State Estimation of Continuous-time Dynamical Systems with Uncertain Inputs with Bounded Variation: Entropy, Bit Rates, and Relation with Switched Systems

Hussein Sibai and Sayan Mitra, senior member

Abstract—We extend the notion of estimation entropy of autonomous dynamical systems proposed by Liberzon and Mitra [1] to nonlinear dynamical systems with uncertain inputs with bounded variation. We call this new notion the ε-estimation entropy of the system and show that it lower bounds the bit rate needed for state estimation. ε-estimation entropy represents the exponential rate of the increase of the minimal number of functions that are adequate for ε-approximating any trajectory of the system. We show that alternative entropy definitions using spanning or separating trajectories bound ours from both sides. On the other hand, we show that other commonly used definitions of entropy, for example the ones in [1], diverge to infinity. Thus, they are potentially not suitable for systems with uncertain inputs. We derive an upper bound on ε-estimation entropy and estimation bit rates, and evaluate it for two examples. We present a state estimation algorithm that constructs a function that approximates a given trajectory up to an ε error, given time-sampled and quantized measurements of state and input. We investigate the relation between ε-estimation entropy and a previous notion for switched systems and derive a new upper bound for the latter, showing the generality of our results on systems with uncertain inputs.

Index Terms—Entropy, State Estimation, Bit Rates, Non-linear Systems, Switched Systems.

I. INTRODUCTION

State estimation is a fundamental problem for controlling and monitoring dynamical systems. In most application scenarios, the estimator has to work with plant state information sent by a sensor over a channel with finite bit rate. If a certain accuracy is required from the estimator, then a natural question is to ask: what is the minimal bit rate of the channel for the estimator to support this accuracy requirement? This question has been investigated for both stochastic and non-stochastic system models and channels. In the non-stochastic setting, the point of view of topological entropy has proven to be a fruitful line of investigation. In particular, it has been used for deriving the minimal necessary bit-rates for closed, i.e., autonomous, non-switched, and switched systems (see, for example [1]–[3]). This paper contributes in this line of investigation and proposes answers to the above question for continuous-time dynamical systems with uncertain inputs with bounded variation. This class includes a wide variety of systems, for example: those with open-loop control, those with external disturbances, and autonomous systems as a special case.

State estimation of systems with uncertain inputs is a more challenging problem than that of autonomous nonlinear switched and non-switched ones of [2], [3]. That is because even if the uncertainty about the state can be made to decrease over time using sensor measurements, the uncertainty about the input may not decrease. The input can change slowly and the continuous effect of the uncertain input prevents the uncertainty about the state from going to zero. In contrast, the uncertainty can be made to decrease exponentially over time for autonomous non-switched systems [1], as well as for autonomous switched ones between their mode switches [3]. We contend this challenge of changing input using a weaker notion of estimation, akin to that in [4], that only requires the error to be bounded by a constant ε > 0.

Given two estimation error parameters ε > 0 and α ≥ 0, we first define a notion of topological entropy h_{est}(ε, α) for systems with uncertain inputs that represents the rate of increase over time of the minimal number of functions needed to approximate all system trajectories up to an exponentially decaying error of εe^{-αt}, for any t ≥ 0. We suggest an alternative entropy notion h_{est}^*(ε, α), where we restrict the approximating functions to be trajectories of the system. We further consider a third notion h_{est}^+(ε, α) that represents the rate of increase over time of the number of trajectories that pairwise violate the approximation error bound εe^{-αt}, at some t ≥ 0. We show that all three notions are close, and thus would be viable alternative options.

Then, we show that requiring exponential decay of estimation error, i.e., an α that is strictly positive, leads to an infinite entropy h_{est}(ε, α), even for a simple linear one-dimensional system with piece-wise constant input (Section V-B). Moreover, we show that h_{est}(ε, α) of the same system grows to infinity as ε goes to zero (Section V-A). On the other hand, we show that h_{est}(ε, 0) for that simple system is finite for any strictly positive ε (Corollary 5). These observations imply that the definition of entropy for autonomous systems (e.g. [4]), that
takes the limit of \( \varepsilon \) as it goes to zero, or that of [1], [2] that requires exponential convergence of error, are not suitable for systems with inputs. We thus use \( h_{est}(\varepsilon) \) as a weaker notion of estimation entropy, which we name \( \varepsilon \)-estimation entropy, from there forward, as well as \( h_{l1}(\varepsilon) \) and \( h_{l2}(\varepsilon) \) for the alternative notions, implicitly fixing \( \alpha \) to zero and \( \varepsilon \) to a strictly-positive real number.

We show that there is no state estimation algorithm with a fixed bit rate smaller than \( h_{est}(\varepsilon) \) (Section VI). While computing \( h_{est}(\varepsilon) \) exactly is generally hard, we compute an upper bound. To do that, we use local discrepancy functions to upper-bound the sensitivity of a trajectory of a nonlinear system with uncertain inputs to changes in its initial state and in its input signal. Then, we present a procedure (Algorithm 1) that, given sampled and quantized states and inputs, constructs a function that estimates the trajectory up to an \( \varepsilon \) error. This procedure is of independent interest, as it can also be used as a state estimation algorithm, provided that the uncertain input signal can be sampled. We count the number of trajectories that can be constructed by this procedure for different initial states and input signals, up to a time bound \( T \). The rate of exponential increase of this number as \( T \) increases gives an upper bound on entropy, and the bit rate needed for state estimation.

The upper bound is presented in terms of the state and the input dimensions \( n \) and \( m \), global bounds on the norms of the Jacobian matrices of the vector field with respect to the state and the input, \( M_{x} \) and \( M_{u} \), and two constants \( \mu \) and \( \eta \) that bound the variation of the input signal over time (Proposition 3). We consider two approaches in bounding the variation of the input signal (Definition 1). The first bounds the standard total variation of the signal over any time interval \([t, t+\tau]\) by \( \mu \tau + \eta \). The other is a less-restrictive one that bounds the pointwise variation of the signal at any two time instants \( t \) and \( t+\tau \) by \( \mu \tau + \eta \). We show that all of our results in the paper hold for both cases. We further show that if the variation of the input, with either definition, goes to zero, the upper bound on entropy approaches \( \frac{M_{u}L_{x}}{\ln 2} \), where \( L_{x} \) is the Lipschitz constant of the vector field with respect to the state, exactly matching that computed in [2] for autonomous systems, with the exponential decay of estimation error parameter \( \alpha \) being equal to zero (Corollary 4).

Finally, we rewrite the upper bound on \( h_{est}(\varepsilon) \) in a similar format to that of autonomous switched systems which we presented in [3]. Such rewriting shows the similarities and differences between the construction of the bounds in the two cases (Section VIII-B). In addition, in Section VIII-C, we over-approximate the solutions of autonomous switched systems with differential inclusions. Then, we model these inclusions as systems with inputs and obtain a new upper bound on entropy of autonomous switched systems, as an alternative to that we presented in [3].

We presented some of the results in this paper in [3], [5], particularly, earlier versions of Proposition 1, Proposition 3, Lemma 8, Lemma 10, and Theorem 4. The main novel contributions of this paper are: (1) Comparing the alternative entropy notions using approximating, spanning, and separated sets; (2) showing that the entropy of the simple system \( \dot{x} = u \) is infinite as \( \varepsilon \to 0 \) and for \( \alpha > 0 \) and a similar result for the simple switched system \( \dot{x} = \sigma x \), where \( \sigma \) switches between two positive reals, showing the necessity for bounded divergence between modes for the finiteness of entropy; (3) correcting the discrepancy function presented in [5] which was assuming that the largest eigenvalue of the symmetric part of the Jacobian of the dynamics with respect to the state is non-negative as well as assuming 2-norm while being applied as an \( \infty \)-norm bound; (4) applying the bound on entropy of systems with uncertain inputs to obtain a bound on entropy of differential inclusions and autonomous switched systems, and thus relating the bounds in [3] and [5]. Finally, we improve the presentation of both papers [3] and [5].

### A. Related Work

Several definitions of topological entropy for control systems have been proposed to bound the data rates necessary for control over limited-bandwidth communication channels [6]–[10] for tasks including stabilization [6], [11]–[13], invariance [14]–[16], and state estimation [1]–[3], [5], and types of systems including discrete-time [14], continuous-time [8], switched [3], [12], [17]–[21], and interconnected [15], [22]. In [1], [2], Liberson and Mita introduce the notion of estimation entropy for autonomous nonlinear systems. They define it in terms of the number of trajectories needed to approximate all other trajectories starting from a compact initial set up to an exponentially decaying error. They establish an upper bound of \( n(Lx + \alpha) \) on estimation entropy, where \( n \) is \( \ln 2 \) the dimension of the system and \( L_{x} \) is the Lipschitz constant of the vector field. In [3], we extend the notion of estimation entropy to switched nonlinear systems, derive a corresponding upper bound, and construct a state estimation algorithm. In this paper, in Section VIII, we relate that entropy definition for autonomous switched systems and its bounds with those of systems with inputs. We first presented the latter upper bound in [5], and restate it here for completeness. We further show in this paper that an assumption we made to derive the upper bound on entropy of autonomous switched systems in [3] is necessary by presenting a simple linear system that violates the assumption, and has an infinite entropy (Section VIII-A).

### II. Preliminaries

We denote the infinity norm of a real vector \( v \in \mathbb{R}^{n} \) by \( \|v\| \), its 2-norm as \( \|v\|_{2} \), and its transpose by \( v^{\top} \). We denote the infinity norm of a real matrix \( A \in \mathbb{R}^{n \times n} \) by \( \|A\| \), its 2-norm by \( \|A\|_{2} \), and its largest eigenvalue by \( \lambda_{\max}(A) \). If \( A \) is symmetric positive definite, then \( \lambda_{\max}(A) = \|A\|_{2} \).

\( B(v, \delta) \) is a \( \delta \)-ball–closed hypercube of radius \( \delta \)–centered at \( v \). For a hyperrectangle \( S \subseteq \mathbb{R}^{n} \) and \( \delta > 0 \), \( \text{grid}(S, \delta) \) is a collection of \( 2\delta \)-separated points along axis parallel planes such that the \( \delta \)-balls around these points cover \( S \). In that case, we say that the grid is of size \( \delta \). Given \( x \in \mathbb{R}^{n} \) and \( C = \text{grid}(S, \delta) \), we define \text{quantize}(x, C) \) to be the nearest point to \( x \) in \( C \).

We denote by \( [n_{1} : n_{2}] \), the set of integers in the interval \( [n_{1} : n_{2}] \), inclusive, and by \( [n_{2}] \) the set \( [1 : n_{2}] \). We denote the cardinality of a finite set \( S \) by \( |S| \). We denote the diameter of a compact set \( S \subseteq \mathbb{R}^{n} \) by \( \text{diam}(S) = \max_{x_{1}, x_{2} \in S} \|x_{1} - x_{2}\| \) for any set \( S \) and positive integer \( m \), \( S^{m} \) is the \( m \)-way
Cartesian product $S \times S \cdots \times S$. Given two sets $S_1$ and $S_2$, we define $S_1 \oplus S_2 := \{x_1+x_2 | x_1 \in S_1, x_2 \in S_2\}$. A continuous function $\gamma : \mathbb{R}_0^2 \rightarrow \mathbb{R}_0^1$ belongs to class-$K$ if it is strictly increasing and $\gamma(0) = 0$.

For all $t \in \mathbb{R}_0^2$, we define the right and left hand limits of a function $u : \mathbb{R}_0^2 \rightarrow \mathbb{R}^n$ at $t$ as follows:

$$u(t^+) = \lim_{\tau \rightarrow t^+} u(\tau) \text{ and } u(t^-) = \lim_{\tau \rightarrow t^-} u(\tau).$$

If $t$ is a point of discontinuity and $u$ is piecewise-right-continuous, we call it a switch of $u$, and define $u(t) = u(t^+)$. In this paper, we will consider dynamical systems with uncertain input signals whose variations are bounded by an affine function of time. We consider two types of variation: pointwise and total. Such signals are defined as follows.

**Definition 1** (Signals with affine-bounded variation). Given $\mu \geq 0$, $\eta \geq 0$, and a compact set $U \subseteq \mathbb{R}^n$, we define $\mathcal{U}^p(\mu, \eta)$, with the $p$-superscript referring to the word “pointwise”, to be the set of all piecewise-right-continuous functions $u : \mathbb{R}_0^2 \rightarrow \mathbb{R}^n$ with affine-bounded pointwise variation, i.e.,

$$u(0) \in U \text{ and } \|u(t + \tau) - u(t)\| \leq \mu \tau + \eta, \quad (1)$$

for all $t$ and $\tau \geq 0$. We define $\mathcal{U}^s(\mu, \eta)$, with the $s$-superscript referring to the word “slow”, to be the set of all piecewise-right-continuous functions $u : \mathbb{R}_0^2 \rightarrow \mathbb{R}^n$ with affine-bounded total variation, i.e.,

$$u(0) \in U \text{ and } \int_t^{t+T} \|du\| \leq \mu \tau + \eta, \quad (2)$$

for all $t$ and $\tau \geq 0$, where $\int_t^{t+T} \|du\|$ is the total variation of $u$ defined as follows:

$$\int_t^{t+T} \|du\| := \int_t^{t+T} \hat{u}(s)ds + \sum_{k=1}^{T} \left[\hat{u}(d_k^-) - \hat{u}(d_k^+)^T\right]ds,$$

where $d_1, d_2, \ldots, d_k$ are the points of discontinuity in $u$ between $t$ and $t + \tau$.

Note that for any $\mu$ and $\eta \geq 0$, $\mathcal{U}^s(\mu, \eta) \subseteq \mathcal{U}^p(\mu, \eta)$. This can be shown as follows: fix any $\mu$, $\eta$, $t$, and $\tau \geq 0$, and consider any $u \in \mathcal{U}^s(\mu, \eta)$, then

$$\|u(t + \tau) - u(t)\| = \int_t^{t+\tau} \|du\| \leq \int_t^{t+\tau} \|du\| \leq \mu \tau + \eta. \quad (3)$$

Thus, $u \in \mathcal{U}^p(\mu, \eta)$.

We abbreviate $\mathcal{U}^p(\mu, \eta)$ and $\mathcal{U}^s(\mu, \eta)$ with $\mathcal{U}^p$ and $\mathcal{U}^s$ when $\mu$ and $\eta$ are clear from context. The set $\mathcal{U}^p(\mu, \eta)$ contains functions that vary arbitrarily fast, while having a finite number of switches in any finite time interval, without an $\eta$ bound. $\mathcal{U}^p(\mu, 0)$ is the class of globally Lipschitz continuous functions with Lipschitz constant $\mu$. In general, knowing that $u$ cannot deviate too much, can be expressed by setting $\mu$ and $\eta$ to smaller values. Roughly, $\eta$ restricts the maximum norm of a jump at any time instant and $\mu$ restricts the distance between the values of a signal at different time instances. However, it is more common in the control and dynamical systems literature to bound the total variation of functions instead of their deviations. Accordingly, in addition to $\mathcal{U}^p$, we introduced $\mathcal{U}^s$ which bounds the total variation. Signals in $\mathcal{U}^s$ are upper-bounded in how many switches or jumps they have within any time interval as well as the norm of their gradients. Thus, we call them slowly-varying signals. We present results for both cases in this paper. When deriving lower bounds on entropy, we assume that the input signals are in $\mathcal{U}^s$, and when deriving upper bounds, we assume they are in $\mathcal{U}^p$. Consequently, since $\mathcal{U}^s \subseteq \mathcal{U}^p$, all the results in the paper hold for both cases.

The slow variation constraint in $\mathcal{U}^s$ is the same to that of Theorem 2 of [23] which was made on the variation of the system matrix of a time-varying linear dynamical system to relate its stability conditions to those of a switched linear dynamical system with slow switching. Also, it is similar to the slow switching assumption made by Hesphana and Morse in [24] to prove the stability of switched systems with stable subsystems. In this paper, we use the bound on the variation of signals in $\mathcal{U}^p$ to derive an upper bound on the entropy of dynamical systems with inputs in $\mathcal{U}^p$ and $\mathcal{U}^s$, since $\mathcal{U}^s \subseteq \mathcal{U}^p$. Further, we use it to relate the upper bound on entropy of autonomous switched systems with slow switching (having minimum dwell-time constraints) with that of dynamical systems with inputs with affine-bounded pointwise variation in Sections VIII-B and VIII-C.

### III. Entropy for Open Dynamical Systems

We consider a dynamical system of the form:

$$\dot{x}(t) = f(x(t), u(t)), \quad (4)$$

where $t \geq 0$ and $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$. The function $f$ is locally Lipschitz, and has piecewise-continuous Jacobian matrices $J_x(x, u) = \frac{\partial f}{\partial x}(x, u)$ and $J_u(x, u) = \frac{\partial f}{\partial u}(x, u)$, with respect to the first and second arguments, respectively. When $f$ is globally Lipschitz, we denote its global Lipschitz constants by $L_x$ and $L_u$.

For any initial state in and measurable input signal we assume that the solution of system (4) exists for any $t \geq 0$, is unique, and depends continuously on the initial state [25]. We denote this solution, or trajectory, by $x_{\xi_0, u}(t) = x(t)$. In the rest of this section, we fix a compact set of initial states $K \subseteq \mathbb{R}^n$ and the set of input functions $\mathcal{U}^p(\mu, \eta)$. The same results hold if the set of input signals was $\mathcal{U}^p(\mu, \eta)$ instead.

The reachable set of states of system (4) starting from $K$, having input signals from $\mathcal{U}$, and running for a time horizon $T$, is defined as follows:

$$\text{Reach}(K, \mathcal{U}, T) := \{x \in \mathbb{R}^n | \exists x_0 \in K, u \in \mathcal{U}, t \in [0, T], x_{\xi_0, u}(t) = x\}. \quad (5)$$

**Example 1** (Dubin’s vehicle). Consider a car moving at a constant speed $v$. We describe its dynamics as follows:

$$\dot{x}_1 = v \cos x_3, \dot{x}_2 = v \sin x_3, \dot{x}_3 = u. \quad (6)$$

where $(x_1, x_2)$ is the position of the car and $x_3$ is its heading angle. $u$ is a control signal defining the steering velocity. The Jacobians of the car dynamics with respect to $x$ and $u$ are as follows:

$$J_x = \begin{bmatrix} 0 & 0 & -v \sin x_3 \\ 0 & 0 & v \cos x_3 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad J_u = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \quad (7)$$
Therefore, the dynamics are globally Lipschitz with Lipschitz constants:
\[ L_x = \| J_x \| \leq v \quad \text{and} \quad L_u = \| J_u \| = 1. \]  

A. Approximating functions and \((\varepsilon, \alpha)\)-estimation entropy

Let us fix the error parameters \(\varepsilon > 0\) and \(\alpha \geq 0\). Given a \( T > 0, x_0 \in K \) and \( u \in \mathcal{U}^p(\mu, \eta)\), we say that a function \( z : [0, T] \to \mathbb{R}^n \) is \((T, \varepsilon, \alpha)\)-approximating for the trajectory \( \xi_{x_0, u} \) over the interval \([0, T]\), if
\[ \| z(t) - \xi_{x_0, u}(t) \| \leq \varepsilon e^{-\alpha t}, \]  
for all \( t \in [0, T]\). We say that a set of functions \( Z := \{ z : [0, T] \to \mathbb{R}^n \} \) is \((T, \varepsilon, \alpha)\)-approximating for system (4), if for every \( x_0 \in K \) and \( u \in \mathcal{U}^p\), there exists a \((T, \varepsilon, \alpha)\)-approximating function \( z \in Z \) for the trajectory \( \xi_{x_0, u} \) over 
\([0, T]\). In this paper, we mostly follow the same notation used by Liberzon and Mitra in [1] in their definition of estimation entropy for autonomous dynamical systems. We denote the minimal cardinality of such an approximating set by \( s^*_{\text{est}}(T, \varepsilon, \alpha) \). The subscript \( \text{est} \) corresponds to the word estimation and the letter \( s \) stands for the word set.

**Definition 2.** The \((\varepsilon, \alpha)\)-estimation entropy of system (4) is defined as follows:
\[ h_{\text{est}}(\varepsilon, \alpha) := \limsup_{T \to \infty} \frac{1}{T} \log s_{\text{est}}(T, \varepsilon, \alpha). \]  

The estimation entropy \( h_{\text{est}} \) represents the exponential growth rate over time of the number of distinguishable trajectories of the system. Hence, \( h_{\text{est}} \) also represents the bit rate need to be sent by the sensor so that the estimator can construct a “good” estimate of the state.

**Remark 1.** The value of \((\varepsilon, \alpha)\)-estimation entropy has the same value if the \(\infty\)-norm in inequality (9) is replaced by any norm. This follows from the equivalence of norms in finite-dimensional spaces. More specifically, since for any two norms \( \| \cdot \|_a \) and \( \| \cdot \|_b \), for all \( x \in \mathbb{R}^n \), there exists two positive constants \( c_1 \) and \( c_2 \) such that \( c_1 \| x \|_a \leq \| x \|_b \leq c_2 \| x \|_a \). \( s_{\text{est}} \) will only be multiplied by a constant factor under change of norms. That factor will be eliminated when divided by \( T \) as \( T \) goes to infinity in (10).

The following remark relates the above notion of approximating sets to the sample complexity of reachable sets. The latter is not a central theme in the current paper but is an important concept in control and verification.

**Remark 2.** Fix a time horizon \( T > 0 \). Assume that there is a method to generate a \((T, \varepsilon, \alpha)\)-approximating function for any given trajectory of system (4). Then, \( s_{\text{est}}(T, \varepsilon, \alpha) \) is also the minimum number of simulations needed to approximate the reachset \( \text{Reach}(K, \mathcal{U}, T) \) of system (4) up to an \( \varepsilon e^{-\alpha t} \) over-approximation error for any \( t \in [0, T] \). Formally, let \( Z \) be the \((T, \varepsilon, \alpha)\)-approximating set with cardinality \( s_{\text{est}}(T, \varepsilon, \alpha) \) and define an over-approximation of the reachable set \( \text{Reach}_{\varepsilon, \alpha}(K, \mathcal{U}, T) := \{ x + c \mid \exists x_0 \in K, u \in \mathcal{U}, t \in [0, T], c \in B_{\varepsilon e^{-\alpha t}}, \xi_{x_0, u}(t) = x \} \) and another over-approximation using \( Z \)
\[ R_{\varepsilon, \alpha} := \{ z(t) + c \mid z \in Z, t \in [0, T], c \in B_{\varepsilon e^{-\alpha t}} \}, \] where \( B_{\varepsilon e^{-\alpha t}} := B(0, \varepsilon e^{-\alpha t}) \). Then,
\[ \text{Reach}(K, \mathcal{U}, T) \subseteq R_{\varepsilon, \alpha} \subseteq \text{Reach}_{\varepsilon, \alpha}(K, \mathcal{U}, T). \]

IV. ALTERNATIVE ENTROPY NOTIONS

In this section, we show that two alternative definitions of entropy are comparable in the sense that a bound on one of them results in bounds on the others.

A. Approximating trajectories instead of functions

In this section, we modify the definition of entropy \( h_{\text{est}} \) by restricting the approximating functions to be trajectories of system (4). Following previous notation (e.g. [1], [2]), we call the resulting restricted approximating sets spanning sets. Formally, we say that a set \( Z^* \subseteq K \times \mathcal{U} \) is \((T, \varepsilon, \alpha)\)-spanning, if for every \( x_1 \in K \) and \( u_1 \in \mathcal{U} \), there exists a pair \((x_2, u_2) \in Z^*\) such that the trajectory \( \xi_{x_2, u_2} \) is a \((T, \varepsilon, \alpha)\)-approximating function for the trajectory \( \xi_{x_1, u_1} \) over \([0, T]\). The minimum cardinality of a spanning set is denoted by \( s^*_\text{span}(T, \varepsilon, \alpha) \). If we propagate this restriction to Definition 2, the resulting definition of entropy would be as follows:

**Definition 3.** The spanning sets-based \((\varepsilon, \alpha)\)-estimation entropy of system (4) is defined as follows:
\[ h_{\text{span}}^*(\varepsilon, \alpha) := \limsup_{T \to \infty} \frac{1}{T} \log s_{\text{span}}^*(T, \varepsilon, \alpha). \]

The following inequality is an application of the fact that any \((T, \varepsilon, \alpha)\)-spanning set is a \((T, \varepsilon, \alpha)\)-approximating set:
\[ s_{\text{est}}(T, \varepsilon, \alpha) \leq s_{\text{span}}^*(T, \varepsilon, \alpha). \]

B. Separated sets instead of approximating ones

In this section, we provide an entropy definition based on the concepts of \((T, \varepsilon, \alpha)\)-separated trajectories and \((T, \varepsilon, \alpha)\)-separated sets.

Fix any error parameters \(\varepsilon > 0\) and \(\alpha \geq 0\). Given \( T \geq 0 \), two initial states \( x_1 \) and \( x_2 \) and two input signals \( u_1 \) and \( u_2 \), the two trajectories \( \xi_{x_1, u_1} \) and \( \xi_{x_2, u_2} \) are \((T, \varepsilon, \alpha)\)-separated iff there exists \( t \in [0, T] \) such that
\[ \| \xi_{x_1, u_1}(t) - \xi_{x_2, u_2}(t) \| > \varepsilon e^{-\alpha t}. \]

A set of pairs of initial states and input signals \( \mathcal{Z}^* \subseteq K \times \mathcal{U}^p \) is called a \((T, \varepsilon, \alpha)\)-separated set if the trajectories corresponding to any two pairs in \( \mathcal{Z}^* \) are \((T, \varepsilon, \alpha)\)-separated. We denote the largest cardinality of a \((T, \varepsilon, \alpha)\)-separated set by \( s_{\text{span}}^*(T, \varepsilon, \alpha) \).

**Definition 4.** The separated sets-based \((\varepsilon, \alpha)\)-estimation entropy of system (4) is defined as follows:
\[ h_{\text{span}}^*(\varepsilon, \alpha) := \limsup_{T \to \infty} \frac{1}{T} \log s_{\text{span}}^*(T, \varepsilon, \alpha). \]

This section is a novel contribution of this paper.
The following two lemmas are analogous to Lemmas 1 and 2 in [2]. They draw the relation between approximating, spanning, and separated sets.

**Lemma 1.** For all $K$, $U^p$, $\varepsilon$, $\alpha$, and $T$,

$$s_{est}^\ast(T, \varepsilon, \alpha) \leq s_{est}^1(T, \varepsilon, \alpha). \quad (15)$$

**Proof.** The lemma follows from the observation that every maximal $(T, \varepsilon, \alpha)$-separated set $Z^\dagger$ is also $(T, \varepsilon, \alpha)$-spanning set. In fact, if there is a pair of initial state $x_1 \in K$ and input signal $u_1 \in U^p$ such that there is no pair $(x_2, u_2) \in Z^\dagger$ where $x_{2i}u_{2i}$ is $(T, \varepsilon, \alpha)$-approximating for $x_{2i}u_{2i}$, then $(x_1, u_1)$ and any pair in $Z^\dagger$ violate (9) for some $t \in [0, T]$. Therefore, $x_{2i}u_{2i}$ can be added to $Z^\dagger$, which contradicts its maximality.

**Lemma 2.** For all $K$, $U^p$, $\varepsilon$, $\alpha$, and $T$,

$$s_{est}^1(T, 2\varepsilon, \alpha) \leq s_{est}^1(T, \varepsilon, \alpha). \quad (16)$$

**Proof.** Consider an arbitrary $(T, \varepsilon, \alpha)$-approximating set $Z$ and an arbitrary $(T, 2\varepsilon, \alpha)$-separated set $Z^\dagger$. Assume, for the sake of contradiction, the cardinality of $Z^\dagger$ is larger than the cardinality of $Z$. Then, since $Z$ is a $(T, \varepsilon, \alpha)$-approximating set, there exist two pairs $(x_1, u_1)$ and $(x_2, u_2) \in Z^\dagger$ such that (9) is satisfied with the same $z \in Z$. Hence,

$$\left\|\xi_{x_1, u_1}(t) - \xi_{x_2, u_2}(t)\right\| \leq \left\|z(t) - \xi_{x_1, u_1}(t)\right\| + \left\|z(t) - \xi_{x_2, u_2}(t)\right\|$$

$$\leq \varepsilon + \varepsilon = 2\varepsilon.$$

Thus, the two trajectories are not $(T, 2\varepsilon, \alpha)$-separated, contradicting the assumption that $Z^\dagger$ is a $(T, 2\varepsilon, \alpha)$-separated set.

The following theorem shows the relation between all of the introduced entropy definitions so far.

**Theorem 1.** For all $K$, $U^p$, $\varepsilon$, $\alpha$, and $T$,

$$h_{est}^1(2\varepsilon, \alpha) \leq h_{est}(\varepsilon, \alpha) \leq h_{est}^1(\varepsilon, \alpha) \leq h_{est}(\varepsilon, \alpha). \quad (17)$$

**Proof.** The first inequality follows from Lemma 2, the second one follows from (12), and the third one from Lemma 1.

V. **INFEASIBLE ALTERNATIVE ENTROPY NOTIONS**

In this section, we show that the $(\varepsilon, \alpha)$-estimation entropy $h_{est}(\varepsilon, \alpha)$ of a simple dynamical system is infinite if $\varepsilon$ approaches zero or if $\alpha > 0$. On the other hand, we show in Corollary 5 that $h_{est}(\varepsilon, 0)$ of that system is finite for any fixed $\varepsilon > 0$. We conclude that it is not meaningful to take the supremum over $\varepsilon > 0$ or choose an $\alpha > 0$, in contrast with the definition of estimation entropy of autonomous systems [2].

Consider the simple dynamical system:

$$\hat{x}(t) = u(t), \quad (18)$$

where $t \geq 0$, the initial state $x(0)$ is fixed to any real number, and the input signal $u$ will be chosen from sets that we will construct for each variant of entropy definition.

It is worth noting that system (18) has been shown earlier to not being exponentially stabilized by any finite number of control signals, and thus having an infinite stabilization entropy [26]. A modification of the stabilization property to include a constant stabilization error, similar to our constant estimation error, is made in [26] to ensure finiteness of entropy.

The sets of input signals that we use in this section share the same structure described in the following definition.

**Definition 5.** Given a sequence of $l$ real numbers $A = a_0, \ldots, a_{l-1}$ and an increasing sequence of $l + 1$ time points $t_{seq} = t_0, \ldots, t_l$, we define the set $SE$ of sequences (SE for $S E$quences) of length $l$ that are of the form: $se[0] = 0$ or $se[0] = a_0$ and for any $i > 0$, $se[i] = se[i-1] + a_i$. Following such a construction, we define a piecewise-constant signal $u : [t_0, t_l] \to \mathbb{R}$ generated by a sequence $se$ as follows: for all $i < l$ and $t \in [t_i, t_{i+1})$, $u_{se,se}(t) = se[i]$. \quad (19)

We denote the set of all such piecewise-right-constant signals, with $t_{seq}$ and $A$ fixed, by $U_{t_{seq}} = \bigcup_{se \in SE} U_{se, se}.$

Observe that the cardinality of $U_{t_{seq}}^A$ is $2^l$. Moreover, each of $U^p(0, \sum_{i=0}^{l-1} a_i)$ and $U^p(0, \sum_{i=0}^{l-1} a_i)$ contains $U_{t_{seq}}^A$. Example signals from $U_{t_{seq}}^A$ with the corresponding trajectories of system (18) are shown in Figure 1.

![Construction of an example set of input signals following Definition 5. Only five such signals are shown: $u_1$ that is equal to 0 at all times, $u_2$ that differs from $u_1$ by being equal to $a_1$ starting from $t_1$, $u_3$ that is equal to $a_0$ at all times, $u_4$ that differs from $u_3$ by being equal to $a_0 + a_1$ starting from $t_1$, and $u_5$ that differs from $u_2$ by being equal to $a_1 + a_2$ starting from $t_2$ (bottom). Corresponding trajectories of system (18) are shown in Figure 1.](image-url)
Hence, it is sufficient to show that the RHS of (20) is infinite. We will construct a $(T, 2\varepsilon, \alpha)$-separated set of size $O(2^{T/\alpha-\tau})$ when substituted in (20) would evaluate to infinity.

To construct the separated set, for any fixed $\varepsilon$ and $T$, we use the family of input signals $U_{\text{seq}}$, where $A = \sqrt{\varepsilon}, \ldots, \sqrt{\varepsilon}$, $t_{\text{seq} \in A} = \{0, \tau, 2\tau, \ldots, [T/\tau], \}$, and $\tau = 3\sqrt{\varepsilon}$. By such choices of $A$ and $\tau$, the total variation of any signal in $U_{\text{seq}}$ is upper-bounded by $\sqrt{T}/(3\sqrt{\varepsilon})$ which is equal to $\sqrt{\varepsilon}[T/(3\sqrt{\varepsilon})] = O(T)$. Therefore, there exist some constants $\mu$ and $\eta$ such that for any $\varepsilon > 0$ and $T > 0$, $U_{\text{seq}}$ is a subset of $U^0(\mu, \eta)$, and consequently, a subset of $U^0(\mu, \eta)$ as well.

**Lemma 3.** Fix $T$ and $\varepsilon > 0$. Consider any initial state $x_0 \in \mathbb{R}$ and the set of input signals $U_{\text{seq}}$. The resulting trajectories of system (18) form a $(T, 2\varepsilon, 0)$-separated set.

**Proof.** Consider any two signals $u_1, u_2 \in U_{\text{seq}}$ with corresponding sequences $s_{u_1}$ and $s_{u_2}$ with the same prefix up till the $j$th value for some $j \in [\lfloor T/\tau \rfloor - 1]$, and with different values in the $(j + 1)^{th}$ index. That is $s_{u_1}(j + 1) \neq s_{u_2}(j + 1)$. The resulting trajectories $x_{u_1, u_0}$ and $x_{u_2, u_0}$ from these input signals and the common initial state $x_0$ are $(T, 2\varepsilon, 0)$-separated. To see this, first note that the solution of the system is: $\xi_{t, u}(t) = \xi_{t, u}(j\tau) + u(j\tau)(t - j\tau), \quad t \in [j\tau, (j + 1)\tau),$ where $u(t)$ is constant over that interval. Then, the distance between the trajectories that start from $x_{u_1, u_0}(j\tau)$, or equivalently, $x_{u_2, u_0}(j\tau)$, and have $s_{u_1}(j + 1)$ and $s_{u_2}(j + 1)$, respectively, as input for $\tau$ time units is $a_i + 1$, which is equal to $3\sqrt{\varepsilon}$. By our choice of $\tau$ to be equal to $3\sqrt{\varepsilon}$, the distance is $3\varepsilon$. Hence, the distance $\|\xi_{t, u_1}(j + 1) - \xi_{t, u_2}(j + 1)\|$ between $x_{u_1, u_0}$ and $x_{u_2, u_0}$ at $t = (j + 1)\tau$ is equal to $3\varepsilon$, strictly larger than $2\varepsilon$. Since it is enough for two trajectories to have a distance larger than $2\varepsilon$ at a single point in time in the interval $[0, T]$ to be considered $(T, 2\varepsilon, 0)$-separated, $\xi_1$ and $\xi_2$ are $(T, 2\varepsilon, 0)$-separated. Hence, all the trajectories resulting from the initial state $x_0$ and $U_{\text{seq}}$ are $(T, 2\varepsilon, 0)$-separated.

**Theorem 2.** $\lim_{x \to 0} h_{\text{est}}(\varepsilon, \alpha)$ of system (18) with the input set $U_{\text{seq}}$ is infinite.

**Proof.** The number of trajectories in the separated set constructed in Lemma 3 is equal to $2^{T/\tau}$, which is $2^{[T/\tau]}$. Moreover, observe that any $(T, 2\varepsilon, 0)$-separated set is also a $(T, 2\varepsilon, \alpha)$-separated set, for any $\alpha > 0$. Hence, for any $\alpha > 0$, $\frac{s_{\text{est}}(T, 2\varepsilon, \alpha)}{s_{\text{est}}(T, 2\varepsilon, 0)} \geq T^\alpha$. Therefore, $\lim_{x \to 0} h_{\text{est}}(\varepsilon, \alpha)$ is infinite by a simple substitution of $s_{\text{est}}(T, 2\varepsilon, \alpha)$ with $s_{\text{est}}^\alpha(T, 2\varepsilon, 0) = 2^{T/\alpha\tau}$, the size of the $(T, 2\varepsilon, 0)$-separated set we constructed in the previous lemma, in inequality (20).

It follows that taking the supremum over $\varepsilon > 0$ would likely result in an infinite entropy for most systems with inputs, while it might be finite for any fixed $\varepsilon > 0$. Accordingly, we parameterize $h_{\text{est}}$ with $\varepsilon$ as in Definition 2. It is worth noting that Matveev and Savkin in Theorem 2.3.17 in [27] presented a sufficient condition for the topological entropy, with $\sup_{\varepsilon > 0}$ in its definition, of a discrete-time system with uncertain inputs to be infinite. Their sufficient condition is a function of the trajectories of the system. In fact, their result can be generalized to our setting of continuous-time systems of the form of system (4) in the following lemma. Before stating it, it is worth mentioning that there is no bound on the variation of the input signals in this lemma.

**Lemma 4.** Fix a set of input signals $U$ and any trajectory $\xi_{t, u}$ of system (4) with an arbitrary $\varepsilon \in \mathbb{R}^+$ and $u \in U$. Assume that $\exists \tau > 0$ and a vector $a_0 \neq 0$ in $\mathbb{R}^+$, such that for any $c_1, c_2, \ldots$ in the interval $[0, 1]$, there exists a trajectory $\xi_{t, u}$, that for any positive integer $j$, is equal to $\xi_{t, u}(j\tau) + c_j a_0$. Then, $\lim_{x \to 0} h_{\text{est}}(\varepsilon, \alpha)$ of system (4) is infinite.

**Proof.** The proof is almost the same proof as that of Theorem 2.3.17 in [27]. Given the assumptions in the lemma, we can construct a $(\tau, \frac{\log M}{\log 2}, 0)$-separated set, for any $M$ and $l \geq 1$, as follows: First, choose $c_l = \frac{\log M}{\log 2}$, for all $i \in [M]$. Then, consider the trajectories of system (4) which satisfy $\xi_{t, u}(j\tau) + c_j a_0$, for $i \in [M]$ and $j \in [l]$. Such trajectories exist by the second assumption in the lemma. Each trajectory in this set is $(\tau, \frac{\log M}{\log 2}, 0)$-separated from the other trajectories in the set. The cardinality of the set is equal to $M!$. Taking the limit of $\varepsilon$ going to zero is equivalent to taking the limit of $M$ going to infinity. By substituting $s_{\text{est}}(T, 2\varepsilon, \alpha)$ by $M!$ and $T$ by $\tau$ in the RHS of inequality (20) results in it being infinite.

The assumption in Lemma 4 would be satisfied if the system is locally reachable along a trajectory that is separated from the origin, as shown in Theorem 2.3.17 in [27]. Extending Definition 2.3.13 in [27] to continuous-time systems, system (4) is said to be locally reachable along a trajectory $\xi_{t, u}$ if two positive constants $\delta$ and $\tau$ exist such that for any positive $t$ and any $a$ and $b \in \mathbb{R}^2$ such that $\|\xi_{t, u}(t) - a\| \leq \delta \|\xi_{t, u}(t)\|$ and $\|\xi_{t, u}(t + \tau) - b\| \leq \delta \|\xi_{t, u}(t + \tau)\|$, another trajectory $\xi_{t, u}$ of system (4) exists and satisfies $\xi_{t, u}(t) = a$ and $\xi_{t, u}(t + \tau) = b$. Extending Definition 2.3.14 in [27] to continuous time systems, a trajectory $\xi_{t, u}$ of system 4 is said to be separated from the origin if there exists a positive constant $\delta_0$ such that $\|\xi_{t, u}(t)\|$ $\geq$ $\delta_0$, for all $t \geq 0$.

**B. Requiring exponentially convergent estimation error**

In this section, we prove that $h_{\text{est}}(\varepsilon, \alpha)$ of system (18) with a particular input set is infinite for any strictly positive $\alpha$. We present two alternative proofs: the first is a corollary of Theorem 2 and the second is done by constructing a new separated set. We start with the first and easier result, the corollary of Theorem 2. This corollary is suggested by one of the anonymous reviewers of the paper.

**Corollary 1.** For any $\varepsilon$ and $\alpha > 0$, $h_{\text{est}}(\varepsilon, \alpha)$ of system (18) with the input set $\cup_{\varepsilon > 0} U_{\text{seq} \in A}$ is infinite, where $U_{\text{seq} \in A}$ is the input set constructed for a specific $\varepsilon$ in Lemma 3.

**Proof.** Recall from (10) that $h_{\text{est}}(\varepsilon, \alpha)$ $= \lim_{T \to \infty} \frac{1}{T} \log s_{\text{est}}(T, \varepsilon, \alpha; K)$. Thus, for any finite time bound $\tau$, $h_{\text{est}}(\varepsilon, \alpha; K)$ is greater than or equal to:

$$\lim_{T \to \infty} \frac{1}{T} \log \left( s_{\text{est}}(T - \tau, \varepsilon, \alpha; K) \cdot \text{Reach}(K, U, [\tau, \tau]) \right)$$
lower bound on the size of the separated set, and consequently provide a lower bound on entropy of (18). We start by bounding $t_i$ with a function of $i$, the index of the switch.

**Lemma 6.** The $i$th time instant in $\text{ts}_0\alpha$ is upper-bounded as follows: $t_i = \sum_{j=1}^{i} v_j \leq \frac{2}{\alpha} \ln \left( \frac{2\sqrt{2c}}{2 - \alpha \sqrt{2c}} + 1 \right)$. Moreover, the cardinality of $\text{ts}_0\alpha$ over the time horizon $[0, T]$ is lower-bounded as follows: $|\text{ts}_0\alpha| \geq \frac{\sqrt{2c}}{2(\alpha^T - 1)}$.

**Proof.** To upper bound the sum of $v_i$’s, we upper bound the sequence $\{v_j\}$ with another sequence $\{w_j\}$, whose sum is easy to compute. Recall that $v_i = \sqrt{2c}$ and $w_i = \frac{v_i}{\alpha} e^{-\alpha c/2}$ for $i \geq 1$. Since $\alpha v_i > 0$, we can lower bound $w_i$ as follows: $w_{i+1} = \frac{w_i}{\alpha + \alpha v_{i+1}^2}$, for all $i > 1$. This change is propagated in the proof below without red markings.

**Claim:** $v_i \leq w_i$, for all $i \geq 1$.

**Proof of claim:** We will proceed by induction. The base case is when $i = 1$ and we know that $v_1 \leq w_1$. Let $y = \frac{1}{\alpha + \alpha v_i^2}$ for $y \in \mathbb{R}$. Then, $\frac{dy}{dy}(y) = \frac{1}{(1 + \alpha y^2)} > 0$ and thus is an increasing function. Hence, assuming that $v_i \leq w_i$, we get that $v_i < w_i$, since we know that $v_{i+1} < \frac{w_i}{\alpha + \alpha v_i^2}$. Hence, by induction, $v_i \leq w_i$ for all $i \geq 1$.

**Proof of claim:** We again proceed by induction: for $i = 1$, $\frac{1}{2(\alpha + \alpha v)^2} = w_1$. Assume that $w_i = \frac{1}{2(\alpha + \alpha v_i^2)}$ and $w_{i+1} = \frac{1}{2(\alpha + \alpha v_{i+1}^2)}$. Hence, $\sum_{j=1}^{i} w_j = \frac{1}{2} \sum_{j=1}^{i} \frac{1}{\alpha + \alpha v_j^2}$. Now we prove to proceed the main lemma. First, note that $c > -1$ since $w_1$ and $\alpha$ are positive. By convexity of $\frac{1}{2}$,

\[
\frac{1}{2} \leq \frac{1}{\alpha v_{i+1}^2} \leq \frac{1}{\alpha v_{i}^2}
\]

Hence, the time of the $i$th switch can be bounded as follows:

\[
\sum_{j=1}^{i} \frac{1}{j + c} < \sum_{j=1}^{i} \frac{dy}{y + c} = \int_{1/c}^{1} \frac{dy}{y + c} = \ln \left( \frac{i + c + 1/2}{c + 1/2} + 1 \right)
\]

\[
\ln \left( \frac{2\alpha v_i}{2 - \alpha v_i^2} + 1 \right) = \ln \left( \frac{4\alpha \sqrt{2c}}{4 - \alpha \sqrt{2c}} + 1 \right).
\]

Therefore, $\sum_{j=1}^{i} v_j \leq \sum_{j=1}^{i} w_j \leq \frac{2}{\alpha} \ln \left( \frac{2\sqrt{2c}}{2 - \alpha \sqrt{2c}} + 1 \right)$.

Now that we have the upper bound of $t_i$ as a function of $i$, we can lower bound the number of switches that an input signal is allowed before a specified time bound $T$ by $\frac{4\alpha \sqrt{2c}}{4 - \alpha \sqrt{2c}} \left( e^{\alpha T/2} - 1 \right)$.

**Lemma 7.** The sequence $\{t_i\}_i$ diverges to infinity, i.e., $\lim_{i \to \infty} t_i = \infty$, and the set of inputs in $U_{\alpha}$ are not Zeno.

**Proof.** The proof is by contradiction. By definition, $t_{i+1} = t_i + v_i e^{-\alpha c/2}$. Assume that $\lim_{i \to \infty} t_i = c$, for some constant $c \in \mathbb{R}_{\geq 0}$. Then, $c = c + v_i e^{-\alpha c/2}$, a contradiction.

**Corollary 2.** The total variation of any signal in $U_{\alpha}$, with number of switches that is $O(e^{\alpha T})$ is $O(T)$.
For any \( \alpha > 0 \), there exists an estimation algorithm that satisfies the \( \epsilon \)-estimation entropy for system (18) with \( |\alpha| < \frac{\epsilon}{2} \). By substituting \( \epsilon = |\alpha| \) into the above inequality, we get that 
\[
\limsup_{T \to \infty} \frac{1}{T} \sum_{i=0}^{\lfloor T/\epsilon \rfloor} \log |\Gamma| \leq \frac{\epsilon}{2}.
\]

By substituting \( \epsilon = |\alpha| \) into the above inequality, we get that 
\[
\limsup_{T \to \infty} \frac{1}{T} \sum_{i=0}^{\lfloor T/\epsilon \rfloor} \log |\Gamma| \leq \frac{\epsilon}{2}.
\]

Hence, for a sufficiently large \( T \), we should have 
\[
\frac{1}{T} \log |\Gamma(T', \epsilon)| < \frac{1}{T} \log s_{est}(T', \epsilon),
\]
where \( l = \lfloor T/\epsilon \rfloor \). Consequently, we get the inequality 
\[
|\Gamma|^{l+1} < s_{est}(T', \epsilon).
\]
However, \( |\Gamma|^{l+1} \) is the number of possible sequences of length \( l + 1 \) that can be sent by the sensor over \( l + 1 \) iterations. There are \( l + 1 \) instead of \( l \) iterations over the interval \([0, T']\) since the sensor starts sending the codewords at \( t = 0 \). Hence, the number of functions that can be constructed by the estimator is upper bounded by \( |\Gamma|^{l+1} \).

Now, let us define the bit rate of an estimation algorithm:
\[
b_r(\epsilon) := \limsup_{T \to \infty} \frac{1}{T} \sum_{i=0}^{\lfloor T/\epsilon \rfloor} \log |\Gamma| = \frac{\log |\Gamma|}{T_p}.
\]

Given any time bound \( T \), the average bit rate of the state estimation algorithm is the total number of bits it sends divided by \( T \). It sends \( \log |\Gamma| \) bits every \( T_p \) time units.

**Proposition 1.** For any \( \epsilon > 0 \), there is no fixed bit rate state estimation algorithm for system (4) with a bit rate smaller than \( h_{est}(\epsilon) \) that for any \( T > 0 \), generates a \((T, \epsilon)\)-approximating function for any input trajectory.

**Proof.** The proof is similar to the proof of Proposition 2 in [3]. It is based on showing that if any estimation algorithm has a bit rate \( b_r(\epsilon) \) less than \( h_{est}(\epsilon) \), then for any time bound \( T \), the set of functions \( z \) (see Figure 2) that can be constructed by that algorithm is a \((T, \epsilon)\)-approximation set with a smaller cardinality than \( s_{est}(T, \epsilon) \), the supposed minimal cardinality, leading to a contradiction.

Specifically, for the sake of contradiction, assume that there exists an estimation algorithm with a bit rate smaller than \( h_{est}(\epsilon) \) and constructs \((T, \epsilon)\)-approximating functions for input trajectories. Recall that 
\[
h_{est}(\epsilon) = \limsup_{T \to \infty} \frac{1}{T} \log s_{est}(T, \epsilon).
\]
Then,
\[
b_r(\epsilon) < h_{est}(\epsilon),
\]
\[
\limsup_{T \to \infty} \frac{1}{T} \sum_{i=0}^{\lfloor T/\epsilon \rfloor} \log |\Gamma| < \limsup_{T \to \infty} \frac{1}{T} \log s_{est}(T, \epsilon),
\]
and
\[
\limsup_{T \to \infty} \frac{1}{T} \sum_{i=0}^{\lfloor T/\epsilon \rfloor} \log |\Gamma| < \limsup_{T \to \infty} \frac{1}{T} \log s_{est}(T, \epsilon).
\]

Hence, for a sufficiently large \( T' \), we should have 
\[
\frac{(l+1) \log |\Gamma|}{T} < \frac{1}{T} \log s_{est}(T', \epsilon),
\]
where \( l = \lfloor T'/\epsilon \rfloor \). Consequently, we get the inequality 
\[
|\Gamma|^{l+1} < s_{est}(T', \epsilon).
\]
However, \( |\Gamma|^{l+1} \) is the number of possible sequences of length \( l + 1 \) that can be sent by the sensor over \( l + 1 \) iterations. There are \( l + 1 \) instead of \( l \) iterations over the interval \([0, T']\) since the sensor starts sending the codewords at \( t = 0 \). Hence, the number of functions that can be constructed by the estimator is upper bounded by \( |\Gamma|^{l+1} \).

For any given trajectory, the output of the estimator is a corresponding \((T', \epsilon)\)-approximating function over the interval \([0, T']\). This is true since the estimator should be able to construct a \((T', \epsilon)\)-approximating function for the corresponding trajectory of the system over the interval \([0, (l+1)T_p] \) given the symbols sent by the sensor in the first \( l+1 \) iterations. Hence, the set of functions that can be constructed by the estimator defines a \((T', \epsilon)\)-approximating set. However, \( s_{est}(T', \epsilon) \) is the minimal cardinality of such a set. Therefore, the set of functions that can be constructed by the algorithm is a \((T', \epsilon)\)-approximating set with a cardinality smaller than \( s_{est} \), the supposed minimal one, which is a contradiction.

**VII. Entropy Upper Bound and Algorithm**

In this section, we derive an upper bound on the entropy \( h_{est}(\epsilon) \) of system (4) in terms of its sensitivity to its initial state and input and the required bound on the estimation error.
A. Local input-to-state discrepancy function construction

We use a modified version of the definition of local input-to-state discrepancy as introduced in [28] in order to upper bound the distance between any two trajectories. We relax the original definition to include piece-wise continuous input signals and piece-wise continuous Jacobian matrices.

**Definition 7. (Local IS Discrepancy).** Fix an arbitrary compact subset \( \mathcal{X} \subset \mathbb{R}^n \), time interval \([t_0, t_1]\) \( \subset \mathbb{R}_{\geq 0} \), and a set of piecewise-right-continuous functions \( \mathcal{U} \) mapping \([t_0, t_1]\) to \( \mathbb{R}^m \). A function \( V : \mathcal{X}^2 \rightarrow \mathbb{R}_{\geq 0} \) is a local input-to-state (IS) discrepancy function for system (4) over \( \mathcal{X} \) and \([t_0, t_1]\) if:

(i) there exist class-K functions \( \alpha, \alpha' \) such that for any \( x, x' \in \mathcal{X} \), \( \alpha(||x - x'||) \leq V(x, x') \leq \alpha'(||x - x'||) \), and

(ii) there exists a class-K function \( \gamma : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0} \) and a class-K function \( \beta : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0} \) and a class-K function \( \gamma : \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0} \) such that for any \( x_0, x'_0 \in \mathcal{X} \) and \( u, u' \in \mathcal{U} \), if \( \xi_{x, u}(t) \) and \( \xi_{x', u'}(t) \) are local input-to-state (IS) functions that can possibly be constructed by Algorithm 1 for all \( t \in [t_0, t_1] \), then for any such \( t \),

\[
V(\xi_{x, u}(t), \xi_{x', u}(t)) \leq \\
\beta(||x_0 - x'_0||, t - t_0) + \int_{t_0}^{t} \gamma(||u(s) - u'(s)||)ds. \tag{22}
\]

The local discrepancy function \( V \) together with \( \beta \) and \( \gamma \) give the sensitivity of the trajectories of the system to changes in the initial state and the input. The functions \( \alpha, \alpha', \beta, \gamma \) are sometimes called witnesses of the local IS discrepancy \( V \). Techniques for computing local discrepancy functions have been presented [28]-[30].

The following is a straight-forward generalization of Lemma 15 of [28] to handle systems with piece-wise continuous inputs, instead of just continuous ones. In contrast with [28], we consider the distance between the trajectories squared as the discrepancy function instead of the distance itself. This is a result we presented earlier in [5], but had two issues: first, it was false when \( G_x \) is negative, and second, it was used as a bound on the squared norm while it was a bound on the squared distance.

We present the proof in the longer version [31].

**Lemma 8.** The function \( V(x, x') := ||x - x'||^2 \) is a local IS discrepancy for system (4) over any compact set of states \( \mathcal{X} \subset \mathbb{R}^n \), time interval \([t_0, t_1]\) \( \subset \mathbb{R}_{\geq 0} \), and piecewise-right-continuous inputs \( \mathcal{U} \) with

\[
\beta(y, t - t_0) := e^{2G_x(t-t_0)}y^2 \quad \text{and} \quad \\
\gamma(y) := G_u\left( \sup_{t \in [t_0, t_1]} e^{2G_x(t-t)}y^2, \right.
\]

and

\[
\begin{align*}
G_x &:= \frac{n}{2} \left( \sup_{t \in [t_0, t_1]} 2M^2 \max_{u \in \mathcal{U}, x \in \mathcal{X}} \left( J_x(x, u(t)) + J_x(x, u(t)) \right)^T + 1 \right) \\
G_u &:= \sqrt{m} \sup_{t \in [t_0, t_1]} \sup_{u \in \mathcal{U}, x \in \mathcal{X}} \|J_u(x, u(t))\|.
\end{align*}
\tag{23}
\]

If \( f \) is globally Lipschitz continuous in both arguments, one can infer that \( G_x \) and \( G_u \) are finite for any compact set of states \( \mathcal{X} \) and time interval \([t_0, t_1]\). In that case, we will denote any global upper bounds on \( G_x \) and \( G_u \) of any \( \mathcal{X} \) and time interval \([t_0, t_1]\), by \( M_x \) and \( M_u \), respectively. An example of such bounds is presented in the following proposition, with the proof in the longer version of this paper [31].

**Proposition 2.** If \( f \) is globally Lipschitz continuous in both arguments, then for any time interval \([t_0, t_1]\) \( \subset \mathbb{R}_{\geq 0} \) and compact set \( \mathcal{X} \subset \mathbb{R}^n \), \( G_x \leq n(nL_x + \frac{1}{2}) \) and \( G_u \leq n\sqrt{m}L_u \), where \( L_x \) and \( L_u \) are the global Lipschitz constants of \( f \).

Further, it is shown in [29] that if \( f \) has a continuous Jacobian, one can get tighter local bounds on \( G_x \) and \( G_u \) that depend on the set of input functions \( \mathcal{U}(\mu, \eta) \), the compact set \( \mathcal{X} \), and the interval \([t_0, t_1]\).

Therefore, for any \( t > 0 \), \( t \in [0, \tau] \), \( x_0, x'_0 \in \mathcal{X} \) and \( u, u' \in \mathcal{U}(\mu, \eta) \), the squared distance \( ||\xi_{x, u}(t) - \xi_{x', u'}(t)||^2 \), is upper bounded by:

\[
e^{2M_x t}||x_0 - x'_0||^2 + M_u^2 e^{2M_u t} \int_0^t \|u(s) - u'(s)\|^2 ds, \tag{24}
\]

where \( M^+_x = \max\{M_x, 0\} \).

**Example 2** (Dubin’s vehicle IS Discrepancy). Consider the vehicle of example 1. Then, its

\[
J_x + J_u^T = \begin{bmatrix} 0 & 0 & -\frac{1}{2} \sin x_3 \\ 0 & 0 & \frac{1}{2} \sin x_3 \\ -\frac{1}{2} \cos x_3 & \frac{1}{2} \cos x_3 & 0 \end{bmatrix}, \tag{25}
\]

and has eigenvalues 0 and \( \pm \frac{\sqrt{2}}{2} \). Hence, \( G_x = \frac{n}{2}(2x^2/2 + 1) = \frac{3(x+1)}{2} \). Moreover, \( G_u = \sqrt{m} \|J_u\| = 1 \). If we use Proposition 2 to upper bound \( G_x \) and \( G_u \), we get \( M_x = n(nL_x + 1/2) = 3(0 + 1/2) \) and \( M_u = m\sqrt{m}L_u = 1 \). Here \( G_x \) and \( G_u \) do not depend on \( \mathcal{X} \) or on the time interval and thus they can replace \( M_x \) and \( M_u \) in inequality (24).

B. Approximating set construction

Let us fix \( \varepsilon > 0 \) throughout this section. We will describe a procedure (Algorithm 1) that, given a time bound \( T > 0 \), an initial state \( x_0 \in \mathcal{K} \) and an input signal \( u \in \mathcal{U}(\mu, \eta) \), constructs a \((T, \varepsilon, \cdot, \cdot)-approximating\) function for the trajectory \( \xi_{x, u} \) over the time interval \([0, T]\). It follows that the set of functions that can possibly be constructed by Algorithm 1 for different \( x_0 \in \mathcal{K} \) and \( u \in \mathcal{U}(\mu, \eta) \) is a \((T, \varepsilon, \cdot, \cdot)-approximating\) set for system (4). We will show in the next section how to obtain an upper bound on entropy by deriving an upper bound on the cardinality of this set. That upper bound on entropy will be an upper bound on the minimal bit rate of state estimation as well, as Algorithm 1 is a state estimation algorithm that requires a fixed bit rate equal to that upper bound.

The procedure (Algorithm 1) is parameterized by a time horizon \( T > 0 \), a sampling period \( T_p > 0 \), and two constants \( \delta_x \) and \( \delta_u > 0 \). The procedure also uses the set of initial states \( \mathcal{K} \), the set of initial inputs \( \mathcal{U} \), a particular initial state \( x_0 \in \mathcal{K} \), and an input signal \( u \in \mathcal{U}(\mu, \eta) \) for system (4). The output is a piece-wise continuous function \( z : [0, T] \rightarrow \mathbb{R}^n \) that is constructed iteratively over each \([iT_p, (i + 1)T_p)\).
Algorithm 1 Construction of a \((T, \varepsilon)\)-approximating function.

1: **input**: \(T, T_p, \delta_x, \delta_u\)
2: \(S_{x,0} \leftarrow K; S_{u,0} \leftarrow U;\)
3: \(C_{x,0} \leftarrow \text{grid}(S_{x,0}, \delta_x);\)
4: \(C_{u,0} \leftarrow \text{grid}(S_{u,0}, \delta_u);\)
5: \(i \leftarrow 0;\)
6: **while** \(i \leq \left\lfloor \frac{T_p}{T} \right\rfloor\) **do**
7: \(x_i \leftarrow \xi_{x,i}(iT_p); u_i \leftarrow u(iT_p);\)
8: \(q_{x,i} \leftarrow \text{quantize}(x_i, C_{x,i});\)
9: \(q_{u,i} \leftarrow \text{quantize}(u_i, C_{u,i});\)
10: \(z_i \leftarrow \xi_{x,i}(q_{x,i}, q_{u,i});\)
11: \(i \leftarrow i + 1;\) \{parameters for next iteration\}
12: \(S_{x,i} \leftarrow B(z_{i-1}(T_p)), \varepsilon;\)
13: \(S_{u,i} \leftarrow B(q_{x,i-1}, \eta + \mu T_p + \delta_u);\)
14: \(C_{x,i} \leftarrow \text{grid}(S_{x,i}, \delta_x);\)
15: \(C_{u,i} \leftarrow \text{grid}(S_{u,i}, \delta_u);\)
16: \(\text{wait}(T_p);\)
17: **end while**
18: **output**: \(\{z_i(0, T_p) : i \in [0 \mid \frac{T_p}{T} - 1]\} \cup \{z_i(T_p) (0, T_p)\}\)

Lemma 9. Given \(\varepsilon > 0\), fix \(\delta_x, \delta_u, \text{ and } T_p\) such that \(g_c(\delta_x, \delta_u, T_p) \leq \varepsilon^2\). For any \(x_0 \in K\) and \(u \in \mathcal{U}^p(\mu, \eta), \text{ for all } t \in [iT_p, (i+1)T_p),\)

(i) \(u_i \in S_{u,i}\),
(ii) \(\|u(t) - q_{u,i}\| \leq \mu(t - iT_p) + \eta + \delta_u,\)
(iii) \(x_i \in S_{x,i},\) and
(iv) \(z_i(t) - \xi_{x,i}(t - iT_p)\) \(\leq \varepsilon,\)

where \(u(i) := u(iT_p)\), the \(i^{th}\) piece of \(u\) of size \(T_p\).

Proof. First, \(\|u_0 - q_{u,0}\| \leq \delta_u\) since the \(u(0) = u_0 \in U\) and \(q_0\) is a quantization of \(u_0\) with respect to a grid of resolution \(\delta_u\). Moreover, \(\|u(t) - u_0\| \leq \mu + \eta\) by Definition 1. But, by triangle inequality, \(\|u(t) - q_{u,0}\| \leq \|u(t) - u_0\| + \|u_0 - q_{u,0}\| \leq \mu + \eta + \delta_u\). Hence, \(u_1 \in S_{u,1}\). Fix \(i \geq 1\) and assume that \(u_i \in S_{u,i}\). Then, \(\|u_i - q_{u,i}\| \leq \delta_u\). Repeating the same analysis for the \(i = 0\) case, results in \(\|u_i(t - iT_p) - q_{u,i}\| \leq \mu(t) + \eta + \delta_u\). Thus, \(u_i \in S_{u,i}\) for all \(i\). Second, \(x_0 \in S_{x,0}\) since \(x_0 \in K\). Also, \(\|x_0(0) - \xi_{x,0}(0)\| = \|x_0, x_0 - x_\delta, \leq \delta_x \leq \varepsilon,\)

where the last inequality follows from the assumption that \(g_c(\delta_x, \delta_u, T_p) \leq \varepsilon^2\). Now, fix \(t \in [0, T]\) and let \(i = \left\lfloor \frac{T}{T_p} \right\rfloor\). Then,

\[
\|z_i(t) - \xi_{x,i}(t - iT_p) - \xi_{x,i}(t - iT_p)\| \leq \varepsilon,
\]

\[
\leq \|x_i - q_{x,i}\| \leq \varepsilon^2.
\]

where the last inequality follows from the assumption in the lemma statement. Therefore, for all \(i \in [0 \mid \frac{T_p}{T} - 1]\) and \(t \in [0, T],\)

\(x_i \in B(z_{i-1}(T_p), \varepsilon) = S_{x,i}.\)

Corollary 3. For any \(\delta_x, \delta_u, \text{ and } T_p\) such that \(g_c(\delta_x, \delta_u, T_p) \leq \varepsilon^2\), for all \(t \in [0, T],\) the output function \(z\) of Algorithm 1 is a \((T, \varepsilon)\)-approximating function to the input trajectory \(\xi_{x,0,i}\), i.e.

\[|z(t) - \xi_{x,0,i}(t)| \leq \varepsilon.\]

One can conclude from Corollary 3 that the set of all functions that can be constructed by Algorithm 1 for any input trajectory \(\xi_{x,0,i}\), where \(x_0 \in K\) and \(u \in \mathcal{U}^p(\mu, \eta),\) is a \((T, \varepsilon)\)-approximating set. In the following lemma, we will compute an upper bound on the number of these functions to obtain an upper bound on \(s_{\text{est}}(T, \varepsilon)\).

Lemma 10. For any \(T \geq 0\) and \(\delta_x, \delta_u, \text{ and } T_p\) such that \(g_c(\delta_x, \delta_u, T_p) \leq \varepsilon^2\), the number of functions that can be constructed by Algorithm 1 for any \(x_0 \in K\) and \(u \in \mathcal{U}^p(\mu, \eta),\) is upper-bounded by:

\[
\frac{\text{diam}(K)}{2\delta_x}[\frac{\text{diam}(U)}{2\delta_u}]^m \left(\frac{\varepsilon}{\delta_x}\right)^n \left(\frac{\|\mu T_p + \eta\|}{\delta_u} + 1\right)^m \left\lfloor \frac{T}{T_p} \right\rfloor^p.
\]
Proof. To construct a $(T, \varepsilon)$-approximating function for a given trajectory $\xi_{x,u}$, at an iteration $i \in [0, \lceil T_p \rceil]$, Algorithm 1 picks one point in $C_{x,i}$ and one in $C_{u,i}$. Hence, the number of different outputs that it can produce is upper bounded by: $\prod_{i=0}^{\lceil T_p \rceil} |C_{x,i}| |C_{u,i}|$.

To construct the grids $C_{x,0}$ and $C_{u,0}$, in each of the $n$ (or $m$) dimensions of the state (or input) space, we partition a segment of length $diam(K)$ (or $diam(U)$) to smaller segments of size $2\delta_x, 2\delta_u$. Then, $|C_{x,0}| \leq \lceil \frac{diam(K)}{2\delta_x} \rceil^n$ and $|C_{u,0}| \leq \lceil \frac{diam(U)}{2\delta_u} \rceil^m$. Similarly, for all $i > 0$, $S_{x,i} = B(z_{i-1}(T_p), \ldots, z_{i}(T_p), \ldots)$ and $S_{u,i} = B(u_{i-1}, u_i, \ldots, u_{T_p})$. Hence, $|C_{x,i}| \leq \lceil \frac{2\delta_x}{\delta_x} \rceil^n = \lceil \frac{n}{\delta_x} \rceil$ since $diam(S_{x,i}) = 2\varepsilon$ and $|C_{u,i}| \leq \lceil \frac{2\delta_u}{\delta_u} \rceil^m = \lceil \frac{m}{\delta_u} \rceil$, since $diam(S_{u,i}) = 2\varepsilon$. Substituting these values in the bound above leads to the upper bound in the lemma.

C. Entropy upper bound

The following proposition gives an upper bound on the entropy of system (4) in terms of $T_p, \delta_x, \delta_u$. It shows the effect of our choice of the parameters of Algorithm 1. It will also help us recover the bound on estimation entropy $h_{est}(\varepsilon)$ of autonomous systems in [1] Corrolary 4.

Proposition 3. For a fixed $T_p, \delta_x, \delta_u$ where $g_c(\delta_x, \delta_u, T_p) \leq \varepsilon^2$, the entropy $h_{est}(\varepsilon)$ of system (4) is upper bounded by $g_o(\delta_x, \delta_u, T_p)$ defined to be equal to:

$$\frac{1}{T_p} \left( n \log \left( \frac{\varepsilon}{\delta_x} \right) + m \log \left( \frac{T_p + \eta}{\delta_u} + 1 \right) \right).$$

Proof. We substitute the upper bound on the cardinality of the minimal approximating set obtained in the previous section in definition of $h_{est}$ in Equation (2) to get: $h_{est}(\varepsilon)$ =

$$\limsup_{T \to \infty} \frac{1}{T} \log s_{est}(T, \varepsilon) = \limsup_{T \to \infty} \frac{1}{T} \log |C_{x,0}||C_{u,0}|(|C_{x,1}| |C_{u,1}|) \left( \frac{T_p}{T} \right)$$

$$\leq \limsup_{T \to \infty} \frac{1}{T} \log \left( \frac{|C_{x,1}| |C_{u,1}|}{|C_{x,0}| |C_{u,0}|} \right) \left( \frac{T_p}{T} \right)$$

$$\leq \limsup_{T \to \infty} \frac{1}{T} \log \left( \frac{\varepsilon}{\delta_x} \right) \left( \frac{T_p + \eta}{\delta_u} + 1 \right) \left( \frac{T_p}{T} \right)$$

$$\leq \limsup_{T \to \infty} \frac{T_p}{T} \frac{1}{n} \log \left( \frac{\varepsilon}{\delta_x} \right) \left( \frac{T_p + \eta}{\delta_u} + 1 \right)$$

$$+ \limsup_{T \to \infty} \frac{T_p}{T} \frac{1}{m} \log \left( \frac{\mu T_p + \eta}{\delta_u} + 1 \right)$$

$$= \frac{1}{T_p} \left( n \log \left( \frac{\varepsilon}{\delta_x} \right) + m \log \left( \frac{\mu T_p + \eta}{\delta_u} + 1 \right) \right).$$

We show that if the bound on the input norm is negligible, we recover the bound $\frac{n\delta_x + m\delta_u}{T_p}$ on entropy derived in [2].

Corrolary 4. Given any $\varepsilon > 0$, $\lim_{\mu, \eta \to 0} h_{est}(\varepsilon) \leq \frac{n\delta_x + m\delta_u}{T_p}$.

Proof. Fix any $\varepsilon > 0$. Fix a $\eta_0$ and $\mu_0$ such that $\eta_0 + \mu_0 < 1$. Then, fix $\delta_x$ to be equal to $\delta_{x,0} = \delta_0 + \mu_0$. After that, while taking the limits of $\mu$ and $\eta$ going to zero, keep $\delta_x$ fixed to $\delta_{x,0}$, choose $\delta_x = \sqrt{1 - (\mu + \eta)} \varepsilon L_x T_p$, and $T_p$ to be such that $g_c(\delta_x, \delta_{x,0}, T_p) \leq \varepsilon^2$. Hence, by Proposition 3, $h_{est}(\varepsilon) \leq \frac{1}{T_p} \left( n \log \left( \frac{\varepsilon L_x T_p}{\delta_x} \right) + m \log \left( \frac{\mu T_p + \eta}{\delta_u} + 1 \right) \right) \leq n L_x \log \left[ \frac{\varepsilon + \mu T_p + \eta}{\delta_x} + 1 \right]$. Moreover, as $\mu$ and $\eta$ decrease to zero, the argument of the second log approaches one and the upper bound approaches $\frac{n L_x}{T_p}$.

Corrolary 5. For any $\varepsilon > 0$, $h_{est}(\varepsilon)$ of system (18), with any set of inputs $U^p(\mu, \eta)$ with finite $\mu$ and $\eta$, is finite.

Proof. The Jacobian $J_x$ of system (18) is equal to zero since the state $x$ does not appear on the RHS of the differential equation. Hence, $M_x = 1/2$ using Equation (23). Moreover, $J_u = 1$ and thus $M_u = 1$. Then, there exist $\delta_x, \delta_u$ and $T_p$ > 0 that satisfy $g_c(\delta_x, \delta_u, T_p) \leq \varepsilon^2$, and $h_{est}(\varepsilon)$ of system (18) is upper bounded by $g_o(\delta_x, \delta_u, T_p)$, by Proposition 3.

Corrolary 5 is the last piece of the argument that other variants of entropy definitions mentioned in Section V are not adequate for systems with uncertain inputs.

Example 3 (Dubin’s car entropy upper bound). Consider the car of example 1 and its IS discrepancy function of example 2. In this example, we will compute its upper bound per Proposition 3. Suppose $\mu = \pi/4$ and $\eta = \pi/4$ and the needed estimation accuracy $\varepsilon = 0.1$. Let us fix $v = 10$ m/s. Then, $L_x = (10 + 1)/2 = 5.5$, $G_x = \frac{3}{2}(10 + 1) = 16.5$, $M_x = 3(3(10 + 1)/2) = 91.5$ and $G_u = 1$.

First, we use $G_x$ and $G_u$ for the discrepancy function as a replacement of $M_x$ and $M_u$ in the bound of Proposition 3 and in the definition of $g_c$ in Equation (26). Let us fix $\delta_x = \eta = \pi/4$, $T_p = 1.9 \times 10^{-3}$, and $\delta_u = \sqrt{2 G_x T_p}$. Such an assignment satisfies $g_c(\delta_x, \delta_u, T_p) = 0.0099 \leq \varepsilon^2 = 0.01$ and results in the bound $g_o(\delta_x, \delta_u, T_p) = 727$. Second, we derive the resulting bound from using the upper bounds on $G_x$ and $G_u$, $M_x$ and $M_u$ of Proposition 2, instead of $G_x$ and $G_u$. The previous choices of $\delta_x, \delta_u$, and $T_p$ would result in $g_c(\delta_x, \delta_u, T_p) = 0.011 > \varepsilon^2 = 0.01$, which violates the condition in Proposition 3. Instead, we use $\delta_x = \sqrt{2 M_x T_p}$ and $T_p = 1.5 \times 10^{-3}$ while keeping $\delta_u = \pi/4$. Such an assignment satisfies $g_c(\delta_x, \delta_u, T_p) = 0.0098 \leq \varepsilon^2 = 0.01$ and results in the bound $g_o(\delta_x, \delta_u, T_p) = 920$, larger than the one we obtained using $G_x$ and $G_u$.

Example 4 (Harrier jet). We study the Harrier “jump jet” model from [32]. The dynamics of the system is given by:

$$\dot{x}_1 = x_4; \dot{x}_4 = -g \sin x_3 - \frac{c_4 x_4}{m'} + \frac{u_1}{m'} \cos x_3 - \frac{u_2}{m'} \sin x_3;$$

$$\dot{x}_2 = x_5; \dot{x}_5 = g \cos x_3 - 1 - \frac{c_5 x_5}{m'} + \frac{u_1}{m'} \sin x_3 + \frac{u_2}{m'} \cos x_3;$$

$$\dot{x}_3 = x_6; \dot{x}_6 = \frac{r}{J u_1};$$

where $(x_1, x_2)$ is the position of the center of mass and $x_3$ is the orientation of the aircraft in the vertical plane, and $(x_4, x_5, x_6)$ are their time derivatives, respectively. The mass of the aircraft is $m'$, the moment of inertia is $J$, the gravitational constant is $g$, and the damping coefficient is $c$. The inputs $u_1$ and $u_2$ are the force vectors generated by the main downward thruster and the maneuvering thrusters. We computed $G_x$ and $G_u$ in the longer version of this paper [31]. Their values are
\[ G_x = n\left(\frac{-c}{m'} + \frac{1}{2}\right) \] and \[ Gu \leq q\left(\frac{r^2}{J^2} + \frac{1}{m'^2}\right) \]. Fixing \( \mu = 10 \) and \( \eta = 20 \) and the needed estimation accuracy \( \varepsilon = 0.5 \). We choose \( \delta_u = \eta = 20 \), \( T_p = 3.95 \times 10^{-3} \) , and \( \delta_x = \frac{1}{\sqrt{2}}\varepsilon e^{-G_x T_p} \). Then, using Proposition 3, \( g_e(\delta_u, \delta_x, T_p) = 0.249 \leq \varepsilon^2 = 0.25 \). Hence, the bound increased significantly as the bound on the variance of the input signal was relaxed. Suppose now that we take the other extreme, where we restrict the allowed size of jumps in the input signal by decreasing \( \mu \) to 0.1 while allowing large continuous variations by increasing \( \mu \) to 20. In that case, the input signals are almost continuous. If we fix \( \delta_u = \eta = 20 \), \( T_p = 3.5 \times 10^{-3} \) , and \( \delta_x = \frac{1}{\sqrt{2}}\varepsilon e^{-G_x T_p} \). Then, \( g_e(\delta_u, \delta_x, T_p) = 0.244 \leq \varepsilon^2 = 0.25 \), which is significantly smaller than the two previous bounds.

VIII. SWITCHED SYSTEMS: ENTROPY, UPPER BOUNDS, AND RELATION WITH SYSTEMS WITH UNCERTAIN INPUTS

In this section, we define the estimation entropy of general autonomous switched systems and present a corresponding upper bound (Theorem 4), which is a restatement of Theorem 1 in [3]. In Section VIII-A, through an example, we present a novel result of this paper showing that an assumption we made in [3] on the boundedness of the divergence between different modes in the derivation of the upper bound on entropy is reasonable. Then, in Section VIII-B, we rewrite the upper bound of Proposition 3 which we derived for systems with inputs in a similar format to that of Theorem 4 which we derived for switched systems to show the similarities and differences of the two bounds. Finally, we show how switched systems can be modeled by systems with inputs in Section VIII-C. This implies that the upper bound on entropy of systems with inputs can be used for autonomous switched systems. All of these results are a novel contribution of this paper.

We denote a switched system with \( N \) modes by:

\[ \dot{x}(t) = f_{sw}(x(t), \sigma(t)), \] (30)

where \( \sigma : \mathbb{R}_{\geq 0} \to [N] \) is a piece-wise-right-constant signal called a switching signal. \( f_{sw} : \mathbb{R}^n \times [N] \to \mathbb{R}^n \) is globally Lipschitz continuous in the first argument, and \( x(t) \in \mathbb{R}^n \). We call \( f_{sw} \) with a fixed second argument \( p \in [N] \) a mode of the system. We define \( L_x = \max_{p \in [N]} L_p \), where \( L_p \) is the global Lipschitz constant with respect to the state of mode \( p \in [N] \). Observe that we do not need any constraint on the behavior of \( f_{sw} \) in different modes. In the case of globally-Lipschitz non-switched systems with inputs, we had the global Lipschitz constant \( f_{sw} \), which governs how the system reacts to changes in its input. Here, in Equation (32), we will use a function \( d(t) \) that plays a similar role in bounding the difference of behavior resulting from running \( f_{sw} \) in different modes.

The points of discontinuities in \( \sigma \) are called switches. The switching signal \( \sigma \) has a minimum dwell time \( T_d > 0 \) if at least \( T_d \) time units elapse between any two consecutive switches. We fix \( T_d \) throughout the section and denote the corresponding set of switching signals as \( \Sigma(T_d) \). The initial state \( x(0) \in K \), where \( K \) is a compact initial set of states in \( \mathbb{R}^n \). The state trajectory of system (30) with \( \sigma \in \Sigma(T_d) \) is denoted by \( \xi_{x_0,\sigma} : \mathbb{R}_{\geq 0} \to \mathbb{R}^n \), which exists globally for all \( t \geq 0 \), unique, and continuously depends on its initial state.

Moreover, we denote the set of all reachable states of system (30) under the family of switching signals \( \Sigma(T_d) \) as:

\[ \text{Reach}_{sw}(K, T_d, T) := \{ x \mid \exists x_0 \in K, \sigma \in \Sigma(T_d), t \in [0, T], \xi_{x_0,\sigma}(t) = x \}. \]

Based on our setup in [3], the entropy of the autonomous switched system (30) is defined as follows: First, three positive error parameters \( \varepsilon, \alpha, \tau \), and the time interval \( T > 0 \) are fixed. Second, we redefine the concept of approximating functions. A function \( z : \mathbb{R}_{\geq 0} \to \mathbb{R}^n \) is called a \( (T, \varepsilon, \alpha, \tau) \)-approximating function of \( \xi_{x_0,\sigma} \) if for all \( j \geq 0 \) and for all \( t \in [s_j, s_{j+1}) \):

\[ \|z(t) - \xi_{x_0,\sigma}(t)\| \leq \varepsilon \quad \text{if } t \in [s_j, s_{j+1} + \tau), \]

\[ \text{otherwise,} \]

(31)

where \( s_0 = 0, s_1, \ldots \) are the switches of \( \sigma \). The norm in inequality (31) can be arbitrary. Third, we define \( s_{sw}(T, \varepsilon, \alpha, \tau) \) to be the minimal cardinality of an approximating set of all trajectories starting from \( K \) and having a switching signal from \( \Sigma(T_d) \) over the time interval \([0, T]\).

**Definition 8.** The estimation entropy of the switched system (30) is:

\[ h_{sw}(\varepsilon, \alpha, \tau) := \limsup_{T \to \infty} \frac{1}{T} \log s_{sw}(T, \varepsilon, \alpha, \tau). \]

The following function \( d(t) \), which we denote as mode divergence, bounds the distance between any two trajectories of system (30) evolving according to two different modes while starting from any reachable state.

\[ d(t) := \max_{p_1, p_2 \in [N]} \sup_{x \in \text{Reach}_{sw}(K, T_d, \infty)} \int_0^t \| f_{sw}(\xi_{x_0,\sigma}(p_1), p_1) - f_{sw}(\xi_{x_0,\sigma}(p_2), p_2) \| \, ds. \] (32)

**Assumption 1.** The mode divergence \( d(T_d) \) is finite.

This assumption can be checked, for example, if the reachset \( \text{Reach}_{sw}(K, T_d, \infty) \) is compact and \( \| f_{sw}(x, p) \| \) is finite for any \( p \) and any finite \( x \). The reason is that \( \| f_{sw}(x_1, p_1) - f_{sw}(x_2, p_2) \| \) will be finite for any \( x \in \text{Reach}_{sw}(K, T_d, \infty) \).

That can be seen by invoking the triangular inequality to get:

\[ \| f_{sw}(x_1, p_1) - f_{sw}(x_2, p_2) \| \leq \| f_{sw}(x_1, p_1) \| + \| f_{sw}(x_2, p_2) \|, \]

which is finite for \( x_1 \) and \( x_2 \).

Now we restate Theorem 1 in [3] that presents an upper bound on the estimation entropy of Definition 8.

**Theorem 4** (Theorem 1 in [3]). The entropy of the switched system (30) has the following upper bound:

\[ h_{sw}(\varepsilon, \alpha, \tau) \leq \frac{(L_x + 10)\alpha n}{\ln 2} + \log N \frac{T_c}{T_c}, \] (33)

where \( T_c \) is the largest real number in \((0, \tau]\) such that:

\[ d(T_c) \leq \varepsilon (1 - e^{-\alpha(T_c-T_d)}). \] (34)
Remark 3. Following our previous argument from Section V-B for systems with inputs and setting \( \alpha = 0 \), results in the condition (34) on \( d(T_e) \) in Theorem 4 becoming \( d(T_e) = 0 \). In that case, there is no non-zero \( T_e \) that satisfies the condition in Equation (34) unless all the modes are identical. Hence, if \( \alpha = 0 \), then \( T_e = 0 \) and the entropy upper bound in Equation (33) becomes infinite, and thus useless. However, with a closer look at the proof of Theorem 1 in [3], this infinite bound is an artifact of the assumption in (34). In fact, the choices of \( \alpha \) and \( T_e \) in (34) and the bound (33) can be replaced by \( \hat{\alpha} \) and \( \hat{T}_e \) such that \( \hat{\alpha} \geq \alpha \) and \( d(\hat{T}_e) \leq 2(1 - e^{-\lambda(T_2 - T_1)}) \). Such a bound may be finite even in the case when \( \alpha = 0 \).

A. Finiteness of mode divergence \( d(t) \)

In this section, similar to Sections V-A and V-B, we present an example system with an infinite estimation entropy. The system we consider is a scalar switched linear system with an unbounded \( d(t) \), giving an intuition as to why Assumption 1 may be necessary. A similar result was derived by Berger and Jungers [33] to show the necessity of dwell time constraints for the finiteness of stabilization entropy of switched linear systems.

We first define separated sets for switched systems. Two trajectories \( \xi_{x,a,1} \) and \( \xi_{x,a,2} \) of switched system (30) are \( (T, \varepsilon, \alpha, \tau) \)-separated if there exists a \( t \in [0, T) \) such that the inequality (31) is violated when both: \( z \) in the LHS is replaced by \( \xi_{x,a,1} \) and \( \xi_{x,a,2} \), respectively, and when \( z \) in the RHS is replaced by \( \xi_{x,a,2} \) and \( \xi_{x,a,1} \), respectively. A set of trajectories \( Z^1 \) is \( (T, \varepsilon, \alpha, \tau) \)-separated if all pairs of trajectories in \( Z^1 \) are \( (T, \varepsilon, \alpha, \tau) \)-separated. The maximal cardinality of such a set is denoted by \( s_{\text{set}}(T, \varepsilon, \alpha, \tau) \).

Lemma 11. For all \( K, T_d, \varepsilon, \alpha, \tau_1, \tau_2, \) and \( T > 0 \),
\[
s_{\text{set}}(T, 2\varepsilon, 0, \tau_1) \leq s_{\text{set}}(T, \varepsilon, 0, \tau_2) \leq s_{\text{set}}(T, \varepsilon, \alpha, \tau_2). \tag{35}
\]

Proof. The first inequality follows from the same proof as Lemma 2 but by replacing the previous definitions of approximating functions and separated trajectories with those of this section. The second inequality follows from the fact that for any \( \alpha \geq 0 \), any \( (T, \varepsilon, \alpha, \tau) \)-approximating set is also a \( (T, \varepsilon, 0, \tau) \)-approximating one.

Fix arbitrary two constants \( a > b > 0 \), and consider the functions \( f_{sw}^1(x, 1) = ax \) and \( f_{sw}^1(x, 2) = bx \), and the resulting switched system (30) having them as its modes: \[
\dot{x}(t) = f_{sw}^1(x(t), \sigma(t)), \tag{36}
\]
where \( x : \mathbb{R}_{\geq 0} \to \{1, 2\} \) is a switching signal with some minimum dwell time \( T_d > 0. \)

Lemma 12. \( \forall t > 0, d(t) \) of system (36) is unbounded.

Proof. Observe that the state of the system \( x(t) \) at any time \( t \) lower bounded by \( |x_0|e^{bt} \) and upper bounded by \( |x_0|e^{at} \). Since \( x_0 \neq 0 \), \( Reach_{sw}(K, T_d, \infty) \) is unbounded. Moreover, the difference between the two modes is as follows: \( |\xi_{x,a}(t) - \xi_{x,b}(t)| = |e^{at} - e^{bt}| \cdot |x| \). Therefore, the integral in Equation (32) is \( \int_0^T |f_{sw}(\xi_{x,a}(s), a) - f_{sw}(\xi_{x,b}(s), b)| ds = |ae^{at} - be^{bt}| \cdot |x| = |be^{bt}x| |\frac{e^{(a-b)t}}{e^{bt}} - 1|, \) which diverges to \( \infty \) with diverging \( x. \)

Thus, when the \( \sup \) over the reachset is taken, the argument of the integral is infinite and \( d(t) \) is unbounded, for any \( t > 0 \). \( \Box \)

We first ignore the minimum dwell time constraint and construct a family of switching signals in a way similar to how we constructed the sets of input signals in Definition 5. However, in contrast with Definition 5, instead of having a sequence \( A \) to define the allowed increments in the input signals, the switching signals we consider here only switch between two fixed values \( a \) and \( b \) at the time instants in \( t_{\text{seq}} \). In a similar construction to that of the input signals in Section V-B, we consider the sequence \( t_{\text{seq}} \) of time instants of the form \( t_i = \sum_{j=1}^i \tau_j \) that are less than \( T \), where \( v_1 = 2\varepsilon/(|x_0|(|a-b|)) \) and \( v_{i+1} = v_i e^{-b\tau_i} \) for \( i \geq 1 \). We denote the resulting set of switching signals by \( \Sigma_{\text{d}} \). By a similar argument to Lemma 7, \( t_i \to \infty \) as \( i \to \infty \), and thus the input signals do not have Zeno behavior. We add the dwell time constraint back in Lemma 14.

The time sequence used here \( t_{\text{seq}} \) is similar to the time sequence \( t_{\text{seq}} \) we considered in Section V-B to prove that the entropy of system (18) is infinite if \( \alpha > 0 \). The first time instance in \( t_{\text{seq}} \) is equal to \( \sqrt{2\varepsilon/(a-b)|x_0|} \). Moreover, the time between time instants of \( t_{\text{seq}} \) is \( v_{i+1} = v_i e^{-a\tau_i/2} \) while that of \( t_{\text{seq}} \) is \( v_{i+1} = v_i e^{-b\tau_i} \). Consequently, we can follow the same proof of Lemma 6, to conclude that \( |t_{\text{seq}}| \geq 2 - b\varepsilon/(4b\varepsilon) (e^{b\tau} - 1) \). \( \Box \)

Lemma 13. Fix the constants \( T, \varepsilon, \tau, \) and \( \alpha > 0 \). Consider any nonzero initial state \( x_0 \in \mathbb{R} \) and the set of the switching signals \( \Sigma_{\text{d}} \). The resulting set of trajectories of system (36) form a \( (T, 2\varepsilon, 0, \tau) \)-separated set.

Proof. First, using a similar proof to that of Lemma 12, \( \xi_{x,a}(t_1) - \xi_{x,b}(v_1) = |e^{at} - e^{bt}| \cdot |x_0| \cdot (a-b) v_1 = 2\varepsilon \). For any switching signal \( \sigma \), the state at time \( t = t_{\text{seq}} \) is \( \xi_{x,a}(t) \geq x_0 e^{bt} \). Consider any two switching signals \( \sigma_1, \sigma_2 \in \Sigma_{\text{d}} \) with corresponding sequences \( s_{\sigma_1} \) and \( s_{\sigma_2} \) with the same prefix up till the \( i \)th entry and differ in the \((i+1)^{th}\) one. Finally, let \( x_i = \xi_{x,a}(t_i) \). Then,
\[
|\xi_{x,a}(t_i+1) - \xi_{x,a}(t_{i+1})| = |x_i| e^{a(t_i+1)} - e^{b(t_i+1)} > |x_i| (a-b) v_{i+1} \\
\geq |x_0| e^{bt_i} (a-b) v_{i+1} = |x_0| e^{bt_i} (a-b) v_i e^{-b\tau_i} = 2\varepsilon.
\]
Therefore, any two trajectories starting from \( x_0 \) that have switching signals corresponding to two different sequences are separated by more than \( 2\varepsilon \) for at least one time instant. Hence, such set of trajectories is \( (T, 2\varepsilon, 0, \tau) \)-separated.

Theorem 5. For any \( \varepsilon, \alpha, \) and \( \tau > 0 \), \( s_{\text{set}}(\varepsilon, \alpha, \tau) \) of switched system (36) is infinite.

Proof. This theorem follows from the same proof of Theorem 3 while replacing \( \alpha \) with \( b \); the cardinality of the separated set constructed in Lemma 13 is equal to \( 2^{t_{\text{seq}}(\varepsilon, \alpha, \tau)} \). Substituting the lower bound on \( t_{\text{seq}} \) in Equation (37), we get that the separated set cardinality is lower bounded by \( 2^{O(e^{2\varepsilon})} \). Hence, the cardinality of the maximum separated set \( s_{\text{set}}(T, 2\varepsilon, 0, \tau) \)
is lower bounded by $2^O(e^{bT})$. From Lemma 11, we get that the minimum cardinality of an approximating set $s_{\text{est}}(T, \varepsilon, \alpha, \tau)$ is lower bounded by $2^O(e^{bT})$. If we substitute this lower bound in Definition 8, we get $h_{\text{est}}(\varepsilon, \alpha, \tau) = \infty$.

Now, we show that even if the minimum dwell time constraint is assumed, the entropy of system (36) is still infinite.

**Lemma 14.** Even if the set of possible of switching signals of system (36) is restricted to have a minimum dwell time of $T_d$, its entropy is infinite.

**Proof.** We will need to modify $t_{seq_d}$ to satisfy the minimum dwell time constraint. We decompose the interval $[0, T]$ to intervals of the form $[jT_d, (j+1)T_d)$ for all integers $j < \lceil \frac{T}{T_d} \rceil$. Then, we define $t_{seq_d}'$ to be the same sequence of time instants of $t_{seq_d}$ in the intervals with odd $js$, i.e., $[T_d, 2T_d], [3T_d, 4T_d]$, etc., and no time instances in the intervals with even $js$, i.e., $[0, T_d], [2T_d, 3T_d]$, etc. Then, we consider the trajectories that result from switching signals that switch at most one time instant in $t_{seq_d}'$ in each $T_d$ interval with odd $js$, i.e., at most one instant in $[T_d, 2T_d]$, at most one instant in $[3T_d, 4T_d]$, etc. Then, we can observe that such switching signals satisfy the minimum dwell time constraint of $T_d$.

Moreover, the trajectories of system (36) resulting from such switching signals are $(T, 2\varepsilon, \alpha, \tau)$-separated, by the same proof of Lemma 13. Hence, a lower bound on the number of such signals would result in a lower bound on the maximum cardinality $s_{\text{est}}(T, 2\varepsilon, \alpha, \tau)$ of $(T, 2\varepsilon, \alpha, \tau)$-separated set, and consequently a lower bound on $h_{\text{est}}(\varepsilon, \alpha, \tau)$ of system (36).

Now, we will lower bound the number of time instants in $t_{seq_d}'$ in each $T_d$ interval in $[0, T]$ with odd index. The number of switching signals that satisfy the minimum dwell time constraint we defined above is lower bounded by the product of these numbers. To that end, fix an interval $[jT_d, (j+1)T_d)$, with $j$ being a positive odd integer. Let $t_k$ be the first time instant in $t_{seq_d}'$ in that interval. Recall that by definition, $t_{k+1} - t_k = v_{k+1} = v_{1} e^{-b\delta t_d} \leq v_1 e^{-\beta b T_d}$. Using the same analogy as in Equation (37), the number of time instants in $t_{seq_d}'$ in $[jT_d, (j+1)T_d)$ is lower bounded by $\left[\frac{2 - b v_{k+1}}{4 b v_{k+1} - 1} e^{b(T/T_d) - 1}\right]$, which is lower bounded by $\left[\frac{2 - b v_{k+1}}{4 b v_{k+1} - 1} e^{b(T/T_d) - 1}\right]$. If we take the product of such lower bounds for each odd $j$, the result is $O(e^{b T_d} + e^{b T_d} + \ldots + e^{b(T/T_d)}) = O(e^{b(T/T_d)})$. Therefore, $s_{\text{est}}(T, \varepsilon, \alpha, \tau)$ is still lower bounded by $O(e^{b(T/T_d)^2})$ and the entropy $h_{\text{est}}(\varepsilon, \alpha, \tau)$ is still infinite.

**B. Relating the upper bounds on entropy of autonomous switched systems and non-switched systems with inputs**

In the following corollary of Proposition 3, we present a novel alternative format of the upper bound on the entropy of non-switched systems with inputs that is similar to the entropy upper bound of autonomous switched systems in Theorem 4, the latter being a previous result from [3].

**Corollary 6.** Fix any $\varepsilon$ and $\rho > 0$, and let $\delta_x = e^{\varepsilon} (M_x + \rho) T_p$ and $\delta_u$ and $T_p$ be such that

$$g_{c,u}(\delta_u, T_p) \leq e^{\varepsilon(1 - e^{-\rho T_p})},$$

where $g_{c,u}$ is as defined in Equation (26). Then, $h_{\text{est}}(\varepsilon)$ of system (4) is upper bounded by:

$$\frac{(M_x + \rho)\eta}{\ln 2} + \log P T_p,$$

where $P = \left[\frac{2^{\varepsilon} + 1}{\delta_u}\right]^m$ represents the number of possible quantized inputs at each sampling time.

**Proof.** By substituting our choice of $\delta_x$ in $g_{c,u}$, we obtain $g_{c,u}(\delta_x, T_p) = e^{\varepsilon} e^{-\rho T_p}$. Hence, the condition on $g_{c,u}$ in Proposition 3 is satisfied. Now, by substituting our choice of $\delta_x$ in the bound in $g_{c,x}(\delta_x, \delta_u, T_p)$, along with replacing $\left[\frac{2^{\varepsilon} + 1}{\delta_u}\right]^m$ with $P$, we get the upper bound.

Note that $P$ in Corollary 6 represents the number of cells in the grid in the input space $\mathbb{R}^m$ that Algorithm 1 chooses from at each sampling time in the construction of approximating functions. The center of the chosen cell would be the input signal for a sampling period of time. In that sense, the quantized input values act as modes of a switched system whose trajectories are the approximation functions constructed by Algorithm 1. Hence, $P$ affects the bound on $h_{\text{est}}$ of non-switched systems with inputs similar to how $N$, the number of modes, affects the bound on $h_{\text{est}}$ of switched systems.

In Theorem 4, $\alpha$ represents the exponential decay in the uncertainty between mode switches. Since the switching signal has a minimum dwell time of $T_d$, the uncertainty decays enough that even after adding the uncertainty caused by the mode switch, the total uncertainty does not exceed $\varepsilon$. The parameter $\rho$ in Corollary 6 has a similar meaning.

**Lemma 15.** If $\mu = \eta = 0$, line 12 in Algorithm 1 is changed to $S_{z,i} \leftarrow B(z_{i-1}(T_p^{-}), \varepsilon e^{-p T_p}), \delta_{z,0} = \varepsilon$, and for all iterations $i \geq 1, \delta_z, i = \delta_z, i-1 - e^{-p T_p}$, then for all $t \geq 0$,

$$||z(t) - \xi_{z,i}(u(t))|| \leq e^{-p t}.$$  

**Proof.** If the input variation represented by $\eta$ and $\mu$ is zero, then $\delta_z$ can be chosen to be zero and $S_{z,i}$, to be equal to $u(0)$ for any $i \geq 0$. Such a choice of $\delta_z$ and $\delta_{z,i}$ would still satisfy Lemma 9. In addition, it would result in $g_{c,u}(\delta_u, T_p) = 0$ implying zero contribution of the input to the uncertainty in the state in Equation (27) and $g_{c,x}(\delta_x, T_p) = e^{\varepsilon} e^{-2p T_p}$, at each iteration $i \geq 0$ of Algorithm 1.

**Lemma 15** shows that Algorithm 1 is able to generate exponentially converging approximating functions in the absence of input variation. This is similar to the exponential decay of the error between switches in switched systems.

**C. Autonomous switched systems modeled as non-switched systems with inputs**

In this section, we model autonomous switched systems as non-switched systems with inputs. This modeling allows applying the upper bound on entropy $h_{\text{est}}(\varepsilon)$ of Proposition 3 to get a novel upper bound on entropy $h_{\text{est}}(\varepsilon, \alpha, \tau)$, with $\alpha = 0$, of the switched system (30). We compare the resulting bound with the bound of Theorem 4.
In our modeling, we embed the $N$ modes of the switched system in a continuous input space and define a new dynamics function $\hat{f}_{sw}$ that takes a piece-wise continuous input signal instead of a piece-wise constant one in its second argument. We first define the input space $U$ to be the standard simplex of dimension $N - 1$, i.e., $U := \{ u \in \mathbb{R}^{N-1} | |u|_{\infty} \leq 1 \}$ and $\mathcal{U}_{sw}$ be the set of all measurable functions mapping time to $U$. Hence, $\hat{f}_{sw} : \mathbb{R}^n \times U \rightarrow \mathbb{R}^n$.

Such a construction is standard in the literature analyzing switched systems (e.g., [34]). Observe that bounding the input set to be the standard simplex instead of the unbounded real space of dimension $N$, would not affect the fact that $g_u$ in Proposition 3 is an upper bound on $h_{sw}$

Now, let us define the RHS of the ODE modeling the dynamics of the new system with input: $\forall x \in \mathbb{R}^n, \forall u \in \mathcal{U}_{sw}$,

$$\hat{f}_{sw}(x(t), u(t)) := f_{sw}(x(t), 1) + \sum_{p \in [2:N]} g(x(t), p)u(t)[p-1],$$

(40)

where $g(x, p) := f_{sw}(x, p) - f_{sw}(x, 1)$. The solutions of system (40) coincide with the solutions of the differential inclusion $\dot{x} = co(f_{sw}(x(t), 1), \ldots, f_{sw}(x(t), N))$, where $co$ denotes the convex hull, by the Filippov’s Selection Lemma (e.g., Lemma 1 in [34]). This means that the trajectories of the constructed system (40) include those of the switched system (30). Thus, applying the bound in Proposition 3 on (40) may lead to a courser upper bound on entropy than that of the switched system (30) derived in Theorem 4. Yet, it is useful for showing the generality of our results on bounding the entropy of non-switched systems with inputs.

Observe that the Jacobian of $f_{sw}$ with respect to the state and input at a certain $x(t)$ and $u(t)$ can be written as follows:

$$J_u = \begin{bmatrix} g(x(t), 2), \ldots, g(x(t), N) \end{bmatrix},$$

$$J_x = J_x,1(x(t)) + \sum_{p \in [2:N]} u(t)[p-1](J_{x,p}(x(t)) - J_{x,1}(x(t))),$$

(41)

where $J_{x,p}$ is the Jacobian of $f_{sw}(x(t), p)$ with respect to $x$.

Given these $J_x$ and $J_u$, we can compute $G_x$ and $G_u$ in Lemma 8. Using a similar argument to Proposition 2, $G_x \leq n(L_x + \frac{1}{2}) \leq n(n \max_{p \in [N]} L_p + \frac{1}{2})$. Moreover, $G_u = \sqrt{m} \| J_u \| = \sqrt{\| J_u \| + \| J_u \|} = \max_{j \in [n]} \sum_{p \in [2:N]} (J_{u,j,p}) \leq \sum_{p \in [2:N]} \| g(x(t), p) \| = \sum_{p \in [2,N]} \| f_{sw}(x(t), p) - f_{sw}(x(t), 1) \|$, $d_{co}$ is an upper bound on $\| J_u \|$ for all $x \in Reach(K, U_{sw}, \infty)$. To use the bound on entropy for non-switched systems with inputs in Proposition 3, we have to represent our input set $\mathcal{U}_{sw}$ in the format of Definition 1. We do that by over-approximating $\mathcal{U}_{sw}$ by the set of input signals $\mathcal{U}_{sw}$ with affine-bounded pointwise variation constructed according to Definition 1 with parameters $\mu = 0, \eta = 1$, i.e., $\mathcal{U} = (0, 1)$, and initial set of inputs $U$ being the unit box $[0, 1]^{N-1}$. The following theorem is a direct application of Corollary 6 to system (40) after setting $\delta_u$ to be equal to $\eta$.

**Theorem 6.** The entropy $h_{sw}(\varepsilon)$ of system (40) is upper bounded by:

$$\frac{(n(n \max_{p \in [N]} L_p + 1/2) + \rho)n}{\ln 2} + \frac{N - 1}{T_p},$$

where $\rho$ and $T_p$ satisfy $(2M_N e^{M_T T_p})^2 T_p \leq \varepsilon^2 (1 - e^{-\rho T_p})$, i.e., $2d_{co} e^{(n(n \max_{p \in [N]} L_p + 1/2) + \rho)n} \leq \varepsilon^2 (1 - e^{-\rho T_p})$.

Note that there is always $\rho$ and $T_p$ that satisfy the inequality in Theorem 6 since although both sides increase with increasing $T_p$ and the LHS is zero for $T_p = 0$, the LHS is independent of $\rho$ and the RHS strictly increases with $\rho$. The bound is on the entropy of the differential inclusion which contains the trajectories of the switched system. It grows as $O(n^2), O(\max_{p \in [N]} L_p)$, and $O(N)$ compared to the $O(n), O(\max_{p \in [N]} L_p)$, and $O(\log N)$ of the bound in Theorem 4. For $T_p$ when compared to $T_e$ in Theorem 4, they have a comparable type of bound as we establish that $d(t) \leq t d_{co} e^{\max_{p \in [N]} |f_u|}$ in the longer version [31]. This shows that our results for non-switched systems with inputs generalize to wide set of applications, including differential inclusions and switched systems.

**IX. Conclusion**

We presented a notion of topological entropy as a lower bound on the bit rate needed to estimate the state of nonlinear dynamical systems with uncertain inputs with variation upper-bounded by an affine function of time. We considered several alternative definitions of entropy and showed that two of the candidates are not meaningful as they diverge to infinity. We computed an upper bound on entropy. We provided another form of the upper bound that matches the form of that of switched systems we presented in Theorem 1 in [3]. Moreover, we modeled switched systems as dynamical systems with uncertain inputs. Consequently, we were able to use the upper bound we computed to get an upper bound on entropy of switched systems. Finally, we showed that an assumption made in Theorem 1 in [3] on the difference between the modes of the switched system is needed for a meaningful entropy definition.

**Acknowledgment**

We thank Daniel Liberzon and the anonymous reviewers for providing detailed and insightful comments that enhanced the results and presentation of this paper.

Hussein Sibai is an Assistant Professor in the Computer Science and Engineering department at Washington University in St. Louis. He was a postdoctoral scholar at UC Berkeley. He obtained his Ph.D. in Electrical and Computer Engineering from the University of Illinois Urbana-Champaign (UIUC) in December 2021. He received his bachelor's degree in Computer and Communication Engineering from the American University of Beirut and a master's degree in Electrical and Computer Engineering from UIUC. His research interests are in formal methods, control theory, and machine learning. Hussein won the best poster award in HSCC 2018 and best paper nominations at HSCC 2017 and ATVA 2019. His work has been recognized by the Rambus fellowship, the Ernest A. Reid fellowship, the MAVIS Future Faculty fellowship, and the ACM SIGBED gold medal for the graduate student research competition in CPS Week 21.
REFERENCES


