Entropy and Minimal Bit Rates for State Estimation and Model Detection

Daniel Liberzon, *Fellow, IEEE*, and Sayan Mitra, *Senior Member, IEEE*

***Abstract*—We study a notion of *estimation entropy* for continuous-time nonlinear systems, formulated in terms of the number of system trajectories that approximate all other trajectories up to an exponentially decaying error. We also consider an alternative definition of estimation entropy, which uses approximating functions that are not necessarily trajectories of the system, and show that the two entropy notions are equivalent. We establish an upper bound on the estimation entropy in terms of the sum of the desired convergence rate and an upper bound on the matrix measure of the Jacobian, multiplied by the system dimension. A lower bound on the estimation entropy is developed as well. We then turn our attention to state estimation and model detection with quantized and sampled state measurements. We describe an iterative procedure that uses such measurements to generate state estimates that converge to the true state at the desired exponential rate. The average bit rate utilized by this procedure matches the derived upper bound on the estimation entropy, and no other algorithm of this type can perform the same estimation task with bit rates lower than the estimation entropy. Finally, we discuss an application of the estimation procedure in determining, from the quantized state measurements, which of two competing models of a dynamical system is the true model. We show that under a mild assumption of "exponential separation" of the candidate models, detection always happens in finite time.**

***Index Terms*—Estimation, nonlinear systems, quantized systems, topological entropy.**

I. Introduction

ENTROPY is a fundamental notion in the theory of dynamical systems. Roughly speaking, it describes the rate at which the uncertainty about the current state of the system grows as time evolves. One can think of this alternatively as the exponential growth rate of the number of system trajectories distinguishable with finite precision, or in terms of the growth rate of the size of reachable sets. Different entropy definitions (notably, topological and measure-theoretic ones) and relationships between them are studied in detail in the book [18] and in many other sources, and continue to be a subject of active research in the dynamical systems community. The concept of entropy of course also plays a central role in thermodynamics and in information theory (as discussed, e.g., in [10]).

In the context of control theory, if entropy describes the rate at which uncertainty is generated by the system (when no measurements are taken), then it should also correspond to the rate at which information about the system needs to be collected by the controller in order to induce a desired behavior (such as invariance or stabilization). This link has not escaped the control community, and suitable entropy definitions for control systems have been proposed and related to minimal data rates necessary for controlling the system over a communication channel. The first such result was obtained by Nair *et al.* in [27], where topological feedback entropy for discrete-time systems was defined in terms of cardinality of open covers in the state space. An alternative definition was proposed later by Colonius and Kawan in [8], who instead counted the number of "spanning" open-loop control functions. The paper [9] summarized the two notions and established an equivalence between them. Colonius subsequently extended the formulation of [8] for continuous-time dynamics from invariance to exponential stabilization in [7]. The survey [28] provides a broader overview of control under data-rate constraints.

In this work, we are concerned with the problem of estimating the state of a continuous-time system when only quantized and sampled measurements of continuous signals are available to the estimator (which happens, e.g., when state measurements are transmitted via a finite-data-rate communication channel). We do not address control problems here, although such observation problems and control problems are known to be closely related (through duality and the fact that state estimates can be used to close a feedback loop; see the brief discussion at the end of Section V). Observability over finite-data-rate channels and its connection to topological entropy has been studied, most notably by Savkin [32] and more recently by Matveev and Pogromsky [26]. The work [11] is also somewhat related, although it uses a different entropy notion (measure-theoretic entropy) and different channel model (erasure channel). Our point of departure in this paper is a synergy of ideas from Savkin [32] and Colonius [7]. As in [32], we focus on state estimation rather than control. However, we follow [7] in that we consider continuous-time dynamics and require that state estimates converge at a prescribed exponential rate. As a result, the entropy notion with which we work here combines some features of the entropy notions used in [32] and [7].

Manuscript received August 6, 2017; revised October 4, 2017 and November 7, 2017; accepted November 28, 2017. Date of publication December 11, 2017; date of current version September 25, 2018. This work was supported by the Air Force Office of Scientific Research under Grant FA9550-17-1-0236. Recommended by Associate Editor Z. Gao. *(Corresponding author: Daniel Liberzon.)*

The authors are with the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Champaign, IL 61801 USA (e-mail: liberzon@illinois.edu; mitras@illinois.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TAC.2017.2782478

Our first contribution is a definition of *estimation entropy*, formulated in terms of the number of system trajectories that approximate all other trajectories up to an exponentially decaying error. We also consider an alternative definition of entropy, which uses approximating functions that are not necessarily trajectories of the system. We show that the two entropy notions turn out to be equivalent (Theorem 1). We proceed to establish an upper bound of $(M+\alpha)n/\ln 2$ for the estimation entropy of an n-dimensional nonlinear dynamical system whose Jacobian matrix f_x has matrix measure bounded by M, when the desired exponential convergence rate of the estimate is α (Proposition 2). When the system's right-hand side is only Lipschitz but not necessarily differentiable everywhere, a Lipschitz constant L can be used in place of M (as we did in [23]); however, for differentiable systems, the upper bound in terms of M is sharper, as we explain below. We also develop a lower bound of $(\inf \mathrm{tr} f_x + \alpha n)/\ln 2$ on the estimation entropy, where the infimum is taken over the reachable states of the system (Proposition 3). For linear systems, the upper and lower bounds can be refined so that they coincide and give an exact expression for the estimation entropy in terms of the eigenvalues of the system matrix.

Next, we propose an iterative procedure that uses quantized and sampled state measurements to generate state estimates that converge to the true state at the desired exponential rate. The main idea in the algorithm, which borrows some elements from [22] and earlier work cited therein, is to exponentially increase the resolution of the quantizer while keeping the number of bits sent in each round constant. This is achieved by using the quantized state measurement at each round to compute a bounding box for the state of the system for the next round. Then, at the beginning of the next round, this bounding box is partitioned to make a new and more precise quantized measurement of the state. We show that the bounding box is exponentially shrinking in time at a rate α when the average bit rate utilized by this procedure matches the upper bound $(M+\alpha)n/\ln 2$ on the estimation entropy (Theorem 4 and Proposition 5). We also show that no other algorithm of this type can perform the same estimation task with bit rates lower than the estimation entropy (Proposition 6). In other words, the "efficiency gap" of our estimation procedure is at most as large as the gap between the estimation entropy of the dynamical system and the above-mentioned upper bound on it.

In the last part of the paper, we present an application of the estimation procedure in solving a model detection problem. Suppose we are given two competing candidate models of a dynamical system, and from the quantized and sampled state measurements, we would like to determine which one is the true model. For example, the different models may arise from different parameter values or they could model "nominal" and "failure" operating modes of the system. This can be viewed as a variant of the standard system identification or model (in)validation problem (see, e.g., [17], [35]) except, unlike in classical results, which rely on input/output data, here we use quantized state measurements and do not apply a probing input to the system.[1] We demonstrate that under a mild assumption of *exponential separation* of the candidate models' trajectories, a modified version of our estimation procedure can always definitively detect the true model in finite time (Theorem 7). We show that the exponential separation property holds over a compact set if the velocity vectors of the two models are not equal anywhere in that set. Our experiments with an implementation of this model detection procedure on randomly generated affine dynamical systems as well as on a nonlinear example suggest that the algorithm always works in practice, and further illustrate the improvement due to using matrix measures instead of Lipschitz constants.

Preliminary versions of the results of this paper appeared in the conference papers [23] and [24]. The former paper used calculations relying on the system's Lipschitz constant, and the latter refined them with the help of the Jacobian's matrix measure. Compared to [23] and [24], this paper contains a lower bound on the estimation entropy not previously reported; an improved analysis of the exponential separation property; an updated and more detailed simulation study; and complete proofs of all results. We also mention that, following up on our conference papers [23] and [24], an extension of the estimation entropy concept and its analysis to a class of stochastic systems was subsequently considered in [2], while a lower bound on the estimation entropy for measure-preserving maps was independently derived in [19].

II. Preliminaries

In this paper, we work with the continuous-time system

$$\dot{x} = f(x), \qquad x(0) \in K \tag{1}$$

where $x \in \mathbb{R}^n$ is the state, $f : \mathbb{R}^n \to \mathbb{R}^n$ is a C^1 (continuously differentiable) function, and $K \subset \mathbb{R}^n$ is a known compact set of initial states. Let $\xi : K \times [0,\infty) \to \mathbb{R}^n$ denote the map that generates the trajectories or solutions of (1), so that $\xi(x,t)$ is the solution from the initial state x evaluated at time t. According to this notation, $\xi(K,t)$ is the set of states reachable from K at some time t, $\xi(K,[0,T])$ is the set of states reachable from K within some time T, $\xi(K,[0,\infty))$ is the set of all states reachable from K in non-negative time, and so on. We assume that solutions are defined globally in time, i.e., the system (1) is forward complete.[2]

We denote by $|\cdot|$ some chosen norm in $\mathbb{R}^n$. In general definitions and results, this norm can be arbitrary, but in specific quantized algorithm implementations, we will find it convenient to use the ∞-norm $\|x\|_\infty := \max_{1\le i\le n} |x_i|$; in those places, the choice of the ∞-norm will be explicitly declared. For any $x \in \mathbb{R}^n$ and $\delta > 0$, $B(x,\delta) \subseteq \mathbb{R}^n$ is the closed ball of radius δ centered at x, that is, $B(x,\delta) = \{y \in \mathbb{R}^n : \ |x-y| \le \delta\}$; for the ∞-norm, this is a hypercube.

Let $\|\cdot\|$ be the induced matrix norm on $\mathbb{R}^{n\times n}$ corresponding to a chosen norm $|\cdot|$ on $\mathbb{R}^n$. Then, the *matrix measure* $\mu : \mathbb{R}^{n\times n} \to \mathbb{R}$ is defined by

$$\mu(A) := \lim_{\varepsilon \searrow 0} \frac{\|I+\varepsilon A\| - 1}{\varepsilon}$$

[1] For a different line of work where system identification problems are studied using (relative) entropy notions, see [40] and the references therein.

[2] We will later impose a condition on the Jacobian of f guaranteeing that the distance between solutions of (1) grows at most exponentially, and this implies forward completeness.

(see, e.g., [38]). For standard norms, there are explicit formulas for the matrix measure; for example, for the ∞-norm, we have

$$\mu(A) = \max_i \left\{ a_{ii} + \sum_{j \neq i} |a_{ij}| \right\}. \tag{2}$$

One of the basic properties of matrix measures is that for every matrix A, we have

$$\mu(A) \leq \|A\| \tag{3}$$

and we note that the left-hand side of (3) may be negative, while the right-hand side is always positive; see also Example 1. The role that matrix measures will play in our analysis of the nonlinear system (1) is enabled by the following assumption, which we impose throughout the paper, and by the well-known fact stated in Lemma 1.

Assumption 1: The matrix measure of the Jacobian matrix

$$f_x(x) := \frac{\partial f}{\partial x}(x)$$

of f is uniformly bounded: for some $\bar{\mu} \in \mathbb{R}$, we have

$$\mu(f_x(x)) \leq \bar{\mu} \qquad \forall\, x \in \mathbb{R}^n. \tag{4}$$

Example 1: Consider an affine system $\dot{x} = Ax + b$ in $\mathbb{R}^2$ with

$$A = \begin{pmatrix} 0 & 1 \\ -2 & -2 \end{pmatrix}$$

and $b \in \mathbb{R}^2$ arbitrary. With respect to the ∞-norm, we have $\bar{\mu} = \mu(A) = 1$, whereas $\|A\| = 4$. We will use this system in our simulation study in Section VI-C, where we will see the advantages of using the matrix measure instead of the induced norm.

Lemma 1: Consider the system (1) satisfying Assumption 1. Then, for every pair of initial states $x_1, x_2 \in \mathbb{R}^n$, the corresponding solutions of (1) satisfy

$$|\xi(x_1, t) - \xi(x_2, t)| \leq e^{\bar{\mu} t} |x_1 - x_2|$$

for all $t \geq 0$.

From the proof of this result (see, e.g., [3], [36]) it can be seen that if only initial conditions in K are used, then instead of requiring the bound (4) to hold globally over $\mathbb{R}^n$, it is enough to know that it holds for all points x reachable from K at some time $t \geq 0$, provided that K is a convex set (otherwise K must be replaced by its convex hull). Moreover, if all solutions of (1) starting from K remain in a bounded invariant set then $\bar{\mu}$ with the indicated property always exists (by continuity of f_x).

For a bounded set $S \subseteq \mathbb{R}^n$ and $\delta > 0$, a *δ-cover* is a finite collection of points[3] $C = \{x_i\}$ such that $\cup_{x_i \in C} B(x_i, \delta) \supseteq S$. For a hyperrectangle $S \subseteq \mathbb{R}^n$ and $\delta > 0$, a *δ-grid* is a special type of δ-cover of S by hypercubes centered at points along axis-parallel lines that are 2δ apart. The boundaries of the δ-hypercubes centered at adjacent δ-grid points overlap. For a given set S, there are many possible ways of constructing specific δ-grids. We can choose any strategy for constructing them without changing the results in this paper. For example, we can construct a special grid on, say, the unit interval. Then, when working with a general interval I (a cross section of S in any given dimension), we map I to the unit interval, mark the chosen grid on it, and then map it back to I. We denote the δ-grid on S by $\mathrm{grid}(S, \delta)$.

By default, the base of all logarithms is 2. When we use the natural logarithm, we write $\ln$. We use the standard notation tr, det, vol, diam for the trace, determinant, volume, and diameter, respectively.

[3]With a slight abuse of terminology, we take the elements of a cover to be the centers of the balls covering S and not the balls themselves.

III. Estimation Entropy

Let us select a number $\alpha \geq 0$ that defines a desired exponential convergence rate, and let $T > 0$ be a time horizon (which is initially fixed but ultimately approaches ∞). For each $\varepsilon > 0$, we say that a finite set of functions $\hat{X} = \{\hat{x}_1(\cdot), \ldots, \hat{x}_N(\cdot)\}$ from $[0, T]$ to $\mathbb{R}^n$ is *$(T, \varepsilon, \alpha, K)$-approximating* if for every initial state $x \in K$, there exists some function $\hat{x}_i(\cdot) \in \hat{X}$ such that

$$|\xi(x, t) - \hat{x}_i(t)| < \varepsilon e^{-\alpha t} \qquad \forall\, t \in [0, T]. \tag{5}$$

Let $s_{\mathrm{est}}(T, \varepsilon, \alpha, K)$ denote the minimal cardinality of such a $(T, \varepsilon, \alpha, K)$-approximating set. We define *estimation entropy* as

$$h_{\mathrm{est}}(\alpha, K) := \lim_{\varepsilon \searrow 0} \limsup_{T \to \infty} \frac{1}{T} \log s_{\mathrm{est}}(T, \varepsilon, \alpha, K).$$

It is easy to see that instead of $\lim_{\varepsilon \searrow 0}$, we could equivalently write $\sup_{\varepsilon > 0}$, because $s_{\mathrm{est}}(T, \varepsilon, \alpha, K)$ grows as $\varepsilon \to 0$ for fixed T, α, K. Intuitively, since s_{est} corresponds to the minimal number of functions needed to approximate the state with the desired accuracy, h_{est} is the average number of bits needed to identify these approximating functions. The inner $\limsup$ extracts the base-2 exponential growth rate of s_{est} with time and the outer limit gives the worst case over $\varepsilon > 0$.

As a special case, further considered below, we can define $\hat{x}_i(\cdot)$ to be trajectories $\xi(x_i, \cdot)$ of the system from different initial states x_i. Then, s_{est} corresponds to the number of quantization points needed to identify the initial states, and h_{est} gives a measure of the long-term bit rate needed for communicating sensor measurements to the estimator. We pursue this connection in more detail in Section V. We note that the norm in (5) can be arbitrary.

A. Alternative Entropy Notion

In the above-mentioned entropy definition, the functions $\hat{x}_i(\cdot)$ are arbitrary functions of time and not necessarily trajectories of the system (1). If we insist on using system trajectories, then we obtain the following alternative definition: a finite set of points $S = \{x_1, \ldots, x_N\} \subset K$ is *$(T, \varepsilon, \alpha, K)$-spanning* if for every initial state $x \in K$, there exists some point $x_i \in S$ such that the corresponding solutions satisfy

$$|\xi(x, t) - \xi(x_i, t)| < \varepsilon e^{-\alpha t} \qquad \forall\, t \in [0, T]. \tag{6}$$

Letting $s^*_{\mathrm{est}}(T, \varepsilon, \alpha, K)$ denote the minimal cardinality of such a $(T, \varepsilon, \alpha, K)$-spanning set, we could define estimation entropy differently as

$$h^*_{\mathrm{est}}(\alpha, K) := \lim_{\varepsilon \searrow 0} \limsup_{T \to \infty} \frac{1}{T} \log s^*_{\mathrm{est}}(T, \varepsilon, \alpha, K).$$

Since every (T,ε,α,K)-spanning set gives rise to a (T,ε,α,K)-approximating set via $\hat{x}_i(t) := \xi(x_i,t)$, and since entropy is determined by the minimal cardinality of such a set, it is clear that

$$s_{\rm est}(T,\varepsilon,\alpha,K) \le s^*_{\rm est}(T,\varepsilon,\alpha,K) \qquad \forall\, T,\varepsilon,\alpha,K \tag{7}$$

and therefore

$$h_{\rm est}(\alpha,K) \le h^*_{\rm est}(\alpha,K) \qquad \forall\, \alpha,K. \tag{8}$$

Although this might not be obvious, the inequality (8) is actually always equality, as we show next. In other words, there is no advantage—as far as estimation entropy is concerned—in using any approximating functions (even possibly discontinuous ones) other than system trajectories.

Theorem 1: For every $\alpha \ge 0$ and every compact set K, we have $h_{\rm est}(\alpha,K) = h^*_{\rm est}(\alpha,K)$.

Our proof of this result is along the lines of [18, Section 3.1.b] (see also [32, Lemma III.1]) and relies on the notion of separated sets, which we now introduce and which will be needed later as well. With T, ε, α, K given as before, let us call a finite set of points $E = \{x_1,\dots,x_N\} \subset K$ a (T,ε,α,K)-*separated* set if for every pair of points $x_1, x_2 \in E$, the solutions of (1) with these points as initial states have the property that

$$|\xi(x_1,t) - \xi(x_2,t)| \ge \varepsilon e^{-\alpha t} \qquad \text{for some } t \in [0,T]. \tag{9}$$

Let $n^*_{\rm est}(T,\varepsilon,\alpha,K)$ denote the *maximal* cardinality of such a (T,ε,α,K)-separated set. The next two lemmas relate $n^*_{\rm est}$ to the previously defined quantities $s^*_{\rm est}$ and $s_{\rm est}$.[4]

Lemma 2: For all T, ε, α, K, we have

$$s^*_{\rm est}(T,\varepsilon,\alpha,K) \le n^*_{\rm est}(T,\varepsilon,\alpha,K). \tag{10}$$

Proof: The inequality (10) follows immediately from the observation that every maximal (T,ε,α,K)-separated set E is also (T,ε,α,K)-spanning; indeed, if E is not (T,ε,α,K)-spanning, then there exists $x \in K$ such that for every $x_i \in E$, the inequality (6) is violated at least for some t, but then we can add this x to E and the separation property will still hold, contradicting maximality. ■

Lemma 3: For all T, ε, α, K, we have

$$n^*_{\rm est}(T,2\varepsilon,\alpha,K) \le s_{\rm est}(T,\varepsilon,\alpha,K).$$

Proof: Let $\hat{X} = \{\hat{x}_1(\cdot),\dots,\hat{x}_N(\cdot)\}$ be an arbitrary (T,ε,α,K)-approximating set of functions, and let $E = \{x_1,\dots,x_{N'}\}$ be an arbitrary $(T,2\varepsilon,\alpha,K)$-separated set of points in K. We claim that $N' \le N$, which would prove the lemma. By the approximating property of $\hat{X}$, for every $x \in K$, there exists some $\hat{x}_i(\cdot) \in \hat{X}$ such that (5) holds. Suppose that $N' > N$. Then, for at least one function $\hat{x}_i(\cdot) \in \hat{X}$, we can find (at least) two points $x_p, x_q \in E$ such that (5) holds both with $x = x_p$ and with $x = x_q$. By the triangle inequality, this implies $|\xi(x_p,t) - \xi(x_q,t)| < 2\varepsilon e^{-\alpha t}$ for all $t \in [0,T]$. But this contradicts the $(T,2\varepsilon,\alpha,K)$-separating property of E, and the claim is established. ■

[4]We do not define a quantity $n_{\rm est}$ corresponding to separation between arbitrary curves (not necessarily system trajectories) as such a notion does not seem to be useful here.

Proof of Theorem 1: Combining Lemmas 2 and 3 and (7), we obtain for all T,ε,α,K

$$\begin{aligned} n^*_{\rm est}(T,2\varepsilon,\alpha,K) &\le s_{\rm est}(T,\varepsilon,\alpha,K) \\ &\le s^*_{\rm est}(T,\varepsilon,\alpha,K) \le n^*_{\rm est}(T,\varepsilon,\alpha,K). \end{aligned}$$

This implies that

$$\begin{aligned} &\limsup_{T\to\infty} \frac{1}{T} \log n^*_{\rm est}(T,2\varepsilon,\alpha,K) \\ &\le \limsup_{T\to\infty} \frac{1}{T} \log s_{\rm est}(T,\varepsilon,\alpha,K) \\ &\le \limsup_{T\to\infty} \frac{1}{T} \log s^*_{\rm est}(T,\varepsilon,\alpha,K) \\ &\le \limsup_{T\to\infty} \frac{1}{T} \log n^*_{\rm est}(T,\varepsilon,\alpha,K) \end{aligned} \tag{11}$$

for all T,ε,α,K. We can now take the limit as $\varepsilon \to 0$ in (11). This limit always exists (but may be infinite) because all quantities in (11) are monotonically nondecreasing as $\varepsilon \to 0$ (so taking the limit is actually equivalent to taking the supremum over $\varepsilon > 0$). In the limit, the first term and the last term in (11) become the same; hence, all inequalities become equalities. This proves that $h_{\rm est}(\alpha,K) = h^*_{\rm est}(\alpha,K)$, as claimed in Theorem 1. ■

Remark 1: The proof of Theorem 1 shows, in addition, that the two entropy quantities appearing in the statement of Theorem 1 are also equal to $\lim_{\varepsilon\searrow 0} \limsup_{T\to\infty} \frac{1}{T} \log n^*_{\rm est}(T,\varepsilon,\alpha,K)$.

By compactness of K and by the property of continuous dependence of solutions of (1) on initial conditions, for given ε,α,T, there exists $\delta > 0$ such that (6) holds whenever x and x_i satisfy $|x - x_i| < \delta$. From this, it immediately follows that $s^*_{\rm est}(T,\varepsilon,\alpha,K)$, and hence also $s_{\rm est}(T,\varepsilon,\alpha,K)$, is finite for every $\varepsilon > 0$. This does not in principle preclude $h^*_{\rm est}(\alpha,K)$ and $h_{\rm est}(\alpha,K)$ from being infinite (the supremum over positive ε could still be ∞). However, we will see next that this does not happen if the system satisfies Assumption 1.

IV. Entropy Bounds

In this section, we establish an upper bound and a lower bound on the estimation entropy of (1). The upper bound is independent of the choice of the initial set K. The lower bound involves taking an infimum over the set of points reachable from K, but can be made independent of K if the infimum is taken over the whole $\mathbb{R}^n$. Without significant loss of generality, we assume in the sequel that K is a set of positive measure and "regular" shape, such as a hypercube, large enough to contain all initial conditions of interest.

A. Upper Bound

The result given below relies on the global bound $\bar{\mu}$ on the matrix measure of the Jacobian of f provided by Assumption 1. While this assumption is restrictive, we note the following points. First, as we commented after Lemma 1, this can be replaced by a bound over the reachable set, which automatically exists if the reachable set is bounded. Second,

we are not assuming that $\bar{\mu} < 0$, i.e., the system need not be contractive as in [36]. (Note, however, that our upper bound is always non-negative even if $\bar{\mu} < 0$; the entropy itself is of course always non-negative as well, by definition.) Finally, it is useful to compare the entropy bound given here to the one established in [23], which applies to globally Lipschitz (but not necessarily C^1) systems and looks similar but has the Lipschitz constant L of f in place of $\bar{\mu}$. When f is C^1, the bound derived here is sharper because the Lipschitz constant is equal to the induced norm of the Jacobian and so, in light of (3), we have $\bar{\mu} \leq L$.

Proposition 2: For the system (1) satisfying Assumption 1, the estimation entropy $h_{\rm est}(\alpha, K)$ is finite and does not exceed $(M+\alpha)n/\ln 2$, where $M := \max\{\bar{\mu}, -\alpha\}$.

Proof: This proceeds along the lines of the proof in [7, Th. 3.3] (see also [3] and the references therein for earlier results along similar lines). We fix the convergence parameters $\varepsilon > 0, \alpha \geq 0$, the initial set K, and the time horizon $T > 0$, and derive a bound on $s_{\rm est}(T, \varepsilon, \alpha, K)$. Let us consider an open cover C of K with balls of radii $\varepsilon e^{-(M+\alpha)T}$ centered at points $x_1, \ldots, x_N$ (N is the cardinality of the set C). Consider any initial state $x \in K$. By the construction of C, we know that there exists an $x_i \in C$ such that $|x - x_i| \leq \varepsilon e^{-(M+\alpha)T}$. For each $t \leq T$, we have $|\xi(x,t) - \xi(x_i,t)| \leq |x - x_i| e^{\bar{\mu} t} \leq \varepsilon e^{-(M+\alpha)T} e^{\bar{\mu} t} \leq \varepsilon e^{-(M+\alpha)t} e^{Mt} = \varepsilon e^{-\alpha t}$, where the first inequality follows from Lemma 1, the second follows from the construction of C, and the third from the definition of M.

It follows that the cover $C = \{x_1, \ldots, x_N\}$ defines a $(T, \varepsilon, \alpha, K)$-approximating set: $\hat{X} = \{\xi(x_1, \cdot), \ldots, \xi(x_N, \cdot)\}$. That is, $s_{\rm est}(T, \varepsilon, \alpha, K)$ is upper bounded by N, which is the minimum cardinality of the cover of $K \subseteq \mathbb{R}^n$ with balls of radii $\varepsilon e^{-(M+\alpha)T}$. Let $c(\delta, S)$ denote the minimal cardinality of a cover of a set S with balls of radius δ. Then, we can write that $s_{\rm est}(T, \varepsilon, \alpha, K) \leq c(\varepsilon e^{-(M+\alpha)T}, K)$. Next, we proceed to compute a bound on $h_{\rm est}$ as follows:

$$
\begin{aligned}
&\limsup_{T\to\infty} \frac{1}{T} \log s_{\rm est}(T, \epsilon, \alpha, K) \\
&\leq \limsup_{T\to\infty} \frac{1}{T} \log c(\varepsilon e^{-(M+\alpha)T}, K) \\
&= (M+\alpha) \limsup_{T\to\infty} \frac{\log c(\varepsilon e^{-(M+\alpha)T}, K)}{T(M+\alpha)} \\
&= \frac{(M+\alpha)}{\ln 2} \limsup_{T\to\infty} \frac{\ln c(\varepsilon e^{-(M+\alpha)T}, K)}{\ln(e^{(M+\alpha)T}/\varepsilon) + \ln \varepsilon} \\
&= \frac{(M+\alpha)}{\ln 2} \limsup_{T\to\infty} \frac{\ln c(\varepsilon e^{-(M+\alpha)T}, K)}{\ln(e^{(M+\alpha)T}/\varepsilon)} \\
&\quad [\ln \varepsilon \text{ does not affect } \limsup] \\
&= \frac{(M+\alpha)}{\ln 2} \limsup_{\delta \searrow 0} \frac{\ln c(\delta, K)}{\ln(1/\delta)} \quad [\text{defining } \delta := \varepsilon e^{-(M+\alpha)T}] \\
&\leq (M+\alpha)n/\ln 2.
\end{aligned}
$$

The last step follows from the fact that for any $K \subseteq \mathbb{R}^n$, the quantity $\limsup_{\delta \searrow 0} \frac{\ln c(\delta, K)}{\ln(1/\delta)}$, also called the upper box dimension of K, is no larger than (and typically equal to) n; cf., [18, Section 3.2.f]. By taking the limit $\varepsilon \to 0$, we obtain the result $h_{\rm est}(\alpha, K) \leq (M+\alpha)n/\ln 2$. ■

Remark 2: In the case when (1) is a linear system

$$\dot{x} = Ax \tag{12}$$

the result of Proposition 2 can be sharpened. Namely, in this case, one can show that the exact expression (not just an upper bound) for the estimation entropy is

$$
\begin{aligned}
h_{\rm est}(\alpha, K) &= \frac{1}{\ln 2} \sum_{\mathrm{Re}\,\lambda_i(A) > -\alpha} (\mathrm{Re}\,\lambda_i(A) + \alpha) \\
&= \frac{1}{\ln 2} \sum_{\mathrm{Re}\,\lambda_i(A+\alpha I) > 0} \mathrm{Re}\,\lambda_i(A + \alpha I)
\end{aligned} \tag{13}
$$

where $\mathrm{Re}\,\lambda_i(A)$ are the real parts of the eigenvalues of A. This follows from results that are essentially well known, although not well documented in the literature (especially for continuous-time systems); for discrete-time systems, this is shown in [4] as well as in [32]. A detailed proof for the continuous-time case is written down in [33], and its basic outline is as follows. Since the flow is $\xi(x,t) = e^{At}x$, the volume of the reachable set at time T from the initial set K is $\det(e^{AT})\mathrm{vol}(K)$, which by Liouville's trace formula equals $e^{(\mathrm{tr}A)T}\mathrm{vol}(K)$. The decaying factor $e^{-\alpha t}$ on the right-hand side of (5) can be canceled by multiplying by $e^{\alpha t}$ on both sides; the effect of doing this on the left-hand side is that of replacing solutions of $\dot{x} = Ax$ by solutions of $\dot{x} = (A + \alpha I)x$ and suitably modifying the approximating functions. Projecting onto the unstable subspace of $A + \alpha I$, we can refine the trace to be the sum of only unstable eigenvalues of this matrix. The number of approximating functions must be at least proportional to the above-mentioned volume (since the ε-balls around their endpoints must have enough volume to cover the reachable set), and after taking the logarithm, dividing by T, and letting $T \to 0$, we obtain (13) as the lower bound. A similar volume-counting argument will appear in Section IV-B. The upper bound is obtained by reducing A to Jordan normal form followed by an argument similar to the proof of Proposition 2 applied to each Jordan block (with the corresponding eigenvalue replacing M), and ends up giving the same expression (13).

B. Lower Bound

We now derive a lower bound for the estimation entropy, along the lines of [7, Th. 3.2], which gives a lower bound for the control version of entropy considered in that paper. As will be made clear by the results in Section V (see, in particular, Proposition 6), this lower bound is also a lower bound on the bit rate necessary for constructing state estimates that converge to the true state of the system (1) with exponential rate α.

Proposition 3: The estimation entropy of the system (1) satisfies

$$h_{\rm est}(\alpha, K) \geq \Big(\inf_{x \in \xi(K, [0,\infty))} \mathrm{tr}\, f_x(x) + \alpha n \Big) / \ln 2.$$

Proof: We will derive a lower bound on the size of any spanning set, from which we will obtain the desired lower bound on the estimation entropy. (An argument based on approximating

sets is completely analogous.) Recall that for a (T,ε,α,K)-spanning set $S=\{x_1,\dots,x_N\}$, the balls $B(\xi(x_i,t),\varepsilon e^{-\alpha t})$ cover $\xi(K,t)$ for each $t\le T$. Thus, a lower bound on the smallest size $s^*_{\rm est}$ of such a spanning set is obtained by dividing the volume of $\xi(K,T)$ by the volume of each of these (identical) balls:

$$s^*_{\rm est}(T,\varepsilon,\alpha,K)\ge\frac{\mathrm{vol}(\xi(K,T))}{\mathrm{vol}(B(\xi(x_i,T),\varepsilon e^{-\alpha T}))}=\frac{\mathrm{vol}(\xi(K,T))}{(2\varepsilon e^{-\alpha T})^n} \tag{14}$$

where in the last step, we used the ∞-norm for concreteness. Now, we proceed to obtain a lower bound on the term in the numerator. We want to know how the volume of K propagates with time along solutions. If we let $y:=\xi(x,T)$, then

$$\mathrm{vol}(\xi(K,T))=\int_{\xi(K,T)}dy$$

and by the well-known formula for change of integration variables, this equals (we denote $\partial\xi/\partial x$ by ξ_x)

$$\int_K|\det\xi_x(x,T)|dx.$$

This can in turn be lower-bounded by

$$\inf_{x\in K}|\det\xi_x(x,T)|\cdot\mathrm{vol}(K)$$

and now we need to know how the determinant in the previous formula evolves with time. We have

$$\dot\xi(x,t)=f(\xi(x,t)),\qquad \xi(x,0)=x$$

or

$$\xi(x,t)=x+\int_0^t f(\xi(x,s))ds.$$

Taking partials with respect to x, we obtain

$$\xi_x(x,t)=I+\int_0^t f_x(\xi(x,s))\xi_x(x,s)ds$$

(recall that f is assumed to be C^1). This shows that ξ_x satisfies the matrix differential equation

$$\frac{d}{dt}\xi_x(x,t)=f_x(\xi(x,t))\xi_x(x,t)$$

which, in view of the initial condition $\xi_x(x,0)=I$, means that $\xi_x(x,t)$ is the state transition matrix for the linear time-varying system

$$\dot z(t)=f_x(\xi(x,t))z(t)$$

(known as the variational equation for (1); see, e.g., [21, Section 4.2.4]). Applying the well-known Abel–Jacobi–Liouville theorem (see, e.g., [5, Theorem 4.1]), we deduce that

$$\det\xi_x(x,t)=e^{\int_0^t\mathrm{tr}f_x(\xi(x,s))ds}. \tag{15}$$

Next, we write

$$\begin{aligned}\mathrm{vol}(\xi(K,T))&\ge\mathrm{vol}(K)\cdot\inf_{x\in K}e^{\int_0^T\mathrm{tr}f_x(\xi(x,s))ds}\\&=\mathrm{vol}(K)\cdot e^{\inf_{x\in K}\int_0^T\mathrm{tr}f_x(\xi(x,s))ds}\\&\ge\mathrm{vol}(K)\cdot e^{T\cdot\inf_{x\in\xi(K,[0,\infty))}\mathrm{tr}f_x(x)}\\&\quad[\xi(K,[0,\infty))\text{ accounts for all }\xi(x,s)].\end{aligned} \tag{16}$$

Thus, from the inequality (14), we conclude that

$$s^*_{\rm est}(T,\varepsilon,\alpha,K)\ge\frac{\mathrm{vol}(K)\cdot e^{T\cdot\inf_{x\in\xi(K,[0,\infty))}\mathrm{tr}f_x(x)}}{(2\varepsilon e^{-\alpha T})^n}.$$

Taking logs and dividing by T gives

$$\frac{1}{T}\log s^*_{\rm est}\ge\frac{1}{T}\log\left(\frac{\mathrm{vol}(K)}{2^n\varepsilon^n}\right)+\left(\inf_{x\in\xi(K,[0,\infty))}\mathrm{tr}f_x(x)+\alpha n\right)/\ln 2.$$

Finally, by taking $\limsup$ as $T\to\infty$ and $\lim$ as $\varepsilon\to 0$, we obtain the stated lower bound on the estimation entropy $h^*_{\rm est}(\alpha,K)=h_{\rm est}(\alpha,K)$. ■

Note that the lower bound becomes $-\infty$ if $\mathrm{tr}f_x(x)$ does not have a finite lower bound over $\xi(K,[0,\infty))$. Our lower bound is also not very useful for conservative or dissipative systems, i.e., when $\mathrm{tr}f_x(x)\le 0$ (at least on some parts of the state space). As shown in [26], in a neighborhood of a hyperbolic equilibrium, it is possible to restrict the analysis to the unstable manifold and use essentially the same volume-counting argument as in the proof of Proposition 3 to obtain a sharper bound. A more advanced lower bound was recently derived in [19], although it is also more difficult to evaluate in general.

V. Estimation Over Infinite Horizon

We will first describe a procedure for state estimation of the system (1) over infinite time horizon. Next, we will show that the output from this estimation procedure exponentially converges to the actual state of the system. Finally, we will prove a bound on the bit rate that is sufficient to achieve this convergence. This is a measure of the rate at which information has to be communicated from the sensors of the plant to the estimator.

A. Estimation Procedure

From this point on in this section, we will discuss a specific estimation procedure based on quantized state measurements. The norm used here will be the infinity norm $\|\cdot\|_\infty$. Accordingly, the $B(x,\delta)$ balls will be the hypercubes and the grids will be sets of hypercubes. We will treat all previous definitions and results related to entropy in terms of the infinity norm.

The estimation procedure computes a function $v:[0,\infty)\to\mathbb{R}^n$ and an exponentially shrinking envelope around $v(t)$ such that the actual state of the system $\xi(x,t)$ is guaranteed to be within this envelope. It has several inputs:

1) a sampling period $T_p>0$,
2) a desired exponential convergence rate $\alpha\ge 0$,

3) an initial set K and an initial partition size $d_0 > 0$,
4) the constant M defined in Proposition 2, and
5) a subroutine for computing solutions of the differential equation (1).

In this paper, we do not distinguish between this subroutine for computing solutions and the actual solutions $\xi(\cdot,\cdot)$. The procedure works in rounds $i = 1, 2, ...$, and each round lasts T_p time units. In each round, a new state measurement q is obtained, and the values of three state variables S, δ, C are updated. We denote these updated values in the ith round as q_i, δ_i, S_i, and C_i. Roughly, $S_i \subseteq \mathbb{R}^n$ is a hypercubic over-approximation of the set of reachable states, δ_i is the radius of the set S_i, and C_i is a grid on S_i, which defines the set of possible state measurements q_{i+1} for the next round. If we are in a situation where the quantized state measurements q_i are being transmitted from the sensors to the estimator via a finite-data-rate communication channel, then the variables δ_i, S_i, and C_i need to be generated independently and synchronously on both sides of the channel.

The initial values of these state variables are $\delta_0 = d_0$; S_0 is a hypercube with center, say, x_c and radius $r_c = \frac{\mathrm{diam}(K)}{2}$, such that $K \subseteq B(x_c, r_c)$; and $C_0 = \mathrm{grid}(S_0, \delta_0 e^{-(M+\alpha)T_p})$. Recall the definition of a grid cover from Section II: C_0 is a specific collection of points in $\mathbb{R}^n$ such that $S_0 \subseteq \cup_{x\in C_0} B(x, \delta_0 e^{-(M+\alpha)T_p})$.

At the beginning of the ith round, the algorithm takes as input (from the sensors) a measurement q_i of the current state of the system with respect to the cover C_{i-1} computed in the previous round. The measurement q_i is obtained by choosing a grid point $c \in C_{i-1}$ such that the corresponding $\delta_{i-1}e^{-(M+\alpha)T_p}$-ball $B(c, \delta_{i-1}e^{-(M+\alpha)T_p})$ contains the current state $\xi(x, iT_p)$ of the system. (If there are multiple grid points satisfying this condition—and this may happen as C_{i-1} is a cover with closed sets having overlapping boundaries—then one is chosen arbitrarily.) Using this measurement, the algorithm computes the following.

1) $v_i : [0, T_p] \to \mathbb{R}^n$, which is an approximation function for the state over the interval spanning this round, defined as the solution of the system (1) from q_i,
2) δ_i is updated as $e^{-\alpha T_p}\delta_{i-1}$,
3) $S_i \subseteq \mathbb{R}^n$ is a set containing the state after T_p time, that is, at the beginning of round $i + 1$, and
4) C_i is a $\delta_i e^{-(M+\alpha)T_p}$-grid on S_i.

Specifically, S_i is computed by first evaluating the solution $v_i(T_p) = \xi(q_i, T_p)$ of the system starting from q_i after time T_p, and then constructing the hypercube $B(v_i(T_p), \delta_i)$. Note that for $\alpha > 0$, the size of this hypercube decays geometrically at the rate $e^{-\alpha T_p}$ with each successive round. Recall Section II where we defined grids and discussed specific ways of constructing them; here, the specific construction is less important than the fact that each C_i can be computed from q_i by translating and scaling C_{i-1}.

Consider the beginning of the ith round for some $i > 0$. From the algorithm, it follows that if the current state x is contained in the set S_{i-1} computed in the last iteration, then the measurement q_i is one of the points in the cover C_{i-1} computed in the last iteration, and further, the error in the measurement $|q_i - x|$ is at most the precision of the cover, which is $\delta_{i-1}e^{-(M+\alpha)T_p}$. This property will be used in the analysis below.

```
1  input: T_p, α, K, d_0, M, ξ(·,·)
2  i = 0;
3  δ_0 = d_0;
4  S_0 = B(x_c, r_c);   // x_c is the center of K
5  C_0 = grid(S_0, δ_0 e^{-(M+α)T_p});
6  do   //loop; at i^th round, i > 0
7     i = i + 1;
8     input q_i ∈ C_{i-1};
9     // measurement of current state
10    v_i(·) = ξ(q_i, ·)|[0, T_p];
11    δ_i = e^{-αT_p} δ_{i-1};
12    S_i = B(v_i(T_p), δ_i);
13    C_i = grid(S_i, δ_i e^{-(M+α)T_p});
14    output S_i ⊆ R^n, C_i, v_i : [0, T_p] → R^n;
15    wait(T_p);
```

Fig. 1. Estimation procedure.

Remark 3: Line 10 of the estimation procedure uses a subroutine for computing numerical solutions of the differential equation (1) from a given quantized initial state q_i over a fixed time horizon T_p. In this paper, we assume that these computations are precise. Extending the algorithms and results to accommodate numerical imprecisions would proceed along the lines of the techniques used in numerical reachability computations (for example, in [12] and [20]). The present case, however, is significantly simpler as the solutions have to be computed from a single initial state and up to a fixed time horizon.

In order to analyze the accuracy of this estimation procedure, we define a piecewise continuous estimation function $v : [0, \infty) \to \mathbb{R}^n$ by $v(0) := v_1(0)$ and

$$v(t) = v_i(t - (i-1)T_p) \quad \forall\, t \in ((i-1)T_p, iT_p],\ i = 1, 2, \ldots \tag{17}$$

The next theorem establishes an exponentially decaying upper bound on the error between the actual state of the system and the computed approximating function.

Theorem 4: For any choice of the parameters $\alpha \geq 0$ and $d_0, T_p > 0$, the procedure in Fig. 1 has the following properties: for $i = 0, 1, 2, \ldots$ and for any initial state $x \in K$

$$(a)\ \xi(x,t) \in S_i \text{ for each } t = iT_p, \text{ and}$$
$$(b)\ \|\xi(x,t) - v(t)\|_\infty \leq d_0 e^{-\alpha t} \quad \forall\, t \in [iT_p, (i+1)T_p).$$

Proof: Part (a): We fix $x \in K$ and proceed to prove the statement by induction on the iteration index i. The base case: $i = 0$, that is, $t = 0$ and $\xi(x, 0) = x$. The required condition follows since $x \in K \subseteq B(x_c, r_c) = S_0$. For the inductive step, we assume that $\xi(x, iT_p) \in S_i$ and have to show that $\xi(x, (i+1)T_p) \in S_{i+1}$. We proceed by establishing an upper bound on the distance between the actual trajectory of the

system at $t = (i+1)T_p$ and the computed approximation $v(t)$:

$$\begin{aligned}
&\|\xi(x,(i+1)T_p) - v((i+1)T_p)\|_\infty \\
&= \|\xi(\xi(x,iT_p),T_p) - v_{i+1}(T_p)\|_\infty \\
&\text{[from (17) defining } v(t)\text{]} \\
&= \|\xi(\xi(x,iT_p),T_p) - \xi(q_{i+1},T_p)\|_\infty \qquad (18) \\
&\text{[from line 10 } v_{i+1}(T_p) = \xi(q_{i+1},T_p)\text{]} \\
&\le e^{MT_p}\|\xi(x,iT_p) - q_{i+1}\|_\infty \quad \text{[from Lemma 1].} \qquad (19)
\end{aligned}$$

The measurement q_{i+1} is the input received at the beginning of round $i+1$ for the actual state $\xi(x,iT_p)$ with respect to the cover C_i of S_i. From the induction hypothesis we know that $\xi(x,iT_p) \in S_i$, and therefore, $q_{i+1} \in C_i$. Since C_i is a $\delta_i e^{-(M+\alpha)T_p}$-cover of S_i, it follows that

$$\|\xi(x,iT_p) - q_{i+1}\|_\infty \le \delta_i e^{-(M+\alpha)T_p}. \qquad (20)$$

We have $\|\xi(x,(i+1)T_p) - v((i+1)T_p)\|_\infty \le \delta_i e^{-(M+\alpha)T_p} e^{MT_p} = \delta_i e^{-\alpha T_p} = \delta_{i+1}$ (by definition of δ_{i+1}). Thus, it follows that $\xi(x,(i+1)T_p) \in B(v((i+1)T_p),\delta_{i+1}) = S_{i+1}$.

Part (b): We fix an iteration index $i \ge 0$ and an initial state $x \in K$. If $t = iT_p$ then the result follows from part (a) because $\delta_i = d_0 e^{-\alpha i T_p}$. For any $t \in (iT_p,(i+1)T_p)$, we establish an upper bound on the distance between the actual trajectory $\xi(x,t)$ of the system at time t and the computed approximation $v(t)$:

$$\begin{aligned}
&\|\xi(x,t) - v(t)\|_\infty = \|\xi(\xi(x,iT_p),t-iT_p) - v_{i+1}(t-iT_p)\|_\infty \\
&\qquad \text{[from (17) defining } v(t)\text{]} \\
&= \|\xi(\xi(x,iT_p),t-iT_p) - \xi(q_{i+1},t-iT_p)\|_\infty \\
&\qquad \text{[from } v_{i+1}(t) = \xi(q_{i+1},t)\text{]} \\
&\le \|\xi(x,iT_p) - q_{i+1}\|_\infty e^{M(t-iT_p)} \quad \text{[from Lemma 1]} \\
&\le \delta_i e^{-(M+\alpha)T_p} e^{M(t-iT_p)} \quad \text{[from (20)]} \\
&= d_0 e^{-\alpha i T_p} e^{-(M+\alpha)T_p} e^{M(t-iT_p)} \quad \text{[from } \delta_i = d_0 e^{-\alpha i T_p}\text{]} \\
&= d_0 e^{-\alpha(i+1)T_p} e^{M(t-(i+1)T_p)} \\
&\le d_0 e^{-\alpha t} \quad \text{[since } iT_p \le t \le (i+1)T_p\text{]}.
\end{aligned}$$

■

B. Bit Rate of Estimation Scheme and Its Relation to Entropy

Now, we estimate the communication bit rate needed by the estimation procedure in Fig. 1. As the states S_{i-1} and C_{i-1} are maintained and updated by the algorithm in each round, the only information that is communicated from the system to the estimation procedure in each round is the measurement q_i. The number of bits needed for that is $\log(\#C_i)$, where $\#$ stands for the cardinality of a set. The long-term average bit rate of the algorithm is given by

$$b(\alpha,d_0,T_p) := \limsup_{j\to\infty} \frac{1}{jT_p}\sum_{i=1}^{j} \log(\#C_{i-1}).$$

We proceed to characterize this quantity from the description of the estimation procedure in Fig. 1. We calculate $\#C_0 = \lceil \frac{\mathrm{diam}(K)}{2d_0 e^{-(M+\alpha)T_p}} \rceil^n$. For each successive iteration i, $\#C_i = \lceil \frac{\delta_i}{\delta_i e^{-(M+\alpha)T_p}} \rceil^n = \lceil e^{(M+\alpha)T_p} \rceil^n$. Thus, $b(\alpha,d_0,T_p) = \lim_{i\to\infty} \frac{1}{T_p}\log(\#C_i) = (M+\alpha)n/\ln 2$ is the bit rate utilized by the procedure for any d_0 and T_p. Since it is independent of d_0 and T_p, we write it as $b(\alpha)$ from now on. We state our conclusion as follows.

Proposition 5: The average bit rate used by the estimation procedure in Fig. 1 is $(M+\alpha)n/\ln 2$, where M is defined in Proposition 2.

By Proposition 2, the bit rate $(M+\alpha)n/\ln 2$ used by the algorithm is an upper bound on the entropy $h_{\mathrm{est}}(\alpha,K)$. We now establish that no other similar algorithm can perform the same task with a bit rate lower than the entropy $h_{\mathrm{est}}(\alpha,K)$. In other words, the "efficiency gap" of the algorithm is at most as large as the gap between the entropy and its upper bound known from Proposition 2. (Incidentally, combining this result with Proposition 5, we can arrive at an alternative proof of Proposition 2.)

In order to state this result, we need to formalize the class of algorithms to which it applies and to which our algorithm also belongs. As before, assumed given are the system (1), the associated constant M, and initial set K, as well as the desired estimation parameters d_0 (initial bound) and α (convergence rate). We also select the sampling period T_p, which we can think of as a design parameter in the algorithm. It is convenient to consider an encoder (collocated with the system) and a decoder (possibly, but not necessarily, residing at a remote location and connected to the encoder via a communication channel.) On the encoder side, at each step i (corresponding to time $t = (i-1)T_p$), a codeword q_i from a finite set (coding alphabet) C_i is generated based on the state behavior history up to this time. On the decoder side, using this codeword and the previously received codewords, an estimate $v(\cdot)$ of the state over the next sampling interval $((i-1)T_p, iT_p]$ is defined. Such encoding–decoding schemes are by now quite standard (cf., [32, Section 2] and the references therein).

The lower bound on the bit rate in terms of entropy is proved below for an algorithm that uses a constant number of bits at each round; since in our estimation algorithm $\#C_0$ may be higher than $\#C_i$ for $i \ge 1$, we can think of this comparison as being valid once the algorithm has reached "steady state."

Proposition 6: Consider an algorithm of the above-mentioned type such that at each step i, the set C_i has the same number of elements: $\#C_i = N \ \forall\, i$ (i.e., the coding alphabet is of fixed size). If this algorithm achieves the properties listed in Theorem 4 for an arbitrary $d_0 > 0$, then its bit rate cannot be smaller than $h_{\mathrm{est}}(\alpha,K)$.

Proof: This proof follows along the same lines as the proof in [32, Statement 1, Th. III.1]. Here, the choice of norm does not matter, so we revert to an arbitrary norm $|\cdot|$ on $\mathbb{R}^n$. Seeking a contradiction, suppose that an algorithm achieves the properties listed in Theorem 4 and has a bit rate $b(\alpha) < h_{\mathrm{est}}(\alpha,K)$. Recall (see the proof of Lemma 3 and Remark 1) that

$$h_{\text{est}}(\alpha,K)=\lim_{\varepsilon\searrow 0}\limsup_{T\to\infty}\frac{1}{T}\log n^*_{\text{est}}(T,2\varepsilon,\alpha,K)$$
$$=\sup_{\varepsilon>0}\limsup_{T\to\infty}\frac{1}{T}\log n^*_{\text{est}}(T,2\varepsilon,\alpha,K).$$

Thus, for some $\varepsilon>0$ small enough and some $\bar{b}>b(\alpha)$, we have

$$\bar{b}<\limsup_{T\to\infty}\frac{1}{T}\log n^*_{\text{est}}(T,2\varepsilon,\alpha,K).$$

Let d_0 be equal to this ε. Next, we can find a sufficiently large T for which

$$\bar{b}<\frac{1}{T}\log n^*_{\text{est}}(T,2\varepsilon,\alpha,K) \tag{21}$$

and, moreover

$$\frac{T}{T+T_p}>\frac{b(\alpha)}{\bar{b}} \tag{22}$$

where T_p is the sampling period in the algorithm [note that the left-hand side of (22) tends to 1 as $T\to\infty$, while the right-hand side is smaller than 1]. Let ℓ be the positive integer such that $T\in(\ell-1)T_p,\ell T_p]$. Then, it is easy to see that

$$b(\alpha)<\frac{1}{\ell T_p}\log n^*_{\text{est}}(T,2\varepsilon,\alpha,K)<\frac{1}{\ell T_p}\log n^*_{\text{est}}(\ell T_p,2\varepsilon,\alpha,K)$$

where the first inequality relies on (21) and (22) and the second inequality follows from the fact that every $(T,2\varepsilon,\alpha,K)$-separated set is also $(\ell T_p,2\varepsilon,\alpha,K)$-separated. Since the bit rate of the algorithm is given by

$$b(\alpha)=\frac{1}{T_p}\log N$$

we obtain

$$N^\ell<n^*_{\text{est}}(\ell T_p,2\varepsilon,\alpha,K).$$

The left-hand side of this inequality is the number of possible sequences of codewords $\{q_i\}$ that can be produced by the algorithm over ℓ rounds, while the right-hand side is the cardinality of a maximal $(\ell T_p,2\varepsilon,\alpha,K)$-separated set. This means that there must exist two different initial conditions x_1,x_2 in this $(\ell T_p,2\varepsilon,\alpha,K)$-separated set such that the corresponding solutions $\xi(x_1,t),\xi(x_2,t)$ will produce the same sequence of q_i, and hence will be approximated within $\varepsilon e^{-\alpha t}$ by the same approximating function $v(t)$:

$$|\xi(x_i,t)-v(t)|<\varepsilon e^{-\alpha t}\qquad\forall\, t\in[0,\ell T_p],\quad i=1,2. \tag{23}$$

On the other hand, by the definition of a $(\ell T_p,2\varepsilon,\alpha,K)$-separated set, it must hold that

$$|\xi(x_1,t)-\xi(x_2,t)|\geq 2\varepsilon e^{-\alpha t}\qquad\text{for some } t\in[0,\ell T_p]$$

which contradicts (23) in view of the triangle inequality. ■

We note that the algorithm described in [32] performs a similar estimation task (with $\alpha=0$ and in discrete time) and operates at an arbitrary bit rate above the entropy. However, that algorithm is quite abstract, since it relies on the existence of a suitable spanning set and performs block coding over a sufficiently large time window using sequences from this spanning set. By contrast, our algorithm given in Section V-A is constructive in that it utilizes a specific quantization procedure and works with an arbitrary fixed sampling period.

Remark 4: For the case of a linear system (12), the algorithm of Section V-A can be modified so that its average bit rate equals the entropy of the linear system given by (13). This can be achieved by aligning the grids C_i used in the algorithm with eigenvectors of the matrix A and replacing the constant M with eigenvalues of A (i.e., using a different number of quantization points for each dimension). Constructions of this type for linear systems are well established in the literature; see, e.g., [16], [37].

Before finishing this section, we briefly mention that, with minor changes, our state estimation algorithm can be adopted for feedback stabilization of a control system $\dot{x}=f(x,u)$, along the lines of [22]. The main idea is as follows. Suppose that a nominal state feedback law $k(\cdot)$ is given such that the system $\dot{x}=f(x,k(x))$ is asymptotically stable. In the absence of precise state measurements, the estimates generated by the state estimation algorithm can be used instead. Namely, at the ith round, we would redefine $v_i(\cdot)$ to be the solution of the system $\dot{x}=f(x,u)$ with initial state q_i and control $u(t)=k(v_i(t))$. The rest of the procedure stays the same, and the bit rate that it uses remains unchanged. The resulting closed-loop system takes the form $\dot{x}=f(x,k(x+e))$, where e is the state estimation error, which, as in Theorem 4, exponentially converges to 0. If this system satisfies a suitable robustness assumption with respect to e (i.e., input-to-state stability as in [22] or its integral version as in [29]), then we can conclude that it is asymptotically stabilized.

VI. Model Detection

In this section, we show that the estimation algorithm in Fig. 1 can be used to distinguish two system models, provided they are in some sense adequately different. More precisely, a slightly modified procedure will solve this model detection problem while at the same time performing the state estimation task in the same way as before.

Consider two continuous-time system models:

$$\dot{x}=f_1(x),\quad x\in\mathbb{R}^n, \tag{24}$$
$$\dot{x}=f_2(x),\quad x\in\mathbb{R}^n \tag{25}$$

where the initial state is in the known compact set $K\subset\mathbb{R}^n$ and f_1 and f_2 are C^1 functions satisfying Assumption 1, with respective constants M_1 and M_2 defined as in Proposition 2 (see also the comments immediately before that proposition). We denote the trajectories of the systems (24) and (25) by $\xi_1(x,t)$ and $\xi_2(x,t)$, respectively. From runtime data consisting of quantized and sampled measurements of x as before, we are interested in distinguishing whether the true dynamics of the system is f_1 or f_2. For example, if f_1 and f_2 correspond to models with different sets of parameter values, then solutions to this problem could be used for model parameter identification. As another example application, consider a scenario where f_1 captures the nominal dynamics of the system and f_2 models a known aberration or failure mode. Then, a solution to the above-mentioned detection problem can be used for failure detection. It is straightforward

```
1   input: T_p, α, K, d_0, M_1, ξ_1(·,·)
2   i = 0;
3   δ_0 = d_0;
4   S_0 = B(x_c, r_c);
5   C_0 = grid(S_0, δ_0 e^{-(M_1+α)T_p});
6   do  //loop; at i^th round, i > 0
7       i = i + 1;
8       if current state ∉ S_{i-1}
9           output ''second model'';
10          break;
11      else
12          input q_i ∈ C_{i-1};
13          v_i(·) = ξ_1(q_i, ·)|[0, T_p];
14          δ_i = e^{-αT_p} δ_{i-1};
15          S_i = B(v_i(T_p), δ_i);
16          C_i = grid(S_i, δ_i e^{-(M_1+α)T_p});
17      wait(T_p);
```

Fig. 2. Procedure for detecting models.

to generalize the solution proposed below to handle multiple competing models.

For the purpose of obtaining a provably correct model distinguishing algorithm, we introduce the following concept. For $M_s, T_s > 0$, we say that the two models are (M_s, T_s)-*exponentially separated* (locally) if there exists a constant $\varepsilon_{\min} > 0$ such that for any $\varepsilon \le \varepsilon_{\min}$ and any two states $x_1, x_2 \in \mathbb{R}^n$ with $|x_1 - x_2| \le \varepsilon$, we have

$$|\xi_1(x_1, T_s) - \xi_2(x_2, T_s)| > \varepsilon e^{M_s T_s}. \tag{26}$$

We describe and analyze our algorithm for distinguishing models in Section VI-A, and postpone a more detailed discussion of the exponential separation property and conditions for checking it until Section VI-B.

A. Distinguishing Algorithm

In the above-mentioned definition of exponential separation, the norm can be arbitrary, but in the algorithm below, we work with the infinity norm. With some modifications, the procedure in Fig. 1 can detect models using quantized state observations. In Fig. 2, we show the procedure for detecting models. First of all, before taking the measurement in each round (T_p time) it makes an additional check. If the current state is not in the set S_i (line 8) computed from the previous round, then the procedure immediately halts by detecting the second model. If the current state is in S_i, then it proceeds as before and records a measurement q_i of the current state as one of the points in the cover C_i. Second, the function v_i (line 13) is now computed as a solution $\xi_1(q_i, \cdot)$ of the system given by (24). Finally, in computing the radius of the elements in the cover C_i (line 16), the constant M_1 of the system (24) is used.

Theorem 7: Suppose that the true system model is either (24) or (25) and that the two models are (M_1, T_p)-exponentially separated. Then, for any choice of the parameters $\alpha, d_0, T_p > 0$, the procedure in Fig. 2 outputs "second model" if and only if the true model is (25).

Proof: For the "if" part, assume that the true model is the second model, that is, given by (25). Fixing an initial state of the system x_0, we have the true trajectory $\xi_2(x_0, \cdot)$. Let us also fix the parameters T_p, d_0, α of the detection algorithm. Since the value of the program variable $\delta_i = d_0 e^{-\alpha i T_p}$ decays geometrically in each iteration (note that here we take $\alpha > 0$), there exists an i^* such that for any iteration $k - 1 \ge i^*$, $\delta_{k-1} e^{-(M_1+\alpha)T_p} \le \varepsilon_{\min}$. We consider the execution of the algorithm at one such iteration $k - 1$ and show that the condition in line 8 will be satisfied at the next iteration k.

We denote the actual state of the system at the beginning of the $(k-1)$st iteration as $x_2 = \xi_2(x_0, (k-1)T_p)$. Assume that the condition in line 8 is not satisfied, i.e., $x_2 \in S_{k-1}$; otherwise, the algorithm would have already produced the correct "second model" output. The measurement q_k of x_2 obtained in this iteration is an element of C_{k-1}. Thus, $\|x_2 - q_k\|_\infty \le \delta_{k-1} e^{-(M_1+\alpha)T_p} \le \varepsilon_{\min}$. By the (M_1, T_p)-separation with the infinity norm, it follows that

$$\begin{aligned}\|\xi_2(x_2, T_p) - \xi_1(q_k, T_p)\|_\infty &> \delta_{k-1} e^{-(M_1+\alpha)T_p} e^{M_1 T_p} \\ &= \delta_{k-1} e^{-\alpha T_p} = \delta_k.\end{aligned}$$

As $v_k(\cdot) = \xi_1(q_k, \cdot)$, from the strict inequality in the previous formula, it follows that $\xi_2(x_0, kT_p) = \xi_2(x_2, T_p) \notin B(v_k(T_p), \delta_k) = S_k$. Thus, at the beginning of the kth iteration, the condition in line 8 will hold.

For the "only if" part, assume that the true model is not the second model (25). Let us fix an initial state of the system x_0. From the hypothesis, we know that the true model is the first model and the true trajectory of the system is $\xi_1(x_0, t)$. From Theorem 4, it follows that at every iteration i, the state of the system at that round $\xi_1(x_0, iT_p) \in S_i$. Thus, the **if**-condition in line 8 is not satisfied at any iteration and consequently the algorithm never outputs "second model." ■

Remark 5: If state measurements are transmitted by a finite-data-rate communication channel, then the variables δ_i, S_i, and C_i are still generated independently and synchronously on both sides of the channel (the encoding side and the decoding side), with the understanding that both the encoder and the decoder work with the first model without knowing whether it is the correct one. Our result also applies to other scenarios where no channel is explicitly present but the detection procedure has access only to finite-resolution state measurements (collected, for example, by digital sensors).

Remark 6: The definition of exponential separation does not imply that the value of the upper bound $\varepsilon_{\min}$ is known, and short of that we cannot conclude for sure at any given time that the true model is the first model even if the state measurements conform with the constructed bound S_i in every round up to that time. However, if we know such an upper bound $\varepsilon_{\min}$ for which the models are (M_1, T_p)-exponentially separated, then the model detection algorithm can be made to decisively halt with the output "first model." For this, the following conditional statement should be inserted after line 10:

else if $\delta_i e^{-M_1 T_p} < \varepsilon_{\min}$
output "first model"; **break**;

This branch is executed by the algorithm at the ith round only if we had $\delta_{i-1}e^{-(M_1+\alpha)T_p} \leq \varepsilon_{\min}$ at the $(i-1)$st round and the measured state was in S_j for each of the preceding rounds $j < i$. At this point, the algorithm can soundly infer "first model" because, according to the proof of Theorem 7, the second model would have already triggered line 8 in the current round or one of the earlier rounds.

Remark 7: It is possible to run two versions of the detection algorithm, one with each of the candidate models, in parallel. While this may speed up detection in practice, in the worst case the two versions would take the same amount of time to reach a decision. This would also double the data rate without guaranteeing faster model detection. We, thus, opted for an approach which, while "asymmetric," works with the minimal needed data rate.

B. Exponential Separation Property

We now examine more closely the (M_s, T_s)-exponential separation property expressed by the inequality (26). As defined, it must hold for all $x_1, x_2 \in \mathbb{R}^n$ within the distance of $\varepsilon_{\min}$ from each other, and as such, it may be difficult to check. However, inspecting the "if" part of the proof of Theorem 7, we see that the exponential separation property is required only for those pairs x_1, x_2 where one of the points (x_2 in the proof) lies on a trajectory of the true system model (in the proof, it is the second model): $x_2 = \xi_2(x_0, t)$ for some initial state x_0 and some time t. Moreover, in practice, we would not run the detection algorithm for infinitely long, so we can take this time t to be bounded. This implies that it is sufficient for our purposes that the exponential separation property hold with the additional quantification that x_2 belong to some compact set D (large enough to contain the reachable set $\xi_2(K, [0, T])$ for some sufficiently large $T > 0$). In what follows, we call this relaxed property (M_s, T_s)-*exponential separation over* D. We do not place an explicit constraint on x_1 but, by definition, exponential separation over D only involves x_1 within distance $\varepsilon_{\min}$ from D. We now write down a simple condition for checking this modified exponential separation property.

Proposition 8: Let $D \subset \mathbb{R}^n$ be compact and suppose that the two models (24), (25) satisfy

$$f_1(x) \neq f_2(x) \qquad \forall\, x \in D. \tag{27}$$

Then, the two models are (M_s, T_s)-exponentially separated over D for small enough T_s and arbitrary M_s.

Proof: Since f_1 and f_2 are continuous and D is compact, (27) implies that there exists $v_{\min} > 0$ such that

$$|f_1(x) - f_2(x)| \geq v_{\min} \qquad \forall\, x \in D. \tag{28}$$

(We can think of $v_{\min}$ as the minimal separation speed between trajectories of the two systems starting from the same state.) Fix arbitrary x_1, x_2 with $x_2 \in D$ and note that, by the triangle inequality, we have

$$\begin{aligned}|\xi_1(x_2,t) - \xi_2(x_2,t)| \leq\ & |\xi_1(x_2,t) - \xi_1(x_1,t)| \\ & + |\xi_1(x_1,t) - \xi_2(x_2,t)|\end{aligned}$$

or, equivalently

$$\begin{aligned}|\xi_1(x_1,t) - \xi_2(x_2,t)| \geq\ & |\xi_1(x_2,t) - \xi_2(x_2,t)| \\ & - |\xi_1(x_1,t) - \xi_1(x_2,t)|.\end{aligned} \tag{29}$$

Since, by (28) applied with $x = x_2$,

$$\left|\frac{d}{dt}\big(\xi_1(x_2,t) - \xi_2(x_2,t)\big)\big|_{t=0}\right| = |f_1(x_2) - f_2(x_2)| \geq v_{\min}$$

and since $\xi_1(x_2, 0) - \xi_2(x_2, 0) = 0$, we can use the first-order Taylor expansion with respect to t to lower-bound the first term on the right-hand side of (29) as

$$|\xi_1(x_2,t) - \xi_2(x_2,t)| \geq v_{\min} t - o(t) \tag{30}$$

where, by definition, the term $o(t)$ has the property that for each $\delta > 0$ there exists $\tau > 0$ such that $|o(t)| \leq \delta t$ for all $t \in [0, \tau]$. *A priori* τ depends not just on δ but also on x_2; however, in view of compactness of D and continuous dependence of solutions on initial conditions, by taking the minimum of τ over x_2, we can find $\tau > 0$ that depends on δ only.

As for the second term on the right-hand side of (29), Lemma 1 applied to the first model gives us the upper bound

$$|\xi_1(x_1,t) - \xi_1(x_2,t)| \leq e^{\bar{\mu}_1 t}|x_1 - x_2|$$

for $\bar{\mu}_1$ satisfying (4) with $f = f_1$. We can rewrite this as

$$|\xi_1(x_1,t) - \xi_1(x_2,t)| \leq (1 + \bar{\mu}_1 t + o(t))|x_1 - x_2| \tag{31}$$

where the term $o(t)$ again has the property that for each $\delta > 0$, there exists $\tau > 0$ [which we can take, with no loss of generality, to be the same as τ for the $o(t)$ term appearing in (30)] such that $|o(t)| \leq \delta t$ for all $t \in [0, \tau]$.

Now, suppose that $|x_1 - x_2| \leq \varepsilon$ for some $\varepsilon > 0$. Plugging the bounds (30) and (31) into (29), we obtain

$$|\xi_1(x_1,t) - \xi_2(x_2,t)| \geq v_{\min} t - o(t) - (1 + \bar{\mu}_1 t + o(t))\varepsilon.$$

Thus, by the above-mentioned properties of the two $o(t)$ terms, for every $\delta > 0$, there is $\tau > 0$ such that

$$\begin{aligned}|\xi_1(x_1,t) - \xi_2(x_2,t)| \geq\ & (v_{\min} - \delta - \bar{\mu}_1\varepsilon - \delta\varepsilon)t - \varepsilon \\ & \forall\, t \in [0, \tau].\end{aligned}$$

Let $a(\varepsilon, \delta) := v_{\min} - \delta - \bar{\mu}_1\varepsilon - \delta\varepsilon$. Picking $\delta < v_{\min}$ and ε small enough, we can ensure that $a(\varepsilon, \delta) > 0$. It is easy to see that for every $M > 0$ and every $t > 0$, we have $at - \varepsilon > \varepsilon e^{Mt}$ if $\varepsilon > 0$ is small enough. From this, the claimed (M_s, T_s)-exponential separation property follows. Indeed, for an arbitrary $M_s > 0$, we can pick $\delta < v_{\min}$, find a corresponding τ, choose $T_s \in [0, \tau]$, and then find $\varepsilon_{\min}$ small enough so that, first, $a(\varepsilon_{\min}, \delta) > 0$ and, second, $aT_s - \varepsilon_{\min} > \varepsilon_{\min} e^{M_s T_s}$ (thereby ensuring that $aT_s - \varepsilon > \varepsilon e^{M_s T_s}$ for all $\varepsilon \leq \varepsilon_{\min}$). ■

We conclude from Proposition 8 that if the condition (27) holds over a compact domain D, then the detection algorithm in Fig. 2 will work as desired if we pick a sufficiently small sampling period T_p and as long as the state trajectory of the true model remains in D. We also believe that this situation is "generic" in the sense that we expect it to happen for typical pairs of systems and typical initial conditions in D; for example, for

affine systems, this claim can be made precise and is confirmed by numerical experiments, as discussed below.

C. Experimental Evaluation of Detection Algorithm

Our implementation of the detection algorithm in Fig. 2 is available online.[5] In this section, we describe some of the experiments we have performed in evaluating the algorithms on affine and general nonlinear models.

All sets in $\mathbb{R}^n$, including the initial set K and S_i, are n-dimensional hyperrectangles and they are represented either by two corner points or by a center point and a radius. The choice of this representation has implications on the efficiency of the algorithms. It enables the implementation of all the necessary operations such as testing membership in S, computing a grid on S, and quantizing a point with respect to a grid, in time that is linear in the number of dimensions n. Specifically, the grid(S, δ) function computes n lists of points in $\mathbb{R}$ where the ith list is generated by uniformly partitioning the ith dimension of S into intervals of length 2δ. This list representation of grid(S, δ) is adequate for quantizing a state with respect to it. The detection algorithm has to compute solutions $\xi_1(\cdot, \cdot)$ of the system (24) over $[0, T_p]$. Moreover, in order to simulate the algorithm, we have to compute the actual trajectories $\xi_2(\cdot, \cdot)$ of the system (25). Our implementation uses numerical ordinary differential equation solvers for both. (For affine systems, an analytical formula for the solutions is also available.)

Affine models: We generate pairs of random affine dynamical systems sys1: $\dot{x} = A_1 x + b_1$, sys2: $\dot{x} = A_2 x + b_2$, and then sys1 is used as the input model for the algorithm, while sys2 is used as the true model of the system. With this set-up, we performed several experiments. The detection algorithm worked in all experiments (unless we deliberately chose $A_2 = A_1$ and $b_2 = b_1$). Illustrations of executions of the detection algorithm are shown in Fig. 3 (top) with $\alpha = 0.5$, $d_0 = 1$, $T_p = 1$, $n = 2$ and the sys1 and sys2 parameters given by

$$A_1 = \begin{pmatrix} 0 & 1 \\ -2 & -2 \end{pmatrix}, \quad b_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix},$$

$$A_2 = \begin{pmatrix} -0.008 & 1.08 \\ -2.01 & -2 \end{pmatrix}, \; b_2 = \begin{pmatrix} 1.5 \\ -1 \end{pmatrix}$$

where we note that the matrix A_1 is taken from Example 1. The detection time depends on several factors. As is expected from the algorithm, it increases with smaller values of α and T_p. For the same system model sys1, using the matrix measure constant M instead of the Lipschitz constant L (as done in [23]) in the detection algorithm *can* lead to faster detection if $M < L$. For example, in the above-mentioned set-up, $M = 1$ and detection occurs after four rounds, whereas $L = 4$ and detection when using L occurs after six rounds. We note that according to Theorem 4 of this paper and [23, Th. 3], the rate at which the bounding sets S_i are decreasing is independent of this choice. The quantization errors and the bit rate, however, do depend on this choice of M versus L. According to Proposition 5 of this paper and [23, Proposition 4], the bit rates needed are $(M + \alpha)/\ln 2$ and $(L + \alpha)/\ln 2$, respectively, and therefore, for small α, the bandwidth requirements when using M are almost four times smaller.

[5]From `https://bitbucket.org/mitras/detection`.

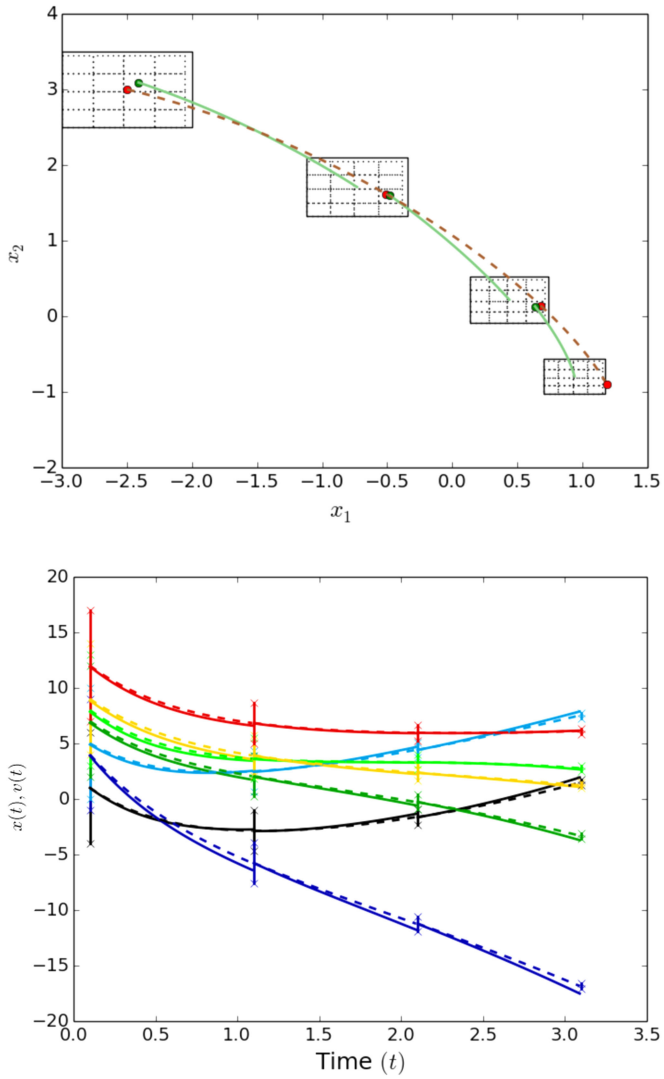

Fig. 3. Top: A sample execution of the detection algorithm on the two-dimensional system. The dashed lines $(--)$ show the trajectories of the actual system sys2, and the solid lines show the estimates $v(t)$ computed by the detection algorithm. The rectangles show decaying envelope of S_i and quantization grids. Eventually, the actual state (red dot) falls outside of S_i, triggering detection. Bottom: A sample execution on a six-dimensional system. State components of the actual system sys2 $(--)$, computed estimates $v(t)$ (solid lines), and decaying envelope S_i (vertical bars). Detection occurs as at least one state component leaves S_i.

The observation that the detection algorithm is always able in practice to distinguish between two randomly generated affine systems is explained by the fact, alluded to earlier, that the condition (27) of Proposition 8 is "generically true" for such systems. For $f_1(x) = A_1 x + b_1$ and $f_2(x) = A_2 x + b_2$, (27) fails at x satisfying

$$(A_1 - A_2)x = b_2 - b_1. \tag{32}$$

If the entries of the matrices A_1 and A_2 are picked at random according to a reasonable (i.e., absolutely continuous) probability distribution, then with probability 1, the matrix $A_1 - A_2$ will have full rank (indeed, the equation $\det(A_1 - A_2) = 0$

identifies a set of measure 0 in the space of the matrix entries). Therefore, there will be a unique $x \in \mathbb{R}^n$ satisfying (32). As long as we do not pick an initial state x_0 from which the true model's trajectory passes through this x (the set of such points x_0 in $\mathbb{R}^n$ is again of measure 0), the detection algorithm will work (see the discussion immediately preceding Proposition 8).

We further experimented with the true affine system sys2 modified with a disturbance term: $\dot{x} = A_2 x + b_2 + k_2\omega_2$, where k_2 is a constant and ω_2 is either a time-varying signal like $\sin t$ or a random noise term taking values in $[-1, 1]$. Keeping all the other parameters the same, we observed that, on the average, the detection time decreases with larger k_2 values (for $k_2 = 1, 0.5, 0.05, 0.005, 0.0005$, the detection took on the average $3, 5, 9, 13, 18$ rounds, respectively). We also experimented with a complementary scenario in which the true model is changed to sys1 but is also affected by a noise term unknown to the algorithm. As expected, the true behavior of the noisy system deviates from the nominal behavior and the algorithm eventually decides, incorrectly, that the true model is sys2. Thus, it can be said that the algorithm has robustness to noise affecting the true model when that model is sys2 but not when it is sys1.

Nonlinear models: As mentioned earlier, our implementation can handle arbitrary nonlinear models. In this section, we discuss several experiments we performed using the Van der Pol oscillator model, which is given by the system equations

$$\begin{aligned} \dot{x}_1 &= x_2, \\ \dot{x}_2 &= p(1 - x_1^2)x_2 - x_1 \end{aligned} \tag{33}$$

where x_1 is the position coordinate and p is a scalar parameter describing the nonlinearity and the strength of the damping. We consider two scenarios with different true system (sys2) models. In scenario 1, the true system dynamics (sys2) is given by (33) with $p = 0.5$; in scenario 2, sys2 has $p = 2$; both lead to a limit cycling behavior with different phase portraits (see, for example, [31]).

For each of these scenarios, we execute the detection algorithm with different internal models (sys1) that have different values of the parameter p. The resulting pairs of systems satisfy the exponential separation criterion of Proposition 8 almost everywhere. We compute an upper bound $\bar{\mu}$ on the matrix measure of the Jacobian of sys1 as follows: first, we derive a symbolic expression for the matrix measure using (2), and then maximize it over the reachable set of the system. We note that this step can be performed automatically for general models described using standard nonlinear functions. We estimated a bounding box containing the reachable states of sys1 using simulations; for more precise estimates, one could use a nonlinear reachability analysis tool [1], [6], [15]. The resulting parameters used for our experiments are as follows: For scenario 1: $p = 0.5, \alpha = T_p = 0.5, \bar{\mu} = 3.5$; for scenario 2: $p = 2, \alpha = 0.5, T_p = 0.1, \bar{\mu} = 13$. The detection times (in terms of the number of iterations i^*) for different sys1 models with different values of the parameter p are shown in Table I. The actual number of iterations is less important than the general observation that, as expected, the detection takes longer as the models become closer. Fig. 4 illustrates a typical run of the algorithm.

TABLE I
DETECTION TIMES FOR VAN DER POL SYSTEMS ($p = 0.5, 2$) WITH DIFFERENT SYS1 MODELS

Scenario 1 ($p = 0.5$)		Scenario 2 ($p = 2$)	
p in sys1	i^*	p in sys1	i^*
2	6	3	26
0.55	17	2.5	69
0.45	17	1.5	62
0.51	23	2.1	73
0.49	21	1.9	73
0.501	32	1.99	100
0.499	32	2.01	100

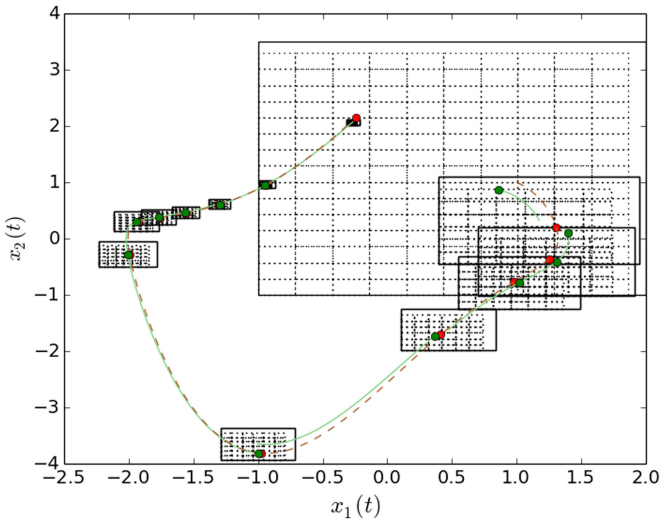

Fig. 4. Sample execution of the detection algorithm for the Van der Pol oscillator. The dashed lines $(--)$ show the trajectories of the actual system sys2, and the solid lines show the estimates $v(t)$ computed by the detection algorithm. The rectangles show decaying envelope of S_i and quantization grids. Eventually, the actual state (red dot) falls outside of S_i (vanishingly small in this plot), triggering detection.

To summarize this section, our experiments show that the proposed algorithm succeeds in performing state estimation and model detection for nonlinear systems. True dynamics of the system is often affected by unknown parameters or disturbances that are unknown to the detection algorithm. In such cases, the algorithm does the reasonable thing, in that it gives exponentially converging state estimates up to a certain time beyond which it detects that the internal model (sys1) has diverged from the true model (sys2). We also empirically observe that this detection time is inversely related to the differences in the models. A more careful analysis of the detection time and some of the above-mentioned empirical conclusions will be the subject of future research.

VII. CONCLUSION AND FUTURE DIRECTIONS

We introduced two different notions of *estimation entropy* and established their equivalence. We derived an upper bound of $O((M + \alpha)n)$ for the estimation entropy of an n-dimensional

nonlinear dynamical system with Jacobian f_x whose matrix measure does not exceed M, where the desired exponential convergence rate of the estimate is α. We also established a lower bound of $O(\inf \mathrm{tr} f_x + \alpha n)$ on the estimation entropy, where the $\inf$ is taken over the reachable states of the system. We developed a procedure for generating exponentially converging state estimates using an average bit rate that matches the upper bound on the entropy, and showed that no other similar state estimation algorithm can work with bit rates lower than the entropy. We presented an application of the estimation procedure in solving a model detection problem where we have to identify one model from a pair of candidate models using quantized measurements. We showed that under a mild assumption of *exponential separation*—which we expect to hold almost surely for randomly chosen model pairs—the algorithm can always detect the true model in finite time. The exponential separation condition was stated in terms of solutions of the candidate models and this concept may be of independent interest. We presented a sufficient condition for exponential separation in terms of the models' vector fields over a compact set.

There are several avenues for future work. Ramifications of Theorem 1 remain to be understood. Computations based on matrix measure bounds can be refined and more general contraction metrics can be exploited (some relevant results that can be leveraged for these purposes include [3], [13], [14], [25], [26], [30], and [36]). In particular, the approach of [26] works with ε-balls with respect to the norm $\sqrt{x^T P x}$ where the matrix P satisfies inequalities in the spirit of Lyapunov's direct method; while not constructive in general, this approach can lead to sharper entropy estimates for special classes of systems, although the improvement is not always significant (see the simulation studies in [26, Section 7]). The procedures in [13] and [14] will be more useful for computing accurate, possibly locally optimal, state estimates at run time, than for obtaining better offline entropy estimates. The exponential separation property and sufficient conditions for it deserve further exploration. The model detection algorithm warrants a more detailed study of its performance (e.g., estimating the number of steps until detection); considering larger families of models and incorporating disturbances, delays, packet losses, etc., are other natural research directions. Entropy for switched and hybrid systems and its role in state estimation and model detection as well as control of such systems is a subject of ongoing work (see [33], [34], and [39] for some recent results).

References

[1] M. Althoff, "An introduction to CORA 2015," in *Proc. 1st 2nd Int. Workshop Appl. Verification Continuous Hybrid Syst.*, Seattle, WA, USA, 2015, pp. 120–151.

[2] A. U. Awan and M. Zamani, "On a notion of estimation entropy for stochastic hybrid systems," in *Proc. 54th Annu. Allerton Conf. Commun., Control, Comput.*, 2016, pp. 780–785.

[3] V. A. Boichenko and G. A. Leonov, "Lyapunov's direct method in estimates of topological entropy," *J. Math. Sci.*, vol. 91, pp. 3370–3379, 1998.

[4] R. Bowen, "Entropy for group endomorphisms and homogeneous spaces," *Trans. Amer. Math. Soc.*, vol. 153, pp. 401–414, 1971.

[5] R. W. Brockett, *Finite Dimensional Linear Systems*. New York, NY, USA: Wiley, 1970.

[6] X. Chen, E. Ábrahám, and S. Sankaranarayanan, "Flow*: An analyzer for non-linear hybrid systems," in *Proc. 25th Int. Conf. Comput. Aided Verification*, Saint Petersburg, Russia, 2013, pp. 258–263.

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

[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] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.

[11] A. Diwadkar and U. Vaidya, "Limitations for nonlinear observation over erasure channel," *IEEE Trans. Automat. Control*, vol. 58, no. 2, pp. 454–459, Feb. 2013.

[12] P. S. Duggirala, S. Mitra, and M. Viswanathan, "Verification of annotated models from executions," in *Proc. Int. Conf. Embedded Softw.*, Montreal, QC, Canada, 2013, pp. 1–10.

[13] C. Fan, J. Kapinski, X. Jin, and S. Mitra, "Locally optimal reach set overapproximation for nonlinear systems," in *Proc. 2016 Int. Conf. Embedded Softw.*, Pittsburgh, PA, USA, 2016, pp. 6:1–6:10.

[14] C. Fan and S. Mitra, "Bounded verification with on-the-fly discrepancy computation," in *Automated Technology for Verification and Analysis*, Lecture Notes in Computer Science, vol. 9364. Cham, Switzerland: Springer, 2015, pp. 446–463.

[15] C. Fan, B. Qi, S. Mitra, M. Viswanathan, and P. S. Duggirala, "Automatic reachability analysis for nonlinear hybrid models with C2E2," in *Proc. 28th Int. Conf. Comput. Aided Verification*, Toronto, ON, Canada, 2016, pp. 531–538.

[16] J. P. Hespanha, A. Ortega, and L. Vasudevan, "Towards the control of linear systems with minimum bit-rate," in *Proc. 15th Int. Symp. Math. Theory Netw. Syst.*, 2002.

[17] P. A. Ioannou and J. Sun, *Robust Adaptive Control*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1996.

[18] A. Katok and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems*. Cambridge, U.K.: Cambridge Univ. Press, 1995.

[19] C. Kawan, "Exponential state estimation, entropy and Lyapunov exponents," 2016, arXiv:1605.03210.

[20] K.-D. Kim, S. Mitra, and P. R. Kumar, "Computing bounded epsilon-reach set with finite precision computations for a class of linear hybrid automata," in *Proc. Hybrid Syst., Comput. Control*, Chicago, IL, USA, 2011, pp. 113–122.

[21] D. Liberzon, *Calculus of Variations and Optimal Control Theory: A Concise Introduction*. Princeton, NJ, USA: Princeton Univ. Press, 2012.

[22] D. Liberzon and J. P. Hespanha, "Stabilization of nonlinear systems with limited information feedback," *IEEE Trans. Automat. Control*, vol. 50, no. 6, pp. 910–915, Jun. 2005.

[23] D. Liberzon and S. Mitra, "Entropy and minimal data rates for state estimation and model detection," in *Proc. Hybrid Syst., Comput. Control*, Vienna, Austria, 2016, pp. 247–256.

[24] D. Liberzon and S. Mitra, "Entropy notions for state estimation and model detection with finite-data-rate measurements," in *Proc. 55th IEEE Conf. Decis. Control*, 2016, pp. 7335–7340.

[25] J. Maidens and M. Arcak, "Reachability analysis of nonlinear systems using matrix measures," *IEEE Trans. Automat. Control*, vol. 60, no. 1, pp. 265–270, Jan. 2015.

[26] A. Matveev and A. Pogromsky, "Observation of nonlinear systems via finite capacity channels: Constructive data rate limits," *Automatica*, vol. 70, pp. 217–229, 2016.

[27] G. N. Nair, R. J. Evans, I. M. Y. Mareels, and W. Moran, "Topological feedback entropy and nonlinear stabilization," *IEEE Trans. Automat. Control*, vol. 49, no. 9, pp. 1585–1597, Sep. 2004.

[28] 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.

[29] C. De Persis, "Nonlinear stabilizability via encoded feedback: The case of integral ISS systems," *Automatica*, vol. 42, pp. 1813–1816, 2006.

[30] A. Yu. Pogromsky and A. S. Matveev, "Estimation of topological entropy via the direct Lyapunov method," *Nonlinearity*, vol. 24, pp. 1937–1959, 2011.

[31] P. Ponzo and N. Wax, "On the periodic solution of the van der Pol equation," *IEEE Trans. Circuit Theory*, vol. CT-12, no. 1, pp. 135–136, Mar. 1965.

[32] A. V. Savkin, "Analysis and synthesis of networked control systems: Topological entropy, observability, robustness and optimal control," *Automatica*, vol. 42, pp. 51–62, 2006.
[33] A. J. Schmidt, "Topological entropy bounds for switched linear systems with Lie structure," Master's thesis, Dept. Elect. Comput. Eng., Univ. Illinois Urbana-Champaign, Champaign, IL, USA, 2016, arXiv:1610.02701.
[34] H. Sibai and S. Mitra, "Optimal data rate for state estimation of switched nonlinear systems," in *Proc. 20th Int. Conf. Hybrid Syst., Comput. Control*, Pittsburgh, PA, USA, 2017, pp. 71–80.
[35] R. S. Smith and J. C. Doyle, "Model validation: A connection between robust control and identification," *IEEE Trans. Automat. Control*, vol. 37, no. 7, pp. 942–952, Jul. 1992.
[36] E. D. Sontag, "Contractive systems with inputs," in *Perspectives in Mathematical System Theory, Control, and Signal Processing*, J. Willems, S. Hara, Y. Ohta, and H. Fujioka, Eds. New York, NY, USA: Springer, 2010, pp. 217–228.
[37] S. Tatikonda and S. K. Mitter, "Control under communication constraints," *IEEE Trans. Automat. Control*, vol. 49, no. 7, pp. 1056–1068, Jul. 2004.
[38] M. Vidyasagar, *Nonlinear Systems Analysis*, 2nd ed. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.
[39] G. Yang, "Switched and hybrid systems with inputs: Small-gain theorems, control with limited information, and topological entropy," Ph.D. dissertation, Dept. Elect. Comp. Eng., Univ. Illinois Urbana-Champaign, Champaign, IL, USA, 2017.
[40] M. Zorzi, "Multivariate spectral estimation based on the concept of optimal prediction," *IEEE Trans. Automat. Control*, vol. 60, no. 6, pp. 1647–1652, Jun. 2015.

Daniel Liberzon (M'98–SM'04–F'13) was born in the former Soviet Union, in 1973. He worked toward the undergraduate studies at the Department of Mechanics and Mathematics, Moscow State University, Moscow, Russia, and received the Ph.D. degree in mathematics from Brandeis University, Waltham, MA, USA, in 1998 (under Prof. Roger W. Brockett of Harvard University).

Following a Postdoctoral Position with the Department of Electrical Engineering, Yale University, from 1998 to 2000, he joined the University of Illinois at Urbana-Champaign, Champaign, IL, USA, where he is currently a Professor with the Electrical and Computer Engineering Department and the Coordinated Science Laboratory. He is the Author of the books *Switching in Systems and Control* (Birkhauser, 2003) and *Calculus of Variations and Optimal Control Theory: A Concise Introduction* (Princeton Univ. Press, 2012). His research interests include nonlinear control theory, switched and hybrid dynamical systems, control with limited information, and uncertain and stochastic systems.

Dr. Liberzon is a Fellow of IFAC and an Editor for *Automatica* (nonlinear systems and control area). He is the recipient of several recognitions, including the 2002 IFAC Young Author Prize and the 2007 Donald P. Eckman Award. He delivered a plenary lecture at the 2008 American Control Conference.

Sayan Mitra (M'01–SM'13) received the B.E.E. degree in electrical engineering from Jadavpur University, Kolkata, India, in 1999, the Master's degree in computer science from the Indian Institute of Science, Bangalore, India, in 2001, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2007.

After one year of postdoctoral work at the Center for Mathematics of Information, California Institute of Technology, he joined the University of Illinois at Urbana-Champaign, Champaign, IL, USA, where he is currently an Associate Professor with the Electrical and Computer Engineering Department. His research interests include formal methods, hybrid systems, distributed systems, and verification of cyber-physical systems and their applications.

Dr. Mitra has served as the Program Co-Chair of the 20th International Conference on Hybrid Systems. He is the recipient of the National Science Foundation's CAREER Award, the Air Force Office of Scientific Research Young Investigator Program Award, IEEE-HKN C. Holmes MacDonald Outstanding Teaching Award, and several best paper awards.