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, IEEE

Abstract—In this article, we extend the notion of estimation entropy of autonomous dynamical systems proposed by Liberzon and Mitra to nonlinear dynamical systems with uncertain inputs with bounded variation. We call this new notion the \( \varepsilon \)-estimation entropy of the system and show that it lower bounds the bit rate needed for state estimation. \( \varepsilon \)-estimation entropy represents the exponential rate of the increase of the minimal number of functions that are adequate for \( \varepsilon \)-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 the work of Liberzon and Mitra (2018), diverge to infinity. Thus, they are potentially not suitable for systems with uncertain inputs. We derive an upper bound on \( \varepsilon \)-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 \( \varepsilon \) error, given time-sampled and quantized measurements of state and input. We investigate the relation between \( \varepsilon \)-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—Bit rates, entropy, nonlinear systems, state estimation, 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:

\[ \text{What is the minimal bit rate of the channel for the estimator to support this accuracy requirement?} \]

This question has been investigated for both stochastic and nonstochastic system models and channels. In the nonstochastic 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, nonswitched, and switched systems (see, for example, [1], [2], [3]). This article contributes in this line of investigation and proposes answers to the abovementioned 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 nonswitched 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 nonswitched 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 \( \alpha \) 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 (see Section V-B). Moreover, we show that \( h_{\text{est}}(\varepsilon, \alpha) \) of the same system grows to infinity...
as $\varepsilon$ goes to zero (see 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 $\varepsilon$ (Corollary 5). These observations imply that the definition of entropy for autonomous systems (e.g., [4]) that takes the limit of $\varepsilon$ as it goes to zero, or that of [1], [2] that requires exponential convergence of error, are not suitable for systems with inputs. We, thus, use $h_{\text{est}}(\varepsilon)$ as a weaker notion of estimation entropy, which we name $\varepsilon$-estimation entropy, from there forward, as well as $h^i_{\text{est}}(\varepsilon)$ and $h^l_{\text{est}}(\varepsilon)$ for the alternative notions, implicitly fixing $\alpha$ to zero and $\varepsilon$ to a strictly-positive real number.

We show that there is no state estimation algorithm with a fixed bit rate smaller than $h_{\text{est}}(\varepsilon)$ (see 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 $\varepsilon$ error. This procedure is of independent interest, as it can also be used as a state estimation algorithm, provided that the uncertain input signal can be sampled. We count the number of trajectories that can be constructed by this procedure for different initial states and input signals, up to a time bound $T$. The rate of exponential increase of this number as $T$ increases gives an upper bound on entropy, and the bit rate needed for state estimation.

The upper bound is presented in terms of the state and the input dimensions $n$ and $m$, global bounds on the norms of the Jacobian matrices of the vector field with respect to the state and the input, $M_x$ and $M_u$, and two constants $\mu$ and $\eta$ that bound the variation of the input signal over time (Proposition 3). We consider two approaches in bounding the variation of the input signal (Definition 1). The first bounds the standard total variation of the signal over any time interval $[t, t + \tau]$ by $\mu \tau + \eta$. The other is a less-restrictive one that bounds the pointwise variation of the signal at any two time instants $t$ and $t + \tau$ by $\mu \tau + \eta$. We show that all of our results in the article 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 $\alpha$ 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 (see Section VIII-B). In addition, in Section VIII-C, we overapproximate 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 article in [3] and [5], particularly, earlier versions of Proposition 1, Proposition 3, Lemma 8, Lemma 10, and Theorem 4. The main novel contributions of this article are as follows:

1) comparing the alternative entropy notions using approximating, spanning, and separated sets;

2) showing that the entropy of the simple system $\dot{x} = u$ is infinite as $\varepsilon \to 0$ and for $\alpha > 0$ and a similar result for the simple switched system $\dot{x} = \sigma x$, where $\sigma$ switches between two positive reals, showing the necessity for bounded divergence between modes for the finiteness of entropy;

3) correcting the discrepancy function presented in [5], which was assuming that the largest eigenvalue of the symmetric part of the Jacobian of the dynamics with respect to the state is nonnegative as well as assuming 2-norm while being applied as an $\infty$-norm bound;

4) applying the bound on entropy of systems with uncertain inputs to obtain a bound on entropy of differential inclusions and autonomous switched systems, and thus relating the bounds in [3] and [5].

Finally, we improve the presentation of both papers [3] and [5].

A. Related Work

Several definitions of topological entropy for control systems have been proposed to bound the data rates necessary for control over limited-bandwidth communication channels [6], [7], [8], [9], [10] for tasks including stabilization [6], [11], [12], [13], invariance [14], [15], [16], and state estimation [1], [2], [3], [5], and types of systems including discrete-time [14], continuous-time [8], switched [3], [12], [17], [18], [19], [20], [21], and interconnected [15], [22]. In [1] and [2], Liberzon and Mitrani 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 $\frac{nL_x}{\ln 2}$ on estimation entropy, where $n$ is 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 article, 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 article 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 (see Section VIII-A).

II. Preliminaries

We denote the infinity norm of a real vector $v \in \mathbb{R}^n$ by $\|v\|_\infty$, its 2-norm as $\|v\|_2$, and its transpose by $v^T$. We denote the infinity norm of a real matrix $A \in \mathbb{R}^{n \times n}$ by $\|A\|_\infty$, its 2-norm by $\|A\|_2$, and its largest eigenvalue by $\lambda_{\text{max}}(A)$. If $A$ is symmetric positive definite, then $\lambda_{\text{max}}(A) = \|A\|_2$.

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

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
We denote by $[n_1; n_2]$, the set of integers in the interval $[n_1, n_2]$, inclusive, and by $[n_2]$ the set $[1; n_2]$. We denote the cardinality of a finite set $S$ by $|S|$. We denote the diameter of a compact set $S \subseteq \mathbb{R}^n$ by \( \text{diam}(S) = \max_{x, y \in S} \| x - y \| \). 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 $S_1 \cup S_2 := \{ x_1 + x_2 \mid x_1 \in S_1, x_2 \in S_2 \}$.

A continuous function $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ belongs to class-$K$ if it is strictly increasing and $\gamma(0) = 0$.

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

\[
u(t^+) = \lim_{t \to t^+} u(\tau) \quad \text{and} \quad \nu(t^-) = \lim_{t \to t^-} u(\tau).
\]

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

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

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

for all $t$ and $\tau \geq 0$.

We define $U^p(\mu, \eta)$, with the $s$-superscript referring to the word “slow,” to be the set of all piecewise-right-continuous functions $u : \mathbb{R}_+ \to \mathbb{R}^m$ with affine-bounded total variation, i.e.,

\[
u(0) \in U \quad \text{and} \quad \int_t^{t+\tau} \| du \| \leq \mu \tau + \eta
\]

for all $t$ and $\tau \geq 0$, where

\[
\begin{align*}
\int_t^{t+\tau} \| du \| &= \int_t^{t+\tau} \| \hat{u}(s) \| ds + \int_t^{t+\tau} \| \hat{\nu}(s) \| ds + \sum_{i=1}^{k-1} \int_{d_i}^{d_{i+1}} \| \hat{u}(s) \| ds \\
&\quad + \| u(t + \tau) - u(t) - \sum_{i=1}^{k} (u(d_i) - u(d_i^-)) \|,
\end{align*}
\]

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

Note that for any $\mu$ and $\eta \geq 0$, $U^p(\mu, \eta) \subseteq U^p(\mu, \eta)$. This can be shown as follows: fix any $\mu$, $\eta$, $t$, and $\tau \geq 0$, and consider any $u \in U^p(\mu, \eta)$, then $\| u(t + \tau) - u(t) \| = \| \int_t^{t+\tau} du \| \leq \int_t^{t+\tau} \| du \| \leq \mu \tau + \eta$. Thus, $u \in U^p(\mu, \eta)$.

We abbreviate $U^p(\mu, \eta)$ and $U^p(\mu, \eta)$ with $U^p$ and $U^s$ when $\mu$ and $\eta$ are clear from context.

The set $U^p(\mu, \eta)$ contains functions that vary arbitrarily fast, while having a finite number of switches in any finite time interval, within an $\eta$ bound. $U^p(\mu, 0)$ is the class of globally Lipschitz continuous functions with Lipschitz constant $\mu$. In general, knowing that $u$ cannot deviate too much, can be expressed by setting $\mu$ and $\eta$ to smaller values. Roughly, $\eta$ restricts the maximum norm of a jump at any time instant and $\mu$ restricts the distance between the values of a signal at different time instances. However, it is more common in the control and dynamical systems literature to bound the total variation of functions instead of their deviations. Accordingly, in addition to $U^p$, we introduced $U^s$, which bounds the total variation. Signals in $U^s$ 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 article. When deriving lower bounds on entropy, we assume that the input signals are in $U^p$, and when deriving upper bounds, we assume they are in $U^s$. Consequently, since $U^s \subseteq U^p$, all the results in this article hold for both cases.

The slow variation constraint in $U^s$ is the same to that of [23, Th. 2], 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 Hespanha and Morse in [24] to prove the stability of switched systems with stable subsystems. In this article, we use the bound on the variation of signals in $U^p$ to derive an upper bound on the entropy of dynamical systems with inputs in $U^s$ and $U^p$, since $U^s \subseteq U^p$. Furthermore, 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

\[
\dot{x}(t) = f(x(t), u(t)),
\]

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 $J_u(x, u) = \partial f/\partial u$ and $J_u(x, u) = \partial f/\partial u$, 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 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 $\xi_{x_0, u} : \mathbb{R}_+ \to \mathbb{R}^n$. In the rest of this section, we fix a compact set of initial states $K \subseteq \mathbb{R}^n$ and the set of input functions $U^p(\mu, \eta)$. The same results hold if the set of input signals was $U^s(\mu, \eta)$ instead.

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

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

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

\[
\begin{align*}
\dot{x}_1 &= v \cos x_3, \\
\dot{x}_2 &= v \sin x_3, \\
\dot{x}_3 &= u,
\end{align*}
\]

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:

\[
J_x = \begin{bmatrix} 0 & 0 & -v \sin x_3 \\ 0 & 0 & v \cos x_3 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad J_u = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}.
\]
Therefore, the dynamics are globally Lipschitz with Lipschitz constants
\[ L_x = \| J_x \| \leq v \text{ and } L_u = \| J_u \| = 1. \] (8)

A. Approximating Functions and \((\varepsilon, \alpha)\)-Estimation Entropy

Let us fix the error parameters \(\varepsilon > 0\) and \(\alpha \geq 0\). Given a
\( T > 0 \), \( x_0 \in K \), and \( u \in U^p(\mu, \nu) \), we say that a function \( z : [0, T] \rightarrow \mathbb{R}^n \) is \((T, \varepsilon, \alpha)\)-approximating for the trajectory \( \xi_{x_0, u} \) over the interval \([0, T]\), if
\[ \| z(t) - \xi_{x_0, u}(t) \| \leq \varepsilon e^{-\alpha t}, \] (9)
for all \( t \in [0, T] \). We say that a set of functions \( Z := \{ z : [0, T] \rightarrow \mathbb{R}^n \} \) is \((T, \varepsilon, \alpha)\)-approximating for system (4), if for every \( x_0 \in K \) and \( u \in U^p \), there exists a \((T, \varepsilon, \alpha)\)-approximating function \( z \) for the trajectory \( \xi_{x_0, u} \) over \([0, T]\). In this article, 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 \( s_{est}(T, \varepsilon, \alpha) \). The subscript \( est \) corresponds to the word \textit{estimation} and the letter \( s \) stands for the word \textit{set}.

Definition 2: The \((\varepsilon, \alpha)\)-estimation entropy of system (4) is defined as follows:
\[ h_{est}(\varepsilon, \alpha) := \limsup_{T \rightarrow \infty} \frac{1}{T} \log s_{est}(T, \varepsilon, \alpha). \] (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 rate at which bits need to be sent by the sensor so that the estimator can construct a “good” estimate of the state.

Remark 1: The value of \((\varepsilon, \alpha)\)-estimation entropy has the same value if the \(\infty\)-norm in inequality (9) is replaced by any norm. This follows from the equivalence of norms in finite-dimensional spaces. More specifically, since for any two norms \( \| \cdot \|_1 \) and \( \| \cdot \|_2 \), for all \( x \in \mathbb{R}^n \), there exists two positive constants \( c_1 \) and \( c_2 \) such that \( c_1 \| x \|_1 \leq \| x \|_2 \leq c_2 \| x \|_1 \), \( s_{est} \) 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 abovementioned notion of approximating sets to the sample complexity of \textit{reachable sets}. The latter is not a central theme in this article but is an important concept in control and verification.

Remark 2: Fix a time horizon \( T > 0 \). Assume that there is a method to generate a \((T, \varepsilon, \alpha)\)-approximating function for any given trajectory of system (4).

Then, \( s_{est}(T, \varepsilon, \alpha) \) is also the minimum number of simulations needed to \textit{approximate} the reachset Reach\((K, U, T)\) of system (4) up to an \(e^{-\alpha t}\) overapproximation error, for any \( t \in [0, T] \). Formally, let \( Z \) be the \((T, \varepsilon, \alpha)\)-approximating set with cardinality \( s_{est}(T, \varepsilon, \alpha) \) and define an overapproximation of the reachable set
\[ \text{Reach}_{\varepsilon, \alpha}(K, U, T) := \{ x + c : \exists x_0 \in K, u \in U, t \in [0, T], c \in B_{\varepsilon e^{-\alpha t}}, \xi_{x_0, u}(t) = x \} \]
and another overapproximation using \( Z \)
\[ R_{\varepsilon, \alpha} := \{ z(t) + c : z \in Z, t \in [0, T], c \in B_{\varepsilon e^{-\alpha t}} \}, \]
where \( B_{\varepsilon e^{-\alpha t}} := B(0, \varepsilon e^{-\alpha t}) \). Then
\[ \text{Reach}(K, U, T) \subseteq R_{\varepsilon, \alpha} \subseteq \text{Reach}_{\varepsilon, \alpha}(K, U, T). \]

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 result the restricted approximating sets \textit{spanning} sets. Formally, we say that a set \( Z^* \subseteq K \times U \) is \((T, \varepsilon, \alpha)\)-spanning, if for every \( x_1 \in K \) and \( u_1 \in U \), there exists a \((x_2, u_2) \in Z^* \) such that the trajectory \( \xi_{x_2, u_2} \) is a \((T, \varepsilon, \alpha)\)-approximating function for the trajectory \( \xi_{x_1, u_1} \) over \([0, T]\). The minimum cardinality of a spanning set is denoted by \( s_{est}^*(T, \varepsilon, \alpha) \). If we propagate this restriction to Definition 2, the resulting definition of entropy would be as follows.

Definition 3: The spanning sets-based \((\varepsilon, \alpha)\)-estimation entropy of system (4) is defined as follows:
\[ h_{est}^*(\varepsilon, \alpha) := \limsup_{T \rightarrow \infty} \frac{1}{T} \log s_{est}^*(T, \varepsilon, \alpha). \] (11)

The following inequality is an application of the fact that any \((T, \varepsilon, \alpha)\)-spanning set is a \((T, \varepsilon, \alpha)\)-approximating set
\[ s_{est}(T, \varepsilon, \alpha) \leq s_{est}^*(T, \varepsilon, \alpha). \] (12)

B. Separated Sets Instead of Approximating Ones

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

Fix any error parameters \(\varepsilon > 0\) and \(\alpha \geq 0\). Given \( T > 0 \), two initial states \( x_1 \) and \( x_2 \) and two input signals \( u_1 \) and \( u_2 \), the two trajectories \( \xi_{x_1, u_1} \) and \( \xi_{x_2, u_2} \) are \((T, \varepsilon, \alpha)\)-separated if there exists \( t \in [0, T] \) such that
\[ \| \xi_{x_1, u_1}(t) - \xi_{x_2, u_2}(t) \| > \varepsilon e^{-\alpha t}. \] (13)

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

Definition 4: The separated sets-based \((\varepsilon, \alpha)\)-estimation entropy of system (4) is defined as follows:
\[ h_{est}^\dagger(\varepsilon, \alpha) := \limsup_{T \rightarrow \infty} \frac{1}{T} \log s_{est}^\dagger(T, \varepsilon, \alpha). \] (14)

The following two lemmas are analogous to [2, Lemmas 1 and 2]. They draw the relation between approximating, spanning, and separated sets.

Lemma 1: For all \( K, U^p, \varepsilon, \alpha, \) and \( T \),
\[ s_{est}^\dagger(T, \varepsilon, \alpha) \leq s_{est}(T, \varepsilon, \alpha). \] (15)

\(^1\)This section is a novel contribution of this article.
any pair in $Z^t$ violate (9) for some $t \in [0, T]$. Therefore, $\xi_{x_1,u_4}$ can be added to $Z^t$, which contradicts its maximality. ■

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

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

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

$$\|\xi_{x_1,u_1}(t) - \xi_{x_2,u_2}(t)\|$$

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

by triangular inequality

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

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

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

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

$$h^1_a(2\varepsilon, \alpha) \leq h_{est}(\varepsilon, \alpha) \leq h^*_{est}(\varepsilon, \alpha) \leq h^1_a(\varepsilon, \alpha).$$

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

V. INFEASIBLE ALTERNATIVE ENTROPY NOTIONS

In this section, we show that the $(\varepsilon, \alpha)$-estimation entropy $h_{est}(\varepsilon, \alpha)$ of a simple dynamical system is infinite if $\varepsilon$ approaches zero or if $\alpha > 0$.

On the other hand, we show in Corollary 5 that $h_{est}(\varepsilon, 0)$ of that system is finite for any fixed $\varepsilon > 0$.

We conclude that it is not meaningful to take the supremum over $\varepsilon > 0$ or choose an $\alpha > 0$, in contrast with the definition of estimation entropy of autonomous systems [2].

Consider the simple dynamical system

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

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

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

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

**Definition 5:** Given a sequence of $l$ real numbers $A = a_0, \ldots, a_{l-1}$ and an increasing sequence of $l + 1$ time points $tseq = t_0, \ldots, t_l$, we define the set $S\Sigma\Sigma$ 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 $u : [t_0, t_l] \rightarrow \mathbb{R}$ generated by a sequence $se$ as follows: for all $i < l$ and $t \in [t_i, t_{i+1})$,

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

The results presented in this section are novel contributions of this article.
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 [27, Th. 2.3.17]. Extending [27, Definition 2.3.13] to continuous-time systems, system (4) is said to be locally reachable along a trajectory $\xi_{x,u}$ if two positive constants $\delta$ and $\tau$ exist such that for any positive $t$ and any $a$ and $b$ in $\mathbb{R}^n$ such that $|\|\xi_{x,u}(t) - a\|_2 \leq \delta(\|\xi_{x,u}(t)\|)$ and $|\|\xi_{x,u}(t + \tau) - b\|_2 \leq \delta(\|\xi_{x,u}(t + \tau)\|)$, another trajectory $\xi_{x',u'}$ of system (4) exists and satisfies $\xi_{x',u'}(t) = a$ and $\xi_{x',u'}(t + \tau) = b$. Extending [27, Definition 2.3.14] 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 $\delta_0$ such that $|\|\xi_{x,u}(t)\|_2 \geq \delta_0$, for all $t \geq 0$.

B. Requiring Exponentially Convergent Estimation Error

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

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

**Proof:** Recall from (10) that

$$h_{\text{est}}(\varepsilon; (T, \varepsilon; \alpha; K)) = \limsup_{T \to \infty} \frac{1}{T} \log s_{\text{est}}(T, \varepsilon; \alpha; K).$$

Thus, for any finite-time bound $\tau$, $h_{\text{est}}(\varepsilon; \alpha; K)$ is greater than or equal to

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

where $\text{Reach}(K, U, [\tau, \tau])$ represents the reachable states exactly at time $\tau$. The inequality follows from the fact that to construct a $(T, \varepsilon; \alpha; K)$-approximating set, one can construct a $(\varepsilon e^{-\alpha \tau}, \alpha; \text{Reach}(K, U, [\tau, \tau]))$-approximating set and a $(T - \tau, \varepsilon e^{-\alpha \tau}, \alpha; \text{Reach}(K, U, [\tau, \tau]))$-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 $K$, $\lim_{\varepsilon \to 0} h_{\text{est}}(\varepsilon; \alpha; \text{Reach}(K, U, [\tau, \tau]))$ 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 $(T, \varepsilon; \alpha)$-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 $U_{\varepsilon,\alpha}$ that results from Definition 5 with the sequence of time instants $\text{seq}_{\varepsilon,\alpha}$, being of the form $t_0 = 0$ and $t_i = \sum_{j=1}^{i} \epsilon_j$, where $i \leq l$, $l$ is the largest index where $v_l$ is less than $T$, $v_l = \sqrt{2\tau}$, and $v_{i+1} = v_i e^{-\alpha \tau, l/2}$.
(or equivalently \( v_{i+1} = v_t e^{-\frac{2}{c} \sum_{j=1}^{i} \xi_j} \)). Observe that for all \( i \), \( v_{i+1} < v_i \), and \( v_t \) is the time interval between \( t_{i-1} \) and \( t_i \). We choose the sequence \( A \) in the definition of \( U_{pa} \) to be of the form \( v_1, \ldots, v_i \), which are the same values used to define the switches above.

**Lemma 5:** Fix \( T, \varepsilon, \alpha \) to be greater than 0. Consider any initial state \( x_0 \in \mathbb{R} \) and the set of input signals \( U_{pa} \). 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 \( v_t \) and \( v_{t+1} \) for the first \( v_t \) time units, respectively. Then, \( x_{v_{t+1}}(v_{t+1}) = x_{v_t}(v_t) = \) \( v_t e^{2\varepsilon/\alpha} \), where the last inequality follows from the fact that \( v_t > 0 \). Now, consider any two signals \( u_1, u_2 \in U_{pa} \) with corresponding sequences \( s_1 \) and \( s_2 \) 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., \( s_1[j+1] \neq s_2[j+1] \). The resulting trajectories \( \xi_{x_0,u_1} \) and \( \xi_{x_0,u_2} \) 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 \( x_0 \in \mathbb{R} \),

\[
\|x_{v_{i+1}}(v_{i+1}) - x_{v_i}(v_i)\| = \|\xi_{x_0,u_{i+1}} - \xi_{x_0,u_i}\| = \|\xi_{x_0,u_i}(v_i) - \xi_{x_0,u_i}(v_{i+1})\| = |x_i + (v_i + u_i(t_i-1))v_i - (v_{i+1} - u_{i+1}(t_i-1))v_i| = v_i = 2\varepsilon e^{-\alpha \sum_{j=1}^{i} v_j} > 2\varepsilon e^{-\alpha \sum_{j=1}^{i} v_j},
\]

where \( x_0 \cdot u_{i+1} = (v_{i+1} - u_{i+1}(t_i-1))v_i \).

Therefore, \( \sum_{j=1}^{i} v_j \leq v_i \leq \frac{2}{\varepsilon} \ln \left( \frac{2\varepsilon}{2\varepsilon - \alpha} \right) = 1 \).

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 \( \frac{4\varepsilon}{2\varepsilon - \alpha} \).

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

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

**Corollary 2:** The total variation of any signal in \( U_{pa} \) with number of switches that is \( O(e^{T}) \) is \( O(T) \).

**Theorem 3:** For any \( \varepsilon \) and \( \alpha > 0 \), \( h_{est}(\varepsilon, \alpha) \) of system (18) with the input set \( U_{pa} \), even when restricted to signals with \( O(e^{\alpha T}) \) and having linearly-bounded total variation, is infinite.

**Proof:** First, there are \( 2^{h_{est}} \) 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 \( \frac{1}{2} \leq \frac{s_{est}(T, 2\varepsilon, \alpha)}{2^{-h_{est}}} \), and hence, by Lemma 2, \( s_{est}(T, \varepsilon, \alpha) \geq 2^O(e^{\varepsilon T}) \). Substituting this lower bound in (10) results in \( h_{est}(\varepsilon, \alpha) = \infty \).

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

**VI. RELATION BETWEEN \( \varepsilon \)-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 $\varepsilon$ estimation error when receiving state estimates at a fixed bit rate smaller than the $\varepsilon$-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 $(S, E)$, where $S : \mathbb{R}^n \times Q_s \rightarrow \Gamma \times Q_s$ represents the sensor, $E : \Gamma \times Q_e \rightarrow ([0, T_p] \rightarrow \mathbb{R}^n) \times Q_e$ represents the estimator, $T_p$ is the sampling time, $\Gamma$ is an alphabet, and $Q_s$ and $Q_e$ are the sets of internal states of the sensor $S$ and estimator $E$, respectively.

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

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

$$b_r(\varepsilon) := \lim_{T \to \infty} \sup_{T_p} \frac{1}{T} \sum_{i=0}^{[T/T_p]} \log |\Gamma| = \frac{\log |\Gamma|}{T_p}.$$ 

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 $h_{\text{est}}(\varepsilon)$ that for any $T > 0$, generates a $(T, \varepsilon)$-approximating function for any input trajectory.

**Proof:** The proof is similar to [3, the Proof of Proposition 2]. It is based on showing that if an 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 $\varepsilon$ (see Fig. 2) that can be constructed by that algorithm is a $(T, \varepsilon)$-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 $(T, \varepsilon)$-approximating functions for input trajectories. Recall that $h_{\text{est}}(\varepsilon) = \lim sup_{T \to \infty} 1/T \log s_{\text{est}}(T, \varepsilon)$. Then

$$b_r(\varepsilon) < h_{\text{est}}(\varepsilon)$$

$$\lim_{T \to \infty} \sup_{T_p} \frac{1}{T} \sum_{i=0}^{[T/T_p]} \log |\Gamma| < \lim_{T \to \infty} \sup_{T_p} \frac{1}{T} \log s_{\text{est}}(T, \varepsilon),$$

$$\lim_{T \to \infty} \sup_{T_p} \frac{1}{T} \sum_{i=0}^{[T/T_p] + 1} \log |\Gamma| < \lim_{T \to \infty} \sup_{T_p} \frac{1}{T} \log s_{\text{est}}(T, \varepsilon).$$

Hence, for a sufficiently large $T'$, we should have $\frac{1}{T'} \log s_{\text{est}}(T', \varepsilon)$, 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 $(T', \varepsilon)$-approximating function over the interval $[0, T']$. This is true since the estimator should be able to construct a $(T', \varepsilon)$-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 $(T', \varepsilon)$-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 $(T', \varepsilon)$-approximating set with a cardinality smaller than $s_{\text{est}}$, the supposed minimal one, which is a contradiction.

**VII. ENTROPY UPPER BOUND AND ALGORITHM**

In this section, we derive an upper bound on the entropy $h_{\text{est}}(\varepsilon)$ 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 (IS) Discrepancy Function Construction**

We use a modified version of the definition of local IS 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 $X \subset \mathbb{R}^m$, time interval $[t_0, t_1] \subset \mathbb{R}$, and a set of piece-wise-right-continuous functions $U$ mapping $[t_0, t_1] \to \mathbb{R}^m$. A function $V : X^2 \to \mathbb{R}$ is a local IS discrepancy function for system (4) over $X$ and $[t_0, t_1]$ if:

i) there exist class-$K$ functions $\hat{a}, \hat{Q}$ such that for any $x, x' \in X$, $\hat{a}(\|x - x'\|) \leq V(x, x') \leq \hat{a}(\|x - x'\|)$;

ii) there exists a class-$K$ function in the first argument and continuous in the second argument $\beta : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ and a class-$K$ function $\gamma : \mathbb{R} \to \mathbb{R}$ such that for any $x_0, x'_0 \in X$ and $u, u' \in U$, if $\xi_{x_0,u}(t)$ and $\xi_{x'_0,u'}(t)$ are trajectories of system (4) for all $t \in [t_0, t_1]$, then for any such $t$

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

The local discrepancy function $V$ together with $\beta$ and $\gamma$ give the sensitivity of the trajectories of the system to changes in the initial state and the input. The functions $\hat{a}, \hat{Q}, \beta, \gamma$ are sometimes called witnesses of the local IS discrepancy $V$. Techniques for computing local discrepancy functions have been presented [28], [29], [30].

The following is a straight-forward generalization of [28, Lemma 15] 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 $\infty$-norm while it was a bound on the squared 2-norm of the difference between the two trajectories. In this article, we correct both errors by adding a $\sup_{\epsilon \in [0, t_1]}$ in the definition of $\gamma$ 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 $u \in \mathcal{U}$ with

$$
\beta(y, t-t_0) := e^{2G_x(t-t_0)} y^2 \quad \text{and}
$$

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

$$
G_x := \frac{n/2}{2} \left( \sup_{t \in [t_0, t_1]} \max_{u \in \mathcal{U}(x)} \frac{J_x(x, u(t)) + J_z(x, u(t))^T}{2} \right) + 1,
$$

$$
G_u := \sqrt{m} \sup_{t \in [t_0, t_1]} \|J_u(x, u(t))\|.
$$

(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 $G_u$ for 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 article [31].

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

Furthermore, 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

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

(24)

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

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

$$
J_x + J_z^T
$$

$$
= \begin{bmatrix}
0 & 0 & -\frac{1}{2} \sin x_3 \\
0 & 0 & \frac{1}{2} \cos x_3 \\
-\frac{1}{2} \sin x_3 & \frac{1}{2} \cos x_3
\end{bmatrix}
$$

(25)

and has eigenvalues 0 and $\pm \frac{1}{2}$. Hence, $G_x = n/2(2v^2 + 1)$, $G_u = \sqrt{m}\|J_u\| = 1$.

If we use Proposition 2 to upper bound $G_x$ and $G_u$, we get

$$
M_x = n(nL_x + \frac{1}{2}) = 3(3v^2 + \frac{1}{2}),
$$

and $M_u = m\sqrt{m}L_u = 1$. Here $G_x$ and $G_u$ do not depend on $x$ 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 $u \in \mathcal{U}(\mu, \eta)$, constructs a $(T, \varepsilon)$-approximating function for the trajectory $\xi_{x_0, u}$ 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 $u \in \mathcal{U}(\mu, \eta)$ is a $(T, \varepsilon)$-approximating set for system (4). We will show in the following section how to obtain an upper bound on entropy by deriving an upper bound on the cardinality of this set. That upper bound on entropy will be an upper bound on the minimal bit rate of state estimation as well, as Algorithm 1 is a state estimation algorithm that requires a fixed bit rate equal to that upper bound.

The procedure (Algorithm 1) is parameterized by a time horizon $T > 0$, a sampling period $T_p > 0$, and two constants $\delta_x$ and $\delta_\eta > 0$, that depends on the initial input $U$, and the initial state $x_0 \in K$, and an input signal $u \in \mathcal{U}(\mu, \eta)$ for system (4). The output is a piece-wise continuous function $z : [0, T] \rightarrow \mathbb{R}^n$ that is constructed iteratively over each $[iT_p, (i + 1)T_p)$ interval for $i \in [0; \lfloor \frac{T}{T_p} \rfloor]$.

Later we will infer several constraints on the parameters such that the output is indeed a $(T, \varepsilon)$-approximating function for the given trajectory $\xi_{x_0, u}$.

Initially, $S_{x_0, 0}$ is assigned to the initial set $K$. Similarly, $S_{u_0, 0}$ is assigned to the initial input set $U$. $C_{x, 0}$ is a grid of size $\delta_x$ over $K$ and $C_{u, 0}$ is a grid of size $\delta_\eta$ over $U$.

At the $i$th iteration, $i \in [0; \lfloor \frac{T}{T_p} \rfloor]$, $x_i$ and $u_i$ store the values $\xi_{x_0,i}(iT_p)$ 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 $u_i$, respectively, and $C_{x,i}$ 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 $z : [0, T] \rightarrow \mathbb{R}^n$.

Algorithm 1: Construction of a $(T, \varepsilon)$-Approximating Function.

1: input: $T, T_p, \delta_x, \delta_\eta$
2: $S_{x,0} \leftarrow K; S_{u,0} \leftarrow U$
3: $C_{x,0} \leftarrow \text{grid}(S_{x,0}, \delta_x)$
4: $C_{u,0} \leftarrow \text{grid}(S_{u,0}, \delta_\eta)$
5: $i \leftarrow 0$
6: while $i \leq \lfloor \frac{T}{T_p} \rfloor$ do
7: $x_i \leftarrow \xi_{x_0,i}(iT_p)$; $u_i \leftarrow u(iT_p)$
8: $q_{x,i} \leftarrow \text{quantize}(x_i, C_{x,i})$
9: $q_{u,i} \leftarrow \text{quantize}(u_i, C_{u,i})$
10: $z_i = \xi_{q_{x,i}, q_{u,i}}$
11: $i \leftarrow i + 1$
12: $S_{x,i} \leftarrow B(z_{i-1}, \delta_x)$
13: $S_{u,i} \leftarrow B(q_{u,i}, \delta_\eta + \mu T_p + \delta_\eta)$
14: $C_{x,i} \leftarrow \text{grid}(S_{x,i}, \delta_x)$
15: $C_{u,i} \leftarrow \text{grid}(S_{u,i}, \delta_\eta)$
16: wait$(T_p)$
17: end while
18: output:

$\{z_i([0, T_p]) : i \in [0; \lfloor \frac{T}{T_p} \rfloor - 1]\} \cup \{z_i([\frac{T}{T_p}], [0, T_p])\}$
Note that for \( i = \frac{T}{T_p} \), 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 \( T_p, \delta_x, \) and \( \delta_u \), are small enough, then the output is a \((T, \varepsilon)\)-approximating function for \( \xi_{x,u} \).

Before stating the lemma, let us define the following function:
\[
g_c(\delta_x, \delta_u, T_p) := \frac{T}{T_p} \
\]
for \( g_c(\delta_x, \delta_u, T_p) := \delta_x e^{\frac{\delta_x}{T_p} \delta_u} \) and \( g_c(\delta_u, T_p) := \delta_u e^{\frac{\delta_u}{T_p} \delta_u} \).

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 \((T, \varepsilon)\)-approximating function.

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

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

1) \( u_i \in S_{u,i} \);
2) \( \|u(t) - q_{u,i}\| \leq \mu(t - iT_p) + \eta + \delta_u \);
3) \( \|x_i - S_{x,i}\| \leq \varepsilon \);
4) \( \|z_i(t) - iT_p\| \leq \xi_{x,u}(t - iT_p) \leq \varepsilon \);

where \( u_i(t) := u(t) + \frac{T}{T_p} T_p \), the \( i \)th piece of \( u(t) \) 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 \( \delta_u \). Moreover, \( \|u(t) - u(0)\| \leq \mu(t) + \eta \) by Definition 1. But, by triangle inequality, \( \|u(t) - q_{u,i}\| \leq \|u(t) - u_0\| + \|u_0 - q_{u,i}\| \leq \mu(t) + \eta + \delta_u \). Hence, \( u_i \in S_{u,i} \). Fix \( i \geq 1 \) and assume \( u_i \in S_{u,i} \). Then, \( \|u_i - q_{u,i}\| \leq \delta_u \). Repeating the same analysis for the \( i = 0 \) case, results in \( \|u(t) - iT_p\| \leq \mu(t) + \eta + \delta_u \). Thus, \( u_i \in S_{u,i} \) for all \( i \) and all \( x_0 \in S_{x,0} \) and \( \delta_u \), since \( x_0 \in K \). Also, \( \xi_{x,0}(0) = \xi_{x,0}(0) = g_c(0,0) \leq \varepsilon \), 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 \( i = \frac{T}{T_p} \).

Then,
\[
\|\xi_{x,u}(t - iT_p) - \xi_{x,i}(t - iT_p)\|^2 \\
\leq \|x_i - q_{u,i}\|^2 |e^{2M_x(t - iT_p)}| \\
+ M_u^2 |e^{2M_t(t - iT_p)}| \int_{iT_p}^t \|u(s) - q_{u,i}\|^2 ds \\
[\text{by (24)}] \\
\leq \delta_x^2 |e^{2M_x(t - iT_p)}| \\
+ M_u^2 |e^{2M_t(t - iT_p)}| \int_{iT_p}^t (\mu(s) + \eta + \delta_u)^2 ds \\
\leq \delta_x^2 |e^{2M_x(t - iT_p)} + M_u^2 |e^{2M_t(t - iT_p)}| \left( \frac{1}{3} \mu^2 T_p^3 + \frac{T}{T_b} \right) \\
+ (\eta + \delta_u)^2 T_p \\
= g_c(\delta_x, \delta_u, T_p) \leq \varepsilon^2 .
\]

The (\( T, \varepsilon \))-approximating function to the input trajectory \( \xi_{x_0,u} \), i.e.,
\[
\|z(t) - \xi_{x_0,u}(t)\| \leq \varepsilon .
\]

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

Lemma 10: For any \( T \geq 0 \), \( \delta_x, \delta_u, \) 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 \( u \in U_p^p(\mu, \eta) \), is upper bounded by
\[
\left( \frac{\text{diam}(K)}{2\delta_x} \right)^n \left( \frac{\text{diam}(U)}{2\delta_u} \right)^m \left( \left( \frac{\varepsilon}{\delta_x} \right)^n \left( \frac{\mu T_p + \eta}{\delta_u} + 1 \right)^{\frac{T}{T_p}} \right).
\]

Proof: To construct a \((T, \varepsilon)\)-approximating function for a given trajectory \( \xi_{x_0,u} \), at an iteration \( i \in [0; \frac{T}{T_p}] \), 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:
\[
\prod_{i=0}^{\frac{T}{T_p}} |C_{x,i}||C_{u,i}|.
\]

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 \( \text{diam}(K) \) (or \( \text{diam}(U) \)) to smaller segments of size \( 2\delta_x \) (or \( 2\delta_u \)). Then,
\[
|C_{x,0}| \leq \left( \frac{\text{diam}(K)}{2\delta_x} \right)^n \quad \text{and} \quad |C_{u,0}| \leq \left( \frac{\text{diam}(U)}{2\delta_u} \right)^m
\]

Similarly, for all \( i > 0, S_{x,i} = B(z_{i-1}(T_p^i), \varepsilon) \) and
\[
S_{u,i} = B(q_{u,i-1},\mu T_p + \eta + \delta_u).
\]

Hence,
\[
|C_{x,i}| \leq \left( \frac{2\mu}{T_p} \right)^n \quad \text{and} \quad |C_{u,i}| \leq \left( \frac{2 \mu T_p + \eta + \delta_u}{T_p} \right)^m.
\]

Substituting these values in the bound above leads to the upper bound in the lemma.

C. Entropy Upper Bound

The following proposition gives an upper bound on the entropy of system \((4)\) in terms of \( T_p, \delta_x, \) and \( \delta_u \).

It shows the effect of our choice of the parameters of Algorithm 1. It will also help us recover the bound on estimation entropy \( h_{\text{est}}(\varepsilon) \) of autonomous systems in \([1]\) in Corollary 4.

Proposition 3: For a fixed \( T_p, \delta_x, \) and \( \delta_u \) where \( g_c(\delta_x, \delta_u, T_p) \leq \varepsilon^2 \), the entropy \( h_{\text{est}}(\varepsilon) \) of system \((4)\) is upper bounded by \( g_c(\delta_x, \delta_u, T_p) \) defined to be equal to
\[
\frac{1}{T_p} \left( n \log \left( \frac{\mu}{\delta_x} \right) + m \log \left( \frac{\mu T_p + \eta}{\delta_u} + 1 \right) \right).
\]

Proof: We substitute the upper bound on the cardinality of the minimal approximating set obtained in the previous section in definition of \( h_{\text{est}}(\varepsilon) \) in \((2)\) to get:
\[
\limsup_{T \to \infty} \frac{1}{T} \log s_{\text{est}}(T, \varepsilon) \\
\leq \limsup_{T \to \infty} \frac{1}{T} \log \left( \frac{C_{x,0}||C_{u,0}||C_{x,1}||C_{u,1}|}{T_p} \right) \\
\leq \limsup_{T \to \infty} \frac{1}{T} \log \left( \frac{C_{x,1}||C_{u,1}|}{T_p} \right) \\
\leq \limsup_{T \to \infty} \frac{1}{T} \log \left( \frac{\varepsilon^n \left( \frac{\mu T_p + \eta}{\delta_u} + 1 \right)^{\frac{T}{T_p}}}{T_p} \right).
\]
limsup_{T \to \infty} \frac{T/T_p}{T} n \log \left( \frac{\varepsilon}{\delta_u} \right) + \limsup_{T \to \infty} \frac{T/T_p}{T} m \log \left( \frac{\mu T_p + \eta u}{\delta_u} + 1 \right)

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

We show that if the bound on the input norm is negligible, we recover the bound $\frac{\delta_u}{\mu \eta > 0}$ on entropy derived in [2].

**Corollary 4:** Given any $\varepsilon > 0$, $\lim_{\mu \eta > 0} h_{\text{est}}(\varepsilon) < \frac{n L_x}{\ln 2}.$

**Proof:** Fix any $\varepsilon > 0$. Fix a $\eta_u$ and $\mu$ such that $\eta_u > 0 < \mu < 1$. Then, fix $\delta_u$ to be equal to $\delta_u = \eta_u > 0$. After that, while taking the limits of $\mu$ and $\eta$ going to zero, keep $\delta_u$ fixed to zero, choose $\delta_u = \sqrt{1 - (\mu + \eta) e^{-L_u T_p}}$, and $T_p$ to be such that $g_\varepsilon(\delta_u, \delta_u, 0) = n L_x \log \left( \frac{\varepsilon}{\delta_u} \right) + m \log \left( \frac{\mu T_p + \eta u}{\delta_u} + 1 \right) \leq n L_x \log \left( \frac{\varepsilon}{\delta_u} \right) + m \log \left( \frac{\mu T_p + \eta u}{\delta_u} + 1 \right).$ Moreover, as $\mu$ and $\eta$ decrease to zero, the argument of the second log approaches one and the upper bound approaches $n L_x \log \frac{\varepsilon}{\delta_u}.$

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

**Proof:** The Jacobian $J_x$ of system (18) is equal to zero since the state $x$ does not appear on the RHS of the differential equation. Hence, $M_x = 1/2$ using (23). Moreover, $J_\delta = 1$, and thus, $M_\delta = 1$. Then, there exist $\delta_x$, $\delta_u$, and $T_p > 0$ that satisfy $g_\varepsilon(\delta_x, \delta_u, T_p) \leq \varepsilon^2$, and $h_{\text{est}}(\varepsilon)$ of system (18) is upper bounded by $g_\varepsilon(\delta_x, \delta_u, T_p)$, 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 $v = 10$. 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 $g_\varepsilon$ in (26). Let us fix $\delta_u = \mu = \pi/4$, $T_p = 1.9 \times 10^{-3}$, and $\delta_x = \varepsilon^2 / \mu$. Such an assignment satisfy $g_\varepsilon(\delta_x, \delta_u, T_p) = 0.0099 < \varepsilon^2 = 0.01$ and results in the bound $g_\varepsilon(\delta_x, \delta_u, T_p) = 727.$

Second, we derive the result bound from using the upper bounds on $G_x$ and $G_u$, $M_x$ and $M_u$ of Proposition 2, instead of $G_x$ and $G_u$. The previous choices of $\delta_x$, $\delta_u$, and $T_p$ would result in $g_\varepsilon(\delta_x, \delta_u, T_p) = 0.011 > \varepsilon^2 = 0.01$, which violates the condition in Proposition 3. Instead, we use $\delta_x = \varepsilon^2 / \mu$, $\delta_u = \delta_u = \varepsilon^2 / \mu$, and $T_p = 1.5 \times 10^{-3}$ while keeping $\delta_u = \varepsilon^2 / \mu$. Such an assignment satisfies $g_\varepsilon(\delta_x, \delta_u, T_p) = 0.0098 < \varepsilon^2 = 0.01$ and results in the bound $g_\varepsilon(\delta_x, \delta_u, T_p) = 920$, larger than the one we obtained using $G_x$ and $G_u$.

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

\[ \dot{x}_1 = x_4; \dot{x}_4 = -g \sin x_3 - g \frac{c_x T_p}{m'} + \frac{u_1}{m'} \sin x_3 - \frac{u_2}{m'} \cos x_3 \]
\[ \dot{\dot{x}}_2 = x_5; \dot{x}_5 = g (\cos x_3 - 1) - \frac{c_x T_p}{m'} + \frac{u_1}{m'} \sin x_3 + \frac{u_2}{m'} \cos x_3 \]
\[ \dot{x}_3 = x_6; \dot{x}_6 = \frac{r}{J} u_1 \]

where $(x_1, x_2)$ is the position of the center of mass and $x_3$ is the orientation of the aircraft in the vertical plane, and $(x_4, x_5, x_6)$ are their time derivatives, respectively. The mass of the aircraft is $m'$, the moment of inertia is $J$, the gravitational constant is $g$, and the damping coefficient is $c$.

We fix $u_1$ 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 article [31]. Their values are:

\[ G_x = n \left( \frac{c_x^2}{m'} + \frac{1}{2} \right) \text{ and } G_u \leq \sqrt{m'(c_x^2 + \frac{1}{2}m c_x^2 r)} \]

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 $\delta_x = \frac{1}{\sqrt{2}} e^{-G_x T_p}$. Then, using Proposition 3, $g_\varepsilon(\delta_x, \delta_u, T_p) = 0.249 < \varepsilon^2 = 0.25$, and thus, $h_{\text{est}}(0.5) \leq g_\varepsilon(\delta_x, \delta_u, T_p) = 0.249$. Moreover, $\delta_x = \frac{1}{\sqrt{2}} e^{-G_x T_p}$. Then, $g_\varepsilon(\delta_x, \delta_u, T_p) = 0.244 < \varepsilon^2 = 0.25$ and hence, $h_{\text{est}}(0.5) \leq 0.3680$.

Hence, the bound decreased 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 $\eta$ to 0.1 while allowing large continuous variations by increasing $\mu$ to 20. In that case, the input signals are almost continuous. If we fix $\delta_u = \eta_u = 0.35$, and $\delta_x = \frac{1}{\sqrt{2}} e^{-G_x T_p}$. Then, $g_\varepsilon(\delta_x, \delta_u, T_p) = 0.247 < \varepsilon^2 = 0.25$ 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 [3, Th. 1]. In Section VIII-A, through an example, we present a novel result of this article 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 novel contributions of this article.

We denote a switched system with $N$ modes by

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

where $\sigma : \mathbb{R}_{\geq 0} \to [N]$, is a piece-wise-right-constant signal called a switching signal, $f_{\text{sw}} : \mathbb{R}^n \times [N] \to \mathbb{R}^n$, is globally Lipschitz continuous in the first argument, and $x(t) \in \mathbb{R}^n$. We call $f_{\text{sw}}$ 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 nonswitched systems with inputs, we had the global Lipschitz constant $L_u$, which governs how the system reacts to changes in its input. Here, in (32), we will use a function $d(t)$ that plays a similar role in bounding the difference of behavior resulting from running $f_{sw}$ in different modes.

The points of discontinuities in $\sigma$ are called switches. The switching signal $\sigma$ has a minimum dwell time $T_d > 0$ if at least $T_d$ time units elapse between any two consecutive switches. We fix $T_d$ throughout the section and denote the corresponding set of switching signals as $\Sigma(T_d)$. The initial state $x(0) \in K$, where $K$ is a compact initial set of states in $\mathbb{R}^n$. The state trajectory of system (30) with $\sigma \in \Sigma(T_d)$ is denoted by $\xi_{x,0,\sigma} : [0, \infty) \to \mathbb{R}^n$, 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 $\varepsilon, \alpha, \tau$, and the time interval $T > 0$ are fixed. Second, we redefine the concept of approximating functions. A function $z : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$ is called a $(T, \varepsilon, \alpha, \tau)$-approximating function of $\xi_{x_0,\sigma}$ if for all $j \geq 0$ and for all $t \in [s_j, s_{j+1})$,

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

(31)

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

$$
\mathcal{H}_{\text{est}}(\varepsilon, \alpha, \tau) := \limsup_{T \to \infty} \frac{1}{T} \log s_{\text{est}}(T, \varepsilon, \alpha, \tau).
$$

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

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

(32)

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

This assumption can be checked, for example, if the reachset $\text{Reach}_{sw}(K, T_d, \infty)$ is compact and $\|f_{sw}(x, p)\|$ is finite for any $p$ and any finite $x$. The reason is that $\|f_{sw}(x_1, p_1) - f_{sw}(x_2, p_2)\|$ will be finite for any $x \in \text{Reach}_{sw}(K, T_d, \infty)$. That can be seen by invoking the triangular inequality to get $\|f_{sw}(x_1, p_1) - f_{sw}(x_2, p_2)\| \leq \|f_{sw}(x_1, p_1)\| + \|f_{sw}(x_2, p_2)\|$, which is finite for finite $x_1$ and $x_2$.

Now we restate [3, Th. 1] 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:

$$
\mathcal{H}_{\text{est}}(\varepsilon, \alpha, \tau) \leq \frac{(L_x + \alpha)n}{\ln 2} + \log N \frac{2^T}{T_d},
$$

(33)

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

$$
d(T_d) \leq \varepsilon(1 - e^{-\alpha(T_d - T_0)}).
$$

(34)

**Remark 3:** Following our previous argument from Section V-B for systems with inputs and setting $\alpha = 0$, results in condition (34) on $d(T_d)$ in Theorem 4 becoming $d(T_d) = 0$. In that case, there is no nonzero $T_d$ that satisfies the condition in (34) unless all the modes are identical. Hence, if $\alpha = 0$, then $T_d = 0$ and the entropy upper bound in (33) becomes infinite and, thus, useless. However, with a closer look at [3, proof of Theorem 1], this infinite bound is an artifact of the assumption in (34). In fact, the choices of $\alpha$ and $T_d$ in (34) and the bound (33) can be replaced by $\tilde{\alpha}$ and $\tilde{T}_d$ such that $\tilde{\alpha} \geq \alpha$ and $d(T_d) \leq \varepsilon(1 - e^{-\alpha(T_d - T_0)})$. Such a bound may be finite even in the case when $\alpha = 0$.

### A. Finiteness of Mode Divergence $d(t)$

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

We first define separated sets for switched systems. Two trajectories $\xi_{x_1,\sigma_1}$ and $\xi_{x_2,\sigma_2}$ of switched system (30) are $(T, \varepsilon, \alpha, \tau)$-separated if there exists a $t \in [0, T]$ such that inequality (31) is violated when both: $z$ in the left-hand side (LHS) is replaced by $\xi_{x_1,\sigma_1}$ and $\xi_{x_2,\sigma_2}$, respectively, and $z$ in the LHS is replaced by $\xi_{x_2,\sigma_2}$ and $\xi_{x_1,\sigma_1}$, respectively. A set of trajectories $Z$ is $(T, \varepsilon, \alpha, \tau)$-separated if all pairs of trajectories in $Z$ are $(T, \varepsilon, \alpha, \tau)$-separated.

**Lemma 11:** For all $K, T_d, T, \varepsilon, \alpha, \tau_1, \tau_2$, and $T > 0$,

$$
\frac{1}{T} \log s_{\text{est}}(T, 2\varepsilon, 0, \tau_1) \leq s_{\text{est}}(T, \varepsilon, 0, \tau_2) \leq s_{\text{est}}(T, \varepsilon, \alpha, \tau_2).
$$

(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 $\alpha > 0$, any $(T, \varepsilon, \alpha, \tau)$-approximating set is also a $(T, \varepsilon, 0, \tau)$-approximating one.

Fix arbitrary two constants $a > b > 0$, and consider the functions $f_{sw}(x, 1) = ax$ and $f_{sw}(x, 2) = bx$, and the resulting switched system (30) having them as its modes

$$
\dot{z}(t) = f_{sw}(z(t), \sigma(t))
$$

(36)

where $z : \mathbb{R}_{\geq 0} \to \{1, 2\}$ 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: $|\xi_{x,\sigma}(t) - \xi_{x,b}(t)| = |e^{at} - e^{bt}| |x|$. 

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
Therefore, the integral in (32) is \( \int_0^t |f_{\xi_{\omega}}(\xi_{\omega}(s), a) - f_{\xi_{\omega}}(\xi_{\omega}(s), b)| ds \), which diverges to \( \infty \) with diverging \( x \). Thus, when the sup over the reachset is taken, the argument of the integral is infinite and \( d(t) \) is unbounded, for any \( t > 0 \).

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 \( \text{tseq} \).

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

\[
\left| \frac{2 - bv_1}{4bv_1} \right| (e^{bv_1} - 1) = \left| \frac{|x_0(a - b) - b\varepsilon}{4\varepsilon} \right| (e^{bv_1} - 1). \tag{37}
\]

**Lemma 13**: Fix the constants \( T, \varepsilon, \alpha, \) and \( \tau \). Consider any nonzero initial state \( x_0 \in \mathbb{R} \) and the set of switching signals \( \Sigma_d \). 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, \( |\xi_{\omega}(a) - \xi_{\omega}(b)| \geq |e^{bv_1} - e^{bv_1}| |x_0| > (a - b)|x_0| = 2\varepsilon \). For any switching signal \( \sigma \), the state at time \( t \) is \( \xi_{\omega}(t) \geq x_0 e^{bt} \). Consider any two switching signals \( \sigma_1, \sigma_2 \in \Sigma_d \), corresponding sequences \( s_{\sigma_1} \) and \( s_{\sigma_2} \) with the same prefix up till the \( i \)th entry and differ in the \( (i + 1) \)th one. Finally, let \( x_i = \xi_{\omega_0, \sigma_2}(t_i) \). Then

\[
|\xi_{\omega_0, \sigma_1}(t_i) - \xi_{\omega_0, \sigma_2}(t_i)| = |x_i| e^{aV_0} - e^{bV_0} > |x_i|(a - b)v_i + 1 \\
\geq |x_0| e^{bV_0} - (a - b)v_i + 1 = |x_0| e^{bV_0} - (a - b)v_i e^{-bV_0} = 2\varepsilon.
\]

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

**Theorem 5**: For any \( \varepsilon, \alpha, \) and \( \tau \), \( h_{\text{est}}(\varepsilon, \alpha, \tau) \) of switched system (36) is infinite.

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

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

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

**Proof**: We will need to modify \( \text{tseq}_d \) 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 \( j < \left\lceil \frac{T}{T_d} \right\rceil \). Then, we define \( \text{tseq}_d \) to be the same sequence of time instants of \( \text{tseq}_1 \) in the intervals with odd \( j \), i.e., \([T_d, 2T_d], [3T_d, 4T_d], \ldots \), and no time instances in the intervals with even \( j \), 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 \( \text{tseq}_d \) in each \( T_d \) interval with odd \( j \), 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 \( (T, 2\varepsilon, 0, \tau) \)-separated, by the same proof of Lemma 13. Hence, a lower bound on the number of such signals would result in a lower bound on the maximum cardinality \( s_{\text{est}}(T, 2\varepsilon, \alpha, \tau) \) of system (36).

Now, we will lower bound the number of time instants in \( \text{tseq}_d \) in each \( T_d \) interval in \([0, T] \) with odd index. The number of switching signals that satisfy the minimum dwell time constraint 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 \( \text{tseq}_d \) in that interval. Recall that by definition, \( t_{k+1} - t_k = v_{k+1} = v_k e^{-bt_k} \leq v_1 e^{-bT_d} \) using the same analogy as in (37), the number of time instants in \( \text{tseq}_d \) in \([jT_d, (j + 1)T_d] \) is lower bounded by \( 2O(bv_1/e^{bV_0}) \), which is lower bounded by \( 2O(bv_1/e^{bV_0}) \).

If we take the product of such lower bounds for each odd \( j \), the result is \( O(e^{bT_d} + e^{bT_d} + \ldots + e^{bT_d}) = O(e^{bT_d}) \).

Therefore, \( s_{\text{est}}(T, \varepsilon, \alpha, \tau) \) is still lower bounded by \( O(e^{bT_d}) \) and the entropy \( h_{\text{est}}(\varepsilon, \alpha, \tau) \) is still infinite.

**B. Relating the Upper Bounds on Entropy of Autonomous Switched Systems and Nonswitched Systems With Inputs**

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

**Corollary 6**: Fix any \( \varepsilon \) and \( \rho > 0 \), and let \( \delta_x = \varepsilon^2 e^{-(M_x + \rho)T_p} \) and \( \delta_u \) and \( T_p \) be such that

\[
g_{c,u}(\delta_u, T_p, \rho) \leq \varepsilon^2 \left(1 - e^{-\rho T_p}\right), \tag{38}
\]
where \( g_{c,u} \) is as defined in (26). Then, \( h_{\text{est}}(\varepsilon) \) of system (4) is upper bounded by

\[
(\varepsilon + \rho)n \frac{\ln 2}{T_p} + \log P \frac{\ln P}{T_p},
\]

where \( P = \left[ \frac{n + \mu T_p}{\delta_0} \right] + 1 \) represents the number of possible quantized inputs at each sampling time.

**Proof:** By substituting our choice of \( \delta_x \) in \( g_{c,u} \), we obtain \( g_{c,u}(\delta_x, T_p) = \varepsilon^2 e^{-\mu T_p} \). Then, \( g_c(\delta_x, \delta_u, T_p) = g_{c,u}(\delta_x, T_p) + g_{c,u}(\delta_u, T_p) = \varepsilon^2 \). Hence, the condition on \( g_c \) in Proposition 3 is satisfied. Now, by substituting our choice of \( \delta_x \) in the bound in \( g_{c,u}(\delta_x, \delta_u, T_p) \), along with replacing \( \left[ \frac{n + \mu T_p}{\delta_0} \right] + 1 \) 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 approximations functions constructed by Algorithm 1. Hence, \( P \) affects the bound on \( \hat{h}_{\text{est}} \) of nonswitched systems with inputs similar to how \( N \), the number of modes, affects the bound on \( h_{\text{est}} \) of switched systems.

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

**Lemma 15:** If \( \mu = \eta = 0 \), line 12 in Algorithm 1 is changed to \( \delta_{x,i} \leftarrow B(\hat{z}_i - 1(T_p), \varepsilon e^{-\mu T_p}) \), \( \delta_{x,0} = \varepsilon \), and for all iterations \( t \geq 1 \), \( \delta_{x,i} = \delta_{x,i-1} e^{-\mu T_p} \), then for all \( t \geq 0 \)

\[
\|\hat{x}(t) - \hat{x}_{\text{est}}(t)\| \leq \varepsilon e^{-\mu t}.
\]

The input variation represented by \( \eta \) and \( \mu \) is zero, then \( \delta_0 \) can be chosen to be zero and \( S_{\delta_0} \) to be equal to \( (0) \) for any \( i \geq 0 \). Such a choice of \( \delta_{x,i} \) and \( \delta_{x,0} \) would still satisfy Lemma 9. In addition, it would result in \( g_{c,u}(\delta_x, T_p) = 0 \) implying zero contribution of the input to the uncertainty in the state in (27) and \( g_{c,u}(\delta_x, T_p) = \varepsilon^2 e^{-\mu T_p} \), at each iteration \( i \geq 0 \) of Algorithm 1.

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

**C. Autonomous Switched Systems Modeled as Nonswitched Systems With Inputs**

In this section, we model autonomous switched systems as nonswitched systems with inputs. This modeling allows applying the upper bound on entropy \( h_{\text{est}}(\varepsilon) \) of Proposition 3 to get a novel upper bound on entropy \( h_{\text{est}}(\varepsilon, \alpha, \tau) \), with \( \alpha = 0 \), of the switched system (30). We compare the resulting bound with the bound of Theorem 4.

In our modeling, we embed the \( N \) modes of the switched system in a continuous input space and define a new dynamics function \( \hat{f}_{sw} \) that takes a piece-wise continuous input signal instead of a piece-wise constant one in its second argument. We first define the input space \( U \) to be the standard simplex of dimension \( N - 1 \), i.e., \( U := \{ u \in \mathbb{R}^{N-1} | u[p-1] \geq 0, \sum_{p=2}^{[2,N]} u[p-1] \leq 1 \} \) and \( U_{sw} \) 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 \( g_{c,u} \) in Proposition 3 is an upper bound on \( h_{\text{est}} \).

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

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

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 \( \dot{x} = \alpha f_{sw}(x(t), 1), \ldots, f_{sw}(x(t), N) \), where \( co \) denotes the convex hull, by the Filippov’s Selection Lemma (e.g., [34, Lemma 1]).

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 nonswitched systems with inputs.

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

\[
J_u = \left[ g(x(t), 2), \ldots, g(x(t), N) \right],
\]

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

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, \( G_x \leq n(n L_x + 1/2) \leq n(n \max_{p \in [2,N]} L_p + 1/2) \). Moreover, \( G_u = \sqrt{n} ||J_u|| = \sqrt{n} N - 1 ||J_u|| \) and \( ||J_x|| = \max_{i \in [n]} \sum_{p=2}^{[2,N]} ||(J_{x,p})_i|| \leq \sum_{p=2}^{[2,N]} \sum_{p=2}^{[2,N]} ||g(x(t), p)|| = \sum_{p=2}^{[2,N]} \sum_{p=2}^{[2,N]} ||f_{sw}(x(t), p) - f_{sw}(x(t), 1)||. \)

\( d_{co} := \sqrt{n} N - 1 \) is an upper bound on \( ||J_x|| \) for all \( x \in \text{Reach}(K, L_{sw}, \infty) \). To use the bound on entropy for nonswitched systems with inputs in Proposition 3, we have to represent our input set \( U_{sw} \) in the format of Definition 1. We do that by overapproximating \( U_{sw} \) by the set of input signals \( U_{sw} \) with affine-bounded pointwise variation constructed according to Definition 1 with parameters \( \mu = 0, \eta = 1, i.e., U_{sw} \in (0, 1) \), and initial set of inputs \( U \) being the unit box \([0, 1]^{N-1}\).

The following theorem is a direct application of Corollary 6 to system (40) after setting \( \delta_{x,0} \) to be equal to \( \eta \).

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

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

where \( \rho \) and \( T_p \) satisfy \( (2 M_u e^{M_x T_p})^2 T_p \leq \varepsilon^2 (1 - e^{-\rho T_p}) \), i.e., \( (2 d_{co} e^{n \max_{p \in [2,N]} L_p + 1/2) T_p \leq \varepsilon^2 (1 - e^{-\rho T_p}) \).
Note that there is always $\rho$ and $T_p$ that satisfy the inequality in Theorem 6 since although both sides increase with increasing $T_p$ and the LHS is zero for $T_p = 0$, the LHS is independent of $\rho$ and the RHS strictly increases with $\rho$. The bound is on the entropy of the differential inclusion, which contains the trajectories of the switched system. It grows as $O(n^3)$, $O(\max_{p\in[N]} L_p)$, 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 $T_n$ in Theorem 4, they have a comparable type of bound as we establish that $d(t) \leq \max_{p\in[N]} L_p e^{T_p} t^p$ in the longer version [31]. This shows that our results for nonswitched systems with inputs generalize to wide set of applications, including differential inclusions and switched systems.

IX. CONCLUSION

In this article, 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 [3, Th. 1]. 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 [3, Th. 1] on the difference between the modes of the switched system is needed for a meaningful entropy definition.

ACKNOWLEDGMENT

The authors would like to thank Daniel Liberzon and the anonymous reviewers for providing detailed and insightful comments that enhanced the results and presentation of this article.

REFERENCES

[1] D. Liberzon and S. Mitra, “Entropy and minimal bit rates for state estimation and model detection,” IEEE Trans. Autom. Control, vol. 63, no. 10, pp. 3330–3344, Oct. 2018.
[2] D. Liberzon and S. Mitra, “Entropy and minimal data rates for state estimation and model detections,” in Proc. 19th Int. Conf. Hybrid Syst.: Comput. Control, 2016, pp. 247–256.
[3] H. Sibai and S. Mitra, “Optimal data rate for state estimation of switched nonlinear systems,” in Proc. 20th Int. Conf. Hybrid Syst., Comput. Control, New York, NY, USA, 2017, pp. 71–80. [Online]. Available: https://doi.org/10.1145/3049797.3049799
[4] A. V. Savkin and I. R. Petersen, “Set-value state estimation via a limited capacity communication channel,” IEEE Trans. Autom. Control, vol. 48, no. 4, pp. 676–680, Apr. 2003.
[5] H. Sibai and S. Mitra, “State estimation of dynamical systems with unknown inputs: Entropy and bit rates,” in Proc. 21st Int. Conf. Hybrid Syst.: Comput. Control (Part of CPS Week), New York, NY, USA, 2018, pp. 217–226. [Online]. Available: https://doi.org/10.1109/3178126.3178150
[6] G. N. Nair, R. J. Evans, I. M. Y. Mareels, and W. Moran, “Topological feedback entropy and nonlinear stabilization,” IEEE Trans. Autom. Control, vol. 49, no. 9, pp. 1585–1597, Sep. 2004.
[7] G. N. Nair, F. Fagnani, S. Zampieri, and R. J. Evans, “Feedback control under data rate constraints: An overview,” Proc. IEEE, vol. 95, no. 1, pp. 108–137, Jan. 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,” Syst. Control Lett., vol. 62, pp. 377–381, 2013.
[10] A. Katok and B. Hasselblatt, Introduction to the Modern Theory of Dynamical Systems. Cambridge, U.K.: Cambridge Univ. Press, 1995.
[11] G. Yang and D. Liberzon, “Stabilizing a switched linear system with disturbance by sampled-data quantized feedback,” in Proc. Amer. Control Conf., 2015, pp. 2193–2198.
[12] G. Yang and D. Liberzon, “Finite data-rate stabilization of a switched linear system with unknown disturbance,” IFAC-PapersOnLine, vol. 49, no. 18, pp. 1085–1090, 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 Trans. Autom. Control, vol. 63, no. 7, pp. 2107–2122, Jul. 2018.
[14] M. Rungger and M. Zamani, “Invariance feedback entropy of nondeterministic control systems,” in Proc. 20th Int. Conf. Hybrid Syst., Comput. Control, New York, NY, USA, 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 Syst. Lett., vol. 4, no. 4, pp. 827–832, Oct. 2020.
[16] M. S. Tomar, C. Kawan, P. Jagtap, and M. Zamani, “Numerical estimation of invariance entropy for nonlinear control systems,” 2020, arXiv:2004.04779.
[17] J. Schmidt, “Topological entropy bounds for switched linear systems with lie structure,” M.S. thesis, Elect. and Comput. Eng., University of Illinois at Urbana-Champaign, Champaign, IL, USA, 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,” Math. Control, Signals, Syst., 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 Proc. IEEE Conf. Decis. Control, 2018, pp. 5682–5687.
[20] G. Yang, J. A. P. Hespanha, and D. Liberzon, “On topological entropy and stability of switched linear systems,” in Proc. 22nd ACM Int. Conf. Hybrid Syst., Comput. Control, New York, NY, USA, 2019, pp. 119–127. [Online]. Available: https://doi.org/10.1145/3302504.3311815
[21] G. S. Vicinanza and D. Liberzon, “Estimation entropy for regular linear switched systems,” in Proc. IEEE 58th Conf. Decis. Control, 2019, pp. 5754–5759.
[22] D. Liberzon, “On topological entropy of interconnected nonlinear systems,” IEEE Control Syst. Lett., vol. 5, no. 6, pp. 2210–2214, Dec. 2021.
[23] X. Gao, D. Liberzon, J. Liu, and T. Basar, “Connections between stability conditions for slowly time-varying and switched linear systems,” in Proc. IEEE 54th Conf. Decis. Control, 2015, pp. 2329–2334.
[24] J. P. Hespanha and A. S. Morse, “Stability of switched systems with average dwell-time,” in Proc. IEEE 38th Conf. Decis. Control, 1999, vol. 3, pp. 2655–2660.
[25] H. K. Khalil, Nonlinear Systems, 3rd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 2002.
[26] F. Colonius, “Minimal bit rates and entropy for exponential stabilization,” SIAM J. Control Optim., vol. 50, pp. 2985–3010, 2012.
[27] A. S. Matveev and A. V. Savkin, “Stabilization of linear multiple sensor systems via limited capacity communication channels,” Boston, MA, USA: Birkaüser, 2009, pp. 1–64. [Online]. Available: https://doi.org/10.1007/978-0-8176-4666-2
[28] Z. Huang, C. Fan, and S. Mitra, “Bounded invariant verification for time-delayed nonlinear networked dynamical systems,” Nonlinear Anal.: Hybrid Syst., vol. 23, pp. 211–229, 2017. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1771575016300279
[29] C. Fan, J. Kapinski, X. Jin, and S. Mitra, “Locally optimal reach set overapproximation for nonlinear systems,” in Proc. 15th Int. Conf. Embedded Softw., New York, NY, USA, 2016, pp. 66–70. [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, Art. no. 21. [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 Univ. 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 *Proc. IEEE 59th Conf. Decis. Control*, 2020, pp. 3823–3828.

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

Hussein Sibai received the bachelor’s degree in computer and communication engineering from the American University of Beirut, Beirut, Lebanon, in 2014, and the master’s and Ph.D. degrees in electrical and computer engineering from the University of Illinois Urbana-Champaign (UIUC), Champaign, IL, USA, in 2017 and 2021, respectively. He is currently an Assistant Professor with the Computer Science and Engineering Department, Washington University in St. Louis, St. Louis, MO, USA. He was a Postdoctoral Scholar with University of California, Berkeley, Berkeley, CA, USA. His research interests include formal methods, control theory, and machine learning. Dr. Sibai was the recipient of 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 (Senior Member, IEEE) received the bachelor’s degree in electrical engineering from Jadavpur University, Kolkata, India, in 1999, the master’s degree in computer science from the Indian Institute of Science, Bengaluru, India, in 2001, and the doctorate degree from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2007. He is currently a Professor of electrical and computer engineering with the University of Illinois at Urbana-Champaign, Champaign, IL, USA. His textbook on verification of cyber-physical systems was published by MIT Press in 2021. His research interests include formal methods, hybrid systems, and safe autonomy. Prof. Mitra was the recipient of the National Science Foundation’s CAREER Award, Air Force Office of Scientific Research Young Investigator Program Award, and several best paper awards.