Optimal Data Rate for State Estimation of Switched Nonlinear Systems

Hussein Sibai
University of Illinois at Urbana-Champaign
1308 W. Main St.
Urbana, IL 61801
sibai2@illinois.edu

Sayan Mitra
University of Illinois at Urbana-Champaign
1308 W. Main St.
Urbana, IL 61801
mitras@illinois.edu

ABSTRACT

State estimation is a fundamental problem for monitoring and controlling systems. Engineering systems interconnect sensing and computing devices over a shared bandwidth-limited channels, and therefore, estimation algorithms should strive to use bandwidth optimally. We present a notion of entropy for state estimation of nonlinear switched dynamical systems, an upper bound for it and a state estimation algorithm for the case when the switching signal is unobservable. Our approach relies on the notion of topological entropy and uses techniques from the theory for control under limited information. We show that the average bit rate used is optimal in the sense that, the efficiency gap of the algorithm is within an additive constant of the gap between estimation entropy of the system and its known upper-bound. We apply the algorithm to two system models and discuss the performance implications of the number of tracked modes.

Keywords

State Estimation; Switched Systems; Data Rate; Entropy

1. INTRODUCTION

This paper deals with monitoring continuous time dynamical systems with optimal usage of network resources. The key problem is to compute approximations of the state of the system from a small number of bits coming from quantized sensor measurements. This is the *state estimation* problem. The related problem of *model detection* arises when the plant dynamics itself is unknown or changing. Contemporary engineering systems interconnect sensing and computing devices over shared communication channel for monitoring and control. For example, more than 70 embedded computing units communicate over shared 1 MBps CAN bus in cars [1]. Large number of machines, conveyor belts, and robotic manipulators need to be monitored in warehouses and factory floors—-again over a shared network backbone [13]. Such bandwidth constraints call for optimal allocation of network resources for estimation and detection.

HSCC '17 April 18–21, 2017, Pittsburgh, PA, USA

DOI: 10.475/123_4

In the stochastic setting, Kalman and particle filtering are used for solving these problems; in some cases using neural networks (see, for example [15, 16, 14]). Our approach relies on the theory of topological entropy for dynamical systems. The measure-theoretic notion of entropy plays a central role in information theory, estimation and detection. In the theory of dynamical systems, the analogous topological notion of entropy plays a fundamental role in describing the rate of growth of uncertainty about system state ([6, 3, 11, 2, 10]). It also relates to the rate at which information about the system should be collected for state estimation. Drawing this connection the notion of *estimation entropy* has been defined in [8, 12] for nonlinear systems. For a dynamical system of the form: $\dot{x}(t) = f(x(t))$, roughly, it is the minimum bit rate needed to construct state estimates from quantized measurements, that converge to the actual state of the system at a desired exponential rate of α. Estimation entropy is in general hard to compute exactly, but can be upper-bounded by $(C+\alpha)n/\ln 2$; where n is the dimension of system and C is either the Lipschitz constant L of f [8] or an upper-bound on the matrix measure of the Jacobian of f [9]. In [8] an algorithm for state estimation is given which uses an average bit rate of $(L+\alpha)n/\ln 2$. This is optimal in the sense that, the efficiency gap of the algorithm is no more than the gap between estimation entropy and its upper-bound.

In this paper, we study state estimation of switched nonlinear dynamical system $\dot{x} = f_{\sigma(t)}(x(t))$ where the switches between N modes or subsystems are brought about by an unknown, exogenous, switching signal $\sigma : \mathbb{R}_{\geq 0} \rightarrow [N]$. Each mode $\dot{x} = f_p(x(t))$, $p \in [N]$, , where $[N]$ is the set of integers from 0 to $N-1$, could capture, for example, uncertainties in the plant, different operating regimes—nominal and failure dynamics, and parameter values.

Since the mode information is not available to the estimator, exponential convergence of state estimates may be impossible immediately after a mode switch. We relax the notion of estimation entropy of [8] by allowing a period of time $\tau > 0$ following a mode switch, during which the estimation error is only bounded by a constant ε; and thereafter the error decays exponentially as usual. We show that for a large enough ε—determined by the minimum dwell time of σ and the difference in the dynamics of the different modes—the estimation entropy is bounded by $\frac{(L+\alpha)n}{\ln 2} + \frac{\log N}{T_e}$. Here L is the largest between the Lipschitz constants of all f_p's and T_e is a positive constant less than or equal to τ.

We present an algorithm for state estimation for switched systems. The interdependence of the uncertainties in the state and the mode requires this algorithm to simultaneously solve the estimation and model detection problems: Unless a mode $f_p, p \in [N]$ is detected, it may be impossible to get exponentially converging estimates, and (b) unless an accurate enough estimate for the state is known, it may not be possible to distinguish between two candidate modes.

Our algorithm keeps track of $\hat{N}$ possible modes of the switched system, where $\hat{N}$ is a parameter between 1 and N. If the actual mode of the system is one of the tracked modes, then, owing to a shrinking quantized measurement strategy, the state estimate converges at the desired exponential rate. If the actual mode is not tracked, then the actual state of the system may *escape* the constructed state estimate bounds. In this case, the algorithm expands the estimate and captures the state. When a mode switch happens, there may be a burst of escapes, but we prove that if the rate of switches is slow enough and the modes are different enough, then the correct mode is detected, and thereafter, the state estimates converge exponentially.

We establish worst case estimation error bounds and time bounds on mode detection. We also show that the average bit rate used is within $\frac{\hat{N}}{T_p} - \frac{\log N}{T_e}$ from the upper bound on the entropy, i.e. the upper bound on the optimal bit-rate; where T_p is the sampling time of the algorithm. We present preliminary experimental results, discuss the application of the algorithm to linear and nonlinear switched systems, and the implications of the choice of the key parameter $\hat{N}$.

2. SWITCHED SYSTEMS AND ENTROPY

A *switched system* is a standard way for describing control systems with several different modes (see, for example, the book [7]). Suppose we are given a family $f_p, p \in [N]$, of functions from $\mathbb{R}^n$ to $\mathbb{R}^n$. Assuming that the functions f_p are sufficiently regular, for example, Lipschitz continuous with Lipschitz constant L_p, the above gives rise to a family of dynamical system modes:

$$\dot{x} = f_p(x), p \in [N] \tag{1}$$

evolving on $\mathbb{R}^n$. The modes could capture structured uncertainty in the system, for example, different changing parameter values, failure conditions, and user inputs. If the mode $p \in [N]$ is known, then the solution of the differential equation is the function $\xi_p : \mathbb{R}^n \times \mathbb{R}_{\geq 0} \to \mathbb{R}^n$. If in addition the initial state x_0 is known, then for any point in time t the state $\xi_p(x_0, t)$ can be approximated using numerical integration. However, for the state estimation problem we are interested in, both the initial state and the mode are unknown.

The time varying mode is modeled as a *switching signal*. This is an (unknown and not observable) piecewise constant function $\sigma : [0, \infty) \to [N]$ which specifies at each time instant t, the index $\sigma(t) \in [N]$ of the function from the family (1) that is currently being followed. The points of discontinuity in σ are called *switching times*. Thus, the switched system with a time-dependent switching signal σ can be described by the equation:

$$\dot{x} = f_{\sigma(t)}(x). \tag{2}$$

For a fixed switching signal σ the solution of the above switched system is defined in the standard way and denoted by the function $\xi_\sigma : \mathbb{R}^n \times \mathbb{R}_{\geq 0} \to \mathbb{R}^n$. Moreover, it is Lipschitz continuous on the vector field with Lipschitz constant $L = \max_{p \in [N]} L_p$.

The switching signal σ models the adversary (or the environment) changing the underlying mode of the system. In general, it may have arbitrary discontinuities, however, to prove stability or in our case correctness of state estimation, typically one assumes bounds on switching speed [7, 5].

Covers, dwell-times, and reachable sets.

A switching signal σ has a *minimum dwell time* $T_d > 0$ if at least T_d time elapses between consecutive switches. For any point $x \in \mathbb{R}^n$ and $\delta > 0$, $B(x, \delta)$ is a closed hypercube of radius δ, in other words, δ-ball, centered at x. For a hyperrectangle $S \subseteq \mathbb{R}^n$ and $\delta > 0$, $grid(S, \delta)$, is a collection of 2δ-separated points along axis parallel planes such that the δ-balls around these points cover S. We denote $\Sigma(T_d)$ the family of switching signals with minimum dwell-time T_d switching between the N modes. Moreover, we define $Reach(\Sigma, K)$ to be the set of *reachable states* by system (2) with any $\sigma \in \Sigma(T_d)$ from the compact initial set K. More formally, $Reach(\Sigma, K) = \{x \in \mathbb{R}^n \mid \exists\ \sigma \in \Sigma(T_d),\ x_0 \in K,\ t \in [0, \infty) : \xi_\sigma(x_0, t) = x\}$. Later in this paper, we will have to bound the error in state estimates when the system evolves according to two *different* dynamics but from the same state. To this end we introduce the quantity:

$$d(t) := \max_{p,r \in [N]} \sup_{x \in Reach(\Sigma, K)} \int_0^t \|f_p(\xi_p(x, s)) - f_r(\xi_r(x, s))\| ds.$$

We assume that the sup exists. A sufficient condition would be the compactness of $Reach(\Sigma, K)$.

2.1 State Estimation, bit-rate, and entropy

Let us fix a compact set K of possible initial states of (2), the family of switching signals $\Sigma(T_d)$, two estimation accuracy related constants $\varepsilon, \alpha > 0$ and a time constant τ ($\tau \leq T_d$). Consider a setup in which a sensor has access to the actual current state of the system $\xi_\sigma(x_0, t)$ (and not the switching signal σ), and it needs to send bits across a bandwidth-constrained channel such that: for any initial state $x_0 \in K$ and for any (unknown) switching signal $\sigma \in \Sigma(T_d)$, the estimator would be able to construct a function $z : \mathbb{R}_{\geq 0} \to \mathbb{R}^n$, where $\forall t \in [s_j, s_{j+1})$,

$$|z(t) - \xi_\sigma(x_0, t)| \leq \begin{cases} \varepsilon & t \in [s_j, s_j + \tau), \\ \varepsilon e^{-\alpha(t-(s_j+\tau))} & otherwise, \end{cases} \tag{3}$$

where $s_0 = 0, s_1, \ldots$ are the switching times in σ. The norm in inequality (3) can be arbitrary. We call such a function $z(.)$ an $(\varepsilon, \alpha, \tau)$-*approximation* of $\xi_\sigma(x_0, \cdot)$. The second bound gives the ideal behavior in which the estimate converges to the actual trajectory $\xi_\sigma(x_0, \cdot)$ exponentially at the rate α as in [8] and [2]. Since this exponential convergence may be unrealistic after a mode switch that may completely change the dynamics, the first condition allows a "lenient" period of duration τ, during which the error is bounded by some positive constant ε.

We would like the number of bits sent over the channel to be minimal and this is made precise by the notion of estimation entropy defined next.

A finite set of functions $\hat{X} = \{\hat{x}_1, \ldots, \hat{x}_M\}$ from $[0, T]$ to $\mathbb{R}^n$ is $(T, \varepsilon, \alpha, \tau)$-*approximating* if for every initial state $x \in K$ and every switching signal $\sigma \in \Sigma(T_d)$ there exists some $\hat{x}_i \in \hat{X}$ such that for all $t \in [0, T]$, $\hat{x}_i$ is an $(\varepsilon, \alpha, \tau)$-approximating function for $\xi_\sigma(x_0, t)$. Note that $\hat{X}$ depends on K, T_d and the N modes but we are suppressing these parameters for brevity.

Let $s_{\rm est}(T,\varepsilon,\alpha,\tau)$ denote the minimal cardinality of such a $(T,\varepsilon,\alpha,\tau)$-approximating set. The *estimation entropy* of the system is defined as

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

Intuitively, since $s_{\rm est}$ corresponds to the minimal number of functions needed to approximate the state with desired accuracy, $h_{\rm est}$ is the average number of bits needed to identify these approximating functions. The lim sup extracts the base-2 exponential growth rate of $s_{\rm est}$ with time.

Then, $s_{\rm est}$ corresponds to the number of different quantization points needed to identify the trajectories, and $h_{\rm est}$ gives a measure of the long-term bit rate needed for communicating sensor measurements to the estimator.

2.2 Entropy upper bound

In this section we will establish an upper-bound on the estimation entropy h_{est} for switched systems. First, we prove an upper bound on the size of the smallest approximating set s_{est} using an inductive construction of approximating functions.

Throughout the paper we will fix the convergence parameters ε, $\alpha > 0$, and $\tau \le T_d$, the initial set K, and the time horizon $T > 0$. Also, we have the following assumption:

Assumption 1 There exists $T_e \in (0,\tau]$ where $(\varepsilon e^{-\alpha T_e} + d(T_e))e^{-\alpha(\tau - T_e)} \le \varepsilon$ and $\varepsilon e^{-\alpha(T_d-\tau)} + d(T_e) \le \varepsilon$.

Let us fix a trajectory $\xi_\sigma(x_0,\cdot)$ of the switched system (2). We define an inductive procedure that constructs a corresponding approximating function $z(\cdot)$. It follows that the set of all functions that can be computed by this procedure is a $(T,\varepsilon,\alpha,\tau)$-*approximating set*. Then, the cardinality of the set of all functions that can be computed by this procedure gives us an upper bound.

Let $s_0 = 0, s_1, \ldots$ be the sequence of switching times in the switching signal σ generating $\xi_\sigma(x_0,\cdot)$. The approximating function $z(\cdot)$ is constructed in time steps of size T_e ($T_e \le \tau$), where T_e is the largest one that satisfy Assumption 1. We start by choosing an open cover C_0 of K with balls of radii $\varepsilon e^{-(L+\alpha)T_e}$. Let q_0 be the center of a ball that contains x_0. We construct $z(t) := \xi_{\sigma(0)}(q_0,t)$ for $t \in [0,T_e]$. Since $\sigma(t) = \sigma(0)$ for $t \in [0,T_e]$ (recall, $T_d \ge \tau$), the estimation error over that interval would be $\|z(t) - \xi_\sigma(x_0,t)\| \le e^{Lt}\|x_0 - q_0\| \le e^{Lt}\varepsilon e^{-(L+\alpha)T_e} \le \varepsilon e^{-\alpha t}$ (by Bellman-Grownall inequality).

Next, for each integer $1 \le i \le \lfloor \frac{T}{T_e} \rfloor$, we compute an n-dimensional ball over-approximating the reachable set of states at $t = iT_e$ given the difference between the actual state x_{i-1} and the quantized one q_{i-1} at $t = (i-1)T_e$, and $\sigma((i-1)T_e)$. Then, we construct a grid with a predefined resolution over that ball. Next, we quantize the actual state at $t = iT_e$ with respect to the grid to get q_i. After that, we compute the trajectory which results from running the actual mode at $t = iT_e$ over the time interval $(iT_e,(i+1)T_e]$ starting from q_i. Finally, we bound the difference between the actual trajectory $\xi_\sigma(x_0,\cdot)$ and the constructed one $z(\cdot)$ and we prove that the ball computed at the $(i+1)^{th}$ iteration does contain the actual state at $t = (i+1)T_e$.

Formally, let s_j be the time of the last switch before iT_e. We construct C_i to be an open cover of $B(z(iT_e),R_i)$, where

$$R_i = \begin{cases} R_{i-1}e^{-\alpha T_e} + d(T_e) & \text{if } s_j \in ((i-1)T_e, iT_e), \\ R_{i-1}e^{-\alpha T_e} & \text{otherwise,} \end{cases}$$

and $R_0 = \varepsilon$, with balls of radii equal to $r_i = R_i e^{-(L+\alpha)T_e}$. Then, we let q_i to be any of the centers of the balls in C_i that contain $\xi_\sigma(x_0,iT_e)$. Note that $\xi_\sigma(x_0,T_e) \in B(z(T_e),R_1)$. Next, we construct $z(t) := \xi_{\sigma(iT_e)}(q_i, t - iT_e)$ for $t \in (iT_e,(i+1)T_e]$.

Lemma 1 *$z(\cdot)$ is an $(\varepsilon,\alpha,\tau)$-approximating function of $\xi_\sigma(x_0,\cdot)$.*

PROOF. Based on where the next switching time s_{j+1} falls with respect to the interval $[iT_e,(i+1)T_e]$, there are two cases here: (a) $s_{j+1} = iT_e$ or $s_{j+1} \ge (i+1)T_e$ and (b) $s_{j+1} \in (iT_e,(i+1)T_e)$. For (a),

$$\begin{aligned}
&\|\xi_\sigma(x_0,t) - z(t)\| = \|\xi_\sigma(x_0,t) - \xi_{\sigma(i\tau)}(q_i,t-i\tau)\| \\
&= \|\xi_{\sigma(iT_e)}(\xi_\sigma(x_0,iT_e),t-iT_e) - \xi_{\sigma(iT_e)}(q_i,t-iT_e)\| \\
&\qquad [\text{since } \sigma(t) = \sigma(iT_e) \text{ for } t \in [iT_e,(i+1)T_e)] \\
&\le e^{L_{\sigma(iT_e)}(t-iT_e)}\|\xi_\sigma(x_0,iT_e) - q_i\| \\
&\qquad [\text{Bellman-Gronwall inequality}] \\
&\le e^{L(t-iT_e)} r_i \\
&\qquad [L_{\sigma(iT_e)} \le L; \text{ by the definition of } q_i \in C_i] \\
&= e^{L(t-iT_e)} R_i e^{-(L+\alpha)T_e} \\
&\qquad [\text{substituting } r_i] \\
&\le e^{L(t-iT_e)} R_i e^{-(L+\alpha)(t-iT_e)} \\
&\qquad [\text{since } t - iT_e \le T_e] \\
&= R_i e^{-\alpha(t-iT_e)}.
\end{aligned}$$

For (b), we can repeat the same steps of part (a) for any $t \in (iT_e, s_{j+1})$ to get $\|z(t) - \xi_\sigma(x_0,t)\| \le R_i e^{-\alpha(t-iT_e)}$. After the switch at s_{j+1}, that is, for any $t \in [s_{j+1},(i+1)T_e]$,

$$\begin{aligned}
&\|\xi_\sigma(x_0,t) - z(t)\| = \|\xi_\sigma(x_0,t) - \xi_{\sigma(iT_e)}(q_i,t-iT_e)\| \\
&= \|\xi_\sigma(\xi_\sigma(x_0,s_{j+1}),t-s_{j+1}) - \xi_{\sigma(iT_e)}(q_i,t-i\tau)\| \\
&\le \|\xi_\sigma(\xi_\sigma(x_0,s_{j+1}),t-s_{j+1}) - \xi_\sigma(q_i,t-iT_e)\| \\
&\qquad + \|\xi_\sigma(q_i,t-iT_e) - \xi_{\sigma(iT_e)}(q_i,t-iT_e)\| \\
&\qquad [\text{by triangular inequality}] \\
&\le e^{L(t-s_{j+1})}\|\xi_\sigma(x_0,s_{j+1}) - \xi_\sigma(q_i,s_{j+1}-iT_e)\| \\
&\qquad + \|\int_0^{t-iT_e} (f_\sigma(\xi_\sigma(q_i,t')) - f_{\sigma(iT_e)}(\xi_{\sigma(iT_e)}(q_i,t')))dt'\| \\
&\qquad [\text{by Bellman-Gronwall inequality}] \\
&\le e^{L(t-s_{j+1})}\|\xi_\sigma(x_0,s_{j+1}) - \xi_\sigma(q_i,s_{j+1}-i\tau)\| \\
&\qquad + \int_0^{t-iT_e} \|f_\sigma(\xi_\sigma(q_i,t')) - f_{\sigma(iT_e)}(\xi_{\sigma(iT_e)}(q_i,t'))\|dt' \\
&\le e^{L(t-s_{j+1})}e^{L(s_{j+1}-iT_e)}\|\xi_\sigma(x_0,iT_e) - q_i\| + d(t-iT_e) \\
&\qquad [\text{using the definition of } d(\cdot)] \\
&\le e^{L(t-iT_e)}R_i e^{-(L+\alpha)T_e} + d(t-iT_e) \\
&\qquad [\text{substituting } \|\xi_\sigma(x_0,iT_e) - q_i\| \text{ with } r_i\text{'s value}] \\
&\le R_i e^{-\alpha(t-iT_e)} + d(T_e) \\
&\qquad [\text{since } d(t) \text{ is an increasing function}].
\end{aligned}$$

In both cases, $\xi_\sigma(x_0,(i+1)T_e) \in B(z((i+1)T_e),R_{i+1})$. Now we want to prove that $z(\cdot)$ is an approximation function to $\xi_\sigma(x_0,\cdot)$. First, we let $i_1 = \lceil s_1/T_e \rceil$ (the first iteration after the first switch) and $i_2 = \lfloor \tau/T_e \rfloor$ (the last iteration before τ). Note that $R_{i_2} = \varepsilon e^{-\alpha i_2 T_e} \leq \varepsilon e^{-\alpha(\tau - T_e)} \leq \varepsilon$, by Assumption 1. Thus, R_i will be less than ε before $i_2 T_e \leq \tau$. Moreover, $R_0 = \varepsilon$ and $R_i \leq R_0 \leq \varepsilon$ for all i before the first switch s_1. Hence, $z(\cdot)$ satisfies inequality (3) between time 0 and s_1. After that, we consider $R_{i_1+i_2}$ which is equal to $R_{i_1} e^{-\alpha i_2 T_e}$. But, we know from the previous argument that $R_{i_1} \leq \varepsilon e^{-\alpha T_d} + d(T_e)$. Thus, $R_{i_1} \leq \varepsilon$ by Assumption 1. Moreover, R_i is decreasing with an $e^{\alpha T_p}$ factor at each iteration before the next switch s_2. Therefore, $R_{i_1+i_2} \leq (\varepsilon e^{-\alpha T_d} + d(T_e)) e^{-\alpha i_2 T_e} \leq \varepsilon e^{-\alpha(\tau - T_e)} \leq \varepsilon$. Then, R_i will be less than or equal to ε before $(i_1+i_2)T_e \leq s_1 + \tau$. For $t \in [(i_1+i_2)T_e]$, $\|z(t) - \xi_\sigma(x_0,t)\| \leq \varepsilon e^{-\alpha(t-(i_1+i_2)T_e)}$ by part (a) of the previous argument. At $i = \lceil s_2/T_e \rceil$, R_i will be less than or equal $\varepsilon e^{-\alpha(i-(i_1+i_2))T_e} + d(T_e)$ which is less than or equal to $\varepsilon e^{-\alpha(T_d - (i_2)T_e} + d(T_e)$ because of the dwell time constraint. Thus, $R_i \leq \varepsilon$ by Assumption 1. Hence, $z(\cdot)$ does satisfy (3) between s_1 and s_2. Finally, by induction on all switches, $z(\cdot)$ satisfy the properties in (3). Therefore, $z(\cdot)$ is an approximating function to $\xi_\sigma(x_0,\cdot)$. □

Lemma 2 *$s_{est}(T,\varepsilon,\alpha,\tau)$ is upper-bounded by $(HN)^{\lfloor T/T_e \rfloor + 1}$, where $H = \lceil e^{(L+\alpha)T_e} \rceil^n$.*

PROOF. We count the number of functions that can be computed by the above procedure. First, note that a function $z(.)$ is defined by the quantization points and the modes chosen at multiples of T_e. Moreover, the cardinality of C_0 is $\#C_0 = \lceil \frac{diam(K)}{2\varepsilon e^{-(L+\alpha)T_e}} \rceil^n$, where $diam(K)$ is the diameter of K. The upper bound on the cardinality of C_i, for $i \geq 1$, is $\#C_i = \lceil \frac{R_i}{R_i e^{-(L+\alpha)T_e}} \rceil^n = \lceil e^{(L+\alpha)T_e} \rceil^n$, which is independent of R_i. At each iteration $0 \leq i \leq \lfloor T/T_e \rfloor$, we are choosing one from the N modes and a quantization point in the cover C_i. We can conclude that the number of functions that can be computed using the above procedure is upper bounded by $(\#C_0)(HN)^{\lfloor T/T_e \rfloor}$. □

Theorem 1 *If Assumption (1) is satisfied, then*

$$h_{est}(\varepsilon,\alpha,\tau) \leq (L+\alpha)n/\ln 2 + (\log N)/T_e.$$

PROOF. This proof follows along the lines of the proof of Proposition 2 in [8].

$$\begin{aligned}
&\limsup_{T\to\infty} \frac{1}{T} \log s_{\text{est}}(T,\varepsilon,\alpha,\tau) \\
&\leq \limsup_{T\to\infty} \frac{1}{T} \log(\#C_0)(HN)^{\lfloor \frac{T}{T_e} \rfloor + 1} \\
&= \limsup_{T\to\infty} \frac{1}{T} \log \#C_0 \\
&\qquad + \limsup_{T\to\infty} \frac{1 + T_e/T}{T_e} (\log \lceil e^{(L+\alpha)T_e} \rceil^n + \log N) \\
&\leq \frac{(L+\alpha)n}{\ln 2} + \frac{\log N}{T_e}.
\end{aligned}$$

The last step is follows from the fact that $\limsup_{T\to\infty} \frac{1}{T} \log \lceil \frac{diam(K)}{2\varepsilon e^{-(L+\alpha)T_e}} \rceil^n = 0$. □

Note that, if $N = 1$, we get the previous bound on entropy given in [8].

2.3 Relation between entropy and the bit rate of estimation algorithms

In the following proposition we prove that no bit rate less than h_{est} can be achieved by any algorithm that constructs an $(\varepsilon,\tau,\alpha)$-approximating function given any trajectory $\xi_\sigma(x_0,t)$ while having a fixed bit rate. Assume that the sampling time of the algorithm is T_p time units. The bit rate of the algorithm is defined as

$$b_r(\varepsilon,\tau,\alpha) := \limsup_{j\to\infty} \frac{1}{jT_p} \sum_{i=0}^{j} \log Q_i$$

where $\log Q_i$ is the number of bits sent at $t = iT_p$. Having a fixed bit rate means $\log Q_i = \log Q$ for all i. Hence, $b_r(\varepsilon,\tau,\alpha) = 1/T_p \log Q$.

Proposition 1 *Consider an algorithm with fixed bit rate at each iteration i. If for each trajectory of the system $\xi_\sigma(x,t)$, the trajectory constructed by the algorithm satisfies the properties in (3) for any ε, τ and $\alpha > 0$, then the algorithm's bit rate cannot be smaller than $h_{est}(\varepsilon,\alpha,\tau)$.*

PROOF. The proof is similar to the proof of Proposition 5 in [8]. Arguing for contradiction, assume that there exists such an algorithm that satisfies the properties and has a bit rate less than $h_{est}(\varepsilon,\alpha,\tau)$. Recall that $h_{\text{est}}(\varepsilon,\alpha,\tau) = \limsup_{T\to\infty} 1/T \log s_{\text{est}}(T,\varepsilon,\alpha,\tau)$. Then, there exists l large enough where $b_r(\varepsilon,\tau,\alpha)$ is less than $1/lT_p \log s_{\text{est}}(lT_p,\varepsilon,\alpha,\tau)$. Substituting $b_r(\varepsilon,\tau,\alpha)$ with $1/T_p \log Q$ leads to the inequality $Q^l < s_{\text{est}}(lT_p,\varepsilon,\alpha,\tau)$. Q^l is the number of possible sequences of quantized states q_i's of length l and the right hand side is the minimal cardinality of an $(lT_p,\varepsilon,\alpha,\tau)$-approximating set. Then, the set of trajectories that can be constructed by the algorithm defines an $(lT_p,\varepsilon,\alpha,\tau)$-approximating set which has a cardinality less than s_{est} which contradicts the assumption that s_{est} has the minimum cardinality. □

2.4 Separation of modes

In order for an algorithm to distinguish two modes $p, q \in [N]$, $p \neq q$, it is necessary for the solutions generated by the two modes to be separable in some sense. The following notion of *exponential separation* is proposed in [8]. For $L_s, T_s > 0$ we say that the two modes $p, q \in [N]$ are (L_s, T_s)-*exponentially separated* if there exists a constant $\epsilon_{\min} > 0$ such that for any $\varepsilon \leq \epsilon_{\min}$, for any two nearby initial states $x_1, x_2 \in \mathbb{R}^n$ with $|x_1 - x_2| \leq \varepsilon$,

$$|\xi_p(x_1,T_s) - \xi_q(x_2,T_s)| > \varepsilon e^{L_s T_s}.$$

That is, trajectories separate out exponentially if they start from a sufficiently small neighborhood. The exponential separation holds if, for example, (1) the two vector fields have a positive separation angle, and (2) at least one of them has a positive velocity. It is believed that this property is generic in the sense that it holds for almost all pairs of systems. We assume that the modes are mutually (L, T_p)-exponentially separated. Also, ϵ_{min} is assumed to be global for all pairs of modes.

3. STATE ESTIMATION

We consider a setup where there is a sensor sampling the state of the switched system each T_p time units without being able to sense the mode. The sensor sends a quantized version of the state along with other few bits over a communication

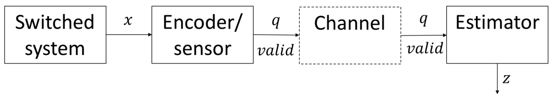

Figure 1: Block diagram showing the flow of information from the switched system to the sensor side algorithm to the estimator algorithm.

channel to the estimator. In turn, the estimator needs to compute $(\varepsilon, \alpha, \tau)$-approximating function of the trajectory of the system using the measurements received from the sensor (see Figure 1).

3.1 Estimation algorithm overview

First, we briefly discuss the basic principle of constant bit-rate state estimation for a single dynamical system (see for example [8]). In this case, the system evolves as $\dot{x}(t) = f_p(x(t))$, for a given $p \in [N]$, $x_0 \in K$, and there is no uncertainty about the mode. Suppose at a given time t the estimator has somehow computed a certain estimate for the state of the system, say represented by a hypercube S. In the absence of any new measurement information, the uncertainty in a state estimate or the size of S blows-up exponentially with time as $e^{L_p t}$, where L_p is the Lipschitz constant of f_p. In order to obtain the required exponentially shrinking state estimates, i.e., S shrinking as $e^{-\alpha t}$, the sensor has to send new measurements to the estimator.

One strategy is for the sensor to send information every $T_p > 0$ time units as follows: it partitions S, which has a radius r, into a grid with cells of radii $re^{-(L_p+\alpha)T_p}$, makes a quantized measurement of the state of the system $\xi(x_0, t)$ according to this grid and sends a few bits to the estimator so that the algorithm running at the estimator can identify the correct cell in which state resides (see Figure 1). At this point, the uncertainty in the state reduces by a factor of $e^{(L_p+\alpha)T_p}$ so that after T_p time units when the uncertainty grows by a multiple of $e^{L_p T_p}$ there is still a net reduction in uncertainty by a factor of $e^{\alpha T_p}$. It can also be seen that the number of bits the sensor needs to send (for identifying one grid cell out of $e^{(L_p+\alpha)T_p n}$) is $O(n(L_p + \alpha)T_p)$ and this gives the average bit rate of ${n(L_p+\alpha)}/{\ln 2}$.

Algorithm 1 which runs on the sensor side extends this strategy to work with switched systems. The basic idea is to track a number $(1 \leq \hat{N} \leq N)$ of possible modes that the system could be in, and run the above algorithm of quantization-based estimation, for each of these $\hat{N}$ modes. The set of tracked modes is stored in the vector m. A mode $m_i[r]$, $r \in [\hat{N}]$, is valid ($valid_i[r] = 1$) if the current state $\xi_\sigma(x_0, iT_p)$ is contained in the corresponding state estimate $S_i[r]$ at line 9 and $m_i[r] \neq -1$. However, it is possible that none of the $\hat{N}$ tracked modes are valid. In particular, the mode may switch and the state may evolve to fall outside of the estimates of the tracked modes or it may be that none of the $\hat{N}$ tracked modes in m_i is the actual mode of the system over $[(i-1)T_p, iT_p]$. This scenario where none of the modes are valid, the state is said to have *escaped* (line 15). In the case of an escape, the algorithm replaces all modes from the vector m and considers a new set of modes from $[N]$. If the rate of actual mode switches is slow enough (Lemma 4) then it is guaranteed to include the actual mode of the system in m before the next switch. And once the actual mode is tracked in m, then once again the state estimation converges exponentially.

In the above description of the algorithm, we suggested that each tracked mode $m_i[r]$ maintains its own corresponding state estimate $S_i[r]$ and quantization grid $C_i[r]$. This not only uses excessive memory, but also implies that $\hat{N}$ different quantized measurements of the state has to be sent by the sensor. In Algorithm 1, at any iteration $i \geq 1$, only a single state estimate S_i is maintained, a single grid C_i is computed according to which a single measurement is sent by the sensor. That is S_i and C_i are actually $S_i[mode_{i-1}]$ and $C_i[mode_{i-1}]$ where $mode_{i-1}$ is some $r \in [\hat{N}]$ agreed on between the sensor and the estimator. In our case we consider it the valid mode with the minimum index in m_i (line 11). In order to check the validity of the other tracked modes in m_i, the actual state is shifted with vectors which are computed according to the dynamics of these modes. That is, $v_i[r]$ represents the center of hypercube $S_i[r]$ which is the state estimate of the system corresponding to the dynamics $\dot{x} = f_{m[r]}(x)$. To check if $x_i \in S_i[r]$, x is shifted with the vector $v_i[mode_{i-1}] - v_i[r]$ and then checked if it belongs to S_i.

If there is an escape at a certain iteration, S_i is constructed as a hyperrectangle centered at $z_i(T_p)$ with radius δ_i plus $d(T_p)$. Recall, that δ_i is the radius used for computing S_i assuming that there is no escape (line 34) and $d(T_p)$ is the additional factor that capture maximum deviation between two trajectories of two different modes in $[N]$ starting from the same state in $Reach(\Sigma, K)$, the reachable states by (2), and running for T_p seconds. Next, q_i will be the quantization of x_i with respect to the new C_i computed in line 19 (see line 20).

The NextMode() function cycles through all the $[N]$ modes in the following two-phase fashion. For a sequence of N calls in phase I, it returns the modes in $[N]$ in some arbitrary order. Then, it returns -1 for the next $\hat{N} - 1$ calls in Phase II and then goes back to Phase I. Phase I is used by the estimation algorithm to cycle through all the modes fairly in discovering the actual mode after a switch. Phase II is used to keep the actual mode as the only mode tracked in m_i while the rest of m_i is equal to -1.

Estimator side algorithm.

On the estimator side, a similar algorithm to 1 is executed with small changes: instead of taking x_i as input (line 7), q_i, a quantized version of x_i, and the $valid_i$ vector are taken. Hence, the estimator knows if $x_i \in S_i[r]$ or not for a certain $r \in [N]$ by examining the $valid_i$ vector sent from the sensor. In addition, line 14 is replaced by “true”. Finally, lines 8 to 10, line 20 and line 22 are omitted. These lines only compute values which are sent by the sensor.

Reading the pseudocode.

$B(x_c, r_c)$ defines an over approximation of the initial set K as a hypercube of radius r_c centered at x_c. The **input** x_i (Line 7) executed at time t, reads the current state of the system $\xi_\sigma(x_0, iT_p)$ into the program variable x_i. In the next line $x_i \in S_i[r]$ is assumed to be computed by checking if $x_i + (v_i[mode_{i-1}] - v_i[r]) \in S_i$ if $i \geq 1$ and $x_i + (v_i[0] - v_i[r]) \in S_i$ if $i = 0$. In Line 11, the minimum index of a valid mode is assigned to $mode_i$ but this could be any arbitrary choice. If there is no valid mode then $mode_i$ is set to $\perp$.

Comparison with upper bound construction.
This algorithm is similar to the construction of an approximating function used in the proof of the upper bound in Section 2.2. However, the mode is known at the sampling times in the upper bound while it is not in the Algorithm. Thus, the construction used in the upper bound knows the iterations where the switch happens. That makes us being able to increase the size of the ball representing the state estimate in the iteration following a switch. However, because it is assumed that the mode is not known, Algorithm 1 needs to wait till the state x_i leaves the state estimate S_i to know that a switch happened or that a mode considered in m_i is different from the actual mode. That required the additional assumption that the modes are exponentially separated to bound the number of iterations needed for the state to leave a state estimate constructed based on a wrong mode. That required us to sample faster ($T_p \leq T_e$) and track several modes in parallel to figure out the actual mode and decrease δ_i to less than ε before τ time units after a switch.

Algorithm 1 Procedure for estimating the state of a switched system (sensor side).

1: input: T_p, α, ε, $K \subset B(x_c, r_c)$, $\hat{N}$
2: $m_0 \leftarrow \langle 0, 1, \dots \hat{N}-1\rangle$; $\delta_0 \leftarrow \varepsilon$;
3: $S_0 \leftarrow B(x_c, r_c)$;
4: $C_0 \leftarrow grid(S_0, \delta_0 e^{-(L+\alpha)T_p})$;
5: $mode_0 \leftarrow 0$; $i \leftarrow 0$;
6: **while** true **do** $\{i^{th}$ iteration$\}$
7: input x_i;
8: **for** $r \in [\hat{N}]$ **do**
9: $valid_i[r] \leftarrow [x_i \in S_i[r]$ *and* $m_i[r] \neq -1]$;
10: **end for**
11: $mode_i \leftarrow \min\{r \mid valid_i[r]\}$;
12: $escape \leftarrow mode_i \neq \bot$;
13: **if** not *escape* **then** {no escape}
14: $q_i \leftarrow quantize(x_i, C_i[mode_i])$;
15: **else** {escape}
16: $mode_i \leftarrow mode_{i-1}$;
17: $\delta_i \leftarrow d(T_p) + \delta_i$;
18: $S_i \leftarrow B(z_i(T_p), \delta_i)$;
19: $C_i \leftarrow grid(S_i, \delta_i e^{-(L+\alpha)T_p})$;
20: $q_i \leftarrow quantize(x_i, C_i[mode_i])$;
21: **end if**
22: send $\langle q_i, valid_i\rangle$;
23: i++; {parameters for next iteration}
24: $m_i \leftarrow m_{i-1}$;
25: **for** $r \in [\hat{N}]$ **do**
26: **if** *escape* or (not $valid_{i-1}[r]$ and $m_i[r] \neq -1$) **then**
27: $m_i[r] \leftarrow NextMode()$;
28: **end if**
29: **if** $m_i[r] \neq -1$ **then**
30: $v_i[r] \leftarrow \xi_{m_i[r]}(q_{i-1}, T_p)$;
31: **end if**
32: **end for**
33: $\delta_i \leftarrow e^{-\alpha T_p}\delta_{i-1}$;
34: $S_i \leftarrow B(v_i[mode_{i-1}], \delta_i)$;
35: $C_i \leftarrow grid(S_i, \delta_i e^{-(L+\alpha)T_p})$;
36: $z_i(.) \leftarrow \xi_{m_i[mode_{i-1}]}(q_{i-1}, \cdot)$;
37: wait(T_p);
38: **end while**

4. ANALYSIS OF ESTIMATION ALGORITHM

In this section, we prove a sequence of error bounds on the state estimate for different cases that arise from considering a mode which is different from the actual mode over a time interval of size T_p. Then in Section 4.2 we establish bounds on the maximum number of possible escapes between switches. The main Theorem in Section 4.3 uses these results together with an upper bound on the speed of mode switches to give detailed bounds on the state estimation error. Finally, in Section 4.4 we analyze the average bit rate and compare it to the upper bound on h_{est} defined in Theorem 1.

Notations.
We fix all the parameters of the algorithm including the sampling period T_p and the mode window size $\hat{N}$. We also fix a particular (unknown) initial state $x_0 \in K$ and a particular (unknown) switching signal σ for the system described by Equation (2). This defines a particular solution $\xi_\sigma(x_0, \cdot)$ of the switched system and the sequence of states $\xi_\sigma(x_0, T_p), \xi(x_0, 2T_p), \dots,$ sampled by Algorithm 1 which runs on the sensor side. We abbreviate $\xi_\sigma(x_0, iT_p)$ as x_i and the quantized measurement of x_i that is sent by the sensor as q_i. Moreover, δ_i, S_i, C_i, etc., denote the valuations of the variables δ, S, C, etc. at line 22 in the i^{th} iteration of the algorithm. However, the modes in m_{i+1} are the modes considered over the interval $(iT_p, (i+1)T_p]$. The switching times in σ are denoted by $s_0 = 0, s_1, \dots$ For a given switching time s_j, we define $last(j) := \lfloor s_j/T_p \rfloor$ and $next(j) := \lceil s_j/T_p \rceil$ as the last iterations of the algorithm before the j^{th} switch and the first iteration after the j^{th} switch respectively.

Recall, that an escape occurs when the state of the system $\xi(x_0, iT_p)$ is not in any of the state estimates $S_i[r]$'s at line 9. In other words, it occurs when the algorithm takes the **else** branch in Line 15 ($mode = \bot$) after s_j.

4.1 Error bounds across a single iteration

In this section, we establish how the error in state estimation, $\|\xi_\sigma(x_0, t) - z(t)\|_\infty$, evolves over a single iteration of the algorithm, that is, over $t \in [iT_p, (i+1)T_p]$. The estimate $z(t)$ over $[iT_p, (i+1)T_p]$ is $\xi_{m_{i+1}}[r](q_i, \cdot)$ for some r, and therefore, we track the error by bounding $\|\xi_\sigma(x_0, t) - \xi_{m_{i+1}[r]}(q_i, t)\|_\infty$, for all $r \in [\hat{N}]$ with $m_{i+1}[r] \neq -1$.

There are several sub-cases to consider based on (a) whether there is a switch, and (b) whether the tracked mode $m_{i+1}[r]$ matches the actual mode at a given time, over the considered interval between the iterations. For each of these cases, we establish a bound on $\|\xi_\sigma(x_0, t) - z(t)\|_\infty$ using (a) Bellman-Gronwall inequality to bound $\|\xi_u(x, t) - \xi_u(x', t)\|_\infty$, and (b) triangular inequality to bound $\|\xi_u(x, t) - \xi_p(x', t)\|_\infty$, where $u \neq p \in [\hat{N}]$ and $x \neq x' \in \mathbb{R}^n$. Recall that $T_p \leq \tau \leq T_d$, so no more than one switch can happen between iT_p and (i+1)T_p.

Each of the following propositions covers one of the above cases. Proposition 2 considers the case when there is a switch between iT_p and $(i+1)T_p$, the considered mode $m_{i+1}[r]$ is the same as the actual mode $\sigma(iT_p)$ at $t = iT_p$, and there exists a state estimate $S_i[p]$ that contains the actual state $\xi_\sigma(x_0, iT_p)$ at $t = iT_p$. It shows that the estimate converges exponentially until the switch, and after that it accumulates an additive factor of $d(T_p)$.

Proposition 2 *Fix an iteration i, a switching time $s_j \in$*

$(iT_p, (i+1)T_p)$, *and an index* $r \in [\hat{N}]$. *If* $m_{i+1}[r] = \sigma(iT_p)$ *and* $x_i \in S_i[p]$ *for some* $p \in [\hat{N}]$, *then for all* $t \in [iT_p, (i+1)T_p]$, $\|\xi_\sigma(x_0, t) - \xi_{m_{i+1}[r]}(q_i, t - iT_p)\|_\infty \leq$

$$\begin{cases} \delta_i e^{-\alpha(t-iT_p)} & \text{if } t < s_j \quad (4) \\ d(T_p) + \delta_i e^{-\alpha(t-iT_p)} & \text{otherwise.} \quad (5) \end{cases}$$

Proof. For (4), $\|x_i - q_i\|_\infty \leq \delta_i e^{-(L+\alpha)T_p}$ since $x_i \in S_i[p]$ for some $p \in [\hat{N}]$ and the boxes in $C_i[p]$ are of radii $\delta_i e^{-(L+\alpha)T_p}$. Then, $\|\xi_\sigma(x_0, t) - \xi_{m_{i+1}[r]}(q_i, t - iT_p)\|_\infty$

$$= \|\xi_\sigma(x_i, t - iT_p) - \xi_{m_{i+1}[r]}(q_i, t - iT_p)\|_\infty$$
[since $\xi_\sigma(x_0, t) = \xi_\sigma(\xi_\sigma(x_0, iT_p), t - iT_p)$]

$$= \|\xi_{m_{i+1}[r]}(x_i, t - iT_p) - \xi_{m_{i+1}[r]}(q_i, t - iT_p)\|_\infty$$
[$\sigma(iT_p) = m_{i+1}[r]$]

$$\leq e^{L_{m_{i+1}[r]}(t-iT_p)}\|x_i - q_i\|_\infty$$
[Bellman-Gronwall inequality]

$$\leq \delta_i e^{L_{m_{i+1}[r]}(t-iT_p)} e^{-(L+\alpha)T_p}$$
[q_i is quantization of x_i]

$$\leq \delta_i e^{-\alpha(t-iT_p)}.$$

The last inequality follows because $L_{m_{i+1}[r]} \leq L$ and $t - iT_p \leq T_p$. For (5), we assume without loss of generality that $m_{i+1}[r] = \sigma(t) = 1$ for $t \in [iT_p, s_j)$, $\sigma(t) = 2$ for $t \in [s_j, (i+1)T_p]$. Then, $\|\xi_\sigma(x_0, t) - \xi_{m_{i+1}[r]}(q_i, t - iT_p)\|_\infty$

$$= \|\xi_2(\xi_1(x_0, s_j), t - s_j) - \xi_1(\xi_1(q_i, s_j - iT_p), t - s_j)\|_\infty$$
$$\leq \|\xi_2(\xi_1(x_0, s_j), t - s_j) - \xi_1(\xi_1(x_0, s_j), t - s_j)\|_\infty$$
$$+ \|\xi_1(\xi_1(x_0, s_j), t - s_j) - \xi_1(\xi_1(q_i, s_j - iT_p), t - s_j)\|_\infty$$
[by triangle inequality]

$$\leq \| \int_0^{t-s_j} (f_2(\xi_2(\xi_1(x_0, s_j), t')) - f_1(\xi_1(\xi_1(x_0, s_j), t')))dt' \|$$
$$+ \|\xi_1(\xi_1(x_0, s_j), t - s_j) - \xi_1(\xi_1(q_i, s_j - iT_p), t - s_j)\|_\infty$$
$$\leq d(t - s_j) + e^{L_1(t-iT_p)}\|x_i - q_i\|_\infty$$
[by Bellman-Gronwall inequality]

$$\leq d(T_p) + e^{L_1(t-iT_p)} e^{-(L+\alpha)T_p}\delta_i \leq d(T_p) + \delta_i e^{-\alpha(t-iT_p)}.$$

□

The next proposition holds under the same conditions as Proposition 2 except that the considered mode $m_{i+1}[r]$ matches the mode of the switched system $\sigma((i+1)T_p)$ at $t = (i+1)T_p$ iteration, but it is not the same as $\sigma(iT_p)$. The proof of (6) is analogous to the proof of (5).

Proposition 3 *Fix an iteration* i, *a switching time* $s_j \in (iT_p, (i+1)T_p)$, *and an index* $r \in [\hat{N}]$. *If* $m_{i+1}[r] \neq \sigma(iT_p)$, $m_{i+1}[r] = \sigma((i+1)T_p)$ *and* $x_i \in S_i[p]$ *for some* $p \in [\hat{N}]$, *then, for all* $t \in [iT_p, (i+1)T_p]$,
$\|\xi_\sigma(x_0, t) - \xi_{m_{i+1}[r]}(q_i, t - iT_p)\|_\infty \leq$

$$\begin{cases} d(T_p) + \delta_i e^{-\alpha(t-iT_p)} & \text{if } t < s_j \quad (6) \\ 2d(T_p) + \delta_i e^{-\alpha(t-iT_p)} & \text{otherwise.} \quad (7) \end{cases}$$

Proof. For (7), $\|\xi_\sigma(x_0, t) - \xi_{m_{i+1}[r]}(q_i, t - iT_p)\|_\infty$

$$\leq \|\xi_\sigma(x_0, t) - \xi_{\sigma(iT_p)}(x_i, t - iT_p)\|_\infty$$
$$+ \|\xi_{\sigma(iT_p)}(x_i, t - iT_p) - \xi_{m_{i+1}[r]}(q_i, t - iT_p)\|_\infty$$
[by triangle inequality]

$$\leq \|\xi_\sigma(\xi_\sigma(x_0, s_j), t - s_j) - \xi_{\sigma(iT_p)}(\xi_\sigma(x_0, s_j), t - s_j)\|_\infty$$
$$+ \|\xi_{\sigma(iT_p)}(x_i, t - iT_p) - \xi_{m_{i+1}[r]}(q_i, t - iT_p)\|_\infty$$

$$\leq d(t - s_j) + d(T_p) + \delta_i e^{-\alpha(t-iT_p)}$$
[by similar argument to (5)]

$$\leq 2d(T_p) + \delta_i e^{-\alpha(t-iT_p)}.$$

□

Proposition 4 also holds under the same conditions as Proposition 2 except that the considered mode $m_{i+1}[r]$, the actual mode $\sigma(iT_p)$ at the i^{th} iteration and $\sigma((i+1)T_p)$ at the $(i+1)^{st}$ iteration are all distinct. Inequality (8) is the same as (6). Also, the proof of (9) is analogous to the proof of (7).

Proposition 4 *Fix an iteration* i, *a switching time* $s_j \in (iT_p, (i+1)T_p)$, *and an index* $r \in [\hat{N}]$. *If* $m_{i+1}[r] \neq \sigma(iT_p)$, $m_{i+1}[r] \neq \sigma((i+1)T_p)$, $m_{i+1} \neq -1$ *and* $x_i \in S_i[p]$ *for some* $p \in [\hat{N}]$, *then, for all* $t \in [iT_p, (i+1)T_p]$,
$\|\xi_\sigma(x_0, t) - \xi_{m_{i+1}[r]}(q_i, t - iT_p)\|_\infty \leq$

$$\begin{cases} d(T_p) + \delta_i e^{-\alpha(t-iT_p)} & \text{if } t < s_j \quad (8) \\ 2d(T_p) + \delta_i e^{-\alpha(t-iT_p)} & \text{otherwise.} \quad (9) \end{cases}$$

From the above Propositions, it follows immediately that if there is no switch between the i^{th} and the $(i+1)^{st}$ iteration, then the bounds given by inequalities (4), (6) and (8) will continue to hold for the entire period between the iterations.

The following assumption will be used to prove several intermediate results about the estimation algorithm detecting the right mode and estimation bounds. Then, in Lemma 4 in Section 4.2, we will establish a lower bound on the dwell-time T_d which guarantees this assumption.

Assumption 2 For each switching time s_j other than $s_0 = 0$, let $i = last(j)$. Then, there exists $r \in [\hat{N}]$ where $m_{i+1}[r]$ is the actual mode of the system $\sigma(iT_p)$ and $m_{i+1}[p] = -1$ for all $p \neq r$ and $\delta_i \leq \min\{\varepsilon, \epsilon_{min}\}$.

Proposition 5 *Under Assumption 2, for each* i *there exists* $r \in [\hat{N}]$ *with* $x_i \in S_i[r]$.

Proof. If there is an escape at iteration i, then the state x_i is not in any of the $S_i[r]$'s at line 9, however, it is still guaranteed to be in all the expanded (corrected) estimates $S_i[r]$'s computed at line 18 based on δ_i and $d(T_p)$. That is because, under Assumption 2, inequalities (7) and (9) in Propositions 3 an 4, are not relevant (they are useful for analyzing the error bounds for faster switching signals). Therefore, line 17 takes care of the worst case scenario in the estimation error over a single iteration. □

4.2 Bounding escapes between switches

Proposition 6 upper bounds the number of escapes that can happen between two consecutive switches to $\lceil N/\hat{N} \rceil$.

Proposition 6 *Under Assumption 2, the maximum number of escapes between two consecutive switches is* $\lceil N/\hat{N} \rceil$.

PROOF. First, note that at an escape, all the $\hat{N}$ invalid modes are dropped from the vector m_i and new candidate modes are added fairly by the *NextMode*() function. Hence, all the N modes would have been considered after $\lceil N/\hat{N} \rceil$ escapes. Thus, the correct mode $\sigma(t)$ would have been in m at some iteration i. Then, let $m_{i+1}[r] = \sigma(iT_p)$. Second, we know that $x_i \in S_i[p]$ for some $p \in [\hat{N}]$ by Proposition 5. Therefore, we can apply the estimation error bound given by (4) in Proposition 2 to conclude that in the next iteration $valid_i[r]$ will be set to 1 and will remain thereafter until a new switch occur. Thus, there will be no more escapes till the next switch. $\square$

Remark 1 If $\exists p, u \in [\hat{N}]$, where $\forall x \in Reach(\Sigma, K)$ and $\forall t \in [0, T_p]$, $\|\xi_p(x_1, t) - \xi_u(x_2, t)\|_\infty \leq \|x_1 - x_2\|_\infty e^{Lt}$, then they will not be exponentially separated. They will behave the same as far as our algorithm is concerned. So, we can remove one of them from the set of possible modes $[N]$ and our algorithm will still have the same correctness guarantees. This may arise, for example, if there are modes that are exponentially stable with convergence exponent larger than or equal to α and a common equilibrium point.

Because of the exponential separation property, we can show that if the dwell time of the switching signal is large enough, then after some maximum number of iterations after a switch, the actual mode $\sigma(t)$ still remains unchanged and the size of the state estimate S_i will be small enough to the point that all incorrect modes in m_i will be invalidated. We define $i_{inv}(\delta)$ to be an upper bound on the number of iterations needed to invalidate a mode when the current radius of the ball representing the state estimate S is δ. Let us define: for any $\delta > 0$,

$$i_{inv}(\delta) := \max\{\lceil \frac{1}{\alpha T_p} \ln(\frac{\delta}{\epsilon_{min}}) - \frac{L}{\alpha} \rceil, 1\}.$$

Proposition 7 *Under Assumption 2, if at a given iteration* $i \geq 0$, $-1 \neq m_{i+1}[r] \neq \sigma(t)$, *then* $m_{i+1}[r]$ *will be replaced with a different mode after a maximum of* $i_{inv}(\delta_i)$ *iterations.*

PROOF. Let $c = \lceil \frac{1}{\alpha T_p} \ln(\frac{\delta_i}{\epsilon_{min}}) - \frac{L+\alpha}{\alpha} \rceil$. First, note that until $m_{i+1}[r]$ is replaced, δ_i will be decreasing by a $e^{\alpha T_p}$ factor in each iteration (because there is no escape if it is not replaced). Then, $\delta_{i+c} e^{-(L+\alpha)T_p} = \delta_i e^{-((i+c)-i)\alpha T_p} e^{-(L+\alpha)T_p} < \epsilon_{min}$. Thus, by the exponential separation property:

$$\begin{aligned}
&\|\xi_\sigma(x_i, (c+1)T_p) - \xi_{m_{i+c+1}[r]}(q_{i+c}, T_p)\|_\infty \\
&= \|\xi_\sigma(x_{i+c}, T_p) - \xi_{m_{i+c+1}[r]}(q_{i+c}, T_p)\|_\infty > \delta_{i+c} e^{-(L+\alpha)T_p} e^{LT_p} \\
&= \delta_{i+c+1}. \qquad \text{[computed at line 33]}
\end{aligned}$$

Thus, the actual state will not belong to $S_{i+c+1}[r]$ computed at line 34 and $m_{i+c+2}[r] \neq m_{i+c+1}[r]$. $\square$

We upper bound the radius δ_i of the state estimate S_i at iteration i with,

$$\delta_{max} := \max_{i \in [1, \lceil N/\hat{N} \rceil]} \{\varepsilon e^{-i\alpha T_p} + d(T_p) \frac{1 - e^{-i\alpha T_p}}{1 - e^{-\alpha T_p}}\}.$$

Note that the first term decays geometrically with i and the second term increases, and the max value could be attained somewhere in the middle.

Proposition 8 *Under Assumption 2,* $\delta_i \leq \delta_{max}$ *for all* i.

PROOF. The radius δ_i of S_i decreases between two escapes and possibly increase at an escape. Therefore, the maximum of δ_i would achieved if some number of escapes (less than or equal to $\lceil N/\hat{N} \rceil$) happened in consecutive iterations immediately after a switch. Assumption 2 is used to make sure that $\delta_i \leq \varepsilon$ at $i = last(j)$. $\square$

The following definitions and two lemmas are used to compute the minimum dwell-time that suffices for Assumption 2 to be true. The following i_{det} represents the maximum number of iterations needed after a switch for the actual mode to be detected, all other modes be invalidated and $\delta_i \leq \epsilon_{min}$.

$$\begin{aligned}
i_{det} &:= 2 + \sum_{i=1}^{\lceil N/\hat{N} \rceil} i_{inv}\big(\varepsilon e^{-i\alpha T_p} + d(T_p) \sum_{j=0}^{i-1} e^{-j\alpha T_p}\big) \\
&\leq \lceil \frac{N}{\hat{N}} \rceil i_{inv}(\delta_{max}) + 2
\end{aligned}$$

Lemma 3 *Under Assumption 2, after a maximum of* i_{det} *iterations of any switch* s_j, $m_{i+1}[r] = \sigma(t)$, *for some* $r \in [\hat{N}]$, $m_{i+1}[u] = -1$ *for all* $u \neq r$ *and* $\delta_i \leq \epsilon_{min}$.

PROOF. (sketch) After a switch, the only mode considered in m_i will no longer be the correct mode. In the worst case, $\sigma(t)$ will be considered in the last set of modes m_{i+1}. Each set of modes m_{i+1} needs a maximum of $i_{inv}(\delta_i)$ iterations to be invalidated. Moreover, there is a maximum of $\lceil N/\hat{N} \rceil$ escapes. The first escape will happen after a maximum of 2 iterations after the switch to invalidate $m_{i+1}[r]$ by the exponential separation assumption since $\delta_i \leq \epsilon_{min}$ before the switch. Since i_{inv} is monotonically increasing w.r.t δ, we summed the values of i_{inv} when evaluated on the $\lceil N/\hat{N} \rceil$ maximum possible values of δ_i. The last $i_{inv}(\delta_{max})$ in i_{det} is to invalidate all wrong modes (and replace them with -1) and keep the actual one in m_i. It will also make $\delta_i \leq \epsilon_{min}$ by the definition of $i_{inv}(\delta_{max})$. $\square$

Finally, we define the following to upper bound the number of iterations, with no escapes, needed to decrease δ_i from ϵ_{min} to less than ε:

$$i_{est} := \max(\lceil \frac{1}{\alpha T_p} \ln(\frac{\epsilon_{min}}{\varepsilon}) \rceil, 0).$$

Lemma 4 *If the minimum dwell-time of the switching signal* σ *is greater than* $(i_{det} + i_{est} + 1)T_p$, *then Assumption 2 is true.*

PROOF. Lemma 3 holds between $s_0 = 0$ and s_1 given the minimum dwell time and the fact that $\epsilon_{min}\ e^{-\alpha T_p (i_{est})} \leq \varepsilon$ without Assumption 2. Then, the argument holds inductively for the rest of the intervals. $\square$

4.3 Estimation error

Combining the above we derive bounds on the estimation error in Theorem 2. It shows that after a switch, the algorithm will be in four possible "phases". The estimation error will increase in the first few iterations after a switch where escapes occur, until the correct mode is found in m, and thereafter, the estimate converges exponentially, provided the dwell time is large enough.

Let the iterations of the algorithm when escapes occur between two consecutive switches s_j and s_{j+1} be numbered $w_1, \ldots w_k$. Fixing j we avoid indexing the w's and k with j.

Theorem 2 *If σ has dwell time $T_d \geq (i_{det} + i_{est} + 1)T_p$, then for any $t \in [s_j, s_{j+1})$, the estimation error* $\|\xi_\sigma(x_0, t) - z(t)\|_\infty \leq$

$$\begin{cases} d(T_p) + \varepsilon e^{-\alpha(t - last(j)T_p)} & \textit{if } t \in [s_j, w_1 T_p] \quad (10) \\ d(T_p) + \delta_{w_h} e^{-\alpha(t - w_h T_p)} & (11) \\ \quad \textit{if } \exists\, h \in \{1, \ldots, k\}, t \in [w_h T_p, w_{h+1} T_p] & \\ d(T_p) + \delta_{w_k} e^{-\alpha(t - w_k T_p)} & (12) \\ \quad \textit{if } t \in [w_k T_p, (w_k + i_{inv}(\delta_{w_k}))T_p] & \\ \delta_{w_k} e^{-\alpha(t - w_k T_p)} & \textit{otherwise.} \quad (13) \end{cases}$$

PROOF. We start by proving (10): By Lemma 4, $\delta_{last(j)} \leq \epsilon_{min}$, $\delta_{last(j)} \leq \varepsilon$ and $z(t) = \xi_\sigma(q_{last(j)}, t - last(j)T_p)$ for $t \in [last(j)T_p, s_j)$. Then, by inequality (5) in Proposition 2 , the inequality is satisfied for $t \in [s_j, next(j)T_p]$. Moreover, if w_1, the first escape after s_j, was not at $next(j)$ then it will be at $next(j)+1$, since, by the exponential separation property, $\|z(t) - \xi_\sigma(x_0, t)\| \geq \varepsilon e^{LT_p}$, so $w_1 = next(j)+1$. If that is the case, then the inequality holds for $t \in [next(j)T_p, (next(j) + 1)T_p]$ as a result of inequality (6) in Proposition 3 and the fact that $\delta_{next(j)} \leq \varepsilon e^{-\alpha T_p} \leq \varepsilon$.

Inequalities (11) and (12) have similar proosf as (10) but instead of ε we have δ_{w_h}. Inequality (13) follows from the fact that at $t = (w_k + i_{inv}(\delta_{w_k})T_p)$ there is $r \in [\hat{N}]$ with $m[r] = \sigma(s_j)$ and $m[p] = -1$ for $p \neq r$, and the repeated application of inequality (4) in proposition 2. □

Corollary 1 summarizes the error bounds in terms of two types of time intervals.

Corollary 1 *Under the assumptions of Lemma 4, consider the time between the two consecutive switches s_j and s_{j+1}. Then, for all $t \in [s_j, s_{j+1})$, $\|\xi_\sigma(x_0, t) - z(t)\|_\infty \leq$*

$$\begin{cases} \delta_{max} + d(T_p) & t \in [s_j, w_k T_p] \quad (14) \\ \delta_{w_k} e^{-\alpha(t - w_k T_p)} & \textit{otherwise.} \quad (15) \end{cases}$$

Thus, for a given ε, τ and α defined as for Theorem 1, we can choose the parameters for the algorithm T_p and $\hat{N}$ to control the variables i_{det}, $d(T_p)$ and δ_{max} so as to achieve inequalities in (3).

4.4 Optimal network usage

We show that the estimation algorithm uses network bandwidth optimally in the following sense: An analysis similar to that of Proposition 4 of [8] shows that the average bit rate used by our algorithm is ${(L+\alpha)n}/{\ln 2} + {\hat{N}}/{T_p}$. The sensor needs to send (a) q_i: the quantization of x_i with respect to one of the $\hat{N}$ $S_i[r]$'s and (b) the $valid_i$ bit vector: for each $r \in [\hat{N}]$ one bit indicating whether or not x_i belongs to $S_i[r]$. The quantized state q_0 requires $\#C_0 = \lceil \frac{diam(K)}{2\varepsilon e^{-(L+\alpha)T_p}} \rceil^n$ bits to be sent. For $i \geq 1$, the number of bits required to represent q_i is $\#C_i = \lceil \frac{\delta_i}{\delta_i e^{-(L+\alpha)T_p}} \rceil^n = \lceil e^{(L+\alpha)T_p} \rceil^n$. Hence, the average bit rate used by the algorithm is $b_r(\varepsilon, \alpha, T_p) = \lim_{i \to \infty} {1}/{T_p} \log(\#C_i \hat{N}) = \frac{(L+\alpha)n}{\ln 2} + \frac{\hat{N}}{T_p}$.

Theorem 3 *Average bit rate of Algorithm 1 is $\frac{(L+\alpha)}{\ln 2} + \frac{\hat{N}}{T_p}$.*

Hence, it follows that the bit-rate used by the estimation algorithm is larger than the upper bound on the estimation entropy by at most $\frac{\hat{N}}{T_p} - \frac{\log N}{T_e}$ bits. Therefore, the efficiency gap between the bit-rate used by our algorithm and the bit rate (h_{est}) used by the best possible algorithm, is at most $\frac{\hat{N}}{T_p} - \frac{\log N}{T_e}$ bits more than the gap between h_{est} and its upper-bound. The unobservability of the switching signal and the switching times contributes to the gap.

5. EXPERIMENTS

Switched affine systems.

In a switched affine system, the dynamics of all the modes are of the form: $\dot{x} = A_p x + B_p u$. We present estimation of a five dimensional switched linear system with five modes. For each $p \in [5] = \{0, \ldots 4\}$ the matrix A_p and the column vector B_p are generated randomly, and the input u is also a random constant. In the presented results, the settling time for the first mode is 11.8914 and the others are unstable. The maximum Lipschitz constant was $L = 28.28$. We work with switching signals that satisfy Assumption 2. We chose the following parameters $\alpha = 1$, $T_p = 0.1s$, $\varepsilon = 2$ and $\hat{N} = 2$. The first two dimension of the results obtained are shown in Figure 2 (Left). Notice how the state estimate (yellow and blue) is enlarged after an escape and how the state and the mode converge to the true ones. $d(T_p)$ was approximated at each escape by computing the distance between all possible pairs of modes starting from the state at that escape. It was around 2. The bit rate used here is $(L+\alpha)n/\ln 2 + \hat{N}/T_p = 231$ bps. The maximum time needed to detect the correct mode and start exponential convergence was 2.2 seconds and the maximum δ was around 3. So, if $\tau \geq 2.2s$ and $\varepsilon \geq 5$, the parameters of the algorithm in this experiment satisfy the properties in (3).

Nonlinear glycemic index model.

Estimating the blood glucose level is an important problem for administering controlling insulin for diabetes patients given [4]. We consider a polynomial switched system model of plasma glucose concentration [1]. The model has nine modes representing different control inputs. The state consists of three variables: G, I and X. In this model, the switching between different modes are brought about by certain threshold based rules depending on the state variables. In the span of $150s$ of each execution, 6 switches happened. Although Assumption 2 was not always satisfied, it was still able to do state estimation. The Lipschitz constant of each of the modes is estimated through sampling. The parameters of the algorithm are chosen as $\alpha = 1$ and $T_p = 1s$. For each value of $\hat{N} \in [1, 9]$, 100 initial states x_0 are drawn randomly and the algorithm is executed on the resulting solutions $\xi_\sigma(x_0, \cdot)$. Two sample executions are shown in Figure 2 and the average results are shown in the table below.

As the number of modes tracked $\hat{N}$ increases, as expected, the number of escapes decreases. In fact, beyond $\hat{N} = 5$, the marginal benefit to sending more bits is small as far as the worst case error estimate (δ_{max}) is concerned. In practice,

[1]Switched system benchmark available from: https://ths.rwth-aachen.de/research/projects/hypro/glycemic-control/

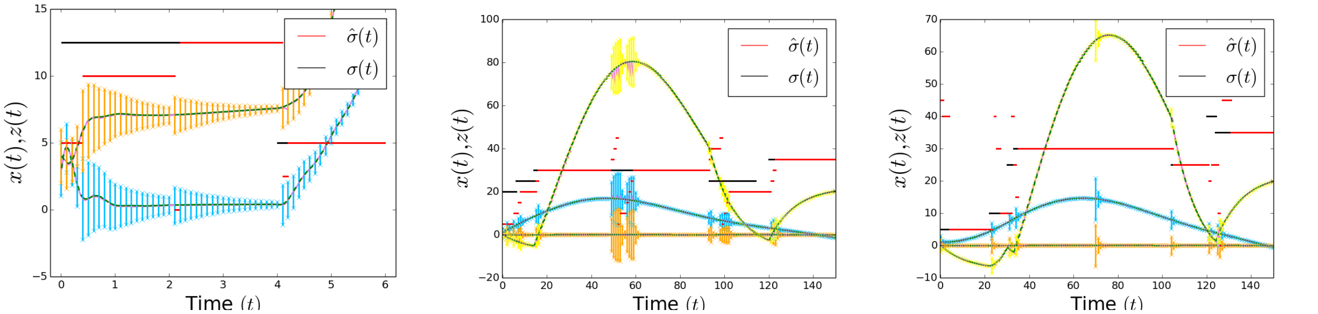

Figure 2: Execution of estimation algorithm. Actual mode (black), mode estimate (red), the values of the other variables are shown by the continuous plots. The vertical cut lines show the error estimates (δ) on those variables. Linear five dimensional system (*left*), Glycemic nonlinear control system, $\hat{N} = 1$ (*center*) and $\hat{N} = 9$ (*right*). Figure with $\hat{N} = 9$ has much less escapes than that with $\hat{N} = 1$.

the choice for $\hat{N}$ and T_p should be chosen to satisfy the convergence parameters specified.

$\hat{N}$	δ_{max}	Escps.
1	14.17	25
2	12.97	12.92
3	12.3	8.95
4	10.16	6.95
5	9.67	6.38
6	10.12	6.5
7	9.67	6.06
8	9.66	6.0
9	9.59	5.81

6. CONCLUSION

We have presented an algorithm for state estimation of switched nonlinear systems with finite number of modes and unobservable switching signal using quantized measurements with optimality guarantees on the number of bits needed to be sent from the sensor to the estimator. These results suggest several future research directions including extensions to hybrid models with partially known switching structure, models with input disturbances, and possibly developing lower-bounds on corresponding notions of estimation entropy.

7. REFERENCES

[1] Bosch. CAN specification version 1.0. Technical report, Robert Bosch Gmbh, Postfach 50, Stuttgart, September 1991.

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

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

[4] S. Furler, E. Kraegen, R. Smallwood, and D. Chisholm. Blood glucose control by intermittent loop closure in the basal mode: computer simulation studies with a diabetic model. *Diabetes Care*, 8:553???–561, 1985.

[5] J. Hespanha and A. Morse. Stability of switched systems with average dwell-time. In *Proceedings of 38th IEEE Conference on Decision and Control*, pages 2655–2660, 1999.

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

[7] D. Liberzon. *Switching in Systems and Control*. Systems and Control: Foundations and Applications. Birkhauser, Boston, June 2003.

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

[9] D. Liberzon and S. Mitra. Entropy notions for state estimation and model detection with finite-data-rate measurements. In *CDC*, Las Vegas, 2016.

[10] G. N. Nair. A nonstochastic information theory for communication and state estimation. *IEEE Transactions on automatic control*, 58(6):1497–1510, 2013.

[11] G. N. Nair and R. J. Evans. Stabilization with data-rate-limited feedback: tightest attainable bounds. *Systems Control Lett.*, 41:49–56, 2000.

[12] A. V. Savkin. Analysis and synthesis of networked control systems: Topological entropy, observability, robustness and optimal control. *Automatica*, 42:51–62, 2006.

[13] C. Steiner. Bot in the delivery:kiva systems. *Forbes Magazine*, March 2009. http://www.forbes.com/forbes/2009/0316/040_bot_time_saves_nine.html.

[14] R. Van Der Merwe, A. Doucet, N. De Freitas, and E. Wan. The unscented particle filter. In *NIPS*, volume 2000, pages 584–590, 2000.

[15] E. A. Wan and A. T. Nelson. Dual kalman filtering methods for nonlinear prediction, smoothing and estimation. In *NIPS 9, Denver, CO, USA, December 2-5, 1996*, pages 793–799, 1996.

[16] R. Wilson and L. Finkel. A neural implementation of the kalman filter. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, editors, *NIPS 22*, pages 2062–2070. Curran Associates, Inc., 2009.