L_f\cdot(1+L_\pi)\\
&\leq \mathbb{E}_{\omega\sim d}\Big[V\Big(f(\tilde{\mathbf{x}},\pi(\tilde{\mathbf{x}}),\omega)\Big)\Big] + \tau\cdot L_V\cdot L_f\cdot(1+L_\pi).
\end{aligned}$$

Hence, by the Nonnegativity condition we also have

$$\begin{aligned} V(\mathbf{x}) &\geq V(\mathbf{x}) - \mathbb{E}_{\omega\sim d}\Big[V\Big(f(\mathbf{x},\pi(\mathbf{x}),\omega)\Big)\Big] \\ &\geq V(\tilde{\mathbf{x}}) - \tau\cdot L_V - \mathbb{E}_{\omega\sim d}\Big[V\Big(f(\tilde{\mathbf{x}},\pi(\tilde{\mathbf{x}}),\omega)\Big)\Big] - \tau\cdot L_V\cdot L_f\cdot(1+L_\pi) \\ &= V(\tilde{\mathbf{x}}) - \tau\cdot K - \mathbb{E}_{\omega\sim d}\Big[V\Big(f(\tilde{\mathbf{x}},\pi(\tilde{\mathbf{x}}),\omega)\Big)\Big] \\ &\geq \delta, \end{aligned} \tag{2}$$

since $K = L_V\cdot(L_f\cdot(L_\pi+1)+1)$, which concludes the proof.

Second, we show that Multiplicative expected decrease condition with the above value of γ holds. Let $\mathbf{x}\in\mathcal{X}\backslash\mathcal{X}_t$ be such that $V_\theta(\mathbf{x})\leq\lambda$. We need to show that $\gamma\cdot V(\mathbf{x})\geq\mathbb{E}_{\omega\sim d}[V(f(\mathbf{x},\pi(\mathbf{x}),\omega))]$. Since we showed in eq. (2) that $V(\mathbf{x})\geq V(\mathbf{x})-\mathbb{E}_{\omega\sim d}[V(f(\mathbf{x},\pi(\mathbf{x}),\omega))]\geq\delta>0$ and since $V_\theta(\mathbf{x})\leq\lambda$, we have

$$\begin{aligned} \frac{\mathbb{E}_{\omega\sim d}[V(f(\mathbf{x},\pi(\mathbf{x}),\omega))]}{V(\mathbf{x})} &= 1 - \frac{V(\mathbf{x})-\mathbb{E}_{\omega\sim d}[V(f(\mathbf{x},\pi(\mathbf{x}),\omega))]}{V(\mathbf{x})} \\ &\leq 1-\frac{\delta}{V(\mathbf{x})}\leq 1-\frac{\delta}{\lambda}. \end{aligned}$$

This concludes the proof. □

F Proof of Theorem 3

Theorem 3. *[Proof in Appendix F] Algorithm 1 is compositional, and if it outputs a policy π, then π guarantees the probabilistic specification (ϕ,p).*

Proof. In order to prove that π guarantees satisfaction of the probabilistic specification (ϕ,p), by Theorem 5 it suffices to show that π satisfies abstract reachability for the abstract graph G with probability at least p.

To prove abstract reachability for G with probability at least p, we show that a random trajectory of the system under policy π satisfies reach-avoid specifications of the edges along the finite path $s=v_{i_0},v_{i_1},\ldots,v_{i_k}=t$ exhibited above with probability at least p. To prove this, we proceed by induction on $0\leq j\leq k$ to show that a random trajectory of the system under policy π satisfies reach-avoid specifications of each edge along a prefix $s=v_{i_0},v_{i_1},\ldots,v_{i_j}$ of this path with probability at least $\mathrm{Prob}[v_{i_j}]$. Recall, Prob is the dictionary computed by Algorithm 1. Abstract reachability for G with probability at least p then follows if we set $j=k$, since $v_{j_k}=t$ and we must have $\mathrm{Prob}[t]\geq p$ for Algorithm 1 to output a policy (lines 15-17).

The base case $j=0$ follows trivially since the system starts in the initial region $\beta(s)=\mathcal{X}_0$ and since $\mathrm{Prob}[v_{i_0}]=\mathrm{Prob}[s]=1$ by line 6. For the inductive step, suppose that $0\leq j\leq k-1$ and that π satisfies reach-avoid specifications of each edge along $s=v_{i_0},v_{i_1},\ldots,v_{i_j}$ with probability at least $\mathrm{Prob}[v_{i_j}]$. The claim for the prefix of length $j+1$ then follows by our construction of the finite path $s=v_{i_0},v_{i_1},\ldots,v_{i_k}=t$, as it implies that Algorithm 1 has successfully learned an edge policy for the edge $(v_{i_j},v_{i_{j+1}})$ that ensures satisfaction of the associated reach-avoid specification with probability at least $p_{(v_{i_j},v_{i_{j+1}})}$ and for which $\mathrm{Prob}[v_{i_{j+1}}]=p_{(v_{i_j},v_{i_{j+1}})}\cdot\mathrm{Prob}[v_{i_j}]$. Since the right-hand-side of this equality is a lower bound on the probability of π satisfying reach-avoid specifications of each edge along $s=v_{i_0},v_{i_1},\ldots,v_{i_j}$ multiplied by a lower bound on the probability of it satisfying the reach-avoid specification of the edge $(v_{i_j},v_{i_{j+1}})$, the claim follows. This concludes the proof by induction. □