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 nonlinear 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, Nonlinear 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 $\varepsilon > 0$.

Given two estimation error parameters $\varepsilon > 0$ and $\alpha \geq 0$, we first define a notion of topological entropy $h_{\text{est}}(\varepsilon, \alpha)$ 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 $\varepsilon e^{-\alpha t}$, for any $t \geq 0$. We suggest an alternative entropy notion $h^*_{\text{est}}(\varepsilon, \alpha)$, where we restrict the approximating functions to be trajectories of the system. We further consider a third notion $h^\dagger_{\text{est}}(\varepsilon, \alpha)$ that represents the rate of increase over time of the number of trajectories that pairwise violate the approximation error bound $\varepsilon e^{-\alpha t}$, at some $t \geq 0$. 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 $h_{\text{est}}(\varepsilon, \alpha)$, even for a simple linear one-dimensional system with piece-wise constant input (Section V-B). Moreover, we show that $h_{\text{est}}(\varepsilon, \alpha)$ of the same system grows to infinity as ε goes to zero (Section V-A). On the other hand, we show that $h_{\text{est}}(\varepsilon, 0)$ 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

Submitted on April 5th, 2021, resubmitted on November 13, 2022,
Hussein Sibai is with the Computer Science and Engineering department at Washington University in St. Louis and Sayan Mitra is with the Coordinated Science Laboratory at the University of Illinois at Urbana-Champaign. Emails: sibai@wustl.edu and mitras@illinois.edu.

takes the limit of ε 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 $h_{\text{est}}(\varepsilon)$ as a weaker notion of estimation entropy, which we name *ε-estimation entropy*, from there forward, as well as $h^*_{\text{est}}(\varepsilon)$ and $h^\dagger_{\text{est}}(\varepsilon)$ for the alternative notions, implicitly fixing α to zero and ε to a strictly-positive real number.

We show that there is no state estimation algorithm with a fixed bit rate smaller than $h_{\text{est}}(\varepsilon)$ (Section VI). While computing $h_{\text{est}}(\varepsilon)$ 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 ε 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 μ and η 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 $\frac{nL_x}{\ln 2}$, 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 α being equal to zero (Corollary 4).

Finally, we rewrite the upper bound on $h_{\text{est}}(\varepsilon)$ 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 σ 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 ∞-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 intercornnected [15], [22]. In [1], [2], Liberzon and Mitra 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 $v^\intercal$. 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 δ-ball–closed hypercube of radius δ–centered at v. For a hyperrectangle $S \subseteq \mathbb{R}^n$ and $\delta > 0$, $grid(S, \delta)$, is a collection of 2δ-separated points along axis parallel planes such that the δ-balls around these points cover S. In that case, we say that the grid is of size δ. Given $x \in \mathbb{R}^n$ and $C = grid(S, \delta)$, we define $quantize(x, C)$ 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 $S \subset \mathbb{R}^n$ by $diam(S) = \max_{x_1,x_2\in S} \|x_1 - x_2\|$. 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 $S_1 \oplus S_2 := \{x_1 + x_2 \mid x_1 \in S_1, x_2 \in S_2\}$. A continuous function $\gamma : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ belongs to class-$\mathcal{K}$ if it is strictly increasing and $\gamma(0) = 0$.

For all $t \in \mathbb{R}_{\geq 0}$, we define the right and left hand limits of a function $u : \mathbb{R}_{\geq 0} \to \mathbb{R}^m$ at t as follows:

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

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 $U \subset \mathbb{R}^m$, 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 $u : \mathbb{R}_{\geq 0} \to \mathbb{R}^m$ with affine-bounded pointwise variation, i.e.,*

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

*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 $u : \mathbb{R}_{\geq 0} \to \mathbb{R}^m$ with affine-bounded total variation, i.e.,*

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

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

$$\int_t^{d_1} \|\dot{u}(s)\| ds + \int_{d_k}^{t+\tau} \|\dot{u}(s)\| ds + \sum_{i=1}^{k-1} \int_{d_i}^{d_{i+1}} \|\dot{u}(s)\| ds$$
$$+ \|u(t+\tau) - u((t+\tau)^-)\| + \sum_{i=1}^{k} \|u(d_i) - u(d_i^-)\|, \tag{3}$$

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

Note that for any μ and $\eta \geq 0$, $\mathcal{U}^s(\mu, \eta) \subseteq \mathcal{U}^p(\mu, \eta)$. This can be shown as follows: fix any μ, η, t, and $\tau \geq 0$, and consider any $u \in \mathcal{U}^s(\mu, \eta)$, then $\|u(t+\tau) - u(t)\| = \| \int_t^{t+\tau} du\| \leq \int_t^{t+\tau} \|du\| \leq \mu\tau + \eta$. 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 μ and η 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, within an η bound. $\mathcal{U}^p(\mu, 0)$ is the class of globally Lipschitz continuous functions with Lipschitz constant μ. In general, knowing that u cannot deviate too much, can be expressed by setting μ and η to smaller values. Roughly, η restricts the maximum norm of a jump at any time instant and μ 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:

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

where $t \geq 0$ and $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$. 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 $\xi_{x_0,u} : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$. In the rest of this section, we fix a compact set of initial states $K \subset \mathbb{R}^n$ and the set of input functions $\mathcal{U}^p(\mu, \eta)$. The same results hold if the set of input signals was $\mathcal{U}^s(\mu, \eta)$ 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:

$$Reach(K, \mathcal{U}, T) := \{x \in \mathbb{R}^n \mid \exists x_0 \in K, u \in \mathcal{U}, t \in [0,T], \\ \xi_{x_0,u}(t) = x\}. \tag{5}$$

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

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

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:

$$J_x = \begin{bmatrix} 0 & 0 & -v\sin x_3 \\ 0 & 0 & v\cos x_3 \\ 0 & 0 & 0 \end{bmatrix} \text{ and } J_u = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}. \tag{7}$$

Therefore, the dynamics are globally Lipschitz with Lipschitz constants:

$$L_x = \|\|J_x\|\| \leq v \text{ and } L_u = \|\|J_u\|\| = 1. \tag{8}$$

A. Approximating functions and (ε, α)-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, ε, α)-*approximating* for the trajectory $\xi_{x_0,u}$ over the interval $[0, T]$, if

$$\|z(t) - \xi_{x_0,u}(t)\| \leq \varepsilon e^{-\alpha t}, \tag{9}$$

for all $t \in [0, T]$. We say that a set of functions $\mathcal{Z} := \{z \mid z : [0, T] \to \mathbb{R}^n\}$ is (T, ε, α)-*approximating* for system (4), if for every $x_0 \in K$ and $u \in \mathcal{U}^p$, there exists a (T, ε, α)-approximating function $z \in \mathcal{Z}$ 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 *est* corresponds to the word *estimation* and the letter s stands for the word *set*.

Definition 2. *The* (ε, α)-estimation entropy *of system (4) is defined as follows:*

$$h_{est}(\varepsilon, \alpha) := \limsup_{T\to\infty} \frac{1}{T} \log s_{est}(T, \varepsilon, \alpha). \tag{10}$$

The estimation entropy h_{est} represents the exponential growth rate over time of the number of distinguishable trajectories of the system. Hence, h_{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 (ε, α)-estimation entropy has the same value if the ∞-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 $\|.\|_b$, for all $x \in \mathbb{R}^n$, there exists two positive constants c_1 and c_2 such that $c_1\|x\|_b \leq \|x\|_a \leq c_2\|x\|_b$, s_{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, ε, α)-approximating function for any given trajectory of system (4). Then, $s_{est}(T, \varepsilon, \alpha)$ is also the minimum number of simulations needed to* approximate *the reachset $Reach(K, \mathcal{U}, T)$ of system (4) up to an $\varepsilon e^{-\alpha t}$ over-approximation error, for any $t \in [0, T]$. Formally, let $\mathcal{Z}$ be the (T, ε, α)-approximating set with cardinality $s_{est}(T, \varepsilon, \alpha)$ and define an over-approximation of the reachable set*

$$Reach_{\varepsilon,\alpha}(K, \mathcal{U}, T) := \{x + c \mid \exists x_0 \in K, u \in \mathcal{U}, t \in [0, T], \\ c \in B_{\varepsilon e^{-\alpha t}}, \xi_{x_0,u}(t) = x\}$$

and another over-approximation using $\mathcal{Z}$

$$R_{\varepsilon,\alpha} := \{z(t) + c \mid z \in \mathcal{Z}, t \in [0, T], c \in B_{\varepsilon e^{-\alpha t}}\},$$

where $B_{\varepsilon e^{-\alpha t}} := B(0, \varepsilon e^{-\alpha t})$. Then,

$$Reach(K, \mathcal{U}, T) \subseteq R_{\varepsilon,\alpha} \subseteq Reach_{2\varepsilon,\alpha}(K, \mathcal{U}, T).$$

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[1].

A. Approximating trajectories instead of functions

In this section, we modify the definition of entropy h_{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 $\mathcal{Z}^* \subseteq K \times \mathcal{U}$ is (T, ε, α)-*spanning*, if for every $x_1 \in K$ and $u_1 \in \mathcal{U}$, there exists a pair $(x_2, u_2) \in \mathcal{Z}^*$ such that the trajectory ξ_{x_2,u_2} is a (T, ε, α)-approximating function for the trajectory ξ_{x_1,u_1} over $[0, T]$. The minimum cardinality of a spanning set is denoted by $s^*_{\text{est}}(T, \varepsilon, \alpha)$. If we propagate this restriction to Definition 2, the resulting definition of entropy would be as follows:

Definition 3. *The spanning sets-based* (ε, α)-estimation entropy *of system (4) is defined as follows:*

$$h^*_{est}(\varepsilon, \alpha) := \limsup_{T\to\infty} \frac{1}{T} \log s^*_{est}(T, \varepsilon, \alpha). \tag{11}$$

The following inequality is an application of the fact that any (T, ε, α)-spanning set is a (T, ε, α)-approximating set:

$$s_{\text{est}}(T, \varepsilon, \alpha) \leq s^*_{\text{est}}(T, \varepsilon, \alpha). \tag{12}$$

B. Separated sets instead of approximating ones

In this section, we provide an entropy definition based on the concepts of (T, ε, α)-*separated* trajectories and (T, ε, α)-*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 ξ_{x_1,u_1} and ξ_{x_2,u_2} are (T, ε, α)-*separated* iff there exists $t \in [0, T]$ such that

$$\|\xi_{x_1,u_1}(t) - \xi_{x_2,u_2}(t)\| > \varepsilon e^{-\alpha t}. \tag{13}$$

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

Definition 4. *The separated sets-based* (ε, α)-estimation entropy *of system (4) is defined as follows:*

$$h^\dagger_{est}(\varepsilon, \alpha) = \limsup_{T\to\infty} \frac{1}{T} \log s^\dagger_{est}(T, \varepsilon, \alpha). \tag{14}$$

[1]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, $\mathcal{U}^p$, ε, α, and T,*

$$s^*_{est}(T,\varepsilon,\alpha) \le s^\dagger_{est}(T,\varepsilon,\alpha). \tag{15}$$

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

Lemma 2. *For all K, $\mathcal{U}^p$, ε, α, and T,*

$$s^\dagger_{est}(T,2\varepsilon,\alpha) \le s_{est}(T,\varepsilon,\alpha). \tag{16}$$

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

$$\begin{aligned}
&\|\xi_{x_1,u_1}(t) - \xi_{x_2,u_2}(t)\| \\
&\le \|z(t) - \xi_{x_1,u_1}(t)\| + \|z(t) - \xi_{x_2,u_2}(t)\| \\
&\qquad\qquad \text{[by triangular inequality]} \\
&\le \varepsilon + \varepsilon = 2\varepsilon.
\end{aligned}$$

Thus, the two trajectories are not $(T,2\varepsilon,\alpha)$-separated, contradicting the assumption that $\mathcal{Z}^\dagger$ 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, $\mathcal{U}^p$, ε, α, and T,*

$$h^\dagger_{est}(2\varepsilon,\alpha) \le h_{est}(\varepsilon,\alpha) \le h^*_{est}(\varepsilon,\alpha) \le h^\dagger_{est}(\varepsilon,\alpha). \tag{17}$$

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 (ε,α)-estimation entropy $h_{\text{est}}(\varepsilon,\alpha)$ of a simple dynamical system is infinite if ε approaches zero or if $\alpha > 0$. On the other hand, we show in Corollary 5 that $h_{\text{est}}(\varepsilon,0)$ 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][2].

Consider the simple dynamical system:

$$\dot{x}(t) = u(t), \tag{18}$$

where $t \ge 0$, 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.

[2]The results presented in this section are novel contributions of this paper.

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 $A = a_0,\dots,a_{l-1}$ and an increasing sequence of $l+1$ time points $tseq = t_0,\dots,t_l$, we define the set SE of sequences (SE for SEquences) of length l that are of the form: $se[0] = 0$ or $se[0] = a_0$ and for any $i > 0$, $se[i] = se[i-1]$ or $se[i] = se[i-1] + a_i$. 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 $t \in [t_i,t_{i+1})$,*

$$u_{tseq,se}(t) = se[i]. \tag{19}$$

We denote the set of all such piecewise-right-constant signals, with $tseq$ and A fixed, by $\mathcal{U}^A_{tseq} = \bigcup_{se \in SE} u_{tseq,se}$.

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

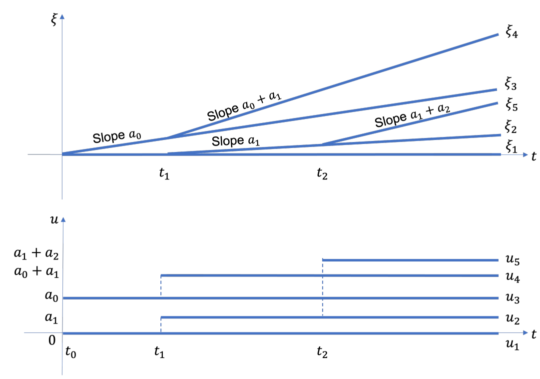

Fig. 1. Construction of an example set of input signals following Definition 5. Only five such signals are shown: u_1 that is equal to 0 at all times, u_2 that differs from u_1 by being equal to a_1 starting from t_1, u_3 that is equal to a_0 at all times, u_4 that differs from u_3 by being equal to $a_0 + a_1$ 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) (top).

A. Adding $\lim_{\varepsilon\to 0}$ to entropy definition

In this section, we will discuss the first variant of entropy definition: taking the limit as ε goes to zero. Such a variant is considered in [1] and represents the case where the allowed estimation error ε is arbitrarily small. We will show that the resulting notion of entropy is infinite for system (18) with inputs constructed using Definition 5.

Using Theorem 1, we get the following inequality:

$$\begin{aligned}
\lim_{\varepsilon\to 0} h_{\text{est}}(\varepsilon,\alpha) &\ge \lim_{\varepsilon\to 0} h^\dagger_{\text{est}}(2\varepsilon,\alpha) \\
&= \lim_{\varepsilon\to 0} \limsup_{T\to\infty} \frac{1}{T} \log s^\dagger_{\text{est}}(T,2\varepsilon,\alpha).
\end{aligned} \tag{20}$$

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 $O(2^{T/3\sqrt{\varepsilon}})$ which when substituted in (20) would evaluate to infinity.

To construct the separated set, for any fixed ε and T, we use the family of input signals $\mathcal{U}^A_{tseq}$, where $A = \sqrt{\varepsilon}, \ldots, \sqrt{\varepsilon}$, $tseq_{z\varepsilon} = \{0, \tau, 2\tau, \ldots \lfloor T/\tau \rfloor \tau\}$, and $\tau = 3\sqrt{\varepsilon}$. By such choices of A and τ, the total variation of any signal in $\mathcal{U}^A_{tseq}$ is upper-bounded by $\sqrt{\varepsilon}\lfloor T/\tau \rfloor$ which is equal to $\sqrt{\varepsilon}\lfloor T/(3\sqrt{\varepsilon})\rfloor = O(T)$. Therefore, there exist some constants μ and η, such that for any $\varepsilon > 0$ and $T > 0$, $\mathcal{U}^A_{tseq}$ is a subset of $\mathcal{U}^s(\mu, \eta)$, and consequently, a subset of $\mathcal{U}^p(\mu, \eta)$ as well.

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

Proof. Consider any two signals $u_1, u_2 \in \mathcal{U}_{tseq}$ with corresponding sequences se_1 and se_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 $se_1[j+1] \neq se_2[j+1]$. The resulting trajectories ξ_{x_0,u_1} and ξ_{x_0,u_2} 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: $\xi_{x,u}(t) = \xi_{x,u}(j\tau) + u(j\tau)(t - j\tau)$, for $t \in [j\tau, (j+1)\tau)$, since $u(t)$ is constant over that interval. Then, the distance $\|\xi_{\xi_{x,u_1}(j\tau), se_1[j+1]}(t - j\tau) - \xi_{\xi_{x,u_2}(j\tau), se_2[j+1]}(t - j\tau)\|$ between the two trajectories that start from $\xi_{x,u_1}(j\tau)$, or equivalently, $\xi_{x,u_2}(j\tau)$, and have $se_1[j+1]$ and $se_2[j+1]$, respectively, as input for τ time units is $a_{j+1}\tau$, which is equal to $\sqrt{\varepsilon}\tau$. By our choice of τ to be equal to $3\sqrt{\varepsilon}$, the distance is 3ε. Hence, the distance $\|\xi_{x_0,u_1}((j+1)\tau) - \xi_{x_0,u_2}((j+1)\tau)\|$ between ξ_{x_0,u_1} and ξ_{x_0,u_2} at $t = (j+1)\tau$ is equal to 3ε, strictly larger than 2ε. Since it is enough for two trajectories to have a distance larger than 2ε at a single point in time in the interval $[0, T]$ to be considered $(T, 2\varepsilon, 0)$-separated, ξ_1 and ξ_2 are $(T, 2\varepsilon, 0)$-separated. Hence, all the trajectories resulting from the initial state x_0 and $\mathcal{U}_{tseq}$ are $(T, 2\varepsilon, 0)$-separated. □

Theorem 2. $\lim_{\varepsilon \to 0} h_{est}(\varepsilon, \alpha)$ *of system (18) with the input set $\mathcal{U}^A_{tseq}$ is infinite.*

Proof. The number of trajectories in the separated set constructed in Lemma 3 is equal to $|\mathcal{U}^A_{tseq}|$ which is $2^{\lfloor T/\tau \rfloor}$. 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 $\alpha > 0$, $s^{\dagger}_{\text{est}}(T, 2\varepsilon, \alpha) \geq s^{\dagger}_{\text{est}}(T, 2\varepsilon, 0)$. Therefore, $\lim_{\varepsilon \to 0} h_{\text{est}}(\varepsilon, \alpha)$ is infinite by a simple substitution of $s^{\dagger}_{\text{est}}(T, 2\varepsilon, \alpha)$ with $2^{|tseq_{z\varepsilon}|} = 2^{\lfloor T/3\sqrt{\varepsilon} \rfloor}$, 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_{est} with ε 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 $\mathcal{U}$ and any trajectory $\xi_{x,u}$ of system (4) with an arbitrary $x \in \mathbb{R}^n$ and $u \in \mathcal{U}$. Assume that $\exists \tau > 0$ and a vector $a_0 \neq 0$ in $\mathbb{R}^n$, such that for any $c_1, c_2, \ldots$ in the interval $[0, 1]$, there exists a trajectory $\xi_{x',u'}$, that for any positive integer j, is equal to $\xi_{x,u}(j\tau) + c_j a_0$. Then, $\lim_{\varepsilon \to 0} h_{est}(\varepsilon, \alpha)$ 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 $(l\tau, \frac{\|a_0\|}{M}, 0)$-separated set, for any M and $l \geq 1$, as follows: First, choose $c_i = \frac{i-1}{M}$, for all $i \in [M]$. Then, consider the trajectories of system (4) which satisfy $\xi_{x,u'}(j\tau) = \xi_{x,u}(j\tau) + c_i a_0$, for $i \in [M]$ and $j \in [l]$. Such trajectories exist by the second assumption in the lemma. Each trajectory in this set is $(l\tau, \frac{\|a_0\|}{M}, 0)$-separated from the other trajectories in the set. The cardinality of the set is equal to M^l. Taking the limit of ε going to zero is equivalent to taking the limit of M going to infinity. By substituting $s^{\dagger}_{\text{est}}(T, 2\varepsilon, \alpha)$ by M^l and T by $l\tau$ 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 $\xi_{x,u}$ if two positive constants δ and τ exist such that for any positive t and any a and $b \in \mathbb{R}^n$ such that $\|\xi_{x,u}(t) - a\| \leq \delta \|\xi_{x,u}(t)\|$ and $\|\xi_{x,u}(t+\tau) - b\| \leq \delta \|\xi_{x,u}(t+\tau)\|$, another trajectory $\xi_{x',u'}$ of system (4) exists and satisfies $\xi_{x',u'}(t) = a$ and $\xi_{x',u'}(t+\tau) = b$. Extending Definition 2.3.14 in [27] to continuous time systems, a trajectory $\xi_{x,u}$ of system 4 is said to be *separated from the origin* if there exists a positive constant δ_0 such that $\|\xi_{x,u}(t)\| \geq \delta_0$, 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 α. 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 ε and $\alpha > 0$, $h_{est}(\varepsilon, \alpha)$ of system (18) with the input set $\cup_{\varepsilon > 0} \mathcal{U}^A_{tseq,\varepsilon}$ is infinite, where $\mathcal{U}^A_{tseq,\varepsilon}$ is the input set constructed for a specific ϵ in Lemma 3.*

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

$$\limsup_{T \to \infty} \frac{1}{T} \log \big(s_{\text{est}}(\tau, \varepsilon, \alpha; K) \\ s_{\text{est}}(T - \tau, \varepsilon e^{-\alpha\tau}, \alpha; \mathit{Reach}(K, \mathcal{U}, [\tau, \tau])) \big)$$

$$= \limsup_{T\to\infty} \frac{1}{T} \log s_{\text{est}}(T-\tau, \varepsilon e^{-\alpha\tau}, \alpha; Reach(K,\mathcal{U},[\tau,\tau]))$$
$$= h_{\text{est}}(\varepsilon e^{-\alpha\tau}, \alpha; Reach(K,\mathcal{U},[\tau,\tau])), \quad (21)$$

where $Reach(K,\mathcal{U},[\tau,\tau])$ represents the reachable states *exactly* at time τ. The inequality follows from the fact that to construct a $(T,\varepsilon,\alpha;K)$-approximating set, one can construct a $(\tau,\varepsilon,\alpha;K)$-approximating set and a $(T-\tau,\varepsilon e^{-\alpha\tau},\alpha;Reach(K,\mathcal{U},[\tau,\tau]))$-approximating one. Then, for each function in the former set, concatenate it with a function in the latter set.

Since $\lim_{\varepsilon\to 0} h_{\text{est}}(\varepsilon,\alpha;K)$ of system (18) is infinity for any K, $\lim_{\tau\to\infty} h_{\text{est}}(\varepsilon e^{-\alpha\tau},\alpha;Reach(K,\mathcal{U},[\tau,\tau]))$ is infinity as well. Therefore, $h_{\text{est}}(\varepsilon,\alpha;K)$ is infinite. □

Our second proof relies on designing a separated set, as in the previous section. Since we require the error to exponentially decrease with rate $\alpha > 0$ in this section, we can decrease the time intervals between switches in the input signals of Definition 5 as well as their values while still ensuring that the resulting set of trajectories is $(T,2\varepsilon,\alpha)$-separated. That way, we can pack more separated trajectories in the same time interval while satisfying the slow-variation constraint.

Specifically, we use the set of input signals $\mathcal{U}_{p\alpha}$ that results from Definition 5 with the sequence of time instants $tseq_{p\alpha}$ being of the form $t_0 = 0$ and $t_i = \sum_{j=1}^{i} v_j$, where $i \leq l$, l is the largest index where v_l is less than T, $v_1 = \sqrt{2\varepsilon}$, and $v_{i+1} = v_i e^{-\alpha v_i/2}$ (or equivalently $v_{i+1} = v_1 e^{-\frac{\alpha}{2}\sum_{j=1}^{i} v_j}$). Observe that for all i, $v_{i+1} < v_i$, and v_i is the time interval between t_{i-1} and t_i. We choose the sequence A in the definition of $\mathcal{U}_{p\alpha}$ to be of the form $v_1,\dots,v_l$, which are the same values used to define the switches above.

Lemma 5. *Fix T, ε, and $\alpha > 0$. Consider any initial state $x_0 \in \mathbb{R}$ and the set of input signals $\mathcal{U}_{p\alpha}$. The resulting trajectories of system (18) form a $(T,2\varepsilon,\alpha)$-separated set.*

Proof. First, consider any two trajectories having initial state x_0 and input signals that are equal to v_1 and 0 for the first v_1 time units, respectively. Then, $|\xi_{x_0,v_1}(v_1) - \xi_{x_0,0}(v_1)| = (v_1 - 0)v_1 = 2\varepsilon > 2\varepsilon e^{-\alpha v_1}$, where the last inequality follows from the fact that $v_1 > 0$. Now, consider any two signals $u_1, u_2 \in \mathcal{U}_{p\alpha}$ with corresponding sequences se_1 and se_2 with the same prefix up till the j^{th} value for some $j \in [l]$, and with different ones in the $(j+1)^{th}$ index, i.e., $se_1[j+1] \neq se_2[j+1]$. The resulting trajectories ξ_{x_0,u_1} and ξ_{x_0,u_2} from these input signals and the common initial state x_0 are $(T,2\varepsilon,\alpha)$-separated. To see this, observe that for any $i \geq 1$, and any $x_0 \in \mathbb{R}$, $|\xi_{x_0,u_1}(t_i) - \xi_{x_0,u_2}(t_i)| = |\xi_{x,a_i+u_1(t_{i-1})}(v_i) - \xi_{x,u_1(t_{i-1})}(v_i)| = |\xi_{x,v_i+u_1(t_{i-1})}(v_i) - \xi_{x,u_1(t_{i-1})}(v_i)| = |x+(v_i+u_1(t_{i-1}))v_i - x - u_1(t_{i-1})v_i| = v_i^2 = 2\varepsilon e^{-\alpha\sum_{j=1}^{i-1} v_j} > 2\varepsilon e^{-\alpha\sum_{j=1}^{i} v_j}$, where $x = \xi_{x_0,u_1}(t_{i-1})$. Hence, any two trajectories resulting from two input signals corresponding to two different strings se_1 and $se_2 \in SE$ are $(T,2\varepsilon,\alpha)$-separated. Therefore, this set of trajectories is $(T,2\varepsilon,\alpha)$-separated set. □

Now that we have proven that the set of trajectories with input signals from $\mathcal{U}_{p\alpha}$ is a $(T,2\varepsilon,\alpha)$-separated set, we compute a lower bound on $|tseq_{p\alpha}|$. This will determine a 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 $tseq_{p\alpha}$ is upper-bounded as follows: $t_i = \sum_{j=1}^{i} v_j \leq \frac{2}{\alpha}\ln(\frac{2\alpha\sqrt{2\varepsilon}i}{2-\alpha\sqrt{2\varepsilon}} + 1)$. Moreover, the cardinality of $tseq_{p\alpha}$ over the time horizon $[0,T]$ is lower-bounded as follows: $|tseq_{p\alpha}| \geq \lfloor \frac{\sqrt{2\varepsilon}-\alpha\varepsilon}{2\alpha\varepsilon}(e^{\alpha T} - 1) \rfloor$.*

Proof. To upper bound the sum of v_i's, we upper bound the sequence $\{v_i\}$ with another sequence $\{w_i\}$, whose sum is easier to compute. Recall that $v_1 = \sqrt{2\varepsilon}$ and $v_{i+1} = v_i e^{-\alpha v_i/2}$ for $i \geq 1$. Since $\alpha v_i > 0$, $v_{i+1} < \frac{v_i}{1+\alpha v_i/2}$. Now, let us define $w_1 = v_1$ and $w_{i+1} = \frac{w_i}{1+\alpha w_i/2}$, 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 $g(y) = \frac{y}{1+\alpha y/2}$ for $y \in \mathbb{R}$. Then, $\frac{d}{dy}g(y) = \frac{1}{(1+\alpha y/2)^2} > 0$ and thus is an increasing function. Hence, assuming that $v_i \leq w_i$, we get that $v_{i+1} < \frac{w_i}{1+\alpha w_i/2} = w_{i+1}$, since we know that $v_{i+1} < \frac{v_i}{1+\alpha v_i/2}$. Hence, by induction, $v_i \leq w_i$ for all $i \geq 1$.

Claim: $w_i = \frac{2}{\alpha(i+c)}$, where $c = \frac{2-\alpha w_1}{\alpha w_1}$, for all $i \geq 1$.

Proof of claim: We again proceed by induction: for $i = 1$, $\frac{2}{\alpha(1+c)} = w_1$. Assume that $w_i = \frac{2}{\alpha(i+c)}$, then $w_{i+1} = \frac{\frac{2}{\alpha(i+c)}}{\frac{2(\alpha/2)}{\alpha(i+c)}+1} = \frac{2}{\alpha((i+1)+c)}$. Hence, $\sum_{j=1}^{i} w_j = \frac{2}{\alpha}\sum_{j=1}^{i}\frac{1}{j+c}$.

Now we proceed to prove the main lemma. First, note that $c > -1$ since w_1 and α are positive. By convexity of $\frac{1}{x}$, $\frac{1}{x} < \int_{x-1/2}^{x+1/2}\frac{ds}{s}$ for $x \geq 1/2$. Thus,

$$\frac{1}{j+c} < \int_{j-1/2}^{j+1/2}\frac{dy}{y+c}, \text{for } j \geq 1/2 - c.$$

Hence, the time of the i^{th} switch can be bounded as follows:

$$\sum_{j=1}^{i}\frac{1}{j+c} < \sum_{j=1}^{i}\int_{j-1/2}^{j+1/2}\frac{dy}{y+c} = \int_{1/2}^{i+1/2}\frac{dy}{y+c}$$
$$= \ln\frac{i+c+1/2}{c+1/2} = \ln(\frac{i}{c+1/2}+1)$$
$$= \ln(\frac{2\alpha w_1 i}{2-\alpha w_1/2}+1) = \ln(\frac{4\alpha\sqrt{2\varepsilon}i}{4-\alpha\sqrt{2\varepsilon}}+1).$$

Therefore, $\sum_{j=1}^{i} v_j \leq \sum_{j=1}^{i} w_j \leq \frac{2}{\alpha}\ln(\frac{4\alpha\sqrt{2\varepsilon}i}{4-\alpha\sqrt{2\varepsilon}}+1)$. .

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 $\lfloor\frac{4-\alpha\sqrt{2\varepsilon}}{4\alpha\sqrt{2\varepsilon}}(e^{\alpha T/2}-1)\rfloor$. □

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 $\mathcal{U}_{p\alpha}$ are not Zeno.*

Proof. The proof is by contradiction. By definition, $t_{i+1} = t_i + v_{i+1} = t_i + v_1 e^{-\alpha t_i/2}$. Assume that $\lim_{i\to\infty} t_i = c$, for some constant $c \in \mathbb{R}_{\geq 0}$. Then, $c = c + v_1 e^{-\alpha c/2}$, a contradiction. □

Corollary 2. *The total variation of any signal in $\mathcal{U}_{p\alpha}$ with number of switches that is $O(e^{\alpha T})$ is $O(T)$.*

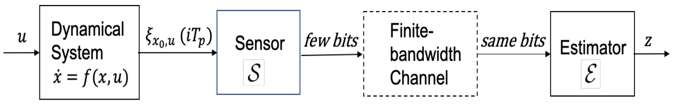

Fig. 2. The flow of information from the system to the sensor to the estimator. For any $t \geq 0$, $\|z(t) - \xi_{x_0,u}(t)\| \leq \varepsilon$.

Proof. The total variation of any signal in $\mathcal{U}_{p\alpha}$ is upper bounded by $\sum_{i=0}^{N-1} a_i = \sum_{i=1}^{N} v_i$, where N is the number of switches in the signal. Thus, we get that the total variation is upper bounded by the same bound on t_i in Lemma 6, i.e., $\frac{2}{\alpha}\ln(\frac{2\alpha\sqrt{2\varepsilon}N}{2-\alpha\sqrt{2\varepsilon}} + 1)$. Thus, if we only consider signals in $\mathcal{U}_{p\alpha}$ with number of switches N that are $O(e^{\alpha T})$, the total variation would be $O(T)$. Therefore, there exist constants μ and η such that those signals belong to $\mathcal{U}^s(\mu,\eta)$ and $\mathcal{U}^p(\mu,\eta)$. $\square$

Theorem 3. *For any ε and $\alpha > 0$, $h_{est}(\varepsilon,\alpha)$ of system (18) with the input set $\mathcal{U}_{p\alpha}$, even when restricted to signals with $O(e^{\alpha T})$ switches and thus having linearly-bounded total variation, is infinite.*

Proof. First, there are $2^{|tseq_{p\alpha}|}$ different sequences with values in $\{a,b\}$. Following the construction of Lemma 5, we can conclude that there is the same number of input signals and $(T, 2\varepsilon, \alpha)$-separated trajectories over the interval $[0,T]$.

By substituting $O(e^{\alpha T})$ in the analysis above, we get that $s^{\dagger}_{\text{est}}(T, 2\varepsilon, \alpha) \geq 2^{O(e^{\alpha T})}$ and hence by Lemma 2, $s_{\text{est}}(T,\varepsilon,\alpha) \geq 2^{O(e^{\alpha T})}$. Substituting this lower bound in (10) results in $h_{\text{est}}(\varepsilon,\alpha) = \infty$. $\square$

We conclude from this section that requiring $\alpha > 0$ or ε approaching zero is not suitable for systems with inputs. Therefore, we assume that $\alpha = 0$ and write $h_{\text{est}}(\varepsilon, 0)$ to be $h_{\text{est}}(\varepsilon)$ and denote the resulting notion by ε-estimation entropy, for conciseness. We drop α as well from all definitions of approximating, spanning, and separated sets, and $h^{\dagger}_{\text{est}}$ and h^{*}_{est}.

VI. Relation between ε-Estimation Entropy and the Bit Rate of Estimation Algorithms

In this section, we show that for any $\varepsilon > 0$, there is no state estimation algorithm for system (4) that can guarantee a maximum of ε estimation error when receiving state estimates at a fixed bit rate smaller than the ε-estimation entropy $h_{\text{est}}(\varepsilon)$ of the system. First, let us define state estimation algorithms.

Definition 6. *A state estimation algorithm for system (4) with a fixed bit rate is a pair of functions $(\mathcal{S},\mathcal{E})$, where $\mathcal{S} : \mathbb{R}^n \times Q_s \to \Gamma \times Q_s$ represents the sensor, $\mathcal{E} : \Gamma \times Q_e \to ([0,T_p] \to \mathbb{R}^n) \times Q_e$ represents the estimator, T_p is the sampling time, Γ is an alphabet, and Q_s and Q_e are the sets of internal states of the sensor $\mathcal{S}$ and estimator $\mathcal{E}$, respectively.*

Given ε and $T_p > 0$, sensor $\mathcal{S}$ samples the state of the system each T_p time units, starting from $t = 0$. For any $i \geq 0$, at $t = iT_p$, $\mathcal{S}$ sends to $\mathcal{E}$ a symbol from Γ representing an approximation of the state $\xi_{x_0,u}(iT_p)$. Finally, $\mathcal{E}$ maps the received symbol to a (T_p,ε)-approximating function of the trajectory for the time interval $[iT_p,(i+1)T_p]$ (see Figure 2).

Now, let us define the bit rate of an estimation algorithm:

$$b_r(\varepsilon) := \limsup_{T\to\infty} \frac{1}{T} \sum_{i=0}^{\lfloor T/T_p \rfloor} \log|\Gamma| = \frac{\log|\Gamma|}{T_p}.$$

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 $\varepsilon > 0$, there is no fixed bit rate state estimation algorithm for system (4) with a bit rate smaller than $h_{est}(\varepsilon)$ that for any $T > 0$, generates a (T,ε)-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 $b_r(\varepsilon)$ less than $h_{\text{est}}(\varepsilon)$, then for any time bound T, the set of functions z (see Figure 2) that can be constructed by that algorithm is a (T,ε)-approximation set with a smaller cardinality than $s_{\text{est}}(T,\varepsilon)$, 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 $h_{\text{est}}(\varepsilon)$ and constructs (T,ε)-approximating functions for input trajectories. Recall that $h_{\text{est}}(\varepsilon) = \limsup_{T\to\infty} {}^1/_T \log s_{\text{est}}(T,\varepsilon)$. Then,

$$b_r(\varepsilon) < h_{\text{est}}(\varepsilon),$$
$$\limsup_{T\to\infty} \frac{1}{T} \sum_{i=0}^{\lfloor T/T_p \rfloor} \log|\Gamma| < \limsup_{T\to\infty} \frac{1}{T} \log s_{\text{est}}(T,\varepsilon), \text{ and}$$
$$\limsup_{T\to\infty} \frac{(\lfloor T/T_p \rfloor + 1)\log|\Gamma|}{T} < \limsup_{T\to\infty} \frac{1}{T} \log s_{\text{est}}(T,\varepsilon).$$

Hence, for a sufficiently large T', we should have $\frac{(l+1)\log|\Gamma|}{T'} < \frac{1}{T'} \log s_{\text{est}}(T',\varepsilon)$, where $l = \lfloor T'/T_p \rfloor$. Consequently, we get the inequality $|\Gamma|^{l+1} < s_{\text{est}}(T',\varepsilon)$. 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',ε)-approximating function over the interval $[0,T']$. This is true since the estimator should be able to construct a (T',ε)-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',ε)-approximating set. However, $s_{\text{est}}(T',\varepsilon)$ is the minimal cardinality of such a set. Therefore, the set of functions that can be constructed by the algorithm is a (T',ε)-approximating set with a cardinality smaller than s_{est}, the supposed minimal one, which is a contradiction. $\square$

VII. Entropy Upper Bound and Algorithm

In this section, we derive an upper bound on the entropy $h_{\text{est}}(\varepsilon)$ 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 $V : \mathcal{X}^2 \to \mathbb{R}_{\geq 0}$ 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-$\mathcal{K}$ functions $\bar{\alpha}, \underline{\alpha}$ such that for any $x, x' \in \mathcal{X}$, $\underline{\alpha}(\|x - x'\|) \leq V(x, x') \leq \bar{\alpha}(\|x - x'\|)$, and

(ii) there exists a class-$\mathcal{K}$ function in the first argument and continuous in the second argument $\beta : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ and a class-$\mathcal{K}$ function $\gamma : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ such that for any $x_0, x_0' \in \mathcal{X}$ and $u, u' \in \mathcal{U}$, if $\xi_{x_0,u}(t)$ and $\xi_{x_0',u'}(t) \in \mathcal{X}$ for all $t \in [t_0, t_1]$, then for any such t,

$$V(\xi_{x_0',u'}(t), \xi_{x_0,u}(t)) \leq \beta(\|x_0 - x_0'\|, t - t_0) + \int_{t_0}^{t} \gamma(\|u(s) - u'(s)\|)ds. \quad (22)$$

The local discrepancy function V together with β and γ give the sensitivity of the trajectories of the system to changes in the initial state and the input. The functions $\bar{\alpha}, \underline{\alpha}, \beta, \gamma$ 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 2-norm of the difference between the two trajectories. In this paper, we correct both errors by adding a $\sup_{\tau \in [t_0,t_1]}$ in the definition of γ to correct the first error and multiply the definition of G_x by n and that of G_u by $\sqrt{m}$ to correct the second one. 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 $[t_0, t_1] \subseteq \mathbb{R}_{\geq 0}$, and piecewise-right-continuous inputs $\mathcal{U}$ with*

$$\beta(y, t - t_0) := e^{2G_x(t-t_0)} y^2 \textit{ and}$$
$$\gamma(y) := G_u^2 (\sup_{\tau \in [t_0,t_1]} e^{2G_x(t_1-\tau)}) y^2, \textit{where } t \in [t_0, t_1],$$

$$G_x := {}^n\!/_2 \Big(\sup_{\substack{t \in [t_0,t_1] \\ u \in \mathcal{U}, x \in \mathcal{X}}} 2\lambda_{max}\big(\frac{J_x(x, u(t)) + J_x(x, u(t))^T}{2}\big) + 1 \Big),$$
$$G_u := \sqrt{m} \sup_{\substack{t \in [t_0,t_1] \\ u \in \mathcal{U}, x \in \mathcal{X}}} |\!|\!| J_u(x, u(t)) |\!|\!|. \quad (23)$$

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 $G_u \leq m\sqrt{m}L_u$, 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 $\tau > 0$, $t \in [0, \tau]$, $x_0, x_0' \in \mathbb{R}^n$, and $u, u' \in \mathcal{U}(\mu, \eta)$, the squared distance $\|\xi_{x_0,u}(t) - \xi_{x_0',u'}(t)\|^2$, is upper bounded by:

$$e^{2M_x t}\|x_0 - x_0'\|^2 + M_u^2 e^{2M_x^+ t} \int_0^t \|u(s) - u'(s)\|^2 ds, \quad (24)$$

where $M_x^+ = \max\{M_x, 0\}$.

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

$$\frac{J_x + J_x^T}{2} = \begin{bmatrix} 0 & 0 & -\frac{v}{2}\sin x_3 \\ 0 & 0 & \frac{v}{2}\cos x_3 \\ -\frac{v}{2}\sin x_3 & \frac{v}{2}\cos x_3 & 0 \end{bmatrix}, \quad (25)$$

and has eigenvalues 0 and $\pm\frac{v}{2}$. Hence, $G_x = {}^n\!/_2({}^{2v}\!/_2 + 1) = \frac{3(v+1)}{2}$. Moreover, $G_u = \sqrt{m}|\!|\!|J_u|\!|\!| = 1$. If we use Proposition 2 to upper bound G_x and G_u, we get $M_x = n(nL_x + {}^1\!/_2) = 3(3v + {}^1\!/_2)$ 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 $x_0 \in K$ and an input signal $u \in \mathcal{U}^p(\mu, \eta)$, constructs a (T, ε)-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 $x_0 \in K$ and $u \in \mathcal{U}^p(\mu, \eta)$ is a (T, ε)-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 δ_x and $\delta_u > 0$. The procedure also uses the set of initial states K, the set of initial inputs U, a particular initial state $x_0 \in K$, and an input signal $u \in \mathcal{U}^p(\mu, \eta)$ for system (4). The output is a piece-wise continuous function $z : [0, T] \to \mathbb{R}^n$ that is constructed iteratively over each $[iT_p, (i+1)T_p)$

Algorithm 1 Construction of a (T,ε)-approximating function.

1: **input:** T, T_p, δ_x, δ_u
2: $S_{x,0} \leftarrow K$; $S_{u,0} \leftarrow U$;
3: $C_{x,0} \leftarrow$ $grid(S_{x,0}, \delta_x)$;
4: $C_{u,0} \leftarrow$ $grid(S_{u,0}, \delta_u)$;
5: $i \leftarrow 0$;
6: **while** $i \leq \lfloor \frac{T}{T_p} \rfloor$ **do**
7: $x_i \leftarrow \xi_{x_0,u}(iT_p)$; $u_i \leftarrow u(iT_p)$;
8: $q_{x,i} \leftarrow quantize(x_i, C_{x,i})$;
9: $q_{u,i} \leftarrow quantize(v_i, C_{u,i})$;
10: $z_i \leftarrow \xi_{q_{x,i},q_{u,i}}$;
11: $i \leftarrow i+1$; {parameters for next iteration}
12: $S_{x,i} \leftarrow B(z_{i-1}({T_p}^-), \varepsilon)$;
13: $S_{u,i} \leftarrow B(q_{u,i-1}, \eta + \mu T_p + \delta_u)$;
14: $C_{x,i} \leftarrow grid(S_{x,i}, \delta_x)$;
15: $C_{u,i} \leftarrow grid(S_{u,i}, \delta_u)$;
16: wait(T_p);
17: **end while**
18: **output:** $\{z_i([0,T_p)) : i \in [0; \lfloor \frac{T}{T_p} \rfloor - 1]\} \cup \{z_{\lfloor \frac{T}{T_p} \rfloor}([0,T_p])\}$

interval for $i \in [0; \lfloor \frac{T}{T_p} \rfloor]$. Later we will infer several constraints on the parameters such that the output z is indeed a (T,ε)-approximating function for the given trajectory $\xi_{x_0,u}$.

Initially, $S_{x,0}$ is assigned to the initial set K. Similarly, $S_{u,0}$ is assigned to the initial input set U. $C_{x,0}$ is a grid of size δ_x over K and $C_{u,0}$ is a grid of size δ_u over U. At the i^{th} iteration, $i \in [0; \lfloor \frac{T}{T_p} \rfloor]$, x_i and u_i store the values $\xi_{x_0,u}(iT_p)$ and $u(iT_p)$, the state and input at the time iT_p, respectively. Then, $q_{x,i}$ and $q_{u,i}$ are set to be the quantization of x_i and u_i with respect to $C_{x,i}$ and $C_{u,i}$, respectively. With slight abuse of notation, we will also denote the function of time that maps the interval $[0, T_p)$ to $q_{u,i}$ by $q_{u,i}$, as in line 10, for example. The variable z_i stores the trajectory that results from running system (4) starting from initial state $q_{x,i}$, with input signal $q_{u,i}$, and running for T_p time units. After that, i is incremented by 1 and the next iteration variables $S_{x,i}$, $S_{u,i}$, $C_{x,i}$, and $C_{u,i}$ are initialized. Finally, the procedure outputs the concatenation of the z_i's, for all $i \in [0; \lfloor \frac{T}{T_p} \rfloor]$ that is denoted by the function $z : [0,T] \rightarrow \mathbb{R}^n$. Note that for $i = \lfloor \frac{T}{T_p} \rfloor$, we consider z_i to have a domain that is closed from the right as well.

In the following lemma, we show that if the parameters of the procedure T_p, δ_x and δ_u, are small enough, then the output is a (T,ε)-approximating function for $\xi_{x_0,u}$.

Before stating the lemma, let us define the following function

$$g_c(\delta_x, \delta_u, T_p) := g_{c,x}(\delta_x, T_p) + g_{c,u}(\delta_u, T_p), \tag{26}$$

where $g_{c,x}(\delta_x, T_p) := \delta_x^2 e^{2M_x T_p}$ and $g_{c,u}(\delta_u, T_p) := M_u^2 e^{2M_x^+ T_p}\big(\frac{1}{3}\mu^2 {T_p}^3 + (\delta_u + \eta)\mu {T_p}^2 + (\delta_u+\eta)^2 T_p\big)$. The subscript c is added since the functions would be used to specify the *constraints* that the parameters should satisfy for the output of Algorithm 1 to be a (T,ε)-approximating function.

Since g_c is continuous in T_p, δ_x, and δ_u, non-negative, , and is equal to zero when $\delta_x = T_p = 0$, then for any $\varepsilon > 0$, there exist T_p, δ_x, and $\delta_u > 0$ such that $g_c(\delta_x, \delta_u, T_p) \leq \varepsilon^2$.

Lemma 9. *Given $\varepsilon > 0$, fix δ_x, δ_u, and T_p such that $g_c(\delta_x, \delta_u, T_p) \leq \varepsilon^2$. For any $x_0 \in K$ and $u \in \mathcal{U}^p(\mu,\eta)$, for all $i \in [0; \lfloor \frac{T}{T_p} \rfloor]$, and 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) $\|z_i(t - iT_p) - \xi_{x_i,u_i}(t - iT_p)\| \leq \varepsilon$,

where $u_i(t) := u(iT_p + t)$, 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 $q_{u,0}$ is its quantization of u_0 with respect to a grid of resolution δ_u. Moreover, $\|u(t) - u(0)\| \leq \mu t + \eta$ by Definition 1. But, by triangle inequality, $\|u(t) - q_{u,0}\| \leq \|u(t) - u_0\| + \|u_0 - q_{u,0}\| \leq \mu t + \eta + \delta_u$. 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 $\|u_i(t - iT_p) - q_{u,i}\| \leq \mu(t - iT_p) + \eta + \delta_u$. Thus, $u_i \in S_{u,i}$ for all i. Second, $x_0 \in S_{x,0}$ since $x_0 \in K$. Also, $\|z_0(0) - \xi_{x_0,u_0}(0)\| = \|q_{x,0} - x_0\| \leq \delta_x \leq \varepsilon$, 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 $i = \lfloor \frac{t}{T_p} \rfloor$. Then,

$$\begin{aligned}
&\|\xi_{x_i,u_i}(t - iT_p) - \xi_{q_{x,i},q_{u,i}}(t - iT_p)\|^2 \\
&\leq \|x_i - q_{x,i}\|^2 e^{2M_x(t - iT_p)} \\
&\qquad + M_u^2 e^{2M_x^+(t - iT_p)} \int_{iT_p}^{t} \|u(s) - q_{u,i}\|^2 ds \\
&\qquad\qquad \text{[by (24)]} \\
&\leq \delta_x^2 e^{2M_x(t - iT_p)} \\
&\qquad + M_u^2 e^{2M_x^+(t - iT_p)} \int_{iT_p}^{t} \big(\mu(s - iT_p) + \eta + \delta_u\big)^2 ds, \\
&\leq \delta_x^2 e^{2M_x T_p} + M_u^2 e^{2M_x^+ T_p}\big(\frac{1}{3}\mu^2 {T_p}^3 + (\delta_u + \eta)\mu {T_p}^2 \\
&\qquad + (\delta_u + \eta)^2 T_p\big) = g_c(\delta_x, \delta_u, T_p) \leq \varepsilon^2,
\end{aligned} \tag{27}$$

where the last inequality follows from the assumption in the lemma statement. Therefore, for all $i \in [0; \lfloor \frac{T}{T_p} \rfloor]$ and $t \in [0,T]$, $x_i \in B(z_{i-1}(T_p), \varepsilon) = S_{x,i}$. □

Corollary 3. *For any δ_x, δ_u, 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,ε)-approximating function to the input trajectory $\xi_{x_0,u}$, i.e.*

$$\|z(t) - \xi_{x_0,u}(t)\| \leq \varepsilon. \tag{28}$$

One can conclude from Corollary 3 that the set of all functions that can be constructed by Algorithm 1 for any input trajectory $\xi_{x_0,u}$, where $x_0 \in K$ and $u \in \mathcal{U}^p(\mu,\eta)$, is a (T,ε)-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 δ_x, δ_u 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:*

$$\lceil \frac{diam(K)}{2\delta_x} \rceil^n \lceil \frac{diam(U)}{2\delta_u} \rceil^m \Big(\lceil \frac{\varepsilon}{\delta_x} \rceil^n \lceil \frac{\mu T_p + \eta}{\delta_u} + 1 \rceil^m \Big)^{\lfloor \frac{T}{T_p} \rfloor}.$$

Proof. To construct a (T, ε)-approximating function for a given trajectory $\xi_{x,u}$, at an iteration $i \in [0; \lfloor \frac{t}{T_p} \rfloor]$, 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: $\prod_{i=0}^{\lfloor T/T_p \rfloor} |C_{x,i}||C_{u,i}|$.

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 $2\delta_x$ (or $2\delta_u$). 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 $i > 0$, $S_{x,i} = B(z_{i-1}(T_p^-), \varepsilon)$ and $S_{u,i} = B(q_{u,i-1}, \mu T_p + \eta + \delta_u)$. Hence, $|C_{x,i}| \leq \lceil \frac{2\varepsilon}{2\delta_x} \rceil^n = \lceil \frac{\varepsilon}{\delta_x} \rceil^n$, since $diam(S_{x,i}) = 2\varepsilon$ and $|C_{u,i}| \leq \lceil \frac{2(\mu T_p + \eta + \delta_u)}{2\delta_u} \rceil^m = \lceil \frac{\mu T_p + \eta}{\delta_u} + 1 \rceil^m$, since $diam(S_{u,i}) = 2(\mu T_p + \eta + \delta_u)$. 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, δ_x and δ_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 $h_{\text{est}}(\varepsilon)$ of autonomous systems in [1] in Corollary 4.

Proposition 3. *For a fixed T_p, δ_x and δ_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:*

$$\frac{1}{T_p}\Big(n \log\lceil \frac{\varepsilon}{\delta_x} \rceil + m \log\lceil \frac{\mu T_p + \eta}{\delta_u} + 1 \rceil\Big). \tag{29}$$

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: $h_{\text{est}}(\varepsilon) =$

$$\begin{aligned}
&\limsup_{T\to\infty} \frac{1}{T} \log s_{\text{est}}(T, \varepsilon) \\
&\leq \limsup_{T\to\infty} \frac{1}{T} \log |C_{x,0}||C_{u,0}|(|C_{x,1}||C_{u,1}|)^{\lfloor \frac{T}{T_p} \rfloor} \\
&\leq \limsup_{T\to\infty} \frac{1}{T} \log(|C_{x,1}||C_{u,1}|)^{\lfloor \frac{T}{T_p} \rfloor} \\
&\leq \limsup_{T\to\infty} \frac{1}{T} \log(\lceil \frac{\varepsilon}{\delta_x} \rceil^n \lceil \frac{\mu T_p + \eta}{\delta_u} + 1 \rceil^m)^{\lfloor \frac{T}{T_p} \rfloor} \\
&\leq \limsup_{T\to\infty} \frac{T/T_p}{T} n \log\lceil \frac{\varepsilon}{\delta_x} \rceil \\
&\qquad + \limsup_{T\to\infty} \frac{T/T_p}{T} m \log\lceil \frac{\mu T_p + \eta}{\delta_u} + 1 \rceil \\
&= \frac{1}{T_p}\Big(n \log\lceil \frac{\varepsilon}{\delta_x} \rceil + m \log\lceil \frac{\mu T_p + \eta}{\delta_u} + 1 \rceil\Big).
\end{aligned}$$

□

We show that if the bound on the input norm is negligible, we recover the bound $\frac{nL_x}{\ln 2}$ on entropy derived in [2].

Corollary 4. *Given any $\varepsilon > 0$, $\lim_{\mu,\eta\to 0} h_{est}(\varepsilon) \leq \frac{nL_x}{\ln 2}$.*

Proof. Fix any $\varepsilon > 0$. Fix a η_0 and μ_0 such that $\eta_0 + \mu_0 < 1$. Then, fix δ_u to be equal to $\delta_{u,0} = \eta_0 + \mu_0$. After that, while taking the limits of μ and η going to zero, keep δ_u fixed to $\delta_{u,0}$, choose $\delta_x = \sqrt{1 - (\mu + \eta)}\varepsilon e^{-L_x T_p}$, and T_p to be such that $g_c(\delta_x, \delta_{u,0}, T_p) \leq \varepsilon^2$. Hence, by Proposition 3, $h_{\text{est}}(\varepsilon) \leq \frac{1}{T_p}\big(n \log\lceil \frac{\varepsilon}{\delta_x} \rceil + m \log\lceil \frac{\eta + \mu T_p}{\delta_u} + 1 \rceil\big) \leq nL_x \log\lceil \frac{e}{\sqrt{1-(\mu+\eta)}} \rceil + m \log\lceil \frac{\eta + \mu T_p}{\delta_{u,0}} + 1 \rceil$. Moreover, as μ and η decrease to zero, the argument of the second log approaches one and the upper bound approaches $\frac{nL_x}{\ln 2}$. □

Corollary 5. *For any $\varepsilon > 0$, $h_{est}(\varepsilon)$ of system (18), with any set of inputs $\mathcal{U}^p(\mu, \eta)$ with finite μ and η, 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 δ_x, δ_u and $T_p > 0$ that satisfy $g_c(\delta_x, \delta_u, T_p) \leq \varepsilon^2$, and $h_{\text{est}}(\varepsilon)$ of system (18) is upper bounded by $g_o(\delta_x, \delta_u, T_p)$, by Proposition 3. □

Corollary 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, $L_x = {}^{(10+1)}/_{2} = 5.5$, $G_x = {}^{3}/_{2}(10 + 1) = 16.5$, $M_x = 3(3(10) + {}^{1}/_{2}) = 91.5$ 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 $\delta_u = \eta = \pi/4$, $T_p = 1.9 \times 10^{-3}$, and $\delta_x = \frac{\varepsilon}{\sqrt{2}} e^{-G_x T_p}$. Such assignment satisfy $g_c(\delta_x, \delta_u, T_p) = 0.0099 \leq \varepsilon^2 = 0.01$ 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 δ_x, δ_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 $\delta_x = \frac{\varepsilon}{\sqrt{2}} e^{-M_x T_p}$ 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:

$$\begin{aligned}
&\dot{x}_1 = x_4; \dot{x}_4 = -g \sin x_3 - \frac{c x_4}{m'} + \frac{u_1}{m'} \cos x_3 - \frac{u_2}{m'} \sin x_3; \\
&\dot{x}_2 = x_5; \dot{x}_5 = g(\cos x_3 - 1) - \frac{c x_5}{m'} + \frac{u_1}{m'} \sin x_3 + \frac{u_2}{m'} \cos x_3; \\
&\dot{x}_3 = x_6; \dot{x}_6 = \frac{r}{J} u_1;
\end{aligned}$$

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

$G_x = n(\frac{-c}{m'} + \frac{1}{2})$ and $G_u \leq \sqrt{m(\frac{r^2}{J^2} + \frac{1}{m'^2})}$. Fixing $m' = 100$, $g = 9.81$, $r = 5$, $c = 100$, and $J = 50$, we get $G_x = -\frac{3}{2}$ and $G_u = 0.14$. Now, suppose that $\mu = 10$ 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, $g_c(\delta_x, \delta_u, T_p) = 0.249 \leq \varepsilon^2 = 0.25$, and thus $h_{\text{est}}(0.5) \leq g_o(\delta_x, \delta_u, T_p) = 699$. However, if $\mu = 0.1$ and $\eta = 45$, and choosing $\delta_u = \eta$, $T_p = 0.75 \times 10^{-3}$, and $\delta_x = \frac{1}{\sqrt{2}}\varepsilon e^{-G_x T_p}$. Then, $g_c(\delta_x, \delta_u, T_p) = 0.244 \leq \varepsilon^2 = 0.25$ and hence $h_{\text{est}}(0.5) \leq 3680$. Hence, the bound increased significantly as the bound on the variation 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 η to 0.1 while allowing large continuous variations by increasing μ to 20. In that case, the input signals are almost continuous. If we fix $\delta_u = \eta$, $T_p = 0.35$, and $\delta_x = \frac{1}{\sqrt{2}}\varepsilon e^{-G_x T_p}$. Then, $g_c(\delta_x, \delta_u, T_p) = 0.247 \leq \varepsilon^2 = 0.25$ and $h_{\text{est}}(0.5) \leq 11$, 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:

$$\dot{x}(t) = f_{sw}(x(t), \sigma(t)), \tag{30}$$

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 yet have 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 L_u, 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 σ are called *switches*. The switching signal σ 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: $Reach_{sw}(K, T_d, T) := \{x \mid \exists x_0 \in K, \sigma \in \Sigma(T_d), t \in [0,T], \xi_{x_0,\sigma}(t) = x\}$.

Based on our setup in [3], the entropy of the autonomous switched system (30) is defined as follows: First, three positive error parameters ε, α, and τ, 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 $t \in [s_j, s_{j+1})$,

$$\|z(t) - \xi_{x_0,\sigma}(t)\| \leq \begin{cases} \varepsilon & t \in [s_j, s_j + \tau), \\ \varepsilon e^{-\alpha(t-(s_j+\tau))} & otherwise, \end{cases} \tag{31}$$

where $s_0 = 0, s_1, \ldots$ are the switches of σ. The norm in inequality (31) can be arbitrary. Third, we define $s_{\text{sest}}(T, \varepsilon, \alpha, \tau)$ 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:*

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

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.

$$d(t) := \max_{p_1,p_2\in[N]} \sup_{x\in Reach_{sw}(K,T_d,\infty)} \int_0^t \|f_{sw}(\xi_{x,p_1}(s), p_1) - f_{sw}(\xi_{x,p_2}(s), p_2)\| ds. \tag{32}$$

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

This assumption can be checked, for example, if the reachset $Reach_{sw}(K, T_d, \infty)$ 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 $x \in Reach_{sw}(K, T_d, \infty)$. 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 finite 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:*

$$h_{sest}(\varepsilon, \alpha, \tau) \leq \frac{(L_x + \alpha)n}{\ln 2} + \frac{\log N}{T_e}, \tag{33}$$

where T_e is the largest real number in $(0, \tau]$ such that

$$d(T_e) \leq \varepsilon(1 - e^{-\alpha(T_d - T_e)}). \tag{34}$$

This article has been accepted for publication in IEEE Transactions on Automatic Control. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TAC.2023.3250510

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 α 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 $d(\hat{T}_e) \leq \varepsilon(1 - e^{-\hat{\alpha}(T_d - \hat{T}_e)})$. Such a bound may be finite even in the case when $\alpha = 0$.*

A. Finiteness of mode divergence $\mathsf{d(t)}$

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 finitness of stabilization entropy of switched linear systems.

We first define separated sets for switched systems. Two trajectories ξ_{x_1,σ_1} and ξ_{x_2,σ_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 ξ_{x_1,σ_1} and $\xi_{x_0,\sigma}$ by ξ_{x_2,σ_2}, respectively, *and* when z in the LHS is replaced by ξ_{x_2,σ_2} and $\xi_{x_0,\sigma}$ by ξ_{x_1,σ_1}, respectively. . A set of trajectories $\mathcal{Z}^\dagger$ is $(T, \varepsilon, \alpha, \tau)$-separated if all pairs of trajectories in $\mathcal{Z}^\dagger$ are $(T, \varepsilon, \alpha, \tau)$-separated. The maximal cardinality of such a set is denoted by $s^\dagger_{\text{sest}}(T, \varepsilon, \alpha, \tau)$.

Lemma 11. *For all K, T_d, ε, α, τ_1, τ_2, and $T > 0$,*

$$s^\dagger_{sest}(T, 2\varepsilon, 0, \tau_1) \leq s_{sest}(T, \varepsilon, 0, \tau_2) \leq s_{sest}(T, \varepsilon, \alpha, \tau_2). \quad (35)$$

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 $f^l_{sw}(x, 1) = ax$ and $f^l_{sw}(x, 2) = bx$, and the resulting switched system (30) having them as its modes:

$$\dot{x}(t) = f^l_{sw}(x(t), \sigma(t)), \quad (36)$$

where $\sigma : \mathbb{R}_{\geq 0} \to \{1, 2\}$ 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 is 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: $|\xi_{x,a}(t) - \xi_{x,b}(t)| = |e^{at} - e^{bt}||x|$. Therefore, the integral in Equation (32) is $\int_0^t |f^l_{sw}(\xi_{x,a}(s), a) - f^l_{sw}(\xi_{x,b}(s), b)| ds = |ae^{at} - be^{bt}||x| = |be^{bt}x||\frac{a}{b}e^{(a-b)t} - 1|$, which diverges to ∞ 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 $t > 0$. □

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 $tseq$. In a similar construction to that of the input signals in Section V-B, we consider the sequence $tseq_d$ of time instants of the form $t_i = \sum_{j=1}^{i} v_j$ that are less than T, where $v_1 = 2\varepsilon/(|x_0|(a-b))$ and $v_{i+1} = v_i e^{-bv_i}$ for $i \geq 1$. We denote the resulting set of switching signals by Σ_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 $tseq_d$ is similar to the time sequence $tseq_{p\alpha}$ we considered in Section V-B to prove that the entropy of system (18) is infinite if $\alpha > 0$. The first time instance in $tseq_{p\alpha}$ is equal to $\sqrt{2\varepsilon}$ while that of $tseq_d$ is $2\varepsilon/((a-b)|x_0|)$. Moreover, the time between time instants of $tseq_{p\alpha}$ is $v_{i+1} = v_i e^{-\alpha v_i/2}$ while that of $tseq_d$ is $v_{i+1} = v_i e^{-bv_i}$. Consequently, we can follow the same proof of Lemma 6, to conclude that $|tseq_d| \geq$

$$\lfloor \frac{2 - bv_1}{4bv_1}(e^{bT} - 1) \rfloor = \lfloor \frac{|x_0|(a-b) - b\varepsilon}{4b\varepsilon}(e^{bT} - 1) \rfloor. \quad (37)$$

Lemma 13. *Fix the constants T, ε, τ, and $\alpha > 0$. Consider any nonzero initial state $x_0 \in \mathbb{R}$ and the set of switching signals Σ_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, $|\xi_{x_0,a}(v_1) - \xi_{x_0,b}(v_1)| = |e^{av_1} - e^{bv_1}||x_0| > (a-b)v_1|x_0| = 2\varepsilon$. For any switching signal σ, the state at time t is $\xi_{x_0,\sigma}(t) \geq x_0 e^{bt}$. Consider any two switching signals $\sigma_1, \sigma_2 \in \Sigma_d$ with corresponding sequences se_1 and se_2 with the same prefix up till the i^{th} entry and differ in the $(i+1)^{th}$ one. Finally, let $x_i = \xi_{x_0,\sigma_2}(t_i)$. Then,

$$\begin{aligned} &|\xi_{x_0,\sigma_1}(t_{i+1}) - \xi_{x_0,\sigma_2}(t_{i+1})| \\ &= |x_i||e^{av_{i+1}} - e^{bv_{i+1}}| > |x_i|(a-b)v_{i+1} \\ &\geq |x_0|e^{bt_i}(a-b)v_{i+1} = |x_0|e^{bt_i}(a-b)v_1 e^{-bt_i} = 2\varepsilon. \end{aligned}$$

Therefore, any two trajectories starting from x_0 that have switching signals corresponding to two different sequences are separated by more than 2ε for at least one time instant. Hence, such set of trajectories is $(T, 2\varepsilon, 0, \tau)$-separated. □

Theorem 5. *For any ε, α, and $\tau > 0$, $h_{sest}(\varepsilon, \alpha, \tau)$ of switched system (36) is infinite.*

Proof. This theorem follows from the same proof of Theorem 3 while replacing α with b: the cardinality of the separated set constructed in Lemma 13 is equal to $2^{|tseq_d|}$. Substituting the lower bound on $|tseq_d|$ in Equation (37), we get that the separated set cardinality is lower bounded by $2^{O(e^{bT})}$. Hence, the cardinality of the maximum separated set $s^\dagger_{\text{sest}}(T, 2\varepsilon, 0, \tau)$

is lower bounded by $2^{O(e^{bT})}$. From Lemma 11, we get that the minimum cardinality of an approximating set $s_{\text{sest}}(T,\varepsilon,\alpha,\tau)$ is lower bounded by $2^{O(e^{bT})}$. If we substitute this lower bound in Definition 8, we get $h_{\text{sest}}(\varepsilon,\alpha,\tau)=\infty$. □

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 $tseq_d$ 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 $j<\lfloor\frac{T}{T_d}\rfloor$. Then, we define $tseq_d'$ to be the same sequence of time instants of $tseq_d$ 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 at one time instant in $tseq_d'$ 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 $s^{\dagger}_{\text{sest}}(T,2\varepsilon,\alpha,\tau)$ of $(T,2\varepsilon,\alpha,\tau)$-separated set, and consequently a lower bound on $h_{\text{sest}}(\varepsilon,\alpha,\tau)$ of system (36).

Now, we will lower bound the number of time instants in $tseq_d'$ 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 $tseq_d'$ in that interval. Recall that by definition, $t_{k+1}-t_k=v_{k+1}=v_1e^{-bt_k}\le v_1e^{-jbT_d}$. Using the same analogy as in Equation (37), the number of time instants in $tseq_d'$ in $[jT_d,(j+1)T_d]$ is lower bounded by $\lfloor\frac{2-bv_{k+1}}{4bv_{k+1}}(e^{bT_d}-1)\rfloor$, which is lower bounded by $\lfloor\frac{2-bv_1e^{-jbT_d}}{4bv_1e^{-jbT_d}}(e^{bT_d}-1)\rfloor$. If we take the product of such lower bounds for each odd j, the result is $O(e^{bT_d}+e^{3bT_d}+\dots e^{b\lfloor T/T_d\rfloor})=O(e^{b\lfloor T/T_d\rfloor^2})$. Therefore, $s_{\text{sest}}(T,\varepsilon,\alpha,\tau)$ is still lower bounded by $O(e^{b\lfloor T/T_d\rfloor^2})$ and the entropy $h_{\text{sest}}(\varepsilon,\alpha,\tau)$ 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 ε and $\rho>0$, and let $\delta_x=\varepsilon^2e^{-(M_x+\rho)T_p}$ and δ_u and T_p be such that*

$$g_{c,u}(\delta_u,T_p)\le\varepsilon^2(1-e^{-\rho T_p}),\tag{38}$$

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

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

where $P=\lceil\frac{\eta+\mu T_p}{\delta_u}+1\rceil^m$ represents the number of possible quantized inputs at each sampling time.

Proof. By substituting our choice of δ_x in $g_{c,x}$, we obtain $g_{c,x}(\delta_x,T_p)=\varepsilon^2e^{-\rho T_p}$. Then, $g_c(\delta_x,\delta_u,T_p)=g_{c,x}(\delta_x,T_p)+g_{c,u}(\delta_u,T_p)=\varepsilon^2$. Hence, the condition on g_c in Proposition 3 is satisfied. Now, by substituting our choice of δ_x in the bound in $g_o(\delta_x,\delta_u,T_p)$, along with replacing $\lceil\frac{\eta+\mu T_p}{\delta_u}+1\rceil^m$ 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_{est} of non-switched systems with inputs similar to how N, the number of modes, affects the bound on h_{sest} of switched systems.

In Theorem 4, α 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 ε. The parameter ρ in Corollary 6 has a similar meaning.

Lemma 15. *If $\mu=\eta=0$, line 12 in Algorithm 1 is changed to $S_{x,i}\leftarrow B(z_{i-1}(T_p^-),\varepsilon e^{-i\rho T_p})$, $\delta_{x,0}=\varepsilon$, and for all iterations $i\ge1$, $\delta_{x,i}=\delta_{x,i-1}e^{-\rho T_p}$, then for all $t\ge0$,*

$$\|z(t)-\xi_{x_0,u}(t)\|\le\varepsilon e^{-\rho t}.\tag{39}$$

Proof. If the input variation represented by η and μ is zero, then δ_u can be chosen to be zero and $S_{u,i}$ to be equal to $u(0)$ for any $i\ge0$. Such a choice of $\delta_{x,i}$ and $\delta_{u,i}$ 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 $g_{c,x}(\delta_x,T_p)=\varepsilon^2e^{-2\rho iT_p}$, at each iteration $i\ge0$ 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 $h_{\text{sest}}(\varepsilon,\alpha,\tau)$, 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., $U := \{u \in \mathbb{R}^{N-1} | u[p-1] \geq 0, \sum_{p\in[2;N]} u[p-1] \leq 1\}$ and $\mathcal{U}_{sw}$ be the set of all measurable functions mapping time to U. Hence, $\hat{f}_{sw} : \mathbb{R}^n \times U \to \mathbb{R}^n$. 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_o in Proposition 3 is an upper bound on h_{est}.

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

$$\hat{f}_{sw}(x(t), u(t)) := f_{sw}(x(t), 1) + \sum_{p\in[2;N]} g(x(t), p) u(t)[p-1], \tag{40}$$

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 $\hat{f}_{sw}$ with respect to the state and input at a certain $x(t)$ and $u(t)$ can be written as follows:

$$\begin{aligned} J_u &= \left[g(x(t),2), \ldots, g(x(t),N)\right], \qquad (41) \\ J_x &= J_{x,1}(x(t)) \\ &\quad + \sum_{p\in[2;N]} u(t)[p-1](J_{x,p}(x(t)) - J_{x,1}(x(t))) \end{aligned}$$

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, $G_x \leq n(nL_x + 1/2) \leq n(n \max_{p\in[N]} L_p + 1/2)$. Moreover, $G_u = \sqrt{m}\|\!|J_u|\!\| = \sqrt{N-1}\|\!|J_u|\!\|$ and $\|\!|J_u|\!\| = \max_{j\in[n]} \sum_{p\in[2;N]} |(J_u)_{j,p}| \leq \sum_{p\in[2;N]} \|g(x(t),p)\| = \sum_{p\in[2;N]} \|f_{sw}(x(t),p) - f_{sw}(x(t),1)\|$. Let us define $d_{co} := \sqrt{N-1}\sup_{x\in Reach(K,\mathcal{U}_{sw},\infty)} \sum_{p\in[2;N]} \|f_{sw}(x,p) - f_{sw}(x,1)\|$. d_{co} is an upper bound on $\|\!|J_u|\!\|$ for all $x \in Reach(K, \mathcal{U}_{sw}, \infty)$. 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 $\hat{\mathcal{U}}_{sw}$ with affine-bounded pointwise variation constructed according to Definition 1 with parameters $\mu = 0$, $\eta = 1$, i.e., $\mathcal{U}^p(0,1)$, 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 δ_u to be equal to η.

Theorem 6. *The entropy $h_{est}(\varepsilon)$ 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 ρ and T_p satisfy $(2M_u e^{M_x T_p})^2 T_p \leq \varepsilon^2(1 - e^{-\rho T_p})$, i.e., $(2d_{co}e^{n(n\max_{p\in[N]} L_p + 1/2)T_p})^2 T_p \leq \varepsilon^2(1 - e^{-\rho T_p})$. .

Note that there is always ρ 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 ρ and the RHS strictly increases with ρ. The bound is on the entropy of the differential inclusion which contains the trajectories of the switched system. It grows as $O(n^3)$, $O(\max_{p\in[N]} L_p)$, 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 $d(t) \leq t d_{co} e^{\max_{p\in[N]} L_p t}$ 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.

Sayan Mitra (M'01–SM'13) received his undergraduate degree in Electrical Engineering from Jadavpur University, Kolkata, India, a Masters Degree in Computer Science from Indian Institute of Science, Bangalore, and a Ph.D. degree in EECS from MIT in 2007. After one year of postdoctoral work at the Center for Mathematics of Information at the California Institute of Technology, he joined the University of Illinois at Urbana-Champaign, where he is now an Professor in the Electrical and Computer Engineering Department. His research interests include formal methods, hybrid systems, autonomous systems, and verification of cyber-physical systems and their applications. His work has been recognized by National Science Foundation's CAREER Award, Air Force Office of Scientific Research Young Investigator Program Award, IEEE-HKN C. Holmes MacDonald Outstanding Teaching Award and several best paper awards. He has served as the Program co-chair of the 20TH INTERNATIONAL CONFERENCE ON HYBRID SYSTEMS.

REFERENCES

[1] D. Liberzon and S. Mitra, "Entropy and minimal bit rates for state estimation and model detection," *IEEE Transactions on Automatic Control*, vol. 63, no. 10, pp. 3330–3344, 2018.

[2] D. Liberzon and S. Mitra, "Entropy and minimal data rates for state estimation and model detections," in *HSCC 2016*. Vienna: ACM, April 2016.

[3] H. Sibai and S. Mitra, "Optimal data rate for state estimation of switched nonlinear systems," in *Proceedings of the 20th International Conference on Hybrid Systems: Computation and Control*, ser. HSCC '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 71–80. [Online]. Available: https://doi.org/10.1145/3049797.3049799

[4] A. V. Savkin and I. R. Petersen, "Set-valued state estimation via a limited capacity communication channel," *IEEE Transactions on Automatic Control*, vol. 48, no. 4, pp. 676–680, April 2003.

[5] H. Sibai and S. Mitra, "State estimation of dynamical systems with unknown inputs: Entropy and bit rates," in *Proceedings of the 21st International Conference on Hybrid Systems: Computation and Control (Part of CPS Week)*, ser. HSCC '18. New York, NY, USA: Association for Computing Machinery, 2018, p. 217–226. [Online]. Available: https://doi.org/10.1145/3178126.3178150

[6] G. N. Nair, R. J. Evans, I. M. Y. Mareels, and W. Moran, "Topological feedback entropy and nonlinear stabilization," *IEEE Trans. Automat. Control*, vol. 49, pp. 1585–1597, 2004.

[7] G. N. Nair, F. Fagnani, S. Zampieri, and R. J. Evans, "Feedback control under data rate constraints: An overview," vol. 95, pp. 108–137, 2007.

[8] F. Colonius and C. Kawan, "Invariance entropy for control systems," *SIAM J. Control Optim.*, vol. 48, pp. 1701–1721, 2009.

[9] F. Colonius, C. Kawan, and G. Nair, "A note on topological feedback entropy and invariance entropy," *Systems Control Lett.*, vol. 62, pp. 377–381, 2013.

[10] A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems*. Cambridge University Press, 1995.

[11] G. Yang and D. Liberzon, "Stabilizing a switched linear system with disturbance by sampled-data quantized feedback," in *2015 American Control Conference (ACC)*, 2015, pp. 2193–2198.

[12] G. Yang and D. Liberzon, "Finite data-rate stabilization of a switched linear system with unknown disturbance**this work was supported by the nsf grants cns-1217811 and eccs-1231196." *IFAC-PapersOnLine*, vol. 49, no. 18, pp. 1085 – 1090, 2016, 10th IFAC Symposium on Nonlinear Control Systems NOLCOS 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2405896316318997

[13] G. Yang and D. Liberzon, "Feedback stabilization of switched linear systems with unknown disturbances under data-rate constraints," *IEEE Transactions on Automatic Control*, vol. 63, no. 7, pp. 2107–2122, 2018.

[14] M. Rungger and M. Zamani, "Invariance feedback entropy of nondeterministic control systems," in *Proceedings of the 20th International Conference on Hybrid Systems: Computation and Control*, ser. HSCC '17. New York, NY, USA: ACM, 2017, pp. 91–100. [Online]. Available: http://doi.acm.org/10.1145/3049797.3049801

[15] M. S. Tomar and M. Zamani, "Compositional quantification of invariance feedback entropy for networks of uncertain control systems," *IEEE Control Systems Letters*, vol. 4, no. 4, pp. 827–832, 2020.

[16] M. S. Tomar, C. Kawan, P. Jagtap, and M. Zamani, "Numerical estimation of invariance entropy for nonlinear control systems," 2020.

[17] J. Schmidt, "Topological entropy bounds for switched linear systems with lie structure," Master's thesis, Electrical and Computer Engineering from the University of Illinois at Urbana-Champaign, 2016.

[18] G. Yang, A. J. Schmidt, D. Liberzon, and J. P. Hespanha, "Topological entropy of switched linear systems: general matrices and matrices with commutation relations," *Mathematics of Control, Signals, and Systems*, vol. 32, no. 3, pp. 411–453, 2020. [Online]. Available: https://doi.org/10.1007/s00498-020-00265-9

[19] G. Yang, A. James Schmidt, and D. Liberzon, "On topological entropy of switched linear systems with diagonal, triangular, and general matrices," in *2018 IEEE Conference on Decision and Control (CDC)*, 2018, pp. 5682–5687.

[20] G. Yang, J. a. P. Hespanha, and D. Liberzon, "On topological entropy and stability of switched linear systems," in *Proceedings of the 22nd ACM International Conference on Hybrid Systems: Computation and Control*, ser. HSCC '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 119–127. [Online]. Available: https://doi.org/10.1145/3302504.3311815

[21] G. S. Vicinansa and D. Liberzon, "Estimation entropy for regular linear switched systems," in *2019 IEEE 58th Conference on Decision and Control (CDC)*, 2019, pp. 5754–5759.

[22] D. Liberzon, "On topological entropy of interconnected nonlinear systems," *IEEE Control Systems Letters*, vol. 5, no. 6, pp. 2210–2214, 2021.

[23] X. Gao, D. Liberzon, J. Liu, and T. Basar, "Connections between stability conditions for slowly time-varying and switched linear systems," in *2015 54th IEEE Conference on Decision and Control (CDC)*, Dec 2015, pp. 2329–2334.

[24] J. P. Hespanha and A. S. Morse, "Stability of switched systems with average dwell-time," in *Proceedings of the 38th IEEE Conference on Decision and Control (Cat. No.99CH36304)*, vol. 3, 1999, pp. 2655–2660 vol.3.

[25] H. K. Khalil, *Nonlinear Systems*, 3rd ed. New Jersey: Prentice Hall, 2002.

[26] F. Colonius, "Minimal bit rates and entropy for exponential stabilization," *SIAM J. Control Optim.*, vol. 50, pp. 2988–3010, 2012.

[27] A. S. Matveev and A. V. Savkin, *Stabilization of Linear Multiple Sensor Systems via Limited Capacity Communication Channels*. Boston: Birkhäuser Boston, 2009, pp. 1–64. [Online]. Available: https://doi.org/10.1007/978-0-8176-4607-3_3

[28] Z. Huang, C. Fan, and S. Mitra, "Bounded invariant verification for time-delayed nonlinear networked dynamical systems," *Nonlinear Analysis: Hybrid Systems*, vol. 23, pp. 211 – 229, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1751570X16300279

[29] C. Fan, J. Kapinski, X. Jin, and S. Mitra, "Locally optimal reach set over-approximation for nonlinear systems," in *Proceedings of the 13th International Conference on Embedded Software*, ser. EMSOFT '16. New York, NY, USA: ACM, 2016, pp. 6:1–6:10. [Online]. Available: http://doi.acm.org/10.1145/2968478.2968482

[30] C. Fan, J. Kapinski, X. Jin, and S. Mitra, "Simulation-driven reachability using matrix measures," *ACM Trans. Embed. Comput. Syst.*, vol. 17, no. 1, Dec. 2017. [Online]. Available: https://doi.org/10.1145/3126685

[31] H. Sibai and S. Mitra, "State estimation of open dynamical systems with slow inputs: Entropy, bit rates, and relation with switched systems," 2020. [Online]. Available: https://arxiv.org/abs/2011.10496

[32] K. J. Astrom and R. M. Murray, *Feedback Systems: An Introduction for Scientists and Engineers*. Princeton, NJ, USA: Princeton University Press, 2008.

[33] G. O. Berger and R. M. Jungers, "Finite data-rate feedback stabilization of continuous-time switched linear systems with unknown switching signal," in *2020 59th IEEE Conference on Decision and Control (CDC)*, 2020, pp. 3823–3828.

[34] M. Margaliot and D. Liberzon, "Lie-algebraic stability conditions for nonlinear switched systems and differential inclusions," *Systems & Control Letters*, vol. 55, no. 1, pp. 8–16, 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167691105000824