2021 60th IEEE Conference on Decision and Control (CDC)
December 13-15, 2021. Austin, Texas

# Temporal-Logic-Based Intermittent, Optimal, and Safe Continuous-Time Learning for Trajectory Tracking

Aris Kanellopoulos[1], Filippos Fotiadis[1], Chuangchuang Sun[2], Zhe Xu[3], Kyriakos G. Vamvoudakis[1], Ufuk Topcu[4], Warren E. Dixon[5]

2021 60th IEEE Conference on Decision and Control (CDC) | 978-1-6654-3659-5/21/$31.00 ©2021 IEEE | DOI: 10.1109/CDC45484.2021.9683309

***Abstract*— In this paper, we develop safe reinforcement-learning-based controllers for systems tasked with accomplishing complex missions that can be expressed as linear temporal logic specifications, similar to those required by search-and-rescue missions. We decompose the original mission into a sequence of tracking sub-problems under safety constraints. We impose the safety conditions by utilizing barrier functions to map the constrained optimal tracking problem in the physical space to an unconstrained one in the transformed space. Furthermore, we develop policies that intermittently update the control signal to solve the tracking sub-problems with reduced burden in the communication and computation resources. Subsequently, an actor-critic algorithm is utilized to solve the underlying Hamilton-Jacobi-Bellman equations. Finally, we support our proposed framework with stability proofs and showcase its efficacy via simulation results.**

## I. Introduction

Assured autonomy is challenging in problems with complex dynamics, unknown models and adversarial environments. For learning-based systems, designing high-confidence, high-performance, and dynamically-configured secure policies for avoiding unsafe operating regions is of paramount importance. Recognizing that reinforcement learning (RL) [1] is an important component of assured autonomy, we focus on learning-enabled systems. A large class of mission objectives can be modeled as a requirement to follow multiple reference trajectories and endpoints sequentially. One such scenario is found in search and rescue missions in which an autonomous vehicle follows specific search trajectories, with possible intermediate stops for recharging, before eventually returning to a specific location after certain conditions have been met [2]. Temporal logic specifications [3] offer a systematic way of describing the different modes of operation of such systems as well as the ways that those modes are interconnected through time. Similarly, to facilitate the use of those methods in high-risk environments, energy expenditure should be minimized by developing strategies that alleviate the burden on the communication and computation resources of the system. This need leads to the introduction of event-triggered mechanisms [4]; techniques that can create another layer of safety by minimizing the opportunities of external signals to affect the system.

The problem of safe learning has been in the forefront in recent years. The authors in [5] combine control barrier functions with Lyapunov control functions to construct safe controllers. In [6], the authors deal with the problem of safety via a dynamic invariance control framework, while the authors of [7] investigate Hamilton-Jacobi-based reachability methods to address the issue. The authors of [8] develop an approximate online adaptive solution to an optimal control problem under safety constraints with the use of barrier functions and sparse learning. The use of temporal logic specifications for safety has been studied in [9], where regulation problems have been solved via RL techniques while temporal logic specifications are guaranteed to be satisfied. In [10], deep Q-learning was leveraged to guarantee given specifications in a Markov decision process framework. The motion-planning problem in a multi-robot system under temporal logic specifications is investigated in [11] where the authors use a library of motion primitives to accomplish the given mission.

The framework of event-triggered control has been extensively investigated in the literature, e.g., in [12]. While the authors in [13] expanded the framework to take into account output feedback, most implementations remained static. Event-triggered control has been used in tandem with RL techniques in various scenarios. In [14], we have brought together intermittent mechanisms to alleviate the burden of an actor-critic framework. This was later extended for systems with unknown dynamics under a Q-learning framework in [15]. The authors of [16] developed a controller with intermittent communication for a multi-agent system whose safety constraints were expressed by metric temporal logic specifications. Finally, event-triggering methods were employed to the problem of autonomous path planning [17].

*Contributions:* The contributions of this work are threefold. First, we formulate a system tasked with accomplishing a mission consisting of regulation and tracking sub-problems.

[1]A. Kanellopoulos, F. Fotiadis, and K. G. Vamvoudakis are with the Daniel Guggenheim School of Aerospace Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA, 30332, USA, e-mail: {ariskan, ffotiadis, kyriakos}@gatech.edu.

[2]Chuangchuang Sun is with the Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA, email: ccsun1@mit.edu.

[3]Zhe Xu is with the School for Engineering of Matter, Transport, and Energy, Arizona State University, Tempe, AZ 85287, email: xzhe1@asu.edu.

[4]Ufuk Topcu are with the Department of Aerospace Engineering and Engineering Mechanics, and the Oden Institute for Computational Engineering and Sciences, University of Texas, Austin, TX 78712, USA, email: utopcu@utexas.edu.

[5]Warren E. Dixon is with the the Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL 32611 − 6250, USA, e-mail: wdixon@ufl.edu

This work was supported in part, by ARO under grant No. W911NF-19 − 1 − 0270, by ONR Minerva under grant No. N00014 − 18 − 1 − 2160, by NSF under grant Nos. CAREER CPS-1851588 and S&AS 1849198, and by the Onassis Foundation-Scholarship ID: F ZQ 064 − 1/2020 − 2021.

We decouple the problems through the use of a finite state automaton (FSA), which also models safety constraints. Secondly, we apply a barrier function-based transformation on the system and the required trajectories, thus allowing us to map the original problem into a series of optimal tracking sub-problems. Finally, we employ an actor-critic framework to solve the underlying tracking problems in a data-driven fashion.

## II. Problem Statement

Consider the time-invariant control-affine nonlinear system

$$\dot{x} = f(x) + g(x)u,\ x(0) = x_0,\ \forall t \geqslant 0, \tag{1}$$

where $x \in \mathbb{R}^n,\ u \in \mathbb{R}$ are the states and the control input, respectively. The system (1) is desired to achieve certain goals, constrained by temporal logic specifications, by following one of the trajectories given by the family of exosystems

$$\dot{z}_i = h_i(z_i),\ z_i(0) = z_{i,0},\ \forall t \geqslant 0,\ i \in \mathcal{I}, \tag{2}$$

where $z_i : \mathbb{R}_+ \to \mathbb{R}^n$ is the $i$-th candidate of the desired trajectories to be tracked, $h_i : \mathbb{R}^n \to \mathbb{R}^n$ is a Lipschitz continuous function with $h_i(0) = 0$, and $\mathcal{I}$ is the set of the trajectories to be tracked.

### A. Linear Temporal Logic Syntax and Semantics

We consider *syntactically co-safe linear temporal logic* (co-safe LTL) and *syntactically safe linear temporal logic* (safe LTL) formulas [18] for the specifications. Let $\mathbb{B} = \{\text{True}, \text{False}\}$ be the Boolean domain. A time set $\mathbb{T}$ is $\mathbb{R}_{>0}$. A set $AP$ is a set of *atomic predicates*, each of which is a mapping $\mathbb{R}^n \times \mathbb{T} \to \mathbb{B}$.

The syntax of a co-safe LTL formulas can be recursively defined as follows

$$\phi := \top \mid p \mid \neg p \mid \phi \wedge \phi \mid \phi \vee \phi \mid \bigcirc \phi \mid \Diamond \phi \mid \phi \mathcal{U} \phi,$$

where $\top$ stands for the Boolean constant True; $p \in AP$ is an atomic predicate; $\neg$ (negation), $\wedge$ (conjunction), and $\vee$ (disjunction) are standard Boolean connectives; $\bigcirc$ (next), $\Diamond$ (eventually), and $\mathcal{U}$ (until) are temporal operators.

The syntax of a safe LTL formulas can be recursively defined as follows,

$$\phi := \top \mid p \mid \neg p \mid \phi \wedge \phi \mid \phi \vee \phi \mid \bigcirc \phi \mid \Box \phi,$$

where $\Box$ (always) is a temporal operator.

We refer the readers to Sec. II-B of [19] for the Boolean semantics of co-safe and safe LTL formulas. For a co-safe LTL formula $\phi$, one can construct an FSA that accepts precisely the proposition sequences (i.e., *words*) that satisfy $\phi$. For a safe LTL formula $\phi$, one can construct an FSA that accepts precisely the proposition sequences (i.e., *words*) that violate $\phi$.

For example, a co-safe LTL specification $\phi_{\mathrm{c}} = \Diamond p_2 \wedge (\neg p_2 \mathcal{U} p_1)$ can express that "an unmanned aerial vehicle (UAV) should track a certain trajectory $z_1$ (see (2)) before tracking another trajectory $z_2$", where $p_1 = (||x(t) - z_1(t)|| \leqslant \epsilon)$, and $p_2 = (||x(t) - z_2(t)|| \leqslant \epsilon)$, and $\epsilon \in \mathbb{R}_{>0}$ is a threshold for tracking error.

*Problem 1:* For the system given in (1) and $\lambda > 0$, find a control policy such that the closed-loop system has a stable equilibrium point, the control input satisfies $\|u\| \leqslant \lambda$, and the trajectory of state $x$ satisfies an LTL specification $\phi = \phi_{\mathrm{c}} \wedge \phi_{\mathrm{s}}$, where $\phi_{\mathrm{c}}$ is a co-safe LTL formula and $\phi_{\mathrm{s}}$ is a safe LTL formula. □

In this paper, for simplicity we consider the safe LTL formula $\phi_{\mathrm{s}}$ to be in the form of $\phi_{\mathrm{s}} = \Box p$, where $p$ is an atomic predicate.

### B. Decomposition of LTL Specifications

Given that $\phi = \phi_{\mathrm{c}} \wedge \phi_{\mathrm{s}}$ and $\phi_{\mathrm{s}} = \Box p$, we construct an FSA that accepts precisely the proposition sequences that satisfy $\phi_{\mathrm{c}}$ and manually divide Problem 1 into a series of sub-problems based on the states of the constructed FSA. For example, consider the LTL specification $\phi = \phi_{\mathrm{c}} \wedge \phi_{\mathrm{s}}$, where $\phi_{\mathrm{c}} = \Diamond p_2 \wedge (\neg p_2 \mathcal{U} p_1)$ and $\phi_{\mathrm{s}} = \Box p_3$, where $p_1$, $p_2$ and $p_3$ are different atomic predicates. Based on $\phi_{\mathrm{c}}$ one can construct an FSA as shown in Figure 1. Let $\mathcal{O}(p)$ denote the set of time-dependent states that satisfy an atomic predicate $p$. We assume that $\mathcal{O}(p_1) \cap \mathcal{O}(p_2) = \varnothing$ if $p_1$ and $p_2$ are different atomic predicates. Then, the only path from the initial FSA state $q_0$ to the final state $q_f$ is $q_0 \to q_1 \to q_f$, where the accepting state $q_f$ indicates that $\phi_{\mathrm{c}}$ is satisfied. There are two resulting two-point boundary-value problem (TPBVP) sub-problems, when the state of the FSA transitions from $q_0$ to $q_1$, and transitions from $q_1$ to $q_f$, respectively. Specifically, when the state of the FSA transitions from $q_0$ to $q_1$, the two boundary conditions for the first TPBVP sub-problem are the initial state $x_0$ and $\mathcal{O}(p_1)$, respectively. In this first TPBVP sub-problem, the safety constraint can be encoded as $x \in \mathcal{O}(\neg p_2) \cap \mathcal{O}(p_3)$, as the state can not reach $\mathcal{O}(p_2)$ until it reaches $\mathcal{O}(p_1)$ according to $\phi_{\mathrm{c}}$, and the state must always be in $\mathcal{O}(p_3)$ according to $\phi_{\mathrm{s}}$. Similarly, when the state of the FSA transitions from $q_1$ to $q_f$, the two boundary conditions for the second TPBVP sub-problem are $\mathcal{O}(p_1)$ and $\mathcal{O}(p_2)$, respectively. In this second TPBVP sub-problem, the safety constraint can be encoded as $x \in \mathcal{O}(p_3)$ according to $\phi_{\mathrm{s}}$. After $q_f$ is reached in the FSA, the state only needs to stay in $\mathcal{O}(p_3)$ until the end of time (according to $\phi_{\mathrm{s}}$).

Note that generally for a complex FSA, it may not be straightforward to select a unique path from an initial FSA state to a final FSA state. Hence, we can view the FSA as a directed graph, (approximately) estimate the edge weights (i.e., distance), and then select a path by solving a shortest path problem. So far, we have finished decomposing the LTL specification into a sequence of sub-problems, that as a whole will eventually satisfy the original LTL specification.

### C. Sub-Problem Tracking a Certain Trajectory

In this paper, given that $\phi = \phi_{\mathrm{c}} \wedge \phi_{\mathrm{s}}$, we consider the predicates in the co-safe LTL formula $\phi_{\mathrm{c}}$ to be in the form of $p = (||x(t) - z_i(t)|| \leqslant \epsilon),\ i \in \mathcal{I}$. In this way, a certain trajectory $z_i,\ i \in \mathcal{I}$ has to be tracked in each sub-problem

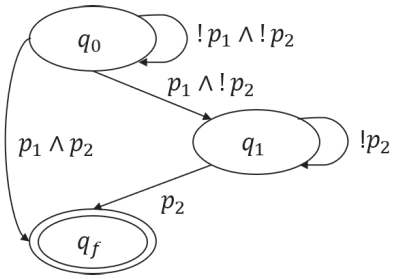


Fig. 1. Finite state automaton generated by a co-safe LTL formula $\phi_c = \Diamond p_2 \ \wedge \ (\neg p_2 \mathcal{U} p_1)$.

decomposed from Problem 1. Moreover, as described in Sec. II-B, the system state $x$ is subject to some safety constraints in each sub-problem, as $x \in \mathcal{Q}$ needs to hold in each sub-problem, where $\mathcal{Q} = \{x \in \mathbb{R}^n | c \leqslant Ax + r \leqslant C\}$, $A = [a_1 \ a_2 \ \dots \ a_m]^{\mathrm{T}} \in \mathbb{R}^{m \times n}$, with $a_i \in \mathbb{R}^n, \ \forall i \in \{1, \dots, m\}$, $r = [r_1, \dots, r_m]^{\mathrm{T}} \in \mathbb{R}^m$, $c = [c_1, \dots, c_m]^{\mathrm{T}} \in \mathbb{R}^m$ and $C = [C_1, \dots, C_m]^{\mathrm{T}} \in \mathbb{R}^m$. The TPBVP sub-problem can thus be summarized as follows.

*Problem 2:* (sub-problem) For the system (1), find a control input $u$ constrained to satisfy $\|u\| \leqslant \lambda$, so that the system state $x$ tracks a certain trajectory $z_i, i \in \mathcal{I}$ (i.e., $(||x(t) - z_i(t)|| \leqslant \epsilon)$ for a given threshold tracking error $\epsilon$) and the state $x$ remains in the set $\mathcal{Q}$, given an initial condition $x(0) = x_0$. □

To guarantee the safety specifications, we transform (1), which is constrained by $\mathcal{Q}$, into an unconstrained system. Thus, we design the following barrier function:

$$b(q, c_0, C_0) = \log\Big(\frac{C_0}{c_0}\frac{c_0 - q}{C_0 - q}\Big), \forall p \in (c_0, C_0), \tag{3}$$

where $c_0 < 0 < C_0$. The barrier $b(q, c_0, C_0)$ is invertible in the interval $(c_0, C_0)$, and its inverse is given by

$$b^{-1}(y, c_0, C_0) = c_0 C_0 \frac{e^{\frac{y}{2}} - e^{-\frac{y}{2}}}{c_0 e^{\frac{y}{2}} - C_0 e^{-\frac{y}{2}}}, \forall y \in \mathbb{R}, \tag{4}$$

with dynamics

$$\frac{\mathrm{d}b^{-1}(y, c_0, C_0)}{\mathrm{d}y} = \frac{C_0 c_0^2 - c_0 C_0^2}{c_0^2 e^y - 2c_0 C_0 + C_0^2 e^{-y}}. \tag{5}$$

We now use (3)-(5) to perform the transformation of (1). In particular, let us define

$$\begin{aligned} s_i &= b(q_i(x), c_i, C_i) \\ q_i(x) &= b^{-1}(s_i, c_i, C_i) \\ q_i(x) &= a_i^{\mathrm{T}} x + r_i, \forall i = 1, \dots, m. \end{aligned} \tag{6}$$

Through the use of the chain rule, we obtain

$$\frac{\mathrm{d}s_i}{\mathrm{d}t} = \frac{1}{\frac{\mathrm{d}b^{-1}(s_i, c_i, C_i)}{\mathrm{d}s_i}} a_i^{\mathrm{T}} \dot{x}. \tag{7}$$

Additionally, from (6), we have that

$$Ax + r = b^{-1}(s, c, C), \tag{8}$$

where $b^{-1}(s, c, C) = [b^{-1}(s_1, c_1, C_1), \dots, b^{-1}(s_m, c_m, C_m)]^{\mathrm{T}} \in \mathbb{R}^m$, hence we conclude that

$$x = (A^{\mathrm{T}} A)^{-1} A^{\mathrm{T}} (b^{-1}(s, c, C) - r). \tag{9}$$

Combining (7) with (1) yields the unconstrained subsystem dynamics, $\forall i = 1, \dots, n$,

$$\frac{\mathrm{d}s_i}{\mathrm{d}t} = \frac{1}{\frac{\mathrm{d}b^{-1}(s_i, c_i, C_i)}{\mathrm{d}s_i}} a_i^{\mathrm{T}} \big(f(x) + g(x)u\big), \ t \geqslant 0.$$

In (10), the constrained state $x$ can be written with respect to the unconstrained state $s = [s_1, \dots, s_n]^{\mathrm{T}}$ as in (9). Therefore, the subsystems (10) can be described in the compact form

$$\dot{s} = F(s) + G(s)u, \ t \geqslant 0. \tag{10}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$, $G : \mathbb{R}^n \to \mathbb{R}^n$.

To achieve optimal tracking while satisfying the required safety constraints, we shall solve Problem 2 in the transformed $s$-domain both for the system—employing the dynamics given by (10)—as well as the image of the target trajectory in the $s$-domain. Thus, let the transformed dynamics of $z \in z_{\mathcal{I}}$ be given by

$$\dot{z}_s(t) = f_d(z_s(t)), \ z_s(0) = z_0, \ t \geqslant 0, \tag{11}$$

where $z_s(t) \in \mathbb{R}^n$ denotes the bounded desired trajectory in the $s$-domain, and $f_d$ is a Lipschitz continuous function, with $f_d(0) = 0$, which yields the dynamics of $z_s$. The function $f_d$ can be derived by following the same procedure that was used to transform the state $x$ into $s$.

We may now define a tracking error $e_s(t) = s(t) - z_s(t) \in \mathbb{R}^n$ in the $s$-domain, $\forall t \geqslant 0$, with dynamics given by $\dot{e}_s(t) = F(e_s(t) + z_s(t)) - f_d(z_s(t))$. Hence, concatenating $e_s$ and $z_s$ into a single state vector $s_{aug} := [e_s^{\mathrm{T}} \ z_s^{\mathrm{T}}]^{\mathrm{T}}$, we derive the concatenated dynamics in the $s$-domain

$$\dot{s}_{\mathrm{aug}} = F_{\mathrm{aug}}(s_{\mathrm{aug}}) + G_{\mathrm{aug}}(s_{\mathrm{aug}})u(t), \ t \geqslant 0, \tag{12}$$

where $F_{\mathrm{aug}}(s_{\mathrm{aug}}) := \begin{bmatrix} F(e_s(t) + z_s(t)) - f_d(z(t)) \\ f_d(z_s(t)) \end{bmatrix}$ and $G_{\mathrm{aug}}(s_{\mathrm{aug}}) := \begin{bmatrix} G(e_s(t) + z_s(t)) \\ 0 \end{bmatrix}$.

## III. Optimal Tracking Sub-Problems

To reduce the communication burden and conserve resources, the system operates under a sampled version of the transformed state

$$\hat{s}_{\mathrm{aug}}(t) = \begin{cases} s_{\mathrm{aug}}(r_j), & \forall t \in (r_j, r_{j+1}] \\ s_{\mathrm{aug}}(t), & t = r_j. \end{cases}$$

The sampling instances constitute a strictly increasing sequence $\{r_j\}_{j=0}^{\infty}$, where $r_j, \ j \in \mathbb{N}$, is the $j$-th consecutive sampling instant, with $r_0 = 0$ and $\lim_{j \to \infty} r_j = \infty$. To decide when to trigger an event, we define the triggering error as the difference between the state $s_{\mathrm{aug}}(t)$ at the current time $t$ and the state $\hat{s}_{\mathrm{aug}}(t)$ that was sampled most recently:

$$e_{\mathrm{trig}}(t) = \hat{s}_{\mathrm{aug}}(t) - s_{\mathrm{aug}}(t), \ \forall t \in (r_j, r_{j+1}], \ j \in \mathbb{N}.$$

Our objective is to find a feedback controller that minimizes a cost functional given by

$$J(s_{\mathrm{aug}}(0); u) = \frac{1}{2} \int_0^{\infty} e^{-\gamma \tau} \big(s_{\mathrm{aug}}^{\mathrm{T}} Q_{\mathrm{aug}} s_{\mathrm{aug}} + R(u)\big) \mathrm{d}\tau,$$

where $\gamma \in \mathbb{R}^+$ is a discount factor, and $Q_{\rm aug} = \begin{bmatrix} Q & 0_{n\times n} \\ 0_{n\times n} & 0_{n\times n} \end{bmatrix}$ is a user defined matrix where $Q > 0$ and $0_{n\times n}$ a square matrix of zeros. Furthermore, to satisfy the magnitude constraint on $u$, i.e., $\|u\| \leqslant \lambda$, we chose $R(u)$ to have following form, adopted from [20],

$$R(u) = \int_0^u \lambda \tanh^{-1}\Big(\frac{v}{\lambda}\Big)\gamma_1 \mathrm{d}v, \ \forall u \in [-\lambda, \lambda], \tag{13}$$

where $\tanh^{-1}(\cdot)$ denotes the inverse of the hyperbolic tangent function and $\gamma_1 > 0$.

Initially, we consider an infinite bandwidth optimal control problem, assuming that the controller has access to the augmented transformed state at all times. We define the optimal value function $V : \mathbb{R}^{2n} \to \mathbb{R}$ given, $\forall s_{\rm aug}$, as

$$V(s_{\rm aug}(t)) = \min_u \frac{1}{2}\int_t^\infty e^{-\gamma(\tau - t)}\big(s_{\rm aug}^{\rm T} Q_{\rm aug} s_{\rm aug} + R(u)\big)\mathrm{d}\tau,$$

and the associated Hamiltonian for the continuously updating controller as

$$H(s_{\rm aug}, u_c, \frac{\partial V}{\partial s_{\rm aug}}) = \frac{\partial V}{\partial s_{\rm aug}}^{\rm T}(F_{\rm aug}(s_{\rm aug}) + G_{\rm aug}(s_{\rm aug})u_c) + \\ + \frac{1}{2}\big(s_{\rm aug}^{\rm T} Q_{\rm aug} s_{\rm aug} + R(u_c) - 2\gamma V(s_{\rm aug})\big), \ \forall s_{\rm aug}, u_c. \tag{14}$$

After employing the stationarity condition, for the Hamiltonian (14), i.e., $\frac{\partial H(\cdot)}{\partial u_c} = 0$, the infinite bandwidth optimal control can be found to be

$$u_c^\star(s_{\rm aug}) = -\lambda \tanh\big(\frac{1}{2\gamma_1\lambda} G_{\rm aug}^{\rm T}(s_{\rm aug})\frac{\partial V}{\partial s_{\rm aug}}\big), \ \forall s_{\rm aug}. \tag{15}$$

By substituting the optimal control (15) into (14) one has the Hamilton-Jacobi-Bellman (HJB) equation given as

$$H(s_{\rm aug}, u_c^\star, \frac{\partial V}{\partial s_{\rm aug}}) = 0, \ \forall s_{\rm aug}. \tag{16}$$

Now, to reduce the communication between the plant and the controller we use an intermittent version of (16) by introducing a sampled-data component that will enforce sparse and aperiodic updates for the controller. Thus, the controller operates with the sampled version of the system states, rather than the actual ones, and (15) becomes

$$u^\star(\hat{s}_{\rm aug}) = -\lambda \tanh\big(\frac{1}{2\gamma_1\lambda} G_{\rm aug}^{\rm T}(\hat{s}_{\rm aug})\frac{\partial V}{\partial \hat{s}_{\rm aug}}\big), \ \forall \hat{s}_{\rm aug}. \tag{17}$$

*Assumption 1:* There exists a positive constant $L$ such that

$$\|u_c(s_{\rm aug}) - u(\hat{s}_{\rm aug})\| \leqslant L\,\|e_{\rm trig}\|, \ \forall s_{\rm aug}, \hat{s}_{\rm aug}. \quad \square$$

*Theorem 1:* Consider the constrained system evolving in the transformed $s$-space (10), following the trajectory given by (11). Let the augmented tracking error system be given by (12), and the intermittent policy by (17). Then, the closed-loop error system has an asymptotically stable equilibrium when $\gamma = 0$, and is ultimately uniformly bounded (UUB) when $\gamma \neq 0$, under the triggering condition given by

$$\|e_{\rm trig}\|^2 \leqslant \frac{(1/2 - \beta^2)\underline{\lambda}(Q)}{L^2\lambda\gamma_1}\|e_s\|^2 + \frac{1}{L^2\lambda\gamma_1}R(u),$$

where $\beta \in (0, \frac{1}{\sqrt{2}})$ is a design parameter and $\underline{\lambda}(Q)$ is the minimum eigenvalue of $Q$. Furthermore, Zeno behavior is excluded via a lower bound on the inter-event times, i.e., $\exists \bar{r} > 0$ such that, $r_{j+1} - r_j > \bar{r}, \ \forall j \in \mathcal{N}$.

*Proof.* The proof closely follows [21]. ∎

## IV. Learning Algorithm

In this section, we employ approximation structures to solve the optimal tracking problem in a data-driven way, while guaranteeing safety constraints.

Initially, we employ a critic approximator, that will be able to estimate the optimal value function that solves the HJB equation. It is known that the optimal value function can be expressed as

$$V^\star(s_{\rm aug}) = \theta_c^{\star\rm T}\phi_c(s_{\rm aug}) + \epsilon_c(s_{\rm aug}), \quad \forall s_{\rm aug}, \tag{18}$$

where $\theta_c^\star \in \mathbb{R}^h$ are unknown ideal weights which are bounded as $\|\theta_c^\star\| \leqslant \theta_{\rm cmax}$. Furthermore, $\phi_c := [\phi_1 \ \phi_2 \ \cdots \ \phi_h] : \mathbb{R}^{2n} \to \mathbb{R}^h$, is a bounded $C^1$ basis function, i.e., with bounded first order derivatives, so that $\|\phi_c\| \leqslant \phi_{\rm cmax}$ and $\left\|\frac{\partial \phi_c}{\partial x}\right\| \leqslant \phi_{\rm dcmax}$, and $h$ is the number of basis. Finally, $\epsilon_c : \mathbb{R}^{2n} \to \mathbb{R}$ is the approximation error. Based on this, the optimal intermittent policy in can be re-written, $\forall t \in (r_j, r_{j+1}]$, as

$$u^\star(\hat{s}_{\rm aug}) = -\lambda \tanh\big(\frac{1}{2\gamma_1\lambda} G_{\rm aug}^{\rm T}(\hat{s}_{\rm aug}) \times \\ \times \big(\frac{\partial \phi(\hat{s}_{\rm aug})}{\partial \hat{s}_{\rm aug}}^{\rm T}\theta_c^\star + \frac{\partial \epsilon_c(\hat{s}_{\rm aug})}{\partial \hat{s}_{\rm aug}}\big)\big). \tag{19}$$

We employ another approximating structure, called an *actor*, to approximate the intermittent controller (19). This is expressed, $\forall t \in (r_j, r_{j+1}]$, as

$$u^\star(\hat{s}_{\rm aug}) = \theta_u^{\star\rm T}\phi_u(\hat{s}_{\rm aug}) + \epsilon_u(\hat{s}_{\rm aug}), \ \forall \hat{s}_{\rm aug}, \tag{20}$$

where $\theta_u^\star \in \mathbb{R}^{h_2}$ are the optimal weights, $\phi_u$ are the basis functions defined similarly to the critic approximator, $h_2$ is the number of basis, and $\epsilon_u$ is the actor approximation error. Thus, the current estimates of the value function and the optimal policy are derived based on estimations of the ideal critic and actor weights, denoted $\hat{\theta}_c$ and $\hat{\theta}_u$, respectively, as

$$\hat{V}(s_{\rm aug}(t)) = \hat{\theta}_c^{\rm T}\phi_c(s_{\rm aug}(t)), \forall s_{\rm aug}, \tag{21}$$

$$\hat{u}(\hat{s}_{\rm aug}) = \hat{\theta}_u^{\rm T}\phi_u(\hat{s}_{\rm aug}), \forall \hat{s}_{\rm aug}. \tag{22}$$

The learning mechanism comprises tuning laws that will allow us to obtain optimal estimates to the critic and actor weights. Towards this, we define the estimation error $e_c \in \mathbb{R}$ based on the Hamiltonian function as

$$e_c = H(s_{\rm aug}, \hat{u}(\hat{s}_{\rm aug}), \frac{\partial \hat{V}}{\partial s_{\rm aug}}) - H(s_{\rm aug}, u_c^\star(s_{\rm aug}), \frac{\partial V^\star}{\partial s_{\rm aug}}) \\ = \hat{\theta}_c^{\rm T}\omega + \hat{r},$$

with $\omega = \frac{\partial \phi_c}{\partial s_{\rm aug}}\big(F_{\rm aug}(s_{\rm aug}) + G_{\rm aug}(s_{\rm aug})\hat{u}(\hat{s}_{\rm aug})\big) - \gamma\phi_c$, $\hat{r} = \frac{1}{2}s_{\rm aug}^{\rm T}Q_{\rm aug}s_{\rm aug} + R(\hat{u}(\hat{s}_{\rm aug}))$ and $H(s_{\rm aug}, u_c^\star(s_{\rm aug}), \frac{\partial V^\star(s_{\rm aug})}{\partial s_{\rm aug}}) = 0$ from (16), where we

omit the dependence of $\phi_c$ on the augmented $s$-state. To drive the error $e_c$ to zero one has to pick the critic weights appropriately. By defining the squared-norm error as $E_c = 1/2e_c^2$ we can apply the normalized gradient descent method to obtain the estimate of the critic weights as

$$\dot{\hat{\theta}}_c = -\alpha \frac{1}{(\omega^{\mathrm{T}}\omega + 1)^2} \frac{\partial E_c}{\partial \hat{\theta}_c} = -\alpha \frac{\omega}{(\omega^{\mathrm{T}}\omega + 1)^2} e_c, \tag{23}$$

where $\alpha \in \mathbb{R}^+$ is a tuning parameter.

By defining the critic error dynamics as $\tilde{\theta}_c = \theta_c^\star - \hat{\theta}_c$ and taking its derivative with respect to time one has

$$\dot{\tilde{\theta}}_c = -\alpha \frac{\omega\ \omega^{\mathrm{T}}}{(\omega^{\mathrm{T}}\omega + 1)^2} \tilde{\theta}_c + \alpha \frac{\omega}{(\omega^{\mathrm{T}}\omega + 1)^2} \epsilon_{\mathrm{Hc}}, \tag{24}$$

where $\epsilon_{\mathrm{Hc}} = -\frac{\partial \epsilon_c}{\partial x}(F + G\hat{u}), \forall x, \hat{u}$, is upper bounded by $\epsilon_{\mathrm{Hcmax}} \in \mathbb{R}^+$ as $\|\epsilon_{Hc}\| \leqslant \epsilon_{\mathrm{Hcmax}}$.

To state stability results on the derived learning system, we can consider the critic error dynamics as a sum of nominal dynamical behavior with a time dependent perturbation due to the approximation error, denoted respectively as $S_{\mathrm{N}}$ and $S_{\mathrm{P}}$, where $\dot{\tilde{\theta}}_c = S_{\mathrm{N}} + S_{\mathrm{P}}$ with $S_{\mathrm{N}} = -\alpha \frac{\omega\ \omega^{\mathrm{T}}}{(\omega^{\mathrm{T}}\omega+1)^2} \tilde{\theta}_c$ and $S_{\mathrm{P}} = \alpha \frac{\omega}{(\omega^{\mathrm{T}}\omega+1)^2} \epsilon_{\mathrm{Hc}}$.

*Theorem 2:* Assume that $M = \frac{\omega}{(\omega^{\mathrm{T}}\omega+1)}$ is persistently exciting, i.e., $\int_t^{t+T} MM^{\mathrm{T}} d\tau \geqslant bI, \forall t \geqslant 0$ for some $b,\ T \in \mathbb{R}^+$, where $I$ is an identity matrix of appropriate dimensions, and $\exists M_b \in \mathbb{R}^+$ such that for all $t \geqslant 0$, $\max\left\{ |M|, \left|\dot{M}\right| \right\} \leqslant M_B$. Then, the nominal system $S_{\mathrm{N}}$ is exponentially stable and its trajectories satisfy $\left\|\tilde{\theta}_c(t)\right\| \leqslant \left\|\tilde{\theta}_c(0)\right\| \kappa_1 e^{-\kappa_2 t}$, for some $\kappa_1,\ \kappa_2 \in \mathbb{R}^+$ and for all $t \geqslant 0$.

*Proof.* The proof follows from [22]. ∎

To derive the tuning laws for the actor approximator, we consider the error $e_u \in \mathbb{R}$ given, $\forall \hat{s}_{\mathrm{aug}}$, by

$$\begin{aligned} e_u =& \hat{u} - u_{\hat{\theta}_c} \\ =& \hat{\theta}_u^{\mathrm{T}} \phi_u(\hat{s}_{\mathrm{aug}}) + \lambda \tanh\Big(\frac{1}{2\gamma_1 \lambda} G_{\mathrm{aug}}^{\mathrm{T}}(\hat{s}_{\mathrm{aug}}) \Big(\frac{\partial \phi(\hat{s}_{\mathrm{aug}})}{\partial \hat{s}_{\mathrm{aug}}}^{\mathrm{T}} \theta_c^\star\Big), \end{aligned}$$

where $u_{\hat{\theta}_c}$ is the controller based on the critic weights $\hat{\theta}_c$.

The objective is to select $\hat{\theta}_u$ such that the error $e_u$ goes to zero. As such, we minimize the quadratic error function $E_u = \frac{1}{2} e_u^{\mathrm{T}} e_u$. In keeping with our objective to avoid over-utilization of the system's resources, we let the actor learn in an aperiodic fashion, by updating the weights only at the triggering instances, and keeping them constant between them. This gives the system an impulsive nature, whose behavior is investigated based on results of [23] and [24].

Then, the update laws are given by

$$\dot{\hat{\theta}}_u(t) = 0,\ \forall t \in \mathbb{R}^+ \backslash \bigcup_{j \in \mathbb{N}} r_j, \tag{25}$$

and the jump dynamics of $\hat{\theta}_u(r_j^+)$ are given, for $t = r_j$, by

$$\begin{aligned} \hat{\theta}_u^+ =& \hat{\theta}_u - \alpha_u \phi_u(x_{\mathrm{aug}}(t)) \Bigg( \hat{\theta}_u^{\mathrm{T}} \phi_u(s_{\mathrm{aug}}(t)) \\ &+ \lambda \tanh\Bigg( \frac{1}{2\gamma_1 \lambda} G_{\mathrm{aug}}^{\mathrm{T}}(\hat{s}_{\mathrm{aug}}) \frac{\partial \phi(\hat{s}_{\mathrm{aug}})}{\partial \hat{s}_{\mathrm{aug}}}^{\mathrm{T}} \theta_c^\star \Bigg)^{\mathrm{T}}. \end{aligned} \tag{26}$$

By defining the actor error dynamics as $\tilde{\theta}_u = \theta_u^\star - \hat{\theta}_u$ and taking the time derivative using the continuous update (25) and by using the jump system (26) updated at the trigger instants one has

$$\dot{\tilde{\theta}}_u(t) = 0,\ \forall t \in \mathbb{R}^+ \backslash \bigcup_{j \in \mathbb{N}} r_j, \tag{27}$$

and, for $t = r_j$,

$$\begin{aligned} \tilde{\theta}_u^+ =& \tilde{\theta}_u - \alpha_u \phi_u(s_{\mathrm{aug}}(t)) \phi_u(s_{\mathrm{aug}}(t))^{\mathrm{T}} \tilde{\theta}_u(t) \\ &- \lambda \tanh\Big( \frac{1}{2\gamma_1 \lambda} G_{\mathrm{aug}}^{\mathrm{T}}(\hat{s}_{\mathrm{aug}}) \Big( \frac{\partial \phi(\hat{s}_{\mathrm{aug}})}{\partial \hat{s}_{\mathrm{aug}}}^{\mathrm{T}} + \frac{\partial \epsilon_c(\hat{s}_{\mathrm{aug}})}{\partial \hat{s}_{\mathrm{aug}}}^{\mathrm{T}} \Big) \theta_c^\star \Big), \end{aligned} \tag{28}$$

respectively. Note that the solution of (27)-(28) is left continuous; that is, it is continuous everywhere except at the resetting times $r_j$

$$\begin{aligned} &\hat{\theta}_u(r_j) = \lim_{\delta \to 0^+} \hat{\theta}_u(r_j - \delta),\ \forall j \in \mathbb{N},\ \text{and} \\ &\hat{\theta}_u^+ = \hat{\theta}_u - \alpha_u \phi_u(x_{\mathrm{aug}}(t)) \Bigg( \hat{\theta}_u^{\mathrm{T}} \phi_u(s_{\mathrm{aug}}(t)) \\ &\qquad + \lambda \tanh\Big( \frac{1}{2\gamma_1 \lambda} G_{\mathrm{aug}}^{\mathrm{T}}(\hat{s}_{\mathrm{aug}}) \frac{\partial \phi(\hat{s}_{\mathrm{aug}})}{\partial \hat{s}_{\mathrm{aug}}}^{\mathrm{T}} \theta_c^\star \Bigg)^{\mathrm{T}},\ t = r_j. \end{aligned}$$

## V. Simulation Results

To validate the effectiveness of the proposed framework, we solve a safety-critical task described as a temporal logic specification, which can be decomposed as a sequence of event-triggered optimal tracking control problems. Same as the example in Sec. II-B, we consider the LTL specification $\phi = \phi_{\mathrm{c}} \wedge \phi_{\mathrm{s}}$, with $\phi_{\mathrm{c}} = \Diamond p_2 \ \wedge\ (\neg p_2 \mathcal{U} p_1)$ and $\phi_{\mathrm{s}} = \Box p_3$, where $p_1$, $p_2$ are to track different target trajectories and $p_3$ is a safe zone to stay in. Given the FSA constructed based on $\phi_{\mathrm{c}}$ as shown in Fig. 1, Problem 1 can be decomposed into two sub-problems (Problem 2). Due to space limitations, we only show the results of the first sub-problem (i.e., reaching $p_1$), and the results for the second sub-problems are omitted as they can be obtained in a similar manner. The safety constraint $p_3$ is defined as $\mathcal{Q} = \{x \in \mathbb{R}^n | c \leqslant Ax + r \leqslant C\}$, where $A = I$ (the identity matrix), $r = \mathbf{0}_{4\times 1}$ and $c = -30 \times \mathbf{1}_{4\times 1}, C = -c$.

We set the predicate $p_1 = (||x(t) - z_1(t)|| \leqslant \epsilon)$, where the trajectory to be tracked is $z = 0.5 \times [\sin 0.5t, \cos 0.5t]^{\mathrm{T}}$, and $\epsilon = 0.6$. The system dynamics are given as $\dot{x} = f(x) + g(x)u$, where we used the drift and input dynamics investigated in [25].

For the learning algorithm, the initial actor and critic weights are picked randomly in $[0, 1]$. The user-defined parameters are selected as $Q = 800I$. The evolution of the tracking error is shown in Fig. 2. In Fig. 3 we present the control input of the system after the exploration noise has been sufficiently decreased. It is validated that the safety constrains are satisfied while the tracking error is decreasing and eventually bounded as proved.

## VI. Conclusion and Future Work

In this paper, we developed an intermittent learning framework for a system tasked with a complex mission while

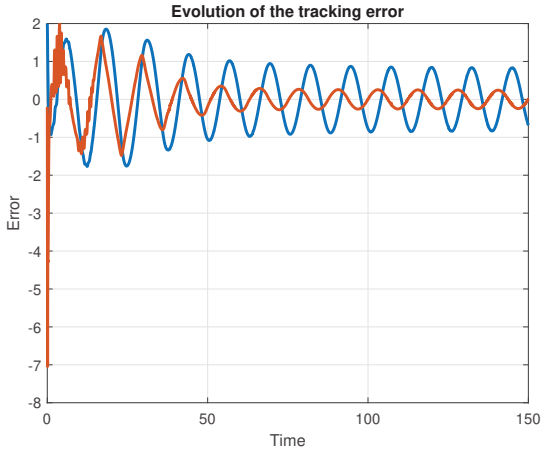


Fig. 2. The evolution of the tracking error of the states.

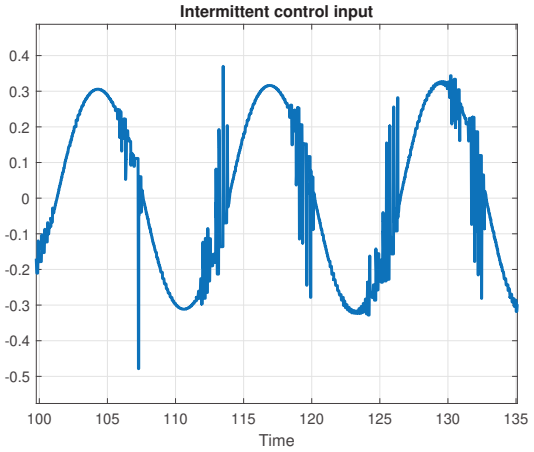


Fig. 3. Intermittent control policy.

guaranteeing safety. We brought together ideas from LTL and control-oriented RL to decompose the mission into a sequence of tracking sub-problems which are constrained by safety specifications. We convert the system using barrier functions, thus, deriving an unconstrained optimal tracking problem in the transformed state space. The tracking problem was tackled via the construction of intermittent policies and guarantees of stability and optimality were presented. Finally, to circumvent the issues arising from the difficulty of solving the underlying HJB equations, we used an RL algorithm to obtain estimated versions of the intermittent safe optimal control policies.

## REFERENCES


[1] R. Kamalapurkar, P. Walters, J. Rosenfeld, and W. Dixon, *Reinforcement learning for optimal feedback control*. Springer, 2018.

[2] L. Lin and M. A. Goodrich, "Uav intelligent path planning for wilderness search and rescue," in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2009, pp. 709–714.

[3] E. A. Emerson, "Temporal and modal logic," in *Formal Models and Semantics*. Elsevier, 1990, pp. 995–1072.

[4] W. Heemels, K. H. Johansson, and P. Tabuada, "An introduction to event-triggered and self-triggered control," in *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*. IEEE, 2012, pp. 3270–3285.

[5] A. D. Ames, X. Xu, J. W. Grizzle, and P. Tabuada, "Control barrier function based quadratic programs for safety critical systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 8, pp. 3861–3876, 2016.

[6] M. Kimmel and S. Hirche, "Active safety control for dynamic human-robot interaction," in *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2015, pp. 4685–4691.

[7] J. F. Fisac, A. K. Akametalu, M. N. Zeilinger, S. Kaynama, J. Gillula, and C. J. Tomlin, "A general safety framework for learning-based control in uncertain robotic systems," *IEEE Transactions on Automatic Control*, vol. 64, no. 7, pp. 2737–2752, 2018.

[8] M. L. Greene, P. Deptula, S. Nivison, and W. E. Dixon, "Sparse learning-based approximate dynamic programming with barrier constraints," *IEEE Control Systems Letters*, vol. 4, no. 3, pp. 743–748, 2020.

[9] C. Sun and K. G. Vamvoudakis, "Continuous-time safe learning with temporal logic constraints in adversarial environments," in *2020 American Control Conference (ACC)*. IEEE, 2020, pp. 4786–4791.

[10] D. Muniraj, K. G. Vamvoudakis, and M. Farhood, "Enforcing signal temporal logic specifications in multi-agent adversarial environments: A deep q-learning approach," in *2018 IEEE Conference on Decision and Control (CDC)*. IEEE, 2018, pp. 4141–4146.

[11] I. Saha, R. Ramaithitima, V. Kumar, G. J. Pappas, and S. A. Seshia, "Automated composition of motion primitives for multi-robot systems from safe ltl specifications," in *2014 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2014, pp. 1525–1532.

[12] W. H. Heemels, M. Donkers, and A. R. Teel, "Periodic event-triggered control for linear systems," *IEEE Transactions on Automatic Control*, vol. 58, no. 4, pp. 847–861, 2012.

[13] M. Donkers and W. Heemels, "Output-based event-triggered control with guaranteed $\mathcal{L}_2$-gain and improved event-triggering," in *49th IEEE Conference on Decision and Control (CDC)*. IEEE, 2010, pp. 3246–3251.

[14] K. G. Vamvoudakis, "Event-triggered optimal adaptive control algorithm for continuous-time nonlinear systems," *IEEE/CAA Journal of Automatica Sinica*, vol. 1, no. 3, pp. 282–293, 2014.

[15] K. G. Vamvoudakis and H. Ferraz, "Event-triggered h-infinity control for unknown continuous-time linear systems using q-learning," in *2016 IEEE 55th Conference on Decision and Control (CDC)*. IEEE, 2016, pp. 1376–1381.

[16] Z. Xu, F. M. Zegers, B. Wu, W. Dixon, and U. Topcu, "Controller synthesis for multi-agent systems with intermittent communication. a metric temporal logic approach," in *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2019, pp. 1015–1022.

[17] G. P. Kontoudis, Z. Xu, and K. G. Vamvoudakis, "Online, model-free motion planning in dynamic environments: An intermittent, finite horizon approach with continuous-time q-learning," in *2020 American Control Conference (ACC)*. IEEE, 2020, pp. 3873–3878.

[18] Z. Xu, M. Ornik, A. A. Julius, and U. Topcu, "Information-guided temporal logic inference with prior knowledge," in *2019 American Control Conference (ACC)*, 2019, pp. 1891–1897.

[19] M. Lahijanian, M. R. Maly, D. Fried, L. E. Kavraki, H. Kress-Gazit, and M. Y. Vardi, "Iterative temporal planning in uncertain environments with partial satisfaction guarantees," *IEEE Transactions on Robotics*, vol. 32, no. 3, pp. 583–599, 2016.

[20] M. Abu-Khalaf and F. L. Lewis, "Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network hjb approach," *Automatica*, vol. 41, no. 5, pp. 779–791, 2005.

[21] Y. Yang, K. G. Vamvoudakis, H. Modares, Y. Yin, and D. C. Wunsch, "Safe intermittent reinforcement learning with static and dynamic event generators," *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[22] K. G. Vamvoudakis, A. Mojoodi, and H. Ferraz, "Event-triggered optimal tracking control of nonlinear systems," *International Journal of Robust and Nonlinear Control*, vol. 27, no. 4, pp. 598–619, 2017.

[23] W. M. Haddad, V. Chellaboina, and S. G. Nersesov, *Impulsive and hybrid dynamical systems: stability, dissipativity, and control*. Princeton University Press, 2006, vol. 49.

[24] J. P. Hespanha, D. Liberzon, and A. R. Teel, "Lyapunov conditions for input-to-state stability of impulsive systems," *Automatica*, vol. 44, no. 11, pp. 2735–2744, 2008.

[25] K. G. Vamvoudakis and F. L. Lewis, "Online actor–critic algorithm to solve the continuous-time infinite horizon optimal control problem," *Automatica*, vol. 46, no. 5, pp. 878–888, 2010.