Quantizer design for switched linear systems with minimal data-rate*

Guilherme S. Vicinansa
Daniel Liberzon
gs16@illinois.edu
liberzon@illinois.edu
Coordinate Science Laboratory, University of Illinois at Urbana-Champaign
Urbana, IL, USA

ABSTRACT

In this paper, we present a quantization scheme that reconstructs the state of switched linear systems with a prescribed exponential decaying rate for the state estimation error. We show how to use the Lyapunov exponents and a geometric object called Oseledets' filtration to design such a quantization scheme. Then, we prove that this algorithm works at an average data-rate close to the estimation entropy of the given system. Furthermore, we can choose the average data-rate to be arbitrarily close to the estimation entropy whenever the switched linear system has the so-called regularity property. We show that, under the regularity assumption, the quantization scheme is completely causal in the sense that it only depends on information that is available at the current time instant. Finally, we present simulation results for a Markov Jump Linear System, a class of systems for which the realizations are known to be regular with probability 1.

CCS CONCEPTS

• **Computing methodologies → Discrete-event simulation**; **Uncertainty quantification**; **Continuous models**; **Systems theory**.

KEYWORDS

Estimation entropy; Switched systems; Quantizer design

ACM Reference Format:
Guilherme S. Vicinansa and Daniel Liberzon. 2021. Quantizer design for switched linear systems with minimal data-rate. In *24th ACM International Conference on Hybrid Systems: Computation and Control (HSCC '21), May 19–21, 2021, Nashville, TN, USA.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3447928.3456645

1 INTRODUCTION

Nowadays, most dynamic systems found in engineering applications have distributed components, such as sensors, controllers, and actuators. For these components to transmit information to each other, we need to use communication channels. Those communication channels, by their turn, impose constraints on the data rate that can be transmitted. Therefore, it is natural to ask what is the minimum data rate needed for us to satisfy the application requirements, such as being able to reconstruct the system's state or stabilize the system.

The answers to the previous questions are invariably related to some definition of entropy. We can understand entropy as the rate at which a system generates information related to the studied problem. Because of that, many authors have proposed several entropy definitions for each different task, see e.g. [5, 8, 12, 13, 15, 17]. In the present paper, we are interested in estimating the state of a switched linear system with a prescribed exponential decay rate of $\alpha \geq 0$ for the estimation error. The entropy concept we use is called estimation entropy, and its description first appeared in [10] for generic autonomous nonlinear systems. We can, therefore, think about the estimation entropy as a rate at which the system generates uncertainty about the state. However, obtaining the value of the estimation entropy is only half of the story, because it does not tell us how to design the coding-estimator scheme to solve the original problem. The goal of the present paper is to address this issue. We show how to construct a coding-estimator scheme that operates with an average data-rate arbitrarily close to the estimation entropy for switched linear systems.

The research in entropy notions for switched systems has drawn the attention of several authors in recent years. Thus, a brief literature review might be helpful to explain the contributions of the present work and its context. The first paper to explicitly present an entropy notion for switched systems, related to the estimation entropy defined in [10], was [18]. Afterward, several distinct methods were developed to obtain bounds for the value of the entropy of switched linear systems, see, for instance, [3, 21–24]. Among these works, [21] provides an inequality that relates Lyapunov exponents with the estimation entropy, and those authors show that that expression holds with equality for a large class of switched linear systems called regular. It should be remarked that a similar relationship appears in several places in the dynamical systems literature, often under the name *Pesin entropy formula* [14, 16, 20], as well as in control and estimation theory on compact manifolds [9, 19]. Another relevant work for our discussion is [4]. There, the authors use entropy notions to describe an algorithm that stabilizes a switched linear system with an average data-rate arbitrarily close to the minimal. However, the algorithm presented in [4] requires us to know an a priori upper bound for the entropy, which might

*This work was supported by the NSF grant CMMI-1662708 and the AFOSR grant FA9550-17-1-0236.

HSCC '21, May 19–21, 2021, Nashville, TN, USA

ACM ISBN 978-1-4503-8339-4/21/05…$15.00
https://doi.org/10.1145/3447928.3456645

not be realistic if we want a causal algorithm, as discussed in the present paper.

In the context above, the current paper can be considered as extending the work in [21] by providing a constructive and causal algorithm that builds a state estimate for a switched linear system with a prescribed exponential decaying rate $\alpha \geq 0$ for the estimation error with an average data-rate as close as desired to the estimation entropy. Moreover, we advocate in favor of the role of regularity because it allows us to build a quantizer using only what is known up to a given time instant. Furthermore, the regularity assumption is fulfilled by several systems of practical interest, such as those modeled as Markov Jump Linear Systems, as shown in [21].

This paper has the following structure: In section 3, we motivate our study through an example where the current methods perform worse than our method presented here. Then, in section 1, we study the concept of estimation entropy, which will be related to our algorithm's average data-rate for a particular choice of the algorithm's parameters, giving it an upper bound. Also, that upper-bound is the exact value of the estimation entropy under the Lyapunov regularity assumption. Further, we study the concepts of Lyapunov exponents and Oseledets' filtration that will be useful when we discuss our quantization algorithm. In section 5, we present our algorithm in its most general framework. Then, by utilizing the Oseledets' filtration and Lyapunov exponents we show that we can operate at an average data-rate close to the estimation entropy when we make specific choices in our algorithm. Furthermore, we present how to make the algorithm reach the minimal average data-rate in a more realistic setting for practical applications. Following, in section 6, we present simulation results for the example-system of section 3. Finally, in section 7, we draw our conclusions and propose future works.

Notations: Unless otherwise stated, we denote by $||\cdot||$ the infinity-norm in a finite dimensional vector space. Let $\mathbb{R} = (-\infty, \infty)$, let $\mathbb{Z}_{\geq 0} = \{0, 1, \dots\}$ the nonegative integers, and let $\mathbb{N} = \{1, 2, \dots\}$ the set of natural numbers. For any set E, we denote by $\#E$ its cardinality. For subsets of $\mathbb{R}^d$ we denote $\mathrm{vol}(E)$ the volume of the set (its Lebesgue measure). Further, we denote by $\mathrm{diam}(E)$, where $E \subset \mathbb{R}^d$ the set's diameter according to the metric induced by the norm $||\cdot||$. We also denote by $\dim(V)$ the dimension of a linear vector space V. Also, for any $x > 0$, $\log x$ is the logarithm with base e and, for $b > 0$, $\log_b x$ is the logarithm with the base b. Additionally, we denote by $\langle v, w\rangle$, where $v \in \mathbb{R}^d$ and $w \in \mathbb{R}^d$, the usual canonical inner product of $\mathbb{R}^d$. Furthermore, we say that a basis $\{v_i\}_{i=1}^{d}$ for a finite dimensional vector space V is orthonormal if for every $i \in \{1, \dots, d\}$ and $j \in \{1, \dots, d\}$ we have that $\|v_i\| = 1$ and $\langle v_i, v_j\rangle = 0$ for $i \neq j$.

We denote by $\mathcal{M}(d, \mathbb{R})$ the set of all $d \times d$ matrices over the reals. We denote $\det(A)$ the determinant of the matrix A. Further, $I_d \in \mathcal{M}(d, \mathbb{R})$ is the identity matrix. Additionally, consider the parallelepiped defined by $\{\kappa_i v_i : \kappa_i \in [0, 1]\}$, where $\{v_i\}_{i=1}^{k} \subset \mathbb{R}^d$ is a linearly independent set of vectors. We denote the k-th volume of the parallelepiped by $\mathrm{vol}\,(v_1, \cdots, v_k)$ and its numerical value is given by $\sqrt{\det\left(V^\top V\right)}$, where V is the $d \times k$ matrix with columns v_i.[1]

[1]Notice that interchanging the order of the columns does not change the k-th volume.

2 PRELIMINARIES

Consider the following switched linear system model

$$\dot{x}(t) = A_{\sigma(t)} x(t), \tag{1}$$

where $x(t) \in \mathbb{R}^d$, $\sigma : \mathbb{R}_{\geq 0} \to \Sigma$ is a switching signal and Σ is a finite cardinality set, and $A_{\sigma(t)} \in \mathcal{M}(d, \mathbb{R})$. We denote by $\Phi(t, t_0)$ the *state-transition matrix* of (1), i.e. the solution of the ODE $\frac{d}{dt}\Phi(t, t_0) = A_{\sigma(t)}\Phi(t, t_0)$ with $\Phi(t_0, t_0) = I_d$ and t_0 being the initial time. Furthermore, we will make the assumption that σ is constant on intervals of the type $[t_i, t_{i+1})$ for $i \in \mathbb{Z}_{\geq 0}$, where $(t_i)_{i\in\mathbb{Z}_{\geq 0}}$ is a strictly increasing sequence of positive times such that $\limsup_{i\to\infty} t_i = \infty$. The elements of the sequence $(t_i)_{i\in\mathbb{Z}_{\geq 0}}$ are called *switching times*. We also need to define an increasing sequence of *sampling times* $(\tau_k)_{k\in\mathbb{Z}_{\geq 0}}$, with $\tau_k = kT_p$ for all $k \in \mathbb{Z}_{\geq 0}$ and some $T_p > 0$.

Then, we can rewrite the model described in equation (1) using its *exact discrete-time model*, defined by:

$$x_{k+1} = A_k x_k, \tag{2}$$

where $(x_k)_{k\in\mathbb{Z}_{\geq 0}}$ is the state at the sampling times τ_k, i.e. $x_k = x(\tau_k)$, and $A_k = \Phi(\tau_{k+1}, \tau_k)$. We are slightly abusing the notation by using A for both the continuous and discrete time matrices, but will make clear which of the models, (1) or (2), we are using in the text.

Consider the following definitions of coder-estimator scheme, see for instance [13, 15]. Let $\{\tau_k\}_{k\in\mathbb{Z}_{\geq 0}}$ be the aforedescribed sequence of sampling times. Also, let $\{C^n\}_{n\in\mathbb{Z}_{\geq 0}}$ be a sequence of alphabets with uniformly bounded cardinality, i.e. $\exists M > 0, \#C^i < M, \forall i \in \mathbb{Z}_{\geq 0}$. We call the elements q of a finite alphabet *symbols*. Furthermore, let $\{\gamma_n\}_{n\in\mathbb{Z}_{\geq 0}}$ be a sequence of functions such that $\gamma_n : \prod_{i=0}^{n-1} C^i \times \mathbb{R}^{d(n+1)} \to C^n$, where γ_n is called the *coder mapping at time n*. We can write the coder mapping in the following more explicit way[2]

$$\begin{aligned}
&\gamma_0 : x(\tau_0) \mapsto q_0,\\
&\gamma_n : (q_0, \dots, q_{n-1}, x(\tau_0), \dots, x(\tau_n)) \mapsto q_n,
\end{aligned}$$

where $q_n \in C^n$ for all $n \in \mathbb{Z}_{\geq 0}$.

The *average data-rate* of a coder-estimator scheme is defined as

$$b := \limsup_{j\to\infty} \frac{1}{t_j} \sum_{i=0}^{j} \log\left(\#C^i\right). \tag{3}$$

3 EXAMPLE

In this section, we motivate our work through a randomly switched system example. In this example, we show that the average data-rate for state estimation taking the switched system dynamics into account is lower than the one obtained by using the optimal quantizer for each mode separately whenever that mode is active.

Example 3.1. Let $B_1 = \begin{bmatrix} 0.9 & 0.03 \\ 0 & 1 \end{bmatrix}$ and $B_2 = \begin{bmatrix} 1.1 & 0.02 \\ 0 & 1 \end{bmatrix}$ be the modes of our discrete-time switched system. Notice that the mode B_2 is unstable. Therefore, applying the conventional quantization scheme [7] that reaches the minimum average data-rate for each

[2]Notice that, since $q_0 = \gamma_0(x(\tau_0))$, one could define $\tilde{\gamma}_1(x(\tau_0), x(\tau_1)) = \gamma_1(\gamma_0(x(\tau_0)), x(\tau_0), x(\tau_1))$. Then, one could define $\tilde{\gamma}_n(x(\tau_0), \dots, x(\tau_n))$ recursively in a similar way. Making the explicit dependence of the quantized value on the previous symbols is a matter of keeping the argumentation clear.

mode separately will use a positive average data-rate. Nonetheless, we will show that, with probability 1, if our switch comes from the Markov chain defined by the matrix of transition probabilities $P = \begin{bmatrix} 0.1 & 0.9 \\ 0.9 & 0.1 \end{bmatrix}$, where P_{ij} is the transition probability from mode i to mode j, then there exists an algorithm that reconstructs the state using an average data-rate equal as close to the estimation entropy as desired with probability 1 in the aforedescribed situation.

In this paper, we will present a quantization scheme that operates at an average data-rate equal to the estimation entropy for a large class of switching signals called regular switchings. It so happens that, with probability 1, the switching signals generated by Markov Jump Linear Systems, like the one in this example, are in this class.

4 ESTIMATION ENTROPY

In this section, we introduce Lyapunov exponents, Lyapunov regularity, estimation entropy, and related concepts. Also, we state a theorem that gives an upper bound for the estimation entropy of discrete-time switched systems using the Lyapunov exponents. Furthermore, the theorem states that the upper bound is the actual value of the estimation entropy when we assume Lyapunov regularity. The definitions presented here were adapted from the references [11], Chapter 2 of [2], and Chapter 3 of [1].

Throughout this document, given a sequence of invertible matrices $(A_n)_{n\in\mathbb{N}} \subset \mathcal{M}(d, \mathbb{R})$, we denote the *discrete-time state-transition matrix* of the system (2) by

$$\Phi_n := A_n \cdots A_1. \tag{4}$$

We assume that $K \subset \mathbb{R}^d$, the set of possible initial conditions, is a compact set with nonempty interior. Further, the solution of (2) at time step n with initial condition $x \in \mathbb{R}^d$ is given by $\xi(x, n) = \Phi_n x$, where the matrix sequence is given by the matrices on the right-hand side of (2).

For the next definition, pick an $\alpha \geq 0$, and let $T \in Z_{\geq 0}$ be the time horizon.

Definition 4.1. For every $\epsilon > 0$, we call a finite set of functions $\hat{X} = \{\hat{x}_1(\cdot), \ldots, \hat{x}_N(\cdot)\}$, from $\{0, \ldots, T-1\}$ to $\mathbb{R}^d$, a (T, ϵ, α, K)-*approximating set* if for every initial condition $x \in K$, there exists $\hat{x}_i \in \hat{X}$ such that $||\xi(x, n) - \hat{x}_i(n)|| < \epsilon e^{-\alpha n}, \quad \forall n \in \{0, \ldots, T-1\}$.

Let $s_{\text{est}}(T, \epsilon, \alpha, K)$ be the minimum cardinality of a (T, ϵ, α, K)-approximating set. We define the estimation entropy as

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

Definition 4.2. A *Lyapunov index* is a function $\lambda : \mathbb{R}^d \to \mathbb{R} \cup \{-\infty\}$ with the following properties:

- $\lambda(\kappa v) = \lambda(v)$, for every real $\kappa \neq 0$
- $\lambda(v + w) = \max\{\lambda(v), \lambda(w)\}$
- $\lambda(0) = -\infty$

A Lyapunov index $\lambda(\cdot)$ can take at most d distinct real values, see e.g. [2]. (Note that $-\infty$, which is the value of $\lambda(0)$, is not a real value.)

Definition 4.3. The *Lyapunov exponent* associated with a sequence of matrices $(A_n)_{n\in\mathbb{N}}$ is the following Lyapunov index[3]:

$$\lambda(v) := \limsup_{n\to\infty} \frac{1}{n} \log(||\Phi_n v||),$$

for $v \in \mathbb{R}^d \setminus \{0\}$. Also, we define $\lambda(0) := -\infty$.

Note that the Lyapunov exponent, $\lambda(\cdot)$, is a particular Lyapunov index, see e.g. [2]. Therefore, it can attain at most d distinct values. We denote these values by χ_i, for $i = 1, \ldots, q$, where $q \leq d$, and we index them according to the increasing order for real numbers, i.e. $\chi_1 < \cdots < \chi_q$. We call χ_i, $i = 1, \ldots, q$ the *Lyapunov exponent values.*

Definition 4.4. A *filtration* (or *flag*) on $\mathbb{R}^d$ is a family of vector subspaces $\mathbb{V} = (E_i)_{i=0}^q$, with $q \leq d$, such that $\{0\} = E_0 \subsetneq E_1 \subsetneq \cdots \subsetneq E_q = \mathbb{R}^d$. Further, we call $\mathcal{V} = \{v_i\}_{i=1}^d$ a *normal basis* of the filtration $\mathbb{V}$ if it is a basis for $\mathbb{R}^d$, and for every $j \geq 1$, the subset of $\mathbb{V}$ given by $\{v_i\}_{i=1}^{\dim(E_j)}$ is a basis for E_j.

A special type of filtration that will be used in the text, and in our quantization algorithm in section 5, is the Oseledets' filtration, which we define next.

Definition 4.5. A filtration $\mathcal{V}_\lambda$ associated with the sequence of invertible matrices $(A_n)_{n\in\mathbb{N}}$ such that $E_i = \left\{v \in \mathbb{R}^d : \lambda(v) \leq \chi_i\right\}$, where $\lambda(\cdot)$ is the Lyapunov exponent for the sequence, and χ_i are the Lyapunov exponent values of the sequence previously defined, is called an *Oseledets' filtration.* Also, the subspaces $E_i \in \mathcal{V}_\lambda$ are called *Oseledets' subspaces.* In addition, the following $\dim(E_i) - \dim(E_{i-1})$ is called the *multiplicity of the Lyapunov exponent value* χ_i. If[4] $\dim(E_i) - \dim(E_{i-1}) = 1$ for every $i \in \{1, \ldots, q\}$, we say that the Lyapunov exponents are *simple.* Finally, define $\Lambda = \left\{\lambda_j\right\}_{j=1}^d$ as an ordered list with repetition where for every $j = 1, \ldots, d$, there exists some $i \in \{1, \ldots, q\}$ such that $\lambda_j = \chi_i$, and for every $i = 1, \ldots, q$, χ_i appears $\dim(E_i) - \dim(E_{i-1})$ times in Λ. The order in Λ can be any total order relation in the set Λ chosen among those for which $\lambda_1 \leq \cdots \leq \lambda_d$. We call the elements $\lambda_i \in \Lambda$ the *Lyapunov exponents with multiplicity* of $(A_n)_{n\in\mathbb{N}}$.

It is important to remark that the Oseledets' filtration depends on the entire sequence $(A_n)_{n\in\mathbb{N}}$. To see that, consider the following example.

Example 4.6. Let $A = \begin{bmatrix} 2 & 0 \\ 0 & 4 \end{bmatrix}$ and $B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ and notice that the sequence $A'_n = A$ for all $n \in \mathbb{N}$, and the sequence $A_n = A$ for $n \in \mathbb{N} \setminus \{N\}$ and $A_N = B$ for some $N \in \mathbb{N}$, have the same Lyapunov exponents, but different Oseledets' filtrations. For the Oseledets's filtration of the first sequence is $E_1 = \text{span}\left\{\begin{bmatrix} 1 & 0 \end{bmatrix}^\top\right\} \subsetneq E_2 = \mathbb{R}^2$ and the filtration of the second is $E_1 = \text{span}\left\{\begin{bmatrix} 0 & 1 \end{bmatrix}^\top\right\} \subsetneq E_2 = \mathbb{R}^2$.

Definition 4.7. A sequence $(A_n)_{n\in\mathbb{N}}$ is called *tempered* if

$$\lim_{n\to\infty} \frac{1}{n} \log ||A_n|| = 0$$

[3]Note that the function does not change if we change the norm.

[4]Equivalently, we could say that $d = q$.

Notice that, if a sequence $(A_n)_{n\in\mathbb{N}}$ belongs to a compact set, then it is tempered. A particular case is the one in which $(A_n)_{n\in\mathbb{N}}$ has finitely many values. It is worth mentioning that temperedness does not imply that the growth rate of Φ_n is sub-exponential. To see why, take $A_n = n$, which is tempered because $\lim_{n\to\infty} \frac{\log(n)}{n} = 0$, and notice that $\Phi_n = n!$, which grows faster than any exponential.

Example 4.8 (Example 3.1 revisited.). This is a good moment for us to revisit our Example 3.1. Denote by $a_{ij}(n)$ the element in the i-th row and j-th column of the matrix A_n, and denote, analogously, by $\phi_{ij}(n)$ the elements of Φ_n. Further, denote by $m_i(n) = \sum_{k=1}^{n} \mathbb{I}_{(A_n)_{n\in\mathbb{N}}:A_k=B_i}((A_n)_{n\in\mathbb{N}})$, where $\mathbb{I}_A(x) = 1$ if $x \in A$ and $\mathbb{I}_A(x) = 0$, otherwise. We should think of $m_i(n)$ as how many time instants mode i was active until time n. Note that, $\phi_{11}(n) = 0.9^{m_1(n)}1.1^{m_2(n)}$, $\phi_{22}(n) = 1$, and $\phi_{12}(n) = a_{11}(n)\phi_{12}(n-1) + a_{12}(n)$ for $n \geq 1$ with initial conditions $\phi_{ii} = 1$ and $\phi_{ij} = 0$ if $i \neq j$. Now, let $\{e_1, e_2\}$ be the canonical basis for $\mathbb{R}^2$. Then, the Lyapunov exponents of the sequence $(A_n)_{n\in\mathbb{N}}$ are given by

$$\begin{aligned}\lambda(e_1) &= \limsup_{n\to\infty} \frac{1}{n} \log(\|\Phi_n e_1\|) = \limsup_{n\to\infty} \frac{1}{n} \log(0.9^{m_1(n)}1.1^{m_2(n)}) \\ &= \limsup_{n\to\infty} \frac{m_1(n)}{n} \log(0.9) + \frac{m_2(n)}{n} \log(1.1).\end{aligned}$$

Recall that the fraction of time that a Markov chain stays on mode i is given, with probability 1, by the probabilities π_i obtained by solving $\pi = \pi P$ and $\sum_{i=1}^{2} \pi_i = 1$, where $(\pi_1, \pi_2) = \pi$. For this example, we get that $\pi_1 = \pi_2 = 1/2$. Thus, with probability 1, a specific realization will have the fractions $\frac{m_i(n)}{n}$ converging to the probabilities π_i, where $i \in \{1, 2\}$. Hence, $\lambda(e_1) = \frac{1}{2}\log(0.99) < 0$. Finally, we notice that $\phi_{12}(n) = a_{11}(n)\phi_{12}(n-1) + a_{12}(n)$ is a scalar linear time-varying system with an input $a_{12}(n)$. Therefore, if $\prod_{j=1}^{n} a_{11}(j) < 1$ and $a_{12}(n)$ are bounded, we prove that $\phi_{12}(n)$ is bounded. Indeed, $a_{12}(n)$ is always bounded and the product $\prod_{j=1}^{n} a_{11}(j) = 0.9^{m_1(n)}1.1^{m_2(n)}$ can be upper bounded 1. To see that, take the logarithm of the product and divide it by n so that we get $\frac{1}{n}\log\left(\prod_{j=1}^{n} a_{11}(j)\right) = \frac{m_1(n)}{n}\log(0.9) + \frac{m_2(n)}{n}\log(1.1) < 0$. From which we conclude that $\prod_{j=1}^{n} a_{11}(j) < 1$ and that ϕ_{12} is bounded with probability 1. Now, we can calculate $\lambda(e_2)$ by noticing that $\|\Phi_n e_2\| = \max\{\phi_{12}(n), 1\}$ is bounded, hence $\lambda(e_2) = 0$ with probability 1.

Furthermore, we notice that the filtration $E_1 = \text{span}\{e_1\} \subsetneq E_2 = \mathbb{R}^2$ is the Oseledets' filtration. Moreover, we see that $\{e_1, e_2\}$ form a normal basis for this filtration.

We remark that, although the sequence $(A_n)_{n\in\mathbb{N}}$ comes from a stochastic process, we calculated the values of the Lyapunov exponents for a generic realization. Thus, we always choose a specific realization, as in the deterministic case. Nonetheless, we use the Markov chain's properties to show that our result holds for almost all realizations of the random process.

Definition 4.9. A sequence $(A_n)_{n\in\mathbb{N}}$ is called (Lyapunov) *regular* if

$$\lim_{n\to\infty} \frac{1}{n} \log\left(|\det(\Phi_n)|\right) = \sum_{i=1}^{d} \lambda_i.$$

We call a system given by equation (2) *regular*, if its associated matrix sequence is regular.

The following examples, 4.10 and 4.11, should help illustrate the concept of regularity.

Example 4.10. Let $\rho > 1$. Also, let $B_1 = \begin{bmatrix} \rho & 0 \\ 0 & \rho^{-1} \end{bmatrix}$ and $B_2 = \begin{bmatrix} \rho^{-1} & 0 \\ 0 & \rho \end{bmatrix}$. Consider the sequence $A_n = B_1$ if $n \in \{2^i, \cdots, 2^{i+1} - 1\}$, for i odd, and $A_n = B_2$ otherwise. Note that $\det(|\Phi_n|) = 1$ for all possible sequences $(A_n)_{n\in\mathbb{N}}$. Denote by $\{e_1, e_2\}$ the canonical basis. Further, consider the subsequence with indices $n_k = 2^k$ for $k \in \mathbb{N}$. Then, one can show by induction that $\left\|\Phi_{n_k}(e_1)\right\| = \rho^{-\sum_{i=1}^{k}(-2)^{i-1} + (-1)^k}$. Thus, $\frac{\log\left(\left\|\Phi_{n_k}(e_1)\right\|\right)}{2^k} = \sum_{\ell=1}^{k}\left((-1)^{\ell+1}(2)^{-\ell} + (-1)^k 2^{-k}\right)\log(\rho)$, after the change of varibles $\ell = -i + k + 1$. Now, looking at the subsequence with indices $n_k = 2^k$ with k even, we show that this subsequence has a positive limit because:

$$\lim_{k\to\infty} \sum_{\ell=1}^{k} (-1)^{\ell+1}(2)^{-\ell}\log(\rho) + (-1)^k 2^{-k}\log(\rho) = \frac{1}{3}\log(\rho) > 0.$$

Hence, by the fact that the limit superior is larger than all sublimits, we conclude that $\lambda(e_1) > 0$, because it is the limit superior. We can show the analogous result $\lambda(e_2) > 0$ by considering the odd values of k. Therefore, the original sequence cannot be regular.

Example 4.11. Let B_1 and B_2 be as in Example 4.10. Consider the sequence $A_n = B_1$ whenever n is divisible by 4, and $A_n = B_2$ otherwise. Also let $\{e_1, e_2\}$ be the canonical basis for $\mathbb{R}^2$. Then one can check that $\lambda(e_1) = -\frac{1}{2}\log\rho$ and $\lambda(e_2) = \frac{1}{2}\log\rho$. Therefore, the sequence is regular and $\{e_1, e_2\}$ is a basis for the Oseledets' filtration.

In Example 4.10, the limit superior in Definition 4.3 of Lyapunov exponent cannot be replaced by a limit, but in Example 4.11, where the matrix sequence is regular, it can. This fact is not a coincidence, as shown by the second bullet of Lemma 4.12, which implies that the limit exists when the sequence is regular.

The following lemma was extracted from Chapters 3 and 7 of [2] and it presents equivalent characterizations for regularity that will be used in this article.

LEMMA 4.12. *Given a tempered sequence $(A_n)_{n\in\mathbb{N}}$ of invertible matrices, let $\{v_1, \cdots, v_d\}$ be any normal basis for the Oseledets' filtration of the sequence $(A_n)_{n\in\mathbb{N}}$, and let $\mathcal{I} \subset \{1, \cdots, d\}$ be any set of indices. Further, let λ_i be the Lyapunov exponents with multiplicity of the sequence $(A_n)_{n\in\mathbb{N}}$. Then, the following conditions are equivalent*

- $\lim_{n\to\infty} \frac{1}{n}\log\left(|\det(\Phi_n)|\right) = \sum_{i=1}^{d} \lambda_i$;
- $\lim_{n\to\infty} \frac{1}{n}\log\left(\text{vol}\left(\{\Phi_n v_i : i \in \mathcal{I}\}\right)\right) = \sum_{i\in\mathcal{I}} \lambda_i$.
- *The matrix* $\lim_{n\to\infty}\left(\Phi_n^\top \Phi_n\right)^{\frac{1}{2n}}$ *exists.*

Now, we state the main Theorem of this section.

THEOREM 4.13. *Let $\alpha \geq 0$. Let $(A_n)_{n\in\mathbb{N}}$ be a tempered sequence of invertible matrices. Let $K \subset \mathbb{R}^d$ be a set of possible initial conditions with a nonempty interior. Denote by λ_i, with $i = 1, \cdots, d$, the Lyapunov exponents with multiplicity of $(A_n)_{n\in\mathbb{N}}$. Then, the estimation entropy of the discrete switched system* (2) *satisfies:*

$$h_{est}(\alpha, K) \leq \sum_{i=1}^{d} \max\{0, \lambda_i + \alpha\}, \tag{5}$$

with equality if the system is regular.

A proof of this theorem can be found in [21].

Remark 4.1. *It seems logical to draw a parallel between the bound presented above and the bounds presented in [24]. The bounds obtained in that paper rely on the individual modes and their activation times. On the other hand, the result in Theorem 4.13 uses information about the entire switching signal given by the Lyapunov exponents. That is why the bounds in Theorem 4.13, although much harder to compute, are generally tighter than the ones presented in [24].*

Example 4.14. [Example 3.1 revisited] Now, we can analyse Example 3.1 again. From our calculations in section 2, we saw that the Lyapunov exponents of our system are $\lambda(e_1) = \frac{1}{2}\log(0.99) < 0$ and $\lambda(e_2) = 0$ with probability 1. From this, we conclude that the system's estimation entropy satisfies the inequality $h_{\text{est}}(\alpha, K) \le \max\left\{\frac{1}{2}\log(0.99) + \alpha, 0\right\} + \max\{\alpha, 0\}$ with probability 1.

5 QUANTIZATION ALGORITHM

In this section, we describe the quantization algorithm. This algorithm's goal is to estimate the state of system (2), with a desired exponential decay rate for the estimation error, using quantized measurements. The algorithm works by giving an over-approximation to the reachable set that depends on a few parameters such as the set of possible initial conditions, the switching signal, and the desired exponential decay for the estimation error. Also, we need to provide a family of bases $\mathcal{V}_j = \left\{v_1^j, \ldots, v_d^j\right\}$, $j \in \mathbb{Z}_{\ge 0}$ for $\mathbb{R}^d$. Using this family, the proposed algorithm generates an over-approximation for the reachable set. Then, we show that by using a proper choice of family $(\mathcal{V}_j)_{j\in\mathbb{Z}_{\ge 0}}$ the algorithm's average data-rate can be made as close to the estimation entropy of our system as desired. Finally, we present a way of generating a family $(\mathcal{V}_j)_{j\in\mathbb{Z}_{\ge 0}}$ that makes the algorithm achieve an average data-rate arbitrarily close to the estimation entropy online, assuming that the switching signal is known. Also, throughout this section, we will let $T_p > 0$ be a sampling time and the sequence $(A_n)_{n\in\mathbb{N}}$ corresponds to the exact discrete-time model of some continuous-time model described by equation (1), i.e. $A_n = \Phi(T_p n, T_p(n-1))$.

5.1 The Algorithm

In this Subsection, we describe a quantization scheme for switched linear systems under the assumption that we know $\sigma(t)$ for all values of $t \in \mathbb{R}_{\ge 0}$. Under the hypothesis that model (2) holds, the previous assumption becomes the hypothesis of knowing the sequence $(A_n)_{n\in\mathbb{N}}$. We also assume that we are given an arbitrary family of orthonormal[5] bases $\mathcal{V}^j$ for $\mathbb{R}^d$. After our scheme's description, we show that, under a particular choice of the family $\mathcal{V}^j$, our algorithm can operate at an average data-rate arbitrarily close to the upper bound for the estimation entropy obtained in Theorem 4.13, i.e. $\sum_{i=1}^d \{\lambda_i + \alpha, 0\}$. Moreover, for the case where our system is known to be regular, again because of Theorem 4.13, the algorithm can operate at an average data-rate arbitrarily close to the estimation entropy.

Before we provide an informal description of the algorithm, we need to define some concepts. First, we define ℓ to be a positive integer that we call *block length*. Second, let j be a positive integer that indexes our algorithm's iteration. Also, we need to mention that our informal description is only valid for time t greater than zero since the initial case is slightly different because of how we initialize the algorithm. Nonetheless, the logic is essentially the same. In words, the algorithm does the following: Let the initial state x be inside the region $\bar{B}^{j-1}$, a parallelepiped in $\mathbb{R}^d$. Given a basis $\left\{v_i^j\right\}_{i=1}^d$ from the family $\mathcal{V}^j$, build a new parallelepiped $\tilde{B}^j$ with sides parallel to the v_i^j's that contains $\bar{B}^{j-1}$. Now, we flow $\tilde{B}^j$ forward using $\Phi_{j\ell+1}$ and denote it by B^j. More preceisely, we define $B^j = \Phi_{j\ell+1}(\tilde{B}^j)$. Note that, since x belongs to $\bar{B}^{j-1}$ and $\bar{B}^{j-1} \subset \tilde{B}^j$, we have that the state at the current time $j\ell+1$, i.e. $\xi(x, j\ell+1)$, belongs to B^j. Inside the set B^j, we have quantization subregions, each corresponding to a distinct quantization symbol. We denote by q^j the quantization symbol corresponding to the quantization subregion that contains $\xi(x, j\ell+1)$. Next, we flow the previous quantization subregion, that corresponds to the symbol q^j, backwards by $\Phi_{j\ell+1}$ and define the result to be $\bar{B}^j$. Finally, we repeat the procedure.

We emphasize that the bases $\left\{v_i^j\right\}_{i=1}^d$ with $j \in \mathbb{Z}_{\ge 0}$ are, in principle, arbitrary. By that, we mean that our quantization algorithm works for any choice of the family of bases at the possible cost of working at a higher average data-rate. However, we show in Corollary 5.2 and Theorem 5.3 how to choose those bases so that the average data-rate will approach the estimation entropy. Further, it is worth emphasizing that we build our estimates using measurements that happen only at time instants $t = j\ell+1$ with $j \in \mathbb{Z}_{\ge 0}$ and at the initial time $t = 0$. The idea of using the block length was borrowed from the block coding approach[6], and it allows the average data-rate to approach the estimation entropy arbitrarily close in some specific cases.

In what follows, we assume that $\mathbb{R}^d$ is endowed with the canonical inner product $\langle \cdot, \cdot \rangle$.

Quantizer algorithm

Initialization: Let K be the set of possible initial conditions, $x \in K$ be the true initial condition, $\epsilon > 0$ a prescribed precision, $T_p > 0$ the sampling time, and $\ell \in \mathbb{N}$ be the block length. Also, consider the sequence $(A_n)_{n\in\mathbb{N}}$, where[7] $A_n = \Phi(T_p n, T_p(n-1))$ and $\Phi_n = A_n \ldots A_1$. Further, let $\mathcal{V}_j = \left\{v_1^j, \ldots, v_d^j\right\}$, $j \in \mathbb{Z}_{\ge 0}$ be a family of orthonormal bases for $\mathbb{R}^d$. We define $\Gamma_i^0 = 1$ for all $i \in \{1, \ldots, d\}$. If the system is known to be regular, set $\Gamma_i^j := \max_{k\in\{0,\ldots,\ell-1\}} \left\|\Phi_{j\ell-k} v_i^j\right\|$, otherwise

$$\Gamma_i^j := \max\left\{\max_{k\in\{0,\ldots,\ell-1\}} \left\|\Phi_{j\ell-k} v_i^j\right\|, e^{T_p(\lambda_i+\delta)j\ell}, e^{T_p(\lambda_i+\delta)((j-1)\ell+1)}\right\}$$

for a prescribed $\delta > 0$ and[8] $\lambda_i := \limsup_{j\to\infty} \frac{1}{j}\log\left(||\Phi_j v_i^j||\right)$. Also, let $\alpha \ge 0$ be the prescribed exponential decay rate for the estimation error.

Step 0:

[5] We omit the orthonormal from this point onward.

[6] See e.g. Chapter 5 of [6].

[7] Note that $(A_n)_{n\in\mathbb{N}} \subset \mathcal{M}(d, \mathbb{R})$ might be an infinite set in general.

[8] Notice that these λ_i's are not the same as the Lyapunov exponents with multiplicity since the v_i^j's are not a normal basis for the Oseledets' filtration in principle.

In this step, we define an estimate $\hat{x}(0)$ for $\xi(x,0) = x$.

- Define $B^0 = \left\{\sum_{i=1}^d \gamma_i v_i^0 : \underline{\kappa}_i^0 \le \gamma_i < \overline{\kappa}_i^0\right\}$, where $\underline{\kappa}_i^0$ and $\overline{\kappa}_i^0$ are such that B^0 is the smallest set of such type that contains the initial set K.
- Write $\xi(x,0) = \sum_{i=1}^d \beta_i^0 v_i^0$. Then, the symbol related to the quantized value of $\xi(x,0)$ is given by $q^0 = \left(q_1^0, \dots, q_d^0\right)$, constructed as follows. Define $C_i^0 := \left\{1, \dots, \left\lceil d \frac{\overline{\kappa}_i^0 - \underline{\kappa}_i^0}{\epsilon} \right\rceil\right\}$. We define q_i^0, for every $i \in \{1, \dots, d\}$, as the $k \in C_i^0$ such that
$$\beta_i^0 \in \left[\underline{\kappa}_i^0 + \frac{\epsilon}{d}(k-1), \underline{\kappa}_i^0 + \frac{\epsilon}{d}k\right)$$
holds true.
- Denote $\hat{\beta}_i^0 := \underline{\kappa}_i^0 + \frac{\epsilon}{d}\left(q_i^0 - 1/2\right)$. Our estimate for the state at the moment $t = 0$ is
$$\hat{x}(0) := \sum_{i=1}^d \left(\underline{\kappa}_i^0 + \frac{\epsilon}{d}\left(q_i^0 - 1/2\right)\right) v_i^0.$$

We could describe this step 0 in words as follows. B^0 is divided into cubic boxes with sides of length ϵ/d; q_i^0 encodes the position of the box in the i-th dimension that contains x; and $\hat{x}(0)$ is the center of this box.

Step 1:

In this step, we define an estimate $\hat{x}(t)$ for $\xi(x,t)$ with $1 \le t \le \ell$. Notice that we generated a box
$$\bar{B}^0 := \left\{\sum_{k=1}^d \mu_k v_k^0 : \underline{\kappa}_k^0 + \frac{\epsilon}{d}(q_k^0 - 1) \le \mu_k < \underline{\kappa}_k^0 + \frac{\epsilon}{d} q_k^0\right\}$$
at the end of Step 0 and that $x \in \bar{B}^0$. Now, in this step, we generate the smallest box aligned with the new basis $\{v_i^1\}_{i=1}^d$ that contains $\bar{B}^0$. This box takes the form
$$\tilde{B}^1 := \left\{\sum_{i=1}^d \gamma_i v_i^1 : \underline{\kappa}_i^1 \le \gamma_i < \overline{\kappa}_i^1\right\}.$$
To compute the bounds $\underline{\kappa}_i^1$ and $\overline{\kappa}_i^1$, let $y = \sum_{k=1}^d \mu_k v_k^0$ be an arbitrary point in $\bar{B}^0$. Thus, its coordinate relative to each v_i^1 is
$$\gamma_i = \langle \sum_{k=1}^d \mu_k v_k^0, v_i^1 \rangle = \sum_{k=1}^d \mu_k \langle v_k^0, v_i^1 \rangle.$$
Hence, to find the smallest such box, we need to take
$$\underline{\kappa}_i^1 := \min\Big\{ \sum_{k=1}^d \mu_k \langle v_k^0, v_i^1 \rangle : \underline{\kappa}_k^0 + \frac{\epsilon}{d}\left(q_k^0 - 1\right) \le \mu_k \le \underline{\kappa}_k^0 + \frac{\epsilon}{d} q_k^0, \quad k = 1, \dots, d\Big\},$$
for every $i \in \{1, \dots, d\}$.Notice that this is a linear programming problem. Therefore, the solution will occur at the boundary. Moreover, this set of inequalities forms a box, and we only need to check its vertices to find the optimal value. The upper bounds, $\overline{\kappa}_i^1$, are defined similarly but with max instead of min. Finally, we define the box
$$B^1 := \left\{\sum_{i=1}^d \gamma_i \Phi_1 v_i^1 : \underline{\kappa}_i^1 \le \gamma_i < \overline{\kappa}_i^1\right\}$$
by flowing the box $\tilde{B}^1$ forward by Φ_1. We can write the procedure of this step in the following itemized way.

- Define $B^1 := \left\{\sum_{i=1}^d \gamma_i \Phi_1 v_i^1 : \underline{\kappa}_i^1 \le \gamma_i < \overline{\kappa}_i^1\right\}$, where $\underline{\kappa}_i^1$ is obtained as described above, and $\overline{\kappa}_i^1$ is obtained in an analogous fashion by changing min by max.
- Write $\xi(x,1) = \sum_{i=1}^d \beta_i^1 \Phi_1 v_i^1$. Then, the symbol related to the quantized value of $\xi(x,1)$ is given by $q^1 = \left(q_1^1, \dots, q_d^1\right)$. Define $C_i^1 := \left\{1, \dots, \left\lceil d \Gamma_i^1 e^{T_p \alpha \ell} \frac{\overline{\kappa}_i^1 - \underline{\kappa}_i^1}{\epsilon} \right\rceil\right\}$. We define q_i^1, for every $i \in \{1, \dots, d\}$, as the $k \in C_i^1$ such that
$$\beta_i^1 \in \left[\underline{\kappa}_i^1 + \frac{\epsilon}{d}\frac{e^{-T_p \alpha \ell}}{\Gamma_i^1}(k-1), \underline{\kappa}_i^1 + \frac{\epsilon}{d}\frac{e^{-T_p \alpha \ell}}{\Gamma_i^1}k\right)$$
holds true.
- Denote by $\hat{\beta}_i^1 = \underline{\kappa}_i^1 + \frac{\epsilon}{d}\frac{e^{-T_p \alpha \ell}}{\Gamma_i^1}(q_i^1 - 1/2)$. Our estimate for the state at the moments $1 \le t \le \ell$ is
$$\hat{x}(t) := \sum_{i=1}^d \hat{\beta}_i^1 \Phi_t v_i^1.$$

Step j+1:

In this step, we define an estimate $\hat{x}(t)$ for $\xi(x,t)$ with $j\ell+1 \le t \le (j+1)\ell$. Notice that we generated a box
$$\bar{B}^j := \Big\{ \sum_{k=1}^d \mu_k v_k^j : \underline{\kappa}_k^j + \frac{\epsilon}{d}\frac{e^{-T_p \alpha j \ell}}{\Gamma_k^j}(q_k^j - 1) \le \mu_k < \underline{\kappa}_k^j + \frac{\epsilon}{d}\frac{e^{-T_p \alpha j \ell}}{\Gamma_k^j} q_k^j \Big\}$$
at the end of Step j and that $x \in \bar{B}^j$. Now, in this step, we generate the smallest box aligned with the new basis $\left\{v_i^{j+1}\right\}_{i=1}^d$ that contains $\bar{B}^j$. We define this smallest box as
$$\tilde{B}^{j+1} = \left\{\sum_{i=1}^d \gamma_i v_i^{j+1} : \underline{\kappa}_i^{j+1} \le \gamma_i < \overline{\kappa}_i^{j+1}\right\},$$
and obtain $\underline{\kappa}_i^{j+1}$ and $\overline{\kappa}_i^{j+1}$ in an analogous manner as we obtained $\underline{\kappa}_i^1$ and $\overline{\kappa}_i^1$ in step 1. Observe that the box $\tilde{B}^{j+1}$ contains the initial state x by construction. Finally, we define the box B^{j+1} as the box obtained after flowing $\tilde{B}^{j+1}$ forward by $\Phi_{j\ell+1}$. We describe the procedure in the following itemized way.

- Define
$$B^{j+1} = \left\{\sum_{i=1}^d \gamma_i \Phi_{j\ell+1} v_i^{j+1} : \underline{\kappa}_i^{j+1} \le \gamma_i < \overline{\kappa}_i^{j+1}\right\},$$
where
$$\underline{\kappa}_i^{j+1} := \min\left\{\sum_{k=1}^d \mu_k \langle v_k^j, v_i^{j+1} \rangle : \underline{\kappa}_k^j + \frac{\epsilon}{d}\frac{e^{-T_p \alpha j \ell}}{\Gamma_k^j}\left(q_k^j - 1\right) \le \mu_k \le \underline{\kappa}_k^j + \frac{\epsilon}{d}\frac{e^{-T_p \alpha j \ell}}{\Gamma_k^j} q_k^j, k = 1, \dots, d\right\},$$

$\overline{\kappa}_i^{j+1}$ is obtained in an analogous fashion by changing min by max.

- Write $\xi(x, j\ell+1) = \sum_{i=1}^d \beta_i^{j+1} \Phi_{j\ell+1} v_i^{j+1}$. Then, the symbol related to the quantized value of $\xi(x, j\ell+1)$ is given by $q^{j+1} = \left(q_1^{j+1}, \ldots, q_d^{j+1}\right)$. Let
$$C_i^{j+1} = \left\{1, \ldots, \left\lceil d e^{T_p \alpha (j+1)\ell} \Gamma_i^{j+1} \frac{\overline{\kappa}_i^{j+1} - \underline{\kappa}_i^{j+1}}{\epsilon} \right\rceil \right\}.$$
We define q_i^{j+1} as the $k \in C_i^{j+1}$ such that
$$\beta_i^{j+1} \in \left[\underline{\kappa}_i^{j+1} + \frac{\epsilon}{d} \frac{e^{-T_p\alpha(j+1)\ell}}{\Gamma_i^{j+1}}(k-1), \underline{\kappa}_i^{j+1} + \frac{\epsilon}{d} \frac{e^{-T_p\alpha(j+1)\ell}}{\Gamma_i^{j+1}} k\right)$$
holds true.
- Denote by $\hat{\beta}_i^{j+1} = \underline{\kappa}_i^{j+1} + \frac{\epsilon}{d} \frac{e^{-T_p\alpha(j+1)\ell}}{\Gamma_i^{j+1}}(q_i^{j+1} - 1/2)$. Then, our state estimate for the time instants $j\ell+1 \le t \le (j+1)\ell$ is $\hat{x}(t) := \sum_{i=1}^d \hat{\beta}_i^{j+1} \Phi_t v_i^{j+1}$.

The following Theorem 5.1 shows that our algorithm from section 5.1 generates a coding scheme that allows us to reconstruct a state estimate with an exponentially decaying error, and gives an upper bound on the average data-rate that the algorithm uses.

Theorem 5.1. *Let $(A_n)_{n\in\mathbb{N}}$ be a sequence of matrices that comes from the exact discretization of the system* (1) *with sampling time $T_p > 0$. Then, the algorithm from section 5.1 gives a sequence of estimates $(\hat{x}(t))_{t\in\mathbb{Z}_{\ge 0}}$ such that $||\hat{x}(t) - \xi(x,t)|| \le \frac{\epsilon}{2} e^{-T_p \alpha t}$. Further, the average data-rate of the agorithm from section 5.1 is given by $b = \limsup_{j\to\infty} \frac{1}{T_p t \ell} \sum_{j=0}^t \log\left(\#C^j\right)$, with $C^j := \prod_{i=1}^d C_i^j$ and $\#C^j := \prod_{i=1}^d \#C_i^j$, where $\#C_i^{j+1} \le \left\lceil e^{T_p\alpha\ell} \frac{\Gamma_i^{j+1}}{\Gamma_i^j} \sum_{k=1}^d \left|\langle v_k^j, v_i^{j+1}\rangle\right|\right\rceil$ for $j \in \mathbb{Z}_{\ge 0}$ and $\#C_i^0 \le \left\lceil d \frac{diam(B^0)}{\epsilon}\right\rceil$.*

Proof. **Step 0**:

Recall that $\left|\hat{\beta}_i^0 - \beta_i^0\right| \le \epsilon/2d$ by construction. Then,
$$||\hat{x}(0) - \xi(x,0)|| = \left\|\sum_{i=1}^d \left(\hat{\beta}_i^0 - \beta_i^0\right) v_i^0\right\| \le \frac{\epsilon}{2}$$
and $\#C_i^0 := \left\lceil d \frac{\overline{\kappa}_i^0 - \underline{\kappa}_i^0}{\epsilon}\right\rceil \le \left\lceil d \frac{\mathrm{diam}(B^0)}{\epsilon}\right\rceil$. Finally, notice that $x \in \bar{B}^0$.

Step 1:

We need to show that
$$\Phi_1\left(\bar{B}^0\right) = \left\{\sum_{i=1}^d \gamma_i \Phi_1 v_i^0 : \underline{\kappa}_i^0 + \frac{\epsilon}{d}(q_i^0 - 1) \le \gamma_i < \overline{\kappa}_i^0 + \frac{\epsilon}{d} q_i^0\right\} \subset B^1.$$
Take $y \in \bar{B}^0$ and write it as $y = \sum_{k=1}^d y_k v_k^0$ and $\underline{\kappa}_k^0 + \frac{\epsilon}{d}(q_k^0 - 1) \le y_k \le \underline{\kappa}_k^0 + \frac{\epsilon}{d} q_k^0$ for $k \in \{1, \ldots, d\}$. Now, rewriting
$$y = \sum_{i=1}^d \left(\sum_{k=1}^d y_k \langle v_k^0, v_i^1\rangle\right) v_i^1,$$
we can check that $\underline{\kappa}_i^1 \le \left(\sum_{k=1}^d y_k \langle v_k^0, v_i^1\rangle\right) \le \overline{\kappa}_i^1$ by definition. This implies that $\Phi_1\left(\bar{B}^0\right) \subset B^1$.

Now, we need to find an estimate for $\#C_i^1$. First, let $(\underline{\gamma}_1^1, \ldots, \underline{\gamma}_d^1)$ be any argument of the minimum corresponding to the minimization used to define $\underline{\kappa}_i^1$, and let $(\overline{\gamma}_1^1, \ldots, \overline{\gamma}_d^1)$ be any argument of the maximum corresponding to the maximization used to define $\overline{\kappa}_i^1$. Next, notice that $\left|\overline{\kappa}_i^1 - \underline{\kappa}_i^1\right| = \left|\sum_{k=1}^d \left(\overline{\gamma}_k^1 - \underline{\gamma}_k^1\right) \langle v_k^0, v_i^1\rangle\right| \le \frac{\epsilon}{d} \sum_{k=1}^d \left|\langle v_k^0, v_i^1\rangle\right|$, because $\left|\overline{\gamma}_k^1 - \underline{\gamma}_k^1\right| \le \epsilon/d$ by definition. Thus,
$$\#C_i^1 \le \left\lceil \Gamma_i^1 e^{T_p\alpha\ell} \sum_{k=1}^d \left|\langle v_k^0, v_i^1\rangle\right|\right\rceil.$$

Further, by the definition of $\hat{\beta}_i^1$ and β_i^1, we have that $\left|\hat{\beta}_i^1 - \beta_i^1\right| \le \frac{\epsilon}{2d} \frac{e^{-T_p\alpha\ell}}{\Gamma_i^1}$. Then, for $1 \le t \le \ell$
$$\begin{aligned} ||\hat{x}(t) - \xi(x,t)|| = \left\|\sum_{i=1}^d \left(\hat{\beta}_i^1 - \beta_i^1\right)\Phi_t v_i^1\right\| &\le \frac{\epsilon}{2d} e^{-T_p\alpha\ell} \left\|\sum_{i=1}^d \frac{\Phi_t v_i^1}{\Gamma_i^1}\right\| \\ &\le \frac{\epsilon}{2} e^{-T_p\alpha t}, \end{aligned}$$
where the last inequality comes from the facts that $\left\|\frac{\Phi_t v_i^1}{\Gamma_i^1}\right\| \le 1$ and $1 \le t \le \ell$. Finally, notice that $x \in \bar{B}^1$ because $\sum_{i=1}^d \beta_i^1 v_i^1 \in \bar{B}^1$ by construction.

Step j+1:

By our induction hypothesis, we have that $x \in \bar{B}^j$. We need to show that
$$\begin{aligned} \Phi_{j\ell+1}\left(\bar{B}^j\right) = \Bigg\{\sum_{i=1}^d \gamma_i \Phi_{j\ell+1} v_i^j : \underline{\kappa}_i^j + \frac{\epsilon}{d}\frac{e^{-T_p\alpha j\ell}}{\Gamma_i^j}(q_i^j - 1) \le \gamma_i \\ < \overline{\kappa}_i^j + \frac{\epsilon}{d}\frac{e^{-T_p\alpha j\ell}}{\Gamma_i^j} q_i^j\Bigg\} \subset B^{j+1}. \end{aligned}$$
Take $y \in \bar{B}^j$ and write it as $y = \sum_{k=1}^d y_k v_k^j$ and $\underline{\kappa}_k^j + \frac{\epsilon}{d}\frac{e^{-T_p\alpha j\ell}}{\Gamma_i^j}(q_k^j - 1) \le y_k \le \underline{\kappa}_k^j + \frac{\epsilon}{d}\frac{e^{-T_p\alpha j\ell}}{\Gamma_i^j} q_k^j$ for $k \in \{1, \ldots, d\}$. Now, rewriting
$$y = \sum_{i=1}^d \left(\sum_{k=1}^d y_k \langle v_k^j, v_i^{j+1}\rangle\right) v_i^{j+1},$$
we can check that $\underline{\kappa}_i^{j+1} \le \left(\sum_{k=1}^d y_k \langle v_k^j, v_i^{j+1}\rangle\right) \le \overline{\kappa}_i^{j+1}$ by definition. This implies that $\Phi_{j\ell+1}\left(\bar{B}^j\right) \subset B^{j+1}$.

Now, we need to find an estimate for $\#C_i^{j+1}$. First, let $(\underline{\gamma}_1^{j+1}, \ldots, \underline{\gamma}_d^{j+1})$ be any argument of the minimum corresponding to the minimization used to define $\underline{\kappa}_i^{j+1}$, and let $(\overline{\gamma}_1^{j+1}, \ldots, \overline{\gamma}_d^{j+1})$ be any argument of the maximum corresponding to the maximization used to define $\overline{\kappa}_i^{j+1}$. Next, notice that $\left|\overline{\kappa}_i^{j+1} - \underline{\kappa}_i^{j+1}\right| = \left|\sum_{k=1}^d \left(\overline{\gamma}_k^{j+1} - \underline{\gamma}_k^{j+1}\right) \langle v_k^j, v_i^{j+1}\rangle\right| \le \frac{\epsilon}{d}\frac{e^{-T_p\alpha\ell}}{\Gamma_i^j} \sum_{k=1}^d \left|\langle v_k^j, v_i^{j+1}\rangle\right|$, because $\left|\overline{\gamma}_k^{j+1} - \underline{\gamma}_k^{j+1}\right| \le \frac{\epsilon}{d}\frac{e^{-T_p\alpha j\ell}}{\Gamma_i^j}$ by definition. Thus, $\#C_i^{j+1} \le \left\lceil e^{T_p\alpha\ell} \frac{\Gamma_i^{j+1}}{\Gamma_i^j} \sum_{k=1}^d \left|\langle v_k^j, v_i^{j+1}\rangle\right|\right\rceil$.

Further, by the definition of $\hat{\beta}_i^{j+1}$ and β_i^{j+1}, we have that

$$\left|\hat{\beta}_i^{j+1} - \beta_i^{j+1}\right| \leq \frac{\epsilon}{2d}\frac{e^{-T_p\alpha(j+1)\ell}}{\Gamma_i^{j+1}}.$$

Then, for $j\ell + 1 \leq t \leq (j+1)\ell$

$$||\hat{x}(t) - \xi(x,t)|| = \left\|\sum_{i=1}^{d}\left(\hat{\beta}_i^{j+1} - \beta_i^{j+1}\right)\Phi_t v_i^{j+1}\right\| \leq$$

$$\frac{\epsilon}{2d}e^{-T_p\alpha(j+1)\ell}\left\|\sum_{i=1}^{d}\frac{\Phi_t v_i^{j+1}}{\Gamma_i^{j+1}}\right\| \leq \frac{\epsilon}{2}e^{-T_p\alpha t},$$

where the last inequality comes from the facts that $\left\|\frac{\Phi_t v_i^{j+1}}{\Gamma_i^{j+1}}\right\| \leq 1$ and $j\ell + 1 \leq t \leq (j+1)\ell$. Finally, notice that $x \in \bar{B}^{j+1}$ because $\sum_{i=1}^{d}\beta_i^{j+1}v_i^{j+1} \in \bar{B}^{j+1}$ by construction. □

It is important to remark that, if $\mathcal{V} = \{v_1,\ldots,v_d\}$ is a normal basis for the Oseledets's filtration of a tempered matrix sequence $(A_j)_{j\in\mathbb{N}}$ and $\mathcal{V}_j = \mathcal{V}$, i.e. $v_i^j = v_i$ for $j \in \mathbb{Z}_{\geq 0}$ and every $i \in \{1,\ldots,d\}$. Then, $\sum_{k=1}^{d}\left|\langle v_k^j, v_i^{j+1}\rangle\right| = 1$ and $\lambda_i = \limsup_{j\to\infty} \frac{1}{j}\log\left(\left\|\Phi_j v_i^j\right\|\right) = \limsup_{j\to\infty}\frac{1}{j}\log\left(\left\|\Phi_j v_i\right\|\right)$, i.e. λ_i's will be the Lyapunov exponents with multiplicity. We know that for every $\eta > 0$, there exists $N \in \mathbb{N}$ such that $\forall j \geq \left\lceil\frac{N-1}{\ell}+1\right\rceil$ and all $i \in \{1,\ldots,d\}$, we have that $\|\phi_t v_i\| \leq e^{T_p(\lambda_i+\eta)t} \leq e^{T_p(\lambda_i+\delta+\eta)t}$ for all $t \geq N$ and this δ is the same as the one used in the definition of Γ_i^j in the algorithm from section 5.1. Further, we know that for $\eta > 0$ sufficiently small, $\lambda_i + \delta + \eta < 0$ for all $\lambda_i + \delta < 0$ with $i \in \{1,\ldots,d\}$. Therefore, for $j \geq \left\lceil\frac{N-1}{\ell}+1\right\rceil$ we have that $\max_{\{0,\ldots,\ell-1\}}\left\{\left\|\phi_{j\ell-k}v_i\right\|\right\} \leq \max\left\{e^{T_p(\lambda_i+\delta+\eta)j\ell}, e^{T_p(\lambda_i+\delta+\eta)((j-1)\ell+1)}\right\}$.

Hence, for all $i \in \{1,\ldots,d\}$, if $\lambda_i + \delta < 0$, we have that $\Gamma_i^j = e^{T_p(\lambda_i+\delta)((j-1)\ell+1)}$, $\forall j \geq \left\lceil\frac{N-1}{\ell}+1\right\rceil$ and $\Gamma_i^j = e^{T_p(\lambda_i+\delta)j\ell}$, $\forall j \geq \left\lceil\frac{N-1}{\ell}+1\right\rceil$, otherwise. Note that for $\lambda_i + \delta \geq 0$, we have

$$e^{T_p(\lambda_i+\delta-\eta)j\ell} \leq \Gamma_i^j \leq e^{T_p(\lambda_i+\delta+\eta)j\ell}$$

and that

$$e^{T_p(\lambda_i+\delta-\eta)((j-1)\ell+1)} \leq \Gamma_i^j \leq e^{T_p(\lambda_i+\delta+\eta)((j-1)\ell+1)}$$

if $\lambda_i + \delta < 0$. Therefore, we have that $\frac{\Gamma_i^{j+1}}{\Gamma_i^j} \leq e^{T_p(\lambda_i+\delta+2\eta)\ell}$ independently of the sign of $\lambda_i + \delta$. Thus, by Theorem 5.1, we have that $\#C_i^{j+1} \leq \left\lceil e^{T_p(\lambda_i+\alpha+\delta+2\eta)\ell}\right\rceil$, $\forall j \geq \left\lceil\frac{N-1}{\ell}+1\right\rceil$ and every $i \in \{1,\ldots,d\}$. We conclude, by showing that the first $\left\lceil\frac{N-1}{\ell}+1\right\rceil + 1$ terms of the sum in the definition of b go to zero and that $\#C^j \leq \prod_{i=1}^{d}\left\lceil e^{T_p(\lambda_i+\alpha+\delta+2\eta)\ell}\right\rceil$ for all $j \geq \left\lceil\frac{N-1}{\ell}+1\right\rceil$, that[9]

$$b \leq \frac{1}{T_p\ell}\sum_{i=1}^{d}\log\left\lceil e^{T_p(\lambda_i+\alpha+\delta+2\eta)\ell}\right\rceil.$$

[9]These steps are similar to those used in the proof of the entropy's upper bound in Theorem 4.13.

Also, because η can be arbitrarily small, we have that

$$b \leq \frac{1}{T_p\ell}\sum_{i=1}^{d}\log\left\lceil e^{T_p(\lambda_i+\alpha+\delta)\ell}\right\rceil.$$

Finally, by choosing ℓ large enough, b can get as close to[10]

$$\sum_{i=1}^{d}\max\{\lambda_i+\alpha+\delta, 0\}$$

as desired.

Following analogous steps, we can prove a similar result for the case when the system is known to be regular. To see this, note that, under the regularity assumption, for every $\eta > 0$ there exists $N \in \mathbb{N}$ such that $e^{T_p(\lambda_i-\eta)t} \leq \|\phi_t v_i\| \leq e^{T_p(\lambda_i+\eta)t}$ for all $t \geq N$. Then, we notice that for $\lambda_i \geq 0$, we have $e^{T_p(\lambda_i-\eta)j\ell} \leq \Gamma_i^j \leq e^{T_p(\lambda_i+\eta)j\ell}$ and that $e^{T_p(\lambda_i-\eta)((j-1)\ell+1)} \leq \Gamma_i^j \leq e^{T_p(\lambda_i+\eta)((j-1)\ell+1)}$ if $\lambda_i < 0$. Next, we get the inequality $\frac{\Gamma_i^{j+1}}{\Gamma_i^j} \leq e^{T_p(\lambda_i+2\eta)\ell}$ independently of the sign of λ_i. Now, we replace this inequality in our previous argument to get that

$$b \leq \frac{1}{T_p\ell}\sum_{i=1}^{d}\log\left\lceil e^{T_p(\lambda_i+\alpha)\ell}\right\rceil,$$

and by choosing ℓ large enough, b can get as close to $\sum_{i=1}^{d}\max\{\lambda_i+\alpha, 0\}$ as desired. These results are summarized in the next Corollary 5.2.

Corollary 5.2. *Let $\delta > 0$, $\alpha \geq 0$, and $\ell \in \mathbb{N}$. If $\mathcal{V}_j = \mathcal{V}$ for all $j \in \mathbb{Z}_{\geq 0}$, where $\mathcal{V}$ is a normal basis for the Oseledets' filtration, then $b \leq \frac{1}{T_p\ell}\sum_{i=1}^{d}\log\left\lceil e^{T_p(\lambda_i+\alpha)\ell}\right\rceil$ if the system is known to be regular and $b \leq \frac{1}{T_p\ell}\sum_{i=1}^{d}\log\left\lceil e^{T_p(\lambda_i+\alpha+\delta)\ell}\right\rceil$, otherwise. Furthermore, b can be made as close as desired to $h_{est}(\alpha, K)$ by choosing ℓ large enough in case the system is known to be regular, or b can be made as close as desired to $\sum_{i=1}^{d}\max\{\lambda_i+\alpha+\delta, 0\}$, otherwise.*

5.2 Finding $(\mathcal{V}_j)_{j\in\mathbb{Z}_{\geq 0}}$ Online

In many practical cases, a priori knowledge of a family $(\mathcal{V}_j)_{j\in\mathbb{Z}_{\geq 0}}$ that gives us an average data-rate close to the estimation entropy, such as normal bases for the Oseledets' filtration as in Corollary 5.2, is unrealistic. Recall that, because of the limit superior in the Definition 4.3 of Lyapunov exponent, we need to know the entire sequence $(A_n)_{n\in\mathbb{N}}$ beforehand to calculate its exponents. Also, notice that a similar thing happens to the Oseledets' filtration. Further, both Examples 4.6 and 4.10 should help making these claims clearer.

Fortunately, one can estimate $(\mathcal{V}_j)_{j\in\mathbb{Z}_{\geq 0}}$ by using the switching signal. However, knowledge of the entire switching signal is also unrealistic. In this Subsection, we assume that only the switching signal's restriction, from the beginning to the current moment, is known and that the system is known to be regular. Based on this new assumption, we show how to estimate the basis $\mathcal{V}_i$. This will give us a causal algorithm to estimate this family and will allow us to work under a more realistic set of hypotheses.

[10]This follows from the fact that $x \leq \frac{1}{\ell}\log\left(\lceil \ell x\rceil\right) \leq \frac{1}{\ell}\log\left(e^{\ell x}+1\right) < x + \frac{\log(2)}{\ell}$ if x is positive, and $0 \leq \frac{1}{\ell}\log\left(\lceil \ell x\rceil\right) \leq \frac{1}{\ell}\log(2)$ if x is negative.

Theorem 5.3. *Assume that $(A_n)_{n\in\mathbb{N}}$ is regular. Let $Q_j := \left(\Phi_j^\top \Phi_j\right)^{\frac{1}{2j}}$ for $j \in \mathbb{Z}_{\geq 0}$ and let its eigenvalues be $e^{\rho_i(j)}$, where $i \in \{1,\ldots,d\}$ and $e^{\rho_1(j)} \leq \cdots \leq e^{\rho_d(j)}$. Also, let $\mathcal{V}_j = \left\{v_1^j,\ldots,v_d^j\right\}$ be an orthonormal basis that diagonalizes Q_j, with an order induced by the order on their corresponding eigenvalues $e^{\rho_i(j)}$. Then the average data-rate of the algorithm from section 5.1 is upper bounded by $\sum_{i=1}^d \max\left\{\alpha+\lambda_i+\frac{1}{T_p\ell},0\right\}$, if the Lyapunov exponents are simple, or $\sum_{i=1}^d \max\left\{\alpha+\lambda_i+\frac{\log(\sqrt{d})+1}{T_p\ell},0\right\}$, otherwise.*

Proof. Our goal is to find an upper bound for $\#C_i^j$ for j large enough. For that purpose, we will use the upper bound obtained in Theorem 5.1. So, we need to find upper bounds or expressions for $\sum_{k=1}^d \left|\langle v_k^j v_i^{j+1}\rangle\right|$ and $\frac{\Gamma_i^{j+1}}{\Gamma_i^j}$.

First, we show that $\lambda_i = \limsup_{j\to\infty}\frac{1}{j}\log\left\|\Phi_j v_i^j\right\|$, which appear in the definition of the algorithm from section 5.1 for $i\in\{1,\ldots,d\}$, are the Lyapunov exponents with multiplicity, and that they are given by $\lambda_i = \lim_{j\to\infty}\rho_i(j)$. To see that, notice that $\left\|Q_j v_i^j\right\| = e^{\rho_i(j)}$ and that

$$\lambda_i = \limsup_{j\to\infty}\frac{1}{j}\log\left\|\Phi_j v_i^j\right\| = \limsup_{j\to\infty}\frac{1}{j}\log((v_i^j)^\top\Phi_j^\top\Phi_j v_i^j)^{1/2} = \limsup_{j\to\infty}\frac{1}{j}\log((v_i^j)^\top Q_j^{2j} v_i^j)^{1/2} = \limsup_{j\to\infty}\rho_i(j),$$

where the second equality comes from the fact that the Euclidean norm and the infinity norm are equivalent. Also, the last equality comes from the fact that any basis that diagonalizes Q_j also diagonalizes Q_j^{2j}.

As a consequence of regularity, by the third bullet of Lemma 4.12, Q_j has a limit. Therefore, its eigenvalues, $e^{\rho_i(j)}$, have a limit as well. Hence, we conclude that $\lambda_i = \lim_{j\to\infty}\rho_i(j)$, because the limit on the right exists.

Second, because the sequence is regular, we have that Q_j converges. We denote this limit by $Q := \lim_{j\to\infty} Q_j$. Because Lyapunov exponents are simple, there exists $N_0\in\mathbb{N}$ such that for all $j\geq N_0$ the eigenvalues of Q_j are simple as well. Now, a symmetric matrix with simple eigenvalues has a unique, up to a change of signs and subject to the order indicated in the theorem statement, orthonormal basis that diagonalizes it. This implies that for any $\eta_1>0$, there exists $N_1\in\mathbb{N}$ such that $\sum_{k=1}^d\left|\langle v_k^j, v_i^{j+1}\rangle\right| \leq 1+\eta_1$ for all $j\geq N_1$ and $i\in\{1,\ldots,d\}$. To see this, denote by $\{v_1,\ldots,v_d\}$ a basis that diagonalizes Q. Now, we can change the signs of $\left\{v_1^j,\ldots,v_d^j\right\}$ if necessary, so that v_i^j converges to v_i, and notice that changing the sign does not change the absolute value of the inner products mentioned above. Because these are orthonormal bases, there exists $N_1\in\mathbb{N}$ such that, for every $i\in\{1,\ldots,d\}$, we have $|\langle v_k^j, v_i^{j+1}\rangle| \leq \eta_1/d$ if $k\neq i$ and $|\langle v_k^j, v_i^{j+1}\rangle| \leq 1+\eta_1/d$ if $k=i$, and we proved this claim. Notice, however, that the inequalities $\sum_{k=1}^d\left|\langle v_k^j, v_i^{j+1}\rangle\right| \leq \sqrt{d}$ for every $i\in\{1,\ldots,d\}$ always hold, even without simplicity.

Third, again because of regularity, for $\eta_2>0$ such that $\lambda_i+\eta_2<0$ for all $\lambda_i<0$, but otherwise arbitrary[11], there exists $N_2\in\mathbb{N}$ such that for all $j\geq N_2$ and all $i\in\{1,\ldots,d\}$ we have that $\lambda_i-\eta_2\leq\rho_i(j)\leq\lambda_i+\eta_2$. Thus, $\Gamma_i^j := \max_{k\in\{0,\ldots,\ell-1\}}\left\|\Phi_{j\ell-k}v_i^j\right\| = \max_{k\in\{0,\ldots,\ell-1\}}\left\|e^{\rho_i(j\ell-k)}\right\|$. Then, we arrive at the inequalities $e^{T_p(\lambda_i-\eta_2)j\ell}\leq\Gamma_i^j\leq e^{T_p(\lambda_i+\eta_2)j\ell}$, if $\lambda_i\geq 0$, and $e^{T_p(\lambda_i-\eta_2)((j-1)\ell+1)}\leq\Gamma_i^j\leq e^{T_p(\lambda_i+\eta_2)((j-1)\ell+1)}$, if $\lambda_i<0$. Then, $\frac{\Gamma_i^{j+1}}{\Gamma_i^j}\leq e^{T_p(\lambda_i+2\eta_2)\ell}$ for $j\geq N_2$ and $i\in\{1,\ldots,d\}$.

Now, recall the definition of average data-rate

$$b = \limsup_{t\to\infty}\frac{1}{T_p t\ell}\sum_{j=0}^t\sum_{i=1}^d\log(\#C_i^j).$$

Denote $N := \max\{N_1,N_2\}$. For $j\geq N$ we have that

$$\#C_i^j \leq \left\lceil e^{T_p(\alpha+\lambda_i+2\eta_2)\ell}(1+\eta_1)\right\rceil.$$

Further, define $M=\sum_{j=0}^{N-1}\sum_{i=1}^d\log\left(\#C_j^i\right)$. We can upper-bound the average data-rate by

$$b\leq\limsup_{t\to\infty}\frac{1}{T_p t\ell}\left(M+\sum_{k=N}^t\sum_{i=1}^d\log(\left\lceil e^{T_p(\alpha+\lambda_i+2\eta_2)\ell}(1+\eta_1)\right\rceil)\right).$$

Notice that $\log(\lceil x\rceil)\leq\max\{\log(x)+1,0\}$. To see that, we study two cases. If $x\geq 1$, then $2x\geq x+1$ and $\log(2x)=\log(2)+\log(x)=1+\log(x)\geq\log(x+1)\geq\log(\lceil x\rceil)$. If $x<1$, then $\log(\lceil x\rceil)=0$. Therefore, we can derive the upper bound

$$\log(\left\lceil e^{T_p(\alpha+\lambda_i+2\eta_2)\ell}(1+\eta_1)\right\rceil)\leq \max\left\{T_p(\alpha+\lambda_i+2\eta_2)\ell(1+\eta_1)+1,0\right\}.$$

Thus,

$$b\leq\limsup_{t\to\infty}\frac{1}{T_p t\ell}\Big(M+(t-N)\sum_{i=1}^d\max\left\{T_p(\alpha+\lambda_i+2\eta_2)\ell+\log(1+\eta_1)+1,0\right\}\Big)$$

and since M and N are constants, we conclude that

$$b\leq\sum_{i=1}^d\max\left\{\alpha+\lambda_i+2\eta_2+\frac{\log(1+\eta_1)}{T_p\ell}+\frac{1}{T_p\ell},0\right\}.$$

Since $\eta_1>0$ and $\eta_2>0$ can be chosen to be arbitrarily small, we have that

$$b\leq\sum_{i=1}^d\max\left\{\alpha+\lambda_i+\frac{1}{T_p\ell},0\right\}.$$

Finally, if we drop the simplicity assumption, we could replace $\log(1+\eta_1)$ by $\log(\sqrt{d})$ and obtain

$$b\leq\sum_{i=1}^d\max\left\{\alpha+\lambda_i+\frac{\log(\sqrt{d})+1}{T_p\ell},0\right\}.$$

[11] Notice that η_2 can be chosen to be as small as desired.

and, therefore, in both cases, by choosing ℓ sufficiently large, the upper bound on b can be made arbitrarily close to the estimation entropy $h_{\text{est}}(\alpha, K)$ as given by the last statement of Theorem 4.13. □

Remark 5.1. *It is important to remark what still holds without regularity and simplicity. First, it is always true that* $\sum_{k=1}^{d} \left|\langle v_k^j, v_i^{j+1}\rangle\right| \leq \sqrt{d}$ *for every* $i \in \{1, \ldots, d\}$*. Second, without regularity, we have that for every* $\eta_2 > 0$*, there exists* $N \in \mathbb{N}$ *such that* $\frac{\Gamma_i^{j+1}}{\Gamma_i^j} \leq e^{T_p(\lambda_i+\delta+2\eta_2)\ell}$ *for all for* $j \geq N$*, where* $\delta > 0$ *is the same that appears in the definition of* Γ_i^j *in the algorithm from section 5.1. Furthermore, from the inequalities, we conclude that* $\#C_i^j \leq \left\lceil e^{T_p(\alpha+\lambda_i+\delta+2\eta_2)\ell}\sqrt{d}\right\rceil$ *for* $j \geq$ *and* $i \in \{1, \ldots, d\}$*. Using this upper bound for* $\#C_i^j$ *and following steps of the proof above, we conclude that*

$$b \leq \sum_{i=1}^{d} \max\left\{(\alpha+\lambda_i+\delta) + \frac{\log(\sqrt{d})+1}{T_p\ell}, 0\right\}.$$

Observe that these λ_i*'s aren't the Lyapunov exponents with multiplicity. These* λ_i*'s are the upper growth rates of the singular values of* Q_j *as* j *goes to infinity, see e.g. Chapter 6 of [2]. Also, it is well-known that these* λ_i*'s are smaller than or equal to the Lyapunov exponents when we don't have regularity. For that reason, this algorithm might work at an average data-rate smaller than the entropy's upper bound obtained in Theorem 4.13.*

Furthermore, note that, without the regularity assumption, we need to have a priori knowledge either of the λ_i's, or an upper bound to them. Both hypothesis are unreasonable if we want to have a completely causal algorithm, since the λ_i's depend on the entire sequence $(A_n)_{n\in\mathbb{N}}$.

Another important observation is that the simplicity of the Lyapunov exponents is a generic property, and we expect that most systems will have it. See e.g. Chapter 8 of [20].

6 SIMULATION RESULTS

In this section, we implement the algorithm from section 5.1 using the family of bases constructed in Theorem 5.3 to reconstruct the state of the Markov Jump Linear System in Example 3.1. It is important to mention that such systems, although random, have realizations that are regular with probability 1, see [21]. Also, we refer to that work for results concerning more general sufficient conditions for regularity.

Example 6.1 (Example 3.1 revisited). Since the realizations of the system presented in Example 3.1 are regular with probability 1, the upper bound found in Example 4.14 was actually the real value of the estimation entropy for our system, i.e. $h_{\text{est}}(\alpha, K) = \max\left\{\frac{1}{2}\log(0.99)+\alpha, 0\right\} + \max\{\alpha, 0\}$ nats/sample or, equivalently, $h_{\text{est}}(\alpha, K) = \log_2(e)\big(\max\left\{\frac{1}{2}\log(0.99)+\alpha, 0\right\} + \max\{\alpha, 0\}\big)$ bits/sample with probability 1. We can now apply the previous algorithm to a randomly chosen realization of our example system. The parameters chosen were $\alpha = 0.05$, $\epsilon = 0.01$, and the time horizon for our simulation was 140 time units. Further, $K = [0.5, 1.5] \times [1.5, 2.5]$, $x(0) = (1.102, 2.104)^\top$. Notice that, for this α, we get $h_{\text{est}}(0.05, K) \approx 0.137$ bits/sample.

One can see the simulation results of the estimation error in Figure 1 for block lengths $\ell = 1$ in blue, $\ell = 3$ in red, and $\ell = 5$ in yellow. We can see that the error is upper bounded by the purple curve $\epsilon e^{-\alpha t}/2$ for all values of ℓ. Further, the empirical average data-rate, i.e. $\frac{1}{t\ell}\sum_{j=1}^{t}\log\left(C_i^j\right)$, is portrayed in Figure 2, where we can see that the data rate decreases as the block length increases, as expected. Nonetheless, the average data-rate is far from the upper bound derived in Theorem 5.3. That happens because the result in Theorem 5.3 is only asymptotic.

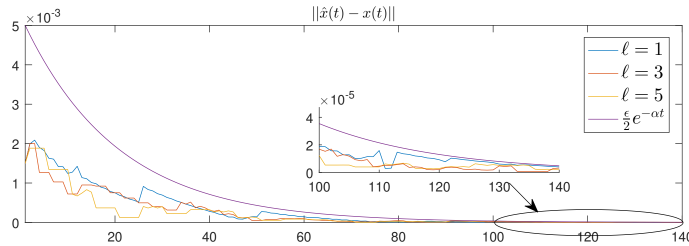

Figure 1: Evolution of error for several block lengths

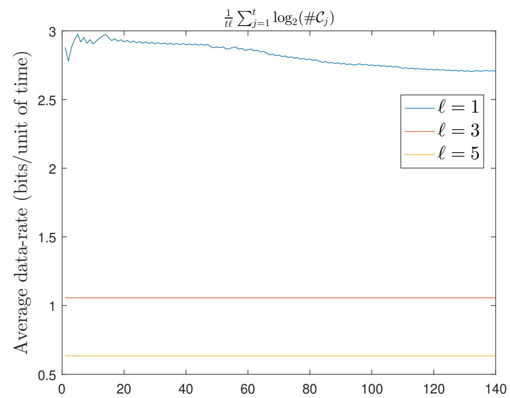

Figure 2: Evolution of the empirical average data-rate for several block lengths

7 CONCLUSION AND FUTURE WORKS

In this paper, we addressed the problem of designing a quantization scheme for exponentially fast state reconstruction that operates at an average data-rate arbitrarily close to the estimation entropy for regular switched systems. Furthermore, we showed how to make the algorithm work only using information that is known up to the current time. Moreover, we showed that our algorithm works even if the underlying system is not regular. As future research directions, we propose to use a modified version of the present algorithm to perform state estimation for nonlinear systems with minimum average data-rate. Also, we plan on addressing the control of switched linear systems with the optimal data rate as well.

REFERENCES

[1] Ludwig Arnold. 1998. *Random Dynamical Systems.* Springer Berlin Heidelberg.
[2] Luís Barreira. 2017. *Lyapunov Exponents.* Springer International Publishing.
[3] Guillaume O. Berger and Raphaël M. Jungers. 2020. Worst-case topological entropy and minimal data rate for state observation of switched linear systems. In *Proceedings of the 23rd International Conference on Hybrid Systems: Computation and Control.*
[4] Guillaume O. Berger and Raphaël M. Jungers. 2021. Quantized Stabilization of Continuous-Time Switched Linear Systems. *IEEE Control Systems Letters* 5, 1 (2021), 319–324.
[5] Fritz Colonius. 2012. Minimal Bit Rates and Entropy for Exponential Stabilization. *SIAM Journal on Control and Optimization* 50, 5 (jan 2012), 2988–3010.
[6] Thomas M. Cover and Joy A Thomas. 1991. *Elements of Information Theory.* Wiley.
[7] João P. Hespanha, Antonio Ortega, and Lavanya Vasudevan. 2002. Towards the Control of Linear Systems with Minimum Bit-Rate. In *In Proc. of the Int. Symp. on the Mathematical Theory of Networks and Syst.*
[8] Christoph Kawan. 2013. *Invariance Entropy for Deterministic Control Systems.* Springer International Publishing.
[9] Christoph Kawan. 2018. Exponential state estimation, entropy and Lyapunov exponents. *Systems & Control Letters* 113 (2018), 78–85.
[10] Daniel Liberzon and Sayan Mitra. 2016. Entropy notions for state estimation and model detection with finite-data-rate measurements. In *2016 IEEE 55th Conference on Decision and Control (CDC).*
[11] Daniel Liberzon and Sayan Mitra. 2018. Entropy and Minimal Bit Rates for State Estimation and Model Detection. *IEEE Trans. Automat. Control* 63, 10 (oct 2018), 3330–3344.
[12] Alexey Matveev and Alexander Pogromsky. 2016. Observation of nonlinear systems via finite capacity channels: Constructive data rate limits. *Automatica* 70 (2016), 217–229.
[13] Alexander S. Matveev and Alexey V. Savkin. 2007. *Estimation and Control over Communication Networks (Control Engineering).* Birkhauser.
[14] Ricardo Mañé. 1981. A proof of Pesin's formula. *Ergodic Theory and Dynamical Systems* 1 (1981), 95–102.
[15] Girish N. Nair, Robin J. Evans, Iven M.Y. Mareels, and William Moran. 2004. Topological Feedback Entropy and Nonlinear Stabilization. *IEEE Trans. Automat. Control* 49, 9 (sep 2004), 1585–1597.
[16] Yakov B. Pesin. 1997. *Dimension Theory in Dynamical Systems: Contemporary Views and Applications.* University of Chicago Press.
[17] Matthias Rungger and Majid Zamani. 2017. Invariance Feedback Entropy of Nondeterministic Control Systems. In *Proceedings of the 20th International Conference on Hybrid Systems: Computation and Control.*
[18] Hussein Sibai and Sayan Mitra. 2017. Optimal Data Rate for State Estimation of Switched Nonlinear Systems. In *Proceedings of the 20th International Conference on Hybrid Systems: Computation and Control.*
[19] Adriano Da Silva and Christoph Kawan. 2019. Lyapunov exponents and partial hyperbolicity of chain control sets on flag manifolds. *Israel Journal of Mathematics* 232, 2 (2019), 947–1000.
[20] Marcelo Viana. 2014. *Lectures on Lyapunov Exponents.* Cambridge University Press.
[21] Guilherme S. Vicinansa and Daniel Liberzon. 2019. Estimation Entropy for Regular Linear Switched Systems. In *2019 IEEE 58th Conference on Decision and Control (CDC).*
[22] Guosong Yang, João P. Hespanha, and Daniel Liberzon. 2019. On topological entropy and stability of switched linear systems. In *Proceedings of the 22nd ACM International Conference on Hybrid Systems: Computation and Control.*
[23] Guosong Yang, A. James Schmidt, and Daniel Liberzon. 2018. On Topological Entropy of Switched Linear Systems with Diagonal, Triangular, and General Matrices. In *2018 IEEE Conference on Decision and Control (CDC).*
[24] Guosong Yang, A. James Schmidt, Daniel Liberzon, and João P. Hespanha. 2020. Topological entropy of switched linear systems: general matrices and matrices with commutation relations. *Mathematics of Control, Signals, and Systems* 32, 3 (2020), 411–453.