Temporal logic control of general Markov decision processes by approximate policy refinement

Sofie Haesaert * Sadegh Soudjani ** Alessandro Abate ***

* California Institute of Technology, United States
** School of Computing, Newcastle University, United Kingdom
*** Computer Science Department, Oxford University, United Kingdom

Abstract: The formal verification and controller synthesis for general Markov decision processes (gMDPs) that evolve over uncountable state spaces are computationally hard and thus generally rely on the use of approximate abstractions. In this paper, we contribute to the state-of-the-art of control synthesis for temporal logic properties by computing and quantifying a less conservative gridding of the continuous state space of linear stochastic dynamic systems and by giving a new approach for control synthesis and verification that is robust to the incurred approximation errors. The approximation errors are expressed as both deviations in the outputs of the gMDPs and in the probabilistic transitions.

1. INTRODUCTION

With the ever more ubiquitous embedding of digital components into physical systems, new computationally efficient verification and control synthesis methods for these cyber-physical systems are needed. Quite importantly, stochastic models of these cyber-physical systems are key to model how computers interact with physical systems such as biological processes, power networks, and smart-grids. In this work, we are interested in the verification and control synthesis for such stochastic models with respect to probabilistic linear temporal logic properties.

Using tools such as PRISM (Kwiatkowska et al., 2011), temporal logic properties defined over finite-state Markov (decision) processes can be verified and policies can be designed to control these Markov decision processes such that the satisfaction of these properties is maximised. For discrete-time stochastic models over uncountable state spaces, the characterisation of these properties is often not be attained analytically (Abate et al., 2008). An alternative is to approximate these models by simpler processes, such as finite-state MDP (Soudjani and Abate, 2013) or continuous-space reduced order models (Safonov and Chiang, 1989) that are prone to be mathematically analysed or algorithmically verified (Soudjani et al., 2015).

In (Haesaert et al., 2017a, 2016), we have proposed (ε, δ)-approximate stochastic similarity relations to bound the deviations between models in both the output signals (ε) and in the transition probabilities (δ). For approximately similar models a control policy synthesised on an abstract model can be refined to an approximately similar model with quantified precision.

2. PROBLEM SET-UP: MODELS AND SPECIFICATIONS

In this work, we focus on Borel measurable spaces (X, B(X)) defined over Polish spaces X (Bogachev, 2007). Together with the measurable space (X, B(X)), a probability measure P defines the probability space, denoted by (X, B(X), P) and has realisations x ∼ P. Let us further denote the set of all probability measures for a given measurable space (X, B(X)) as P(X, B(X)). For a given set X, we denote a metric or distance function on X as dX : X × X → R≥0. For the Euclidean space R^n, we define the weighted two-norm of a vector as \[ \|x\|_M := \sqrt{x^TMx} \] with positive definite matrix M, and \[ \|x\| := \sqrt{x^Tx} \] for any \( x \in \mathbb{R}^n \). For the sets A and B a relation \( R \subset A \times B \) is a subset of the Cartesian product \( A \times B \). The relation \( R \) relates \( x \in A \) with \( y \in B \) if \( (x, y) \in R \), denoted as \( xRy \).

2.1 General Markov decision processes and control strategies

General Markov decision processes extend upon Markov decision processes (Bertsekas and Shreve, 1996) and are formalised next.

Definition 1. (general Markov decision process (gMDP)).

A discrete-time gMDP is a tuple \( \{X, \pi, T, U, h, Y \} \) with \( X \), an (uncountable) Polish state space with states \( x \in X \) as its elements; \( U \), the set of controls, which is a Polish space; \( \pi \), the initial probability measure \( \pi : B(X) \rightarrow [0,1] \); \( T : \mathbb{X} \times \mathbb{X} \times B(X) \rightarrow [0,1] \), a conditional stochastic kernel assigning to each state \( x \in X \) and control \( u \in U \) a probability measure \( T(\cdot | x, u) \) over \( (X, B(X)) \); \( Y \), the output space decorated with metric dY; and \( h : \mathbb{X} \rightarrow b \), a measurable output map.

Given a string of inputs \( \{u(t)\}_{t \leq N} := u(0), u(1), \ldots, u(N) \) over a finite time horizon \([0, N]\), and an initial condition \( x_0 \) sampled from \( \pi \), the state at the \((t+1)\)-st time instant, \( x(t+1) \),
is obtained as a realisation of the controlled Borel-measurable stochastic kernel $\mathbb{T} \cdot (\cdot | x(t), u(t))$ – these semantics induce paths (or executions) of the gMDP. Further, output traces of a gMDP is obtained by applying the output map $h(\cdot)$ to the paths of the gMDP, namely \( \{y(t)\}_{t \in \mathbb{N}} := y(0), y(1), \ldots, y(N) \) with $y(t) = h(x(t))$ for all $t \in [0, N]$. Denote the class of all gMDPs with the same metric output space $\mathcal{Y}$ as $\mathcal{M}_Y$.

When the control inputs are selected based only on the current states, this is referred to as a Markov policy. A Markov policy $\mu$ is a sequence $\mu = \{(\mu_0, \mu_1, \mu_2, \ldots)\}$ of universally measurable maps $\mu_t : X \to \mathcal{P}(U, B(U))$, $t \in \mathbb{N} := \{0, 1, 2, \ldots\}$, from the state space $X$ to the set of controls. A Markov policy $\mu$ is stationary or time homogeneous if $\mu = (\mu_1, \mu_2, \ldots)$ for some $\mu_s$. For control inputs chosen according to a probability measure $\mu_s : B(U) \to [0, 1]$, denote the transition kernel as \( \mathbb{T}(\cdot | x, \mu_s) = \int_0^1 \mathbb{T}(\cdot | x, u) \mu_s(du) \in \mathcal{P}(\mathbb{X}, B(\mathbb{X})) \).

A more general set of control policies are those that depend on the past history of states and controls. For this we introduce the notion of a control strategy, and define it as a broader, memory-dependent version of the above Markov policy.

**Definition 2. (Control strategy).** A control strategy \( C = (X_C, x_0, X, \tau_C, h_C) \) for a gMDP \( M = (X, \pi, T, U, h, \mathcal{Y}) \) is a gMDP with state space $X_C$; initial state $x_0$; input space $X$; universally measurable kernel $\tau_C : X_C \times X \times B(X_C) \to [0, 1]$; and universally measurable output map $h_C : X_C \to \mathcal{P}(U, B(U))$.

The control strategy is formulated as a gMDP that takes as its input the state of the to-be-controlled gMDP. As in Figure 1, the execution \( \{(x(t), x_C(t))\}_{t \in \mathbb{N}} \) of a gMDP $M$ controlled with strategy $C$ (denoted by $C \times M$) is defined on the canonical sample space $\Omega := \{(X \times X_C)^{\mathbb{N}+1}\}$ endowed with its product topology $\mathbb{B}(\Omega)$ and with a unique probability measure $\mathbb{P}_{C \times M}$.

**2.2 Probabilistic path properties of controlled gMDPs**

Consider a set of atomic propositions $AP$ that defines the alphabet $\Sigma := 2^{AP}$ for which each letter of the alphabet evaluates a subset of the atomic propositions as true. Infinite words are strings composed of letters from $\Sigma$, $\omega = \omega(0), \omega(1), \omega(2), \ldots \in \Sigma^{\mathbb{N}}$. Of interest are atomic propositions that are connected to the gMDP via a measurable labelling function $L : \mathcal{Y} \to \Sigma$ from the output space to the alphabet $\Sigma$. Via a trivial extension, output traces $\{y(t)\}_{t \geq 0} \in \Sigma^{\mathbb{N}}$ are mapped to the set of infinite words $\Sigma^{\mathbb{N}}$, as $\omega = L(\{y(t)\}_{t \geq 0}) := \{L(y(t))\}_{t \geq 0}$. It is over the atomic propositions of these words that we define the desired temporal behaviour. Consider properties defined in a fragment of linear-time temporal logic (LTL) known as syntactically co-safe temporal logic (scLTL) (Kupferman and Vardi, 2001).

**Definition 3.** A scLTL over a set of atomic propositions $AP$ has syntax

\[
\psi ::= \text{true} \mid \neg \psi \mid \psi_1 \wedge \psi_2 \mid \psi_1 \vee \psi_2 \mid \bigcirc \psi \mid \psi_1 \psi_2 \mid \psi_1 \psi_2 \mid \psi_2 \psi_2 \mid \phi \psi_2 \mid \phi_2 (1)
\]

with $p \in AP$.

Let $\omega_t = \omega(t), \omega(t+1), \omega(t+2), \ldots$ be a postfix of the word $\omega$, then the satisfaction relation between $\omega$ and a property $\psi$ is denoted by $\omega \models \psi$ (or equivalently $\omega_0 \models \psi$).

The semantics of the satisfaction relation are defined recursively over $\omega_t$ as follows. An atomic proposition $p \in AP$ is satisfied by $\omega_t$, i.e., $\omega_t \models p$, iff $p \in \omega(t)$. Furthermore, $\omega_t \not\models \neg p$ if $\omega_t \not\models p$, and we say that $\omega_t \models \psi_1 \wedge \psi_2$ if $\omega_t \models \psi_1$ and $\omega_t \models \psi_2$. Similarly $\omega_t \models \psi_1 \lor \psi_2$ holds if $\omega_t \models \psi_1$ or if $\omega_t \models \psi_2$.

The next operator $\omega_t \models \bigcirc \psi$ holds if the property holds at the next time instance $\omega_{t+1} \models \psi$. The temporal until operator $\omega_t \models \psi_1 \psi_2$ holds if $3i \in \mathbb{N} : \omega_{t+i} \models \psi_2$ and $\forall j \in \mathbb{N} : 0 \leq j < i, \omega_{t+j} \models \psi_1$. Furthermore, the satisfaction of eventually $\psi_2$ or reach $\psi_2$, i.e., $\psi_2$ can be inferred from its equivalent representation $\bigcirc \neg \psi_2$. Similarly, we often denote the time-bounded reachability of $\psi$ as $\bigcirc_t N \psi_2$.

With respect to a scLTL property $\psi$, we say that a gMDP $M$ satisfies $\psi$ for a given control strategy $C$ with probability at least $p$ iff $\mathbb{P}_{C \times M}(L(\{y(t)\}_{t \geq 0}) \models \psi) \geq p$, or equivalently, iff $\mathbb{P}_{C \times M}(\omega \models \psi) \geq p$. This allows us to define the control synthesis problem tackled in this paper as follows.

**Problem 1. (Temporal logic control).** Given a gMDP $M$, a scLTL property $\psi$ and a labelling function $L$, compute a control strategy $C$ that maximises the probability that the controlled Markov process $C \times M$ satisfies $\psi$, i.e.,

\[
\max_{C} \mathbb{P}_{C \times M}(L(\{y(t)\}_{t \geq 0}) \models \psi).
\]

The verification of scLTL properties is formulated using deterministic finite-state automata (DFAs), as defined next.

**Definition 4. (DFA).** A DFA is a tuple $A = (Q, q_0, \Sigma, F, t)$, where $Q$ is a finite set of locations, $q_0 \in Q$ is the initial location, $\Sigma$ is a finite set, $F \subseteq Q$ is a set of accepting locations, and $t : Q \times \Sigma \rightarrow Q$ is a transition function.

A word $\omega$ is accepted by a DFA $A$ if there exists a finite run $q = (q(0), \ldots, q(n)) \in Q^{n+1}$ such that $q(0) = q_0, q(i+1) = t(q(i), \sigma(i))$ for all $0 \leq i \leq n$ and $q(n) \in F$. The accepted language of $A$, denoted by $L(A)$, is the set of all words accepted by $A$. For every scLTL property $\psi$ as in Def. 3, there exists a DFA $A_\psi$ such that $\omega \models \psi \iff \omega \in L(A_\psi)$. As a result, the satisfaction of the property $\psi$ now becomes equivalent to reaching the accepting states in the DFA. Thus in Eq. (2), the probability that the controlled Markov process $C \times M$ satisfies a scLTL property $\psi$, is equal to

\[
\mathbb{P}_{C \times M}(\omega \models \psi) = \mathbb{P}_{C \times M}(L(\{y(t)\}_{t \geq 0}) \in L(A_\psi)).
\]

We can reduce the computation of $\mathbb{P}_{C \times M}(\omega \in L(A_\psi))$ over the traces $\omega$ of $M$ to a reachability problem over another gMDP $M \otimes A_\psi$, which is a product of the gMDP $M$ and the automaton $A_\psi$. This was originally derived in (Tkachev et al., 2013) for MDPs. We give a similar definition of the product construction as follows.

**Definition 5. (Product between automaton and gMDP).** Given a gMDP $M = (X, \pi, T, U, h, \mathcal{Y})$, a DFA $A_\psi = (Q, q_0, \Sigma, F, t)$, and a labelling function $L : \mathcal{Y} \rightarrow \Sigma$, we define the product of $M$ and $A_\psi$ to be another gMDP denoted as

\[
M \otimes A_\psi = (X, \pi, T, U, h, \mathcal{Y}, \mathcal{X}, \hat{h}(x,q) = h(x) \text{ for any } (x,q) \in \mathcal{X}, \text{ and } \hat{T}(A \times \{q\})[x,u] = \int_{x \in A} 1(q = t(q_0, L(h(x)))) \cdot \mathbb{T}(dx|x,u), \text{ and }.
\]

initialised with $\hat{n}(dx,q) = 1(q = t(q_0, L(h(x)))) \cdot \pi(dx)$.
The quantity $\mathbb{P}_{C \times M} (\omega \in L(A_0))$ can be related to the reachability probability over the gMDP $M \otimes A$ with goal states $F$, as it was shown to be the case for MDPs in (Tkachev et al., 2013). Moreover, given a Markov policy $\mu$ on the product space of $A_0 \otimes M$, a control strategy for $M$, denoted by $C(\mu, \psi)$, can be computed such that

$$\mathbb{P}_{C(\mu, \psi) \times M} (\omega \in L(A_0)) = \mathbb{P}_{\mu \times (A_0 \otimes M)} (\omega \in F).$$

(4)

2.3 Problem statement and approach

Since the temporal logic control problem in Problem 1 is computationally hard to solve, we split it up into two subproblems:

1. For a given concrete model $M$ find an abstract model $\overline{M}$ with quantified deviations (Sec. 3).

2. Find a robust solution method for Problem 1, such that a robust control strategy of Problem 1 computed for $\overline{M}$ automatically yields a controller for $M$ (Sec. 4).

3. SIMULATION RELATIONS AND ABSTRACTIONS

3.1 Approximate simulation relations for gMDPs

Consider two gMDPs $M_i = (X_i, \pi_i, T_i, U_i, h_i, Y_i)$, $i = 1, 2$, that share an output space $Y$ with metric $d_Y$. Given state-action pair $x_1 \in X_1, u_1 \in U_1$ and $x_2 \in X_2, u_2 \in U_2$, we want to relate the corresponding transition kernels, namely the probability measures $\mathbb{T}_1 (\cdot | x_1, u_1) \in P(X_1, B(X_1))$ and $\mathbb{T}_2 (\cdot | x_2, u_2) \in P(X_2, B(X_2))$. As in (Haesaert et al., 2017a), we introduce the concept of $\delta$-lifting as follows.

**Definition 6.** ($\delta$-lifting for general state spaces). Let $X_1, X_2$ be two sets with associated measurable spaces $(X_1, B(X_1)), (X_2, B(X_2))$, and let $\mathcal{R} \subseteq X_1 \times X_2$ be a relation for which $\mathcal{R} \in B(X_1 \times X_2)$. We denote by

$$\mathcal{R}_\delta \subseteq P(B(X_1)) \times P(B(X_2))$$

the corresponding lifted relation so that $\Delta \mathcal{R}_\delta \Theta$ holds if there exists a probability space $(X_1 \times X_2, B(X_1 \times X_2), \mathcal{W})$ (equivalently, a lifting $\mathcal{W}$) satisfying

**L1.** For all $X_1 \in B(X_1)$; $\mathcal{W}(X_1 \times X_2) = \Delta (X_1)$;

**L2.** For all $X_2 \in B(X_2)$; $\mathcal{W}(X_1 \times X_2) = \Theta (X_2)$;

**L3.** For the probability space $(X_1 \times X_2, B(X_1 \times X_2), \mathcal{W})$ it holds that $x_1 \mathcal{R}_\delta x_2$ with probability at least $1 - \delta$, or equivalently that $\mathcal{W} (\mathcal{R}) \geq 1 - \delta$.

We will use a notion of approximate stochastic simulation relations that naturally leads to control refinement. For this, we require the notion of an interface function (Girard and Pappas, 2009) that refines control actions as follows

$$U_\epsilon : U_1 \times X_1 \times X_2 \rightarrow P(U_2, B(U_2)).$$

Intuitively, an interface function implements (or refines) any action control synthesized over the abstract model to an action for the concrete model.

**Definition 7.** ($\epsilon, \delta$-stochastic simulation relation). Let $M_i = (X_i, \pi_i, T_i, U_i, h_i, Y_i)$, $i = 1, 2$, be two gMDPs that share an output space $Y$ with metric $d_Y$. We say that $M_1$ is ($\epsilon, \delta$)-stochastically simulated by $M_2$ if there exists a measurable interface function $U_\epsilon$ and a relation $\mathcal{R}_\epsilon \subseteq X_1 \times X_2$, for which there exists a Borel measurable stochastic kernel $\mathcal{W}_\Gamma \subseteq U_1 \times X_1 \times X_2$ on $X_1 \times X_2$ given $U_1 \times X_1 \times X_2, \mathcal{W}$, such that:

**APS1.** $\forall (x_1, x_2) \in \mathcal{R}, d_Y (h_1 (x_1), h_2 (x_2)) \leq \epsilon;$

**APS2.** $\forall (x_1, x_2) \in \mathcal{R}, \forall u \in U_1 :$

$$(T_1 (\cdot | x_1, u_1) \mathcal{R}_\epsilon T_2 (\cdot | x_2, U_\epsilon (u_1, x_1, x_2)))$$

with lifted probability measure $\mathcal{W}_\Gamma (\cdot | u_1, x_1, x_2);$}

**APS3.** $\pi_1 R_\delta \pi_2.$

The simulation relation is denoted as $M_1 \preceq_M^\delta M_2$. □

For bounded safety and reachability properties, we can leverage this approximate simulation relation to refine computations performed on an abstract model back to the original model. This builds on the following proposition of Haesaert et al. (2017a).

**Proposition 8.** If $M_1 \preceq_M^\delta M_2$, then for all control strategies $C_i$ there exists a control strategy $C_2$ such that, for all measurable events $A \in \mathcal{Y}^{N+1}$,

$$\mathbb{P}_{\mathcal{O}, \mathcal{M}} (\{y(t)\}_{t \leq N} \in A) - \gamma \leq \mathbb{P}_{\mathcal{O}, \mathcal{M}} (\{y_2(t)\}_{t \leq N} \in A) - \gamma \leq \mathbb{P}_{\mathcal{O}, \mathcal{M}} (\{y_1(t)\}_{t \leq N} A) + \gamma,$$

with constant $1 - \gamma := (1 - \delta)^{N+1}$, and with the $\epsilon$-expansion of $A$ defined as

$$A_\epsilon := \{\{y(t)\}_{t \leq N} | \exists \{y_{t}\}_{t \leq N} \in A : \max_{0 \leq t \leq N} d_Y (y(t), y_{t}) \leq \epsilon \}$$

and similarly the $\epsilon$-contraction of $A$ defined as

$$A_{-\epsilon} := \{\{y(t)\}_{t \leq N} | \exists \{y_{t}\}_{t \leq N} \in A : \max_{0 \leq t \leq N} d_Y (y(t), y_{t}) \leq \epsilon \}$$

where $\{\{y(t)\}_{t \leq N}\}$ is the point-wise $\epsilon$-expansion of the discrete set $\{y(t)\}_{t \leq N}$.

For small values of $\epsilon$, the probability deviation can be approximated linearly as $\gamma \approx (N + 1) \delta$. Clearly, $\gamma$ is composed of the probabilistic deviation incurred in $N$-transitions, together with the deviation in the initial probability measures.

3.2 Abstraction of linear gMDPs

Existing results on formal controller synthesis for the class of linear stochastic models either rely on model order reduction (Lavaei et al., 2017) or use abstraction techniques as finite-state MDPs (Soudjani and Abate, 2013). In this section, we present an approach that combines model order reduction with an abstraction to a finite-state model. In contrast to the standard abstraction accuracy, we quantify the gridding error via the disturbance it induces in the state trajectory.

**Concrete model.** Consider the following linear gMDP $M_2$:

$$x_2(t + 1) = A_2 x_2(t) + B_2 u_2(t) + B_{u2} w(t), \quad t = 0, 1, 2, \ldots, y_2(t) = C_2 x_2(t), \quad x_2(0) = x_2_0 \in X_2,$$

(5)

where $x_2(t) \in X_2 \subset \mathbb{R}^n, u_2(t) \in U_2 \subset \mathbb{R}^p$, and $y_2(t) \in \mathcal{Y} \subset \mathbb{R}^q$. Matrices $A_2, B_2, B_{u2}$, and $C_2$ have appropriate dimensions and $w(t)$ is iid with standard Gaussian distributions.

**Constructing the abstract model.** For the concrete model $M_2$, we compute a lower order model with state space $X_\delta$. Then the construction of the abstract model relies on partitioning this new space $X_\delta \subset \mathbb{R}^{n_\delta}$, where $n_\delta < n$, as $\bigcup_{i=1}^{n_\delta} A_{\delta i} = X_{\delta i}$. Over this partition, we select representative points $\{z_i \subset A_{\delta i}, i = 1, 2, \ldots, \}$, and we call this set $X_{\delta i}$, which becomes the state space of the abstract model $M_1$. Introduce the operator $I_\delta : X_{\delta i} \rightarrow X_i$ that assigns to any $x_i \in A_{\delta i}, i \in \{1, \ldots, \}$ the representative point of $A_{\delta i}, z_i = I_\delta (x_i)$. Next we provide a dynamical characterisation of $M_1$. The state evolution of $M_1$ is written as

$$x_1(t + 1) = I_\delta (A_1 x_1(t) + B_1 u_1(t) + B_{u1} w(t)), \quad y_1(t) = C_1 x_1(t), \quad x_1(0) = x_1 \in X_1, \quad t = 0, 1, 2, \ldots$$

(6)

with state $x_1(t) \in X_1$, input $u_1(t) \in U_1$, and output $y_1(t) \in \mathcal{Y}$, and matrices $A_1, B_1, B_{u1}, C_1$ of appropriate dimensions. Note that the noise term $w(t)$ in $M_1$ is the same as the one in $M_2$.\]
The probability of satisfying a scLTL property can be quantified in (5), Theorem 9. The attenuation of the disturbance inputs $i$ are reached within $\bar{x}$, $w$, $\beta$ such that $\|w\| \leq c_w$, $\|u\| \leq c_u$, $\|\bar{x}\|_M \leq \epsilon$, $|\beta| \leq \delta$. Matrices in (8) are defined as $A := A_2 + B_{\bar{x}}K$, $B := B_{\bar{x}}R - PB_1$, $B_w := B_{\bar{x}}w - PB_w$. Vector $\delta$ is the diameter of the partition $\{A_i, i = 1, \ldots, l\}$, which satisfies $|x_i - x_i'| \leq \delta$ component-wise for any $x_i, x_i' \in A_i$ and any $i \in \{1, 2, \ldots, l\}$. Notice that the output deviation $\epsilon$ depends on the attenuation of the disturbance inputs $B_{\bar{x}}u + B_w w + \beta$. When $B_w = 0$, for instance if there is no order reduction, the resulting approximate simulation relation does not have a deviation in probability $\delta = 0$. Condition (8) can be checked using LMIs and S-procedure (Boyd and Vandenberghe, 2004).

Theorem 9. $M_I$ in (6) is $(\epsilon, \delta)$-stochastically simulated by $M_2$ in (5), $M_I \subseteq^\delta M_2$, with interface function (7) if $C_I^T C_I \leq M$, condition (8) is satisfied, and for a given state initialization $x_0$, $\|\|_M \leq \epsilon$ with $P \hat{=} (P^T M P)^{-1} P T M$.

4. ROBUST TEMPORAL LOGIC CONTROL FOR $(\epsilon, \delta)$-DEVIATIONS

4.1 Computing satisfaction probability of scLTL properties

The probability of satisfying a scLTL property can be quantified as the probability that the set of accepting states $F$ is reached over the product gMDP $M \otimes \mathcal{A}_\psi$ as in Eq. (4). For a given time horizon $N$ and Markov policy $\mu$, define time-dependent value functions $V^\mu_{N-i}$, $i \in \{0, N\}$, as the probability that the set of accepting states $F$ is reached within $i$ time steps, i.e.,

$$V^\mu_{N-i}(x, q) = \mathbb{E}\left[\sum_{i=0}^{N} 1_F(q_i) \prod_{j=0}^{i-1} 1_{Q^i}(q_j) \mid x_0 = x, q_0 = q\right],$$

with the expectation defined over the state transitions $(x, q)$ of the process controlled with the Markov policy $\mu$, denoted as $\mu \times (\mathcal{A}_\psi \otimes \mathcal{M})$. These value functions can be computed via backward recursions, initialised with $V_0 = 0$, and iterated for $k = N - 1, \ldots, 0$ as

$$V^\mu_k(x, q) = T^\mu_{k+1}(x, q), \quad \text{with} \quad (9) \quad T^\mu_{k+1}(x, q) = \int_{x, \xi, \xi'} V^\mu_{k+1}(x, q) \|d\xi, \xi' |x, q, \mu_k(x, q))$$

Based on the final value function after $N$ iterations, we have that the $N$-horizon reachability probability is given as

$$P_{\mu \times (\mathcal{A}_\psi \otimes \mathcal{M})}(\diamond^N F) = \int_{x, q} \max (1_F(q), V^\mu_0(x, q)) \#(dx, q).$$

Furthermore, the optimal value functions $V^*_k(x, q)$, $k \in \{0, N\}$ are computed as

$$V^*_k(x, q) = T^*(V^*_{k+1})(x, q), \quad \text{with} \quad \mu^*_k(x) = \arg \sup_{\mu} T^\mu(x, q)$$

with the optimal Bellman operator $T^*(\cdot) := \sup_{\mu} T^\mu(\cdot)$ and they give the optimal $N$-horizon reachability probability

$$\max_{\mu^*}(\mathcal{A}_\psi \otimes \mathcal{M})(\diamond^N F) = \int_{x, q} \max (1_F(q), V^*_0(x, q)) \#(dx, q).$$

Using $V^*_k(x, q)$, the elements $\mu^*_k(x)$ of the optimal Markov policy $\mu^*$ are computed as

$$\mu^*_k(x) = \arg \sup_{\mu_k} T^\mu(x, q)$$

Based on Eq. (4), the satisfaction probability is computed as the unbounded optimal reachability probability, i.e., with $N \to \infty$ as

$$P(C_\psi \times M_\mu(w \in \mathcal{L}(A_\psi)) = \lim_{N \to \infty} P(\mu \times (\mathcal{A}_\psi \otimes \mathcal{M}) (\diamond^N F)).$$

More specifically, the optimal value functions are strictly increasing with the time horizon and converge to the fixed point solution $V^*(x, q) = T^*(V^*(x, q))$ with

$$V^*(x, q) = \lim_{N \to \infty} (T^*)^N(V^*_N(x, q), V_N = 0.$$
The proof of Lemma 12 requires the existence of a refined control strategy as given in Prop. 8. Unlike the result in Prop. 8, for the δ-robust computation, the probabilistic deviation is now relative to the effective length of satisfying traces.

Before tackling unbounded reachability properties, we first analyse the behaviour of $T'_\phi$ and $T_\phi$. Suppose that $W_1(x,q) \geq W_2(x,q)$ for all $(x,q)$, then for a given map $\nu : \mathbb{N} \times Q \to \mathcal{P}(U_1,\mathcal{B}(U_1))$, we have $T'_\phi(W_1)(x,q) \geq T'_\phi(W_2)(x,q)$, hence $T_\phi(W_1)(x,q) \geq T_\phi(W_2)(x,q)$. Therefore, for a given stationary Markov policy $\mu$, the series of functions $\{T'_\phi(W)(x,q)\}_{q \geq 0}$ is initialised with $V = 0$ is point-wise increasing, since it is monotonic and upper bounded. Additionally, the same holds for functions $\{T_\phi(W)(x,q)\}_{q \geq 0}$. For a given stationary Markov policy $\mu$, we can now extend Eq. (12) to the $(0, \delta)$-robust computation as follows

$$R_{\mu \times (A_\psi \otimes M_1)}(\{F\}) := \mathbf{L}\left(\int_{x \times q} \max_{F_\phi} (1 F_\phi, V_\psi)(x,q) \pi(dx, x, q) - \delta\right)$$

with $V_\psi : Y_1 \to [0,1]$, the solution of $V_\psi = T'_\phi(V_\psi)$, computed as the limit of the sequence $\{T'_\phi(V_\psi)\}_{q \geq 0}$ that is initialised with $V = 0$. If $V_\psi$ is computed similarly as the solution of $V_\psi = T'_\phi(V_\psi)$ and $\mu^* \in \arg \sup T'_\phi(V_\psi)$ then we call $\mu^*$ the optimal $(0,\delta)$-robust policy. As in Eq. (14), for every stationary Markov policy $\mu$ for $A_\psi \otimes M_1$, there exists a control strategy $C_\phi(\mu, \psi)$ that preserves the $(0,\delta)$-robustness, i.e.,

$$R_{\mu \times (A_\psi \otimes M_1)}(\{F\}) = R_{C_\phi(\mu, \psi)}.$$

We formalise this next.

**Theorem 13.** Given a gMDP $M_1$ and a scLTL specification $\psi$, a control strategy $C_\phi(\mu, \psi)$ computed as (16) satisfies the specification $(0,\delta)$-robustly as $R_{C_\phi(\mu, \psi)}$. Moreover we can refine $C_\phi(\mu, \psi)$ to $C_\phi(\mu, \psi)$ such that $\psi$ is satisfied by $C_\phi(\mu, \psi) \times M_2$ with a probability $\rho \geq R_{C_\phi(\mu, \psi)}$. 

### 4.3 $(\epsilon, \delta)$-Robust satisfaction of scLTL properties

We now integrate the error $\epsilon$ in the output space into the robust satisfaction problem. Define the expansion of elements of the output space as $\{y_\epsilon \in \mathbb{Y}, d_y(y, y_\epsilon) \leq \epsilon\}$. A robustified version of the labelling can now be defined as

$$L_\epsilon(y) := \{\alpha \in \Sigma | \exists y_\epsilon \in \{y_\epsilon \epsilon \in \mathbb{Y}, d_y(y, y_\epsilon) \leq \epsilon\} \epsilon \in L(\epsilon)\}.$$

Consider $M_1 \subseteq M_2$ with $R_\epsilon$, then for all $(x_1, x_2) \in R_\epsilon$, it holds that $L(\epsilon_2(x_2)) \subseteq L(\epsilon_1(x_1))$. Instead of integrating this relaxed labelling into the product construction of a given gMDP, we will immediately adapt the $(\epsilon,\delta)$-robust reachability computations in Eq. (14) to deal with this non-determinism. Consider the $(\epsilon,\delta)$-robust operator $T_{\epsilon,\delta}(V)(x_1, q_1)$ defined as

$$T_{\epsilon,\delta}(V)(x_1, q_1) := \mathbf{L}\left(\int_{x_1, q_1} \min_{y_1 \in \{y_1 \epsilon \in \mathbb{Y}, d_y(y, y_1) \leq \epsilon\}} \max_{F_\phi} (1 F_\phi, V_{\log_{\delta}}(x_1, q_1)) \times \mathbb{T}(dx_1 | x_1, \nu(x_1, q_1)) - \delta\right)$$

with $t_{\delta, \epsilon}(q, x_1) := \{t(q, \alpha) \in L_\epsilon(h_1(x_1))\}$. For a stationary Markov policy $\mu$ and $V(x_1, q_1)$ satisfying $V(x_1, q_1) = T_{\epsilon,\delta}(V)(x_1, q_1)$, the $(\epsilon,\delta)$-robust reachability probability is defined as

$$L\left(\int_{x_1, q_1} \min_{y_1 \in \{y_1 \epsilon \in \mathbb{Y}, d_y(y, y_1) \leq \epsilon\}} \max_{F_\phi} (1 F_\phi, V(x_1, q_1)) \pi(dx_1) - \delta\right).$$

Consider a scLTL property $\psi$ and the corresponding $A_\psi$ with goal states $F$. If $F$ is $\delta$-robustly reachable with probability $r$, then we can refine $\mu$ to $C_\delta(\mu, \psi)$ such that $\psi$ is satisfied by $C_\delta(\mu, \psi) \times M_2$ with a probability $p \geq r$. Of course the apparent non-determinism – due to the relaxed labelling – will be resolved in the refined control strategy by selecting the labels of the concrete model.

We can also maximise the $(\epsilon,\delta)$-robust probability using $T_{\epsilon,\delta}$, defined as

$$T_{\epsilon,\delta}(V)(x_1, q_1) := \sup_{\mu} T_{\epsilon,\delta}^\mu(V)(x_1, q_1),$$

which yields an optimised robust stationary Markov policy as

$$T_{\epsilon,\delta}(V^*)(x_1, q_1) = \sup_{\mu} T_{\epsilon,\delta}^\mu(V^*)(x_1, q_1)$$

for $T_{\epsilon,\delta}(V^*)(x_1, q_1) = V^*(x_1, q_1)$ if $\delta > 0$. In conclusion, we have shown that we can leverage approximate stochastic simulation relations to use approximate models for controller synthesis and for verification of scLTL properties.

### 4.4 $(\epsilon, \delta)$-Optimistic satisfaction of scLTL properties

We now investigate whether we can quantify an upper bound on the satisfaction probability of a scLTL property using an approximate model $M_1$.

Consider the $(\epsilon,\delta)$-optimistic operator $T_{\epsilon,\delta}(V)(x_1, q_1)$ defined as

$$T_{\epsilon,\delta}(V)(x_1, q_1) := \sup_{\mu} \mathbf{L}\left(\int_{x_1, q_1} \min_{y_1 \in \{y_1 \epsilon \in \mathbb{Y}, d_y(y, y_1) \leq \epsilon\}} \max_{F_\phi} (1 F_\phi, V_{\log_{\delta}}(x_1, q_1)) \times \mathbb{T}(dx_1 | x_1, \nu(x_1, q_1)) + \delta\right),$$

with $t_{\delta, \epsilon}(q, x_1) := \{t(q, \alpha) \in L_\epsilon(h_1(x_1))\}$. 

**Definition 14.** $(\epsilon,\delta)$-Optimistic satisfaction. Consider a gMDP $M_1 \subseteq M_2$. We say that a control strategy $C_\phi$ for $M_1$ $(\epsilon,\delta)$-optimistically satisfies $\psi$ with probability $p$ if for all $M_2 \in M_Y$ with $M_1 \subseteq \epsilon$ $M_2$ and for all controllers $C_2$ for $M_2$ it holds that $\mathbb{P}_{C_\phi,M_{c,\mu}}(\omega \models \psi) \leq p$.

**Theorem 15.** Given a gMDP $M_1$ and a scLTL specification $\psi$, a control strategy $C_\phi$ computed based on the $(\epsilon,\delta)$-optimistic operator $T_{\epsilon,\delta}$ satisfies $\psi(\epsilon,\delta)$-optimistically.

### 5. CASE STUDIES

**Toy example.** We consider specification $\psi = \Box \Box_{\leq n_2} x \in K$ which encodes reach and stay over bounded time intervals. The associated DFA is given in Figure 2, together with an illustration of a potential application of this toy example. Consider the original model $M_2$, which is a 3-dimensional model with output $y_1(t) = x_a$ and $y_2(t + 1) = x_a(t) + a_1 \cdot x_b(t) + a_2 u(t) + a_3 w(t)$ $x_3(t + 1) = b x_b(t) + u(t)$ $x_3(t + 1) = c_1 x_c(t) + c_2 w(t)$ (18) with $a_1 = 0.3, a_2 = 0.03, a_3 = 0.006, b = c_1 = 0.8$ and $c_2 = 0.1$. For the game we select $n_2 = 3$. We follow Section 3.2 and compute the lower dimensional model with state $x_1$ via

![Fig. 2. Game of tag: $\Box \Box_{\leq n_2} \{x_a \in K\}$ with the DFA (right).](image-url)
satisfying the specification is computed based on the abstract
ψ to space discretization δ. We abstract the model without order reduction (P).

Fig. 3. On the left: (ε, δ)-robust satisfaction probability of
\( \Diamond \Box \psi \{ y \in [-2, 2] \} \) with \( \epsilon = 1.2266 \) and \( \delta = 0.03 \). On the right: simulation runs for the original model and the abstract model with the composed robust controller.

Fig. 4. Left: Environment of the robot with obstacles (●), a package (●●), and a client collection point (●). Closed-loop executions of robot fulfills the specification ψ in (19). Right: Robust probabilities computed based on the abstract model.

In this paper, we have introduced a new robust way of synthesising control strategies and verifying probabilistic temporal logic properties. Beyond this theoretical contribution, future work will focus on the computational aspects of this approach to prepare for application on more realistic problems.

REFERENCES

Abate, A., Prandini, M., Lygeros, J., and Sastry, S. (2008). Probabilistic Reachability and Safety for Controlled Discrete Time Stochastic Hybrid Systems. *Automatica*, 44(11), 2724–2734.

Bertsekas, D. and Shreve, S.E. (1996). *Stochastic Optimal control : The discrete time case*. Athena Scientific.

Bogachev, V.I. (2007). *Measure theory*. Springer Science & Business Media.

Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. CUP, Cambridge.

Girard, A. and Pappas, G.J. (2009). Hierarchical control system design using approximate simulation. *Automatica*, 45(2), 566–571.

Haesaert, S., Abate, A., and Van den Hof, P.M.J. (2016). Verification of general Markov decision processes by approximate similarity relations and policy refinement. In *QEST*, 227–243.

Haesaert, S., Soudjani, S., and Abate, A. (2017a). Verification of general Markov decision processes by approximate similarity relations and policy refinement. *SIAM Journal on Control and Optimization*, 55(4), 2333–2367.

Haesaert, S., Soudjani, S., and Abate, A. (2017b). Temporal logic control of general markov decision processes by approximate policy refinement. *CoRR*, abs/1712.07622.

Kupferman, O. and Vardi, M.Y. (2001). Model checking of safety properties. *Formal Methods in System Design*, 291–314.

Kwiatkowska, M., Norman, G., and Parker, D. (2011). PRISM 4.0: Verification of probabilistic real-time systems. In *CAV*, volume 6806 of LNCS, 585–591. Springer.

Lavaei, A., Soudjani, S., Majumdar, R., and Zamani, M. (2017). Compositional abstractions of interconnected discrete-time stochastic control systems. In *CDC*, 3551–3556.

Safonov, M.G. and Chiang, R. (1989). A Schur method for balanced-truncation model reduction. *IEEE Transactions on Automatic Control*, 34(7), 729–733.

Soudjani, S., Gevaerts, C., and Abate, A. (2015). FAUST²: Formal Abstractions of Uncountable-State Stochastic Processes. In *TACAS*, LNCS, 272–286. Springer Berlin.

Soudjani, S. and Abate, A. (2013). Adaptive and sequential gridding procedures for the abstraction and verification of stochastic processes. *SIAM Journal on Applied Dynamical Systems*, 12(2), 921–956.

Tkachev, I., Mereacre, A., Katoen, J.P., and Abate, A. (2013). Quantitative automata-based controller synthesis for non-autonomous stochastic hybrid systems. In *HSCC*, 293–302.