SIAM J. CONTROL OPTIM.
Vol. 61, No. 1, pp. 198-224

ESTIMATION ENTROPY, LYAPUNOV EXPONENTS, AND QUANTIZER DESIGN FOR SWITCHED LINEAR SYSTEMS*

GUILHERME S. VICINANSA† AND DANIEL LIBERZON†

Abstract. In this paper, we study connections between the estimation entropy of a switched linear system and its Lyapunov exponents. We prove lower and upper bounds for the estimation entropy in terms of the Lyapunov exponents and show that, under the so-called regularity assumption, those bounds coincide. To do that, we use a geometric object called Oseledets' filtration of the system. Further, we show how to use the exponents and the Oseledets' filtration to design a quantization scheme for state estimation of switched linear systems. Then, we prove that we can make this algorithm work at an average data-rate arbitrarily close to the upper bound we provided for 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 is regular. We show that, under the regularity assumption, the quantization scheme is completely causal in the sense that it depends only on information that is available up to the current time instant. We show that regularity is a natural property of many practical systems, such as Markov jump linear systems, and give sufficient conditions for it.

Key words. estimation entropy, minimum data-rate, switched systems

MSC codes. 93C30, 94A17

DOI. 10.1137/21M1411871

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 among one another, we need to use communication channels. Those communication channels, in 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 to stabilize the system.

The answers to the above 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., [8, 12, 17, 18, 21, 23]. 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 [15] for generic autonomous nonlinear systems. We can, therefore, think of 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. One of the goals of the present paper is to address this issue. We

*Received by the editors April 13, 2021; accepted for publication (in revised form) September 5, 2022; published electronically February 15, 2023. Some partial preliminary versions of the results in this paper were presented at the 58th Conference on Decision and Control and at the 24th ACM International Conference on Hybrid Systems: Computation and Control.
https://doi.org/10.1137/21M1411871
Funding: This work was supported by NSF grants CMMI-1662708 and CMMI-2106043 and by AFOSR grant FA9550-17-1-0236.
†Coordinated Science Laboratory, University of Illinois Urbana-Champaign, Urbana, IL 61801 USA (gs16@illinois.edu, liberzon@illinois.edu).

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. Plus, we present some of its properties related to its data-rate and robustness. Another goal of this work is to show a relationship between the estimation entropy of a switched linear system and its Lyapunov exponents. This result has been proved by the authors in [29]. However, the present proof is different, and it makes it easier to see an interesting relationship between the Lyapunov exponents, which are geometric objects called filtrations (that play a role similar to eigenspaces in the linear time-invariant case), and the quantizer design.

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 works to explicitly present an entropy notion for switched systems, related to the estimation entropy defined in [15], were [31, 24, 25, 32, 10]. We remark that the authors of [24] studied switched nonlinear systems with unknown switching signals—but that satisfy a minimum dwell-time restriction—using a nonstandard modification of the entropy notion defined in [15]. The same authors extended their previous work on their version of entropy by considering systems with inputs in [25]. The authors of [32] presented upper and lower bounds, under some structural assumptions on the modes, for the topological entropy of switched linear systems, which can be seen as a particular case of the estimation entropy of [15] when $\alpha = 0$. The bounds from [32] were improved in [10]; see also [33]. However, all of these bounds relied on individual modes and their active rates, and no other features of the switching signal were assumed. The authors of [10] concluded that no better bounds could be achieved without further knowledge of the switching signal structure. This issue was addressed in [29], where the authors presented bounds that were tight for a large class of systems, called regular switched systems, and those bounds improved because they rely on the knowledge of the entire switching signal. Furthermore, [29] presented an upper bound for the estimation entropy of switched linear systems that is related to the Lyapunov exponents. It is worth mentioning that in [20], the authors obtained the same minimum average data-rate as in [29], with $\alpha = 0$, but for the mean square stabilization problem of scalar Markov jump linear systems. Also, a very similar relationship between the Lyapunov exponents and the entropy appears in several places in the dynamical systems literature, often under the name *Pesin entropy formula* [19, 22, 28], as well as in the formula for the invariance entropy of partially hyperbolic control systems [26], and in a lower bound for the estimation entropy of a class of differential dynamics on compact manifolds [13]. It is important to note that most of those proofs rely on the differentiability of the flow and on the compactness of the space, which is different from what is done in [29] and the present work.

More recently, in [4] the authors provided a way of computing the maximum topological entropy, over all possible switching signals, that a switched linear system can have. Also, those authors proved in [3] that the topological entropy of a linear time-varying system equals the minimum average data-rate for the state observation with bounded estimation error. Finally, in [5] the authors provide an algorithm that stabilizes a switched linear system with an average data-rate arbitrarily close to the minimum. An important remark is that, although seemingly different, the minimum data-rate for stabilization obtained in [5], in terms of the Lyapunov exponent of exterior products, has the same value as the estimation entropy lower bound obtained in [29] utilizing the usual Lyapunov exponents of linear systems, with[1] $\alpha = 0$. Nonetheless, the algorithm presented in [5] requires us to know an a priori upper bound for the entropy, which might 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 [29], providing a connection between the Lyapunov exponents and the estimation entropy, and giving a constructive and causal algorithm that builds a state estimate for a switched linear system with a prescribed exponential decay 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. We show that the regularity assumption is fulfilled by several systems of practical interest, such as those modeled as Markov jump linear systems. Furthermore, we address the role that the regularity property plays regarding the causality of our algorithm. For instance, we can ensure the exponential decay of the error without knowledge of the switching signal at all times, as required in [5].

This paper has the following structure: In section 2, we explain the problem and its formulation. Also, we motivate this study through an example where the current quantization methods perform worse than the method presented here regarding the use of a data-rate. Then, in section 3 we study the concepts of Lyapunov exponents, Oseledets' filtration, and estimation entropy. In our analysis of the estimation entropy, we give it an upper bound and show that this upper bound is the exact value under the Lyapunov regularity assumption. In section 4, we present our quantization 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 the algorithm's parameters. Furthermore, we present how to make the algorithm operate at an average data-rate arbitrarily close to the entropy in a causal way. Next, in section 5, we advocate in favor of regularity, showing that sufficient conditions for it are natural for many systems. Finally, in section 6 we draw our conclusions and propose future works.

Notation. We denote by $||\cdot||_P$ the norm in $\mathbb{R}^d$ induced by the inner product $\langle x, y\rangle_P = x^\top P y$, with P positive definite. We denote by $||\cdot||$ the infinity-norm in a finite dimensional vector space. Let $\mathbb{R} = (-\infty, \infty)$, let $\mathbb{Z}_{\geq 0} = \{0, 1, \dots\}$ be the nonnegative integers, and let $\mathbb{N} = \{1, 2, \dots\}$ be the set of natural numbers. For a real number x, we denote by $\lceil x \rceil$ the smallest integer number y such that $x \leq y$. For any set E, we denote by $\#E$ its cardinality. For subsets of $\mathbb{R}^d$ we denote by $\mathrm{vol}(E)$ the volume of the set (its Lebesgue measure). Further, we denote $\mathrm{diam}(E)$, where $E \subset \mathbb{R}^d$ is 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. Furthermore, we define by $\mathbb{B}(x, r)$ with $x \in \mathbb{R}^d$ and $r > 0$ the infinity-norm ball (hypercube) with center x and radius r.

We denote by $\mathcal{M}(d, \mathbb{R})$ the set of all $d \times d$ matrices over the reals, and we denote by $\mathrm{GL}(\mathbb{R}, d)$ the set of all $d \times d$ invertible matrices over the reals. We denote $\det(A)$ and $\mathrm{tr}(A)$ as the determinant and the trace of the matrix A, respectively. 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 kth volume of the

[1]See, e.g., Chapter 6 of [2].

parallelepiped by $\mathrm{vol}(\{v_1, \ldots, v_k\})$, and its numerical value is given by $\sqrt{\det(V^\top V)}$, where V is the $d \times k$ matrix with columns v_i.[2]

2. Preliminaries. Consider the switched linear system model

$$\dot{x}(t) = \mathcal{A}_{\sigma(t)} x(t), \tag{2.1}$$

where $x(t) \in \mathbb{R}^d$, $\sigma : \mathbb{R}_{\geq 0} \to \Sigma$ is a switching signal, Σ is a finite cardinality set, and $\mathcal{A}_{\sigma(t)} \in \mathcal{M}(d, \mathbb{R})$. We denote by $\Phi(t, t_0)$ the *state-transition matrix* of (2.1), i.e., the solution of the ODE $\frac{d}{dt}\Phi(t, t_0) = \mathcal{A}_{\sigma(t)}\Phi(t, t_0)$, 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 $\lim_{i\to\infty} t_i = \infty$. The elements of the sequence $(t_i)_{i\in\mathbb{N}}$ 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 (2.1) using its *exact discrete-time model*, defined by

$$x_{k+1} = \tilde{A}_k x_k, \tag{2.2}$$

where $(x_k)_{k\in\mathbb{Z}_{\geq 0}}$ is the state at the sampling times τ_k, i.e., $x_k = x(\tau_k)$ for $k \in \mathbb{Z}_{\geq 0}$, and $\tilde{A}_k := \Phi(\tau_{k+1}, \tau_k)$ for $k \in \mathbb{Z}_{\geq 0}$.

Consider the following definitions of coder-estimator scheme; see, for instance, [18, 21]. Let $\{\tau_k\}_{k\in\mathbb{Z}_{\geq 0}}$ be the above-mentioned sequence of sampling times. Also, let $\{\mathcal{C}^n\}_{n\in\mathbb{Z}_{\geq 0}}$ be a sequence of alphabets with uniformly bounded cardinality, i.e., $\exists M > 0$, $\#\mathcal{C}^i < M$, for all $i \in \mathbb{Z}_{\geq 0}$. 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 $\gamma_n : \prod_{i=0}^{n-1}\mathcal{C}^i \times \mathbb{R}^{d(n+1)} \to \mathcal{C}^n$, where γ_n is called the *coder mapping at time n*. We can write the coder mapping in the more explicit way[3]

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

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

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

$$b := \limsup_{j\to\infty} \frac{1}{t_j} \sum_{i=0}^{j} \log\left(\#\mathcal{C}^i\right). \tag{2.3}$$

2.1. Example. In this subsection, 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 2.1. Let $B_1 = \left[\begin{smallmatrix} 0.9 & 0.03 \\ 0 & 1 \end{smallmatrix}\right]$ and $B_2 = \left[\begin{smallmatrix} 1.1 & 0.02 \\ 0 & 1 \end{smallmatrix}\right]$ be the modes of our discrete-time switched system, i.e., $\tilde{A}_k \in \{B_1, B_2\}$ for $k \in \mathbb{Z}_{\geq 0}$. Notice that the mode B_2

[2] Notice that interchanging the order of the columns does not change the kth volume.

[3] Notice that, since $q_0 = \gamma_0(x(\tau_0))$, one could make an alternative definition of the coder mapping at time 1 as $\tilde{\gamma}_1(x(\tau_0), x(\tau_1)) = \gamma_1(\gamma_0(x(\tau_0)), x(\tau_0), x(\tau_1))$. Then, one could alternatively define the coder mapping at time n as $\tilde{\gamma}_n(x(\tau_0), \ldots, x(\tau_n))$ recursively in a similar way. Defining the coder mappings with the explicit dependence of the quantized value on the previous symbols is a matter of keeping the argumentation clear.

is unstable. Therefore, if we apply the conventional quantization scheme [11], this guarantees that our estimation error is uniformly bounded by some arbitrary constant $\epsilon > 0$ and utilizes the minimum average data-rate; for each mode separately, the average data-rate used will be positive. Nonetheless, we will show that if our switching signal comes from the Markov chain defined by the matrix of transition probabilities $P = \left[\begin{smallmatrix} 0.1 & 0.9 \\ 0.9 & 0.1 \end{smallmatrix}\right]$, where P_{ij} is the transition probability from mode i to mode j, then there exists an algorithm that reconstructs the state with error uniformly bounded by some arbitrary $\epsilon > 0$ using an average data-rate as close to zero as desired with probability 1 in the previous situation. We will see that this latter fact follows since the estimation entropy with $\alpha = 0$ of our switched system is 0 almost surely.

In this paper, we present a quantization scheme that operates at an average data-rate arbitrarily close 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.

3. Estimation entropy.

In this section, we introduce Lyapunov exponents, Lyapunov regularity, estimation entropy, and related concepts. We derive an upper bound for the estimation entropy of discrete-time switched systems in terms of the Lyapunov exponents. Moreover, we show that under the assumption of regularity, this upper bound is the actual value of the estimation entropy. The definitions presented here were adapted from the [16], Chapter 2 of [2], and Chapter 3 of [1].

Throughout this article, 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 (2.2) by

$$\Phi_n := A_n \cdots A_1. \tag{3.1}$$

Here and in the rest of the paper, we denote $A_n := \tilde{A}_k$ for $n = k+1$ with[4] $k \in \mathbb{Z}_{\geq 0}$. 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 (2.2) 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 (2.2).

For the next two definitions, pick an $\alpha \geq 0$, and let $T \in \mathbb{N}$ be the time horizon.

Definition 3.1. *For every $\epsilon > 0$, we call a finite set of functions $\hat{X} = \{\hat{x}_1(\cdot),\dots,\hat{x}_N(\cdot)\}$, from $\{0,\dots,T-1\}$ 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 $||\xi(x,n) - \hat{x}_i(n)|| < \epsilon e^{-\alpha n}$* for all $n \in \{0,\dots,T-1\}$. *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 3.2. *For every $\epsilon > 0$, we call a finite set of points $S = \{x_1,\dots,x_N\} \subset K$ a (T,ϵ,α,K)-spanning set if for every initial state $x \in K$, there exists $x_i \in S$ such that $||\xi(x,n) - \xi(x_i,n)|| < \epsilon e^{-\alpha n}$* for all $n \in \{0,\dots,T-1\}$. *Let $s^*_{\text{est}}(T,\epsilon,\alpha,K)$ be the minimum cardinality of a (T,ϵ,α,K)-spanning set.*

It is important to note that, by Theorem 1 from [16], we know that $h_{\text{est}}(\alpha,K) = \lim_{\epsilon\to 0}\limsup_{T\to\infty}\frac{1}{T}\log s^*_{\text{est}}(T,\epsilon,\alpha,K)$. This equivalent characterization of the estimation entropy will be needed for proving Theorem 3.14.

[4]Notice that $A_1 = \tilde{A}_0$. Thus, $A_n = \Phi(T_p n, T_p(n-1))$ for $n \in \mathbb{N}$. This relabeling is done for consistency with the majority of the literature.

Definition 3.3. *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* $\kappa \neq 0$,
- $\lambda(v + w) \leq \max\{\lambda(v), \lambda(w)\}$,
- $\lambda(0) = -\infty$.

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 3.4. *The Lyapunov exponent associated with a sequence of matrices* $(A_n)_{n\in\mathbb{N}}$ *is the following Lyapunov index:*[5]

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

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 3.5. *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 4, is the Oseledets' filtration, which we define next.

Definition 3.6. *A filtration* $\mathcal{V}_\lambda$ *associated with the sequence of invertible matrices* $(A_n)_{n\in\mathbb{N}}$ *such that* $E_i = \{v \in \mathbb{R}^d : \lambda(v) \leq \chi_i\}$, *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,* $\dim(E_i) - \dim(E_{i-1})$ *in the following is called the multiplicity of the Lyapunov exponent value* χ_i*: If*[6] $\dim(E_i) - \dim(E_{i-1}) = 1$ *for every* $i \in \{1, \ldots, q\}$, *we say that the Lyapunov exponents are simple. Finally, define* $\Lambda = \{\lambda_j\}_{j=1}^d$ *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 note that the Oseledets' filtration depends on the entire sequence $(A_n)_{n\in\mathbb{N}}$. To see this, consider the following example.

Example 3.7. Let $A = \left[\begin{smallmatrix} 2 & 0 \\ 0 & 4 \end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ and note 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. Since the Oseledets' filtration of the first sequence is $E_1 = \operatorname{span}\{[\ 1 \quad 0\ ]^\top\} \subsetneq E_2 = \mathbb{R}^2$ and the filtration of the second is $E_1 = \operatorname{span}\{[\ 0 \quad 1\ ]^\top\} \subsetneq E_2 = \mathbb{R}^2$.

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

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

Definition 3.8. *A sequence* $(A_n)_{n\in\mathbb{N}}$ *is called tempered if*

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

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 subexponential. To see why, take $A_n = n$, which is tempered because $\lim_{n\to\infty}\frac{\log(n)}{n}=0$, and note that $\Phi_n = n!$, which grows faster than any exponential.

Example 3.9 (Example 2.1 revisited). This is a good moment for us to revisit our Example 2.1. Denote by $a_{ij}(n)$ the element in the ith row and jth column of the matrix A_n, and denote, analogously, by $\phi_{ij}(n)$ the elements of Φ_n. Further, denote $m_i(n)=\sum_{k=1}^{n}\mathbb{I}_{\{(A_n)_{n\in\mathbb{N}}:A_k=B_i\}}((A_n)_{n\in\mathbb{N}})$, 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 the number of time instants at which 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 $\lambda(e_1)=\limsup_{n\to\infty}\frac{1}{n}\log(\|\Phi_n e_1\|)=\limsup_{n\to\infty}\frac{1}{n}\log(0.9^{m_1(n)}1.1^{m_2(n)})=\limsup_{n\to\infty}\frac{m_1(n)}{n}\log(0.9)+\frac{m_2(n)}{n}\log(1.1)$. 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 this, 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 this 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 noting 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 $E_1=\operatorname{span}\{e_1\}\subsetneq E_2=\mathbb{R}^2$ 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 3.10. *A sequence* $(A_n)_{n\in\mathbb{N}}$ *is called (Lyapunov) regular if*

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

We call a system given by (2.2) *regular if its associated matrix sequence is regular.*

The following Examples 3.11 and 3.12 should help illustrate the concept of regularity.

Example 3.11. Let $\rho > 1$. Also, let $B_1 = \left[\begin{smallmatrix} \rho & 0 \\ 0 & \rho^{-1} \end{smallmatrix}\right]$ and $B_2 = \left[\begin{smallmatrix} \rho^{-1} & 0 \\ 0 & \rho \end{smallmatrix}\right]$. Consider the sequence $A_n = B_1$ if $n \in \{2^i, \dots, 2^{i+1} - 1\}$, 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 $\|\Phi_{n_k}(e_1)\| = \rho^{-\sum_{i=1}^{k}(-2)^{i-1}+(-1)^k}$. Thus, $\frac{\log(\|\Phi_{n_k}(e_1)\|)}{2^k} = \sum_{\ell=1}^{k} \left((-1)^{\ell+1}(2)^{-\ell} + (-1)^k 2^{-k}\right) \log(\rho)$ after the change of variables $\ell = -i + k + 1$. 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 3.12. Let B_1 and B_2 be as in Example 3.11. 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 3.11, the limit superior in Definition 3.4 of the Lyapunov exponent cannot be replaced by a limit, but in Example 3.12, where the matrix sequence is regular, it can. This fact is not a coincidence, as shown by the second bullet of Lemma 3.13 for $\mathcal{I}$ being a singleton, which implies that the limit exists when the sequence is regular.

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

LEMMA 3.13. *Given a tempered sequence $(A_n)_{n\in\mathbb{N}}$ of invertible matrices, let $\{v_1, \dots, v_d\}$ be any normal basis for the Oseledets' filtration of the sequence $(A_n)_{n\in\mathbb{N}}$, and let $\mathcal{I} \subset \{1, \dots, d\}$ 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:*

- $\lim_{n\to\infty} \frac{1}{n} \log\left(|\det(\Phi_n)|\right) = \sum_{i=1}^{d} \lambda_i$.
- $\lim_{n\to\infty} \frac{1}{n} \log\left(\mathrm{vol}\left(\{\Phi_n v_i : i \in \mathcal{I}\}\right)\right) = \sum_{i\in\mathcal{I}} \lambda_i$.
- *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 3.14. *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 compact set of possible initial conditions with a nonempty interior. Denote by λ_i, with $i = 1, \dots, d$, the Lyapunov exponents with multiplicity of $(A_n)_{n\in\mathbb{N}}$. Then, the estimation entropy of the discrete switched system* (2.2) *satisfies*

$$h_{\mathrm{est}}(\alpha, K) \leq \sum_{i=1}^{d} \max\{0, \lambda_i + \alpha\}, \tag{3.2}$$

with equality if the system is regular.

Proof. For the proof of the upper bound, we build a (T, ϵ, α, K)-approximating set $\mathcal{F}_T$ and calculate its cardinality. First, denote by $\{v_1, \dots, v_d\}$ a normal basis for the Oseledets' filtration associated with the sequence $(A_n)_{n\in\mathbb{N}}$. Then, pick an $\epsilon > 0$. Further, choose an arbitrary block length $\ell \in \mathbb{N}$ and a time horizon $T \in \mathbb{N}$ such that $T > \ell$. Also, for a fixed but arbitrary $\delta > 0$, we define

$$\Gamma_i^j := \max\left\{\max_{k\in\{0,\dots,\ell-1\}} \|\Phi_{j\ell-k}v_i\|, e^{(\lambda_i+\delta)j\ell}, e^{(\lambda_i+\delta)((j-1)\ell+1)}\right\} \tag{3.3}$$

for $i \in \{1,\dots,d\}$ and $j \in \{1,\dots,\lceil (T-1)/\ell\rceil\}$, and $\Gamma_i^0 := 1$ for $i \in \{1,\dots,d\}$.

Consider the box $B^0 := \{\sum_{i=1}^d \gamma_i v_i : \underline{\kappa}_i^0 \le \gamma_i < \overline{\kappa}_i^0\}$, where $\underline{\kappa}_i^0$ and $\overline{\kappa}_i^0$ are such that $K \subset B^0$ and $\mathrm{diam}(B^0) < \infty$. Further, consider the following sets: $\mathcal{C}_i^0 := \left\{1,\dots,\lceil d\frac{\overline{\kappa}_i^0-\underline{\kappa}_i^0}{\epsilon}\rceil\right\}$ for $i \in \{1,\dots,d\}$, and $\mathcal{C}_i^{j+1} := \left\{1,\dots,\lceil \frac{\Gamma_i^{j+1}}{\Gamma_i^j}e^{\alpha\ell}\rceil\right\}$ for $i \in \{1,\dots,d\}$ and $j \in \{0,\dots,\lceil (T-1)/\ell\rceil\}$. Now, define the set $\mathcal{Q}$ of all ordered tuples $(q^0,\dots,q^{\lceil (T-1)/\ell\rceil})$ with $q^j = (q_1^j,\dots,q_d^j)$ and $q_i^j \in \mathcal{C}_i^j$. For a given $q = (q^0,\dots,q^{\lceil (T-1)/\ell\rceil}) \in \mathcal{Q}$, we build a function $\hat{x}_q(\cdot)$ such that the value of the function at time $t \in \{0,\dots,T-1\}$, i.e., $\hat{x}_q(t)$, depends only on $(q^0,\dots,q^{\lceil t/\ell\rceil})$. Before presenting the function's construction, we consider the recursive definitions

$$\underline{\kappa}_i^{j+1}(q) := \underline{\kappa}_i^j(q) + \frac{\epsilon}{d}\left(\Gamma_i^j e^{\alpha j\ell}\right)^{-1}\left(q_i^j - 1\right), \tag{3.4}$$

$$\overline{\kappa}_i^{j+1}(q) := \underline{\kappa}_i^j(q) + \frac{\epsilon}{d}\left(\Gamma_i^j e^{\alpha j\ell}\right)^{-1} q_i^j, \tag{3.5}$$

where $i \in \{1,\dots,d\}$, $j \in \{0,\dots,\lceil (T-1)/\ell\rceil\}$, and $q \in \mathcal{Q}$.

Now, define for $j \in \{0,\dots,\lceil (T-1)/\ell\rceil\}$, $i \in \{1,\dots,d\}$, and $q \in \mathcal{Q}$ the quantity

$$\hat{\beta}_i^j(q) := \underline{\kappa}_i^j(q) + \frac{\epsilon}{d}\left(\Gamma_i^j e^{\alpha j\ell}\right)^{-1}(q_i^j - 1/2). \tag{3.6}$$

Finally, for $t \in \{0,\dots,T-1\}$ and a given $q \in \mathcal{Q}$, define the function $\hat{x}_q(t) := \sum_{i=1}^d \hat{\beta}_i^j \Phi_t v_i$, where $j = \lceil t/\ell\rceil - 1$, i.e., j is such that $j\ell + 1 \le t \le (j+1)\ell$. In words, we are using the same β_i estimate $\hat{\beta}_i^j$ for all t such that $j = \lceil t/\ell\rceil - 1$ holds true. Further note that any such t satisfies $t = (j+1)\ell - k$ for some $k \in \{0,\dots,\ell-1\}$.

Notice that, for given $q \in \mathcal{Q}$, $i \in \{1,\dots,d\}$, and $j \in \{1,\dots,\lceil (T-1)/\ell\rceil\}$ the estimate $\hat{\beta}_i^j(q)$ is the midpoint of $[\underline{\kappa}_i^{j+1}(q), \overline{\kappa}_i^{j+1}(q))$ by (3.6), (3.4), and (3.5). Also, note that for any given $\beta \in [\underline{\kappa}_i^{j+1}(q), \overline{\kappa}_i^{j+1}(q))$, we have that $|\hat{\beta}_i^j(q) - \beta| < \frac{\epsilon}{2d}(\Gamma_i^j e^{\alpha j\ell})^{-1}$ again by (3.6), (3.4), and (3.5). Now, let $\mathcal{F}_T$ be the set of functions $\hat{x}_q(\cdot)$ for $q \in \mathcal{Q}$.

We claim that $\mathcal{F}_T$ is a (T,ϵ,α,K)-approximating set. To see this, let $x \in K$, and write it as $x = \sum_{i=1}^d \beta_i v_i$. We proceed by induction over $j \in \{0,\dots,\lceil (T-1)/\ell\rceil\}$ to show that there exists a $q \in \mathcal{Q}$ such that the corresponding $\hat{\beta}_i^j(q)$ satisfies[7] $\left|\hat{\beta}_i^j(q) - \beta_i\right| < \frac{\epsilon}{2d}\left(\Gamma_i^j e^{\alpha j\ell}\right)^{-1}$. Consequently, we conclude that the corresponding $\hat{x}_q(\cdot)$ satisfies $\|\hat{x}_q(t) - \xi(x,t)\| < \epsilon e^{-\alpha t}$ for $t \in \{0,\dots,T-1\}$.

Step 0: We have that $\beta_i \in [\underline{\kappa}_i^0, \overline{\kappa}_i^0)$ for $i \in \{1,\dots,d\}$ by definition of B^0. Let $q^0 = (q_1^0,\dots,q_d^0)$, with $q_i^0 \in \mathcal{C}_i^0$, be such that $\beta_i \in [\underline{\kappa}_i^1(q), \overline{\kappa}_i^1(q))$ for every $i \in \{1,\dots,d\}$. Note that $\underline{\kappa}_i^1(q)$ and $\overline{\kappa}_i^1(q)$ depend only on $\underline{\kappa}_i^0$, $\overline{\kappa}_i^0$, and q_i^0. By equations (3.6), (3.4), and (3.5), we have that $\left|\beta_i - \hat{\beta}_i^0(q)\right| \le \frac{\epsilon}{2d}$. Thus, for any $\hat{x}_q(\cdot) \in \mathcal{F}_T$, with $q \in \mathcal{Q}$ and with q^0 as described at the beginning of this step, we have that $\|\hat{x}_q(0) - \xi(x,0)\| = \left\|\sum_{i=1}^d \left(\beta_i - \hat{\beta}_i^0(q)\right) v_i\right\| \le \frac{\epsilon}{2d}\left\|\sum_{i=1}^d v_i\right\| \le \frac{\epsilon}{2}$, where the last inequality comes from the fact that $\|v_i\| = 1$.

Step $\mathbf{j+1}$: From our induction hypothesis, $\beta_i \in [\underline{\kappa}_i^j(q), \overline{\kappa}_i^j(q))$ for $i \in \{1,\dots,d\}$. Now, let $q^j = (q_1^j,\dots,q_d^j)$ with $q_i^j \in \mathcal{C}_i^j$, be such that $\beta_i \in [\underline{\kappa}_i^{j+1}(q), \overline{\kappa}_i^{j+1}(q))$ for every $i \in \{1,\dots,d\}$. Notice that $\underline{\kappa}_i^{j+1}(q)$ and $\overline{\kappa}_i^{j+1}(q)$ depend only on $\underline{\kappa}_i^j(q)$, $\overline{\kappa}_i^j(q)$, and q_i^j. By (3.6), (3.4), and (3.5), we have that $\left|\beta_i - \hat{\beta}_i^j(q)\right| \le \frac{\epsilon}{2d}\left(\Gamma_i^j e^{\alpha j\ell}\right)^{-1}$.

[7] Notice that this is equivalent to $\beta_i \in [\underline{\kappa}_i^{j+1}(q), \overline{\kappa}_i^{j+1}(q))$.

Thus, for $(j-1)\ell+1 \leq t \leq j\ell$ and for any $\hat{x}_q(\cdot) \in \mathcal{F}_T$, with $q \in \mathcal{Q}$ and with $(q^0,\ldots,q^j)$ as the tuple inductively described here, we have that $\|\hat{x}_q(t)-\xi(x,t)\| = \left\|\sum_{i=1}^d \left(\hat{\beta}_i^j(q)-\beta_i\right)\Phi_t v_i\right\| \leq \frac{\epsilon}{2d}e^{-\alpha j\ell}\left\|\sum_{i=1}^d \frac{\Phi_t v_i}{\Gamma_i^j}\right\| \leq \frac{\epsilon}{2}e^{-\alpha t}$, where the last inequality comes from the fact that[8] $\|\Phi_t v_i\| \leq \Gamma_i^j$ and the fact that $e^{-\alpha j\ell} \leq e^{-\alpha t}$ for $t \in \{(j-1)\ell+1,\ldots,j\ell\}$. With this, we conclude the induction.[9]

Since there exists a one-to-one correspondence between elements of $\mathcal{Q}$ and $\mathcal{F}_T$, the cardinality of $\mathcal{F}_T$ is given by $\prod_{j=0}^{\lceil (T-1)/\ell\rceil}(\prod_{i=1}^d \#\mathcal{C}_i^j)$. Also, because $\mathcal{F}_T$ is a (T,ϵ,α,K)-approximating set, its cardinality is an upper bound for $s_{\text{est}}(T,\epsilon,\alpha,K)$, the minimum cardinality of any (T,ϵ,α,K)-approximating set. Therefore, we conclude that $\frac{1}{T}\log s_{\text{est}}(T,\epsilon,\alpha,K) \leq \frac{1}{T}\sum_{j=0}^{\lceil (T-1)/\ell\rceil}\sum_{i=1}^d \log\left(\#\mathcal{C}_i^j\right)$.

Recall that, by Definition 3.4 of the Lyapunov exponent, for any given $\delta > 0$, $\exists N_i \in \mathbb{N}$ such that for all $t \geq N_i$ we have that $\frac{1}{t}\log(\|\Phi_t v_i\|) \leq \lambda_i+\delta$ for a given $i \in \{1,\ldots,d\}$, from which we get that $\|\Phi_t v_i\| \leq e^{(\lambda_i+\delta)t}$ for all $t \geq N_i$. We restrict our choice of δ to be such that $\lambda_i+\delta < 0$ for all $\lambda_i < 0$ with $i \in \{1,\ldots,d\}$. However, we can choose $\delta > 0$ arbitrarily small. Now, from (3.3), we have that $\Gamma_i^j = e^{(\lambda_i+\delta)j\ell}$ if $\lambda_i \geq 0$, and $\Gamma_i^j = e^{(\lambda_i+\delta)((j-1)\ell+1)}$ if $\lambda_i < 0$, with both equalities being valid for all j such that $(j-1)\ell+1 \geq N_i$. For simplicity denote $M := \max\left\{\lceil \frac{N_i-1}{\ell}+1\rceil, i=1,\ldots,d\right\}$. Therefore, it is true that $\frac{\Gamma_i^{j+1}}{\Gamma_i^j} = e^{(\lambda_i+\delta)\ell}$ for $j \geq M$ and $i \in \{1,\ldots,d\}$. From our previous discussion, with our previously fixed δ, we have the inequality $h_{\text{est}}(\alpha,K) \leq \lim_{\epsilon\to 0}\limsup_{T\to\infty}\frac{1}{T}\sum_{j=0}^M\sum_{i=1}^d \log\left(\#\mathcal{C}_i^j\right) - \frac{(M+1)}{T}\sum_{i=1}^d \log\left(\lceil e^{(\lambda_i+\alpha+\delta)\ell}\rceil\right) + \frac{1}{\ell}\sum_{i=1}^d \log\left(\lceil e^{(\lambda_i+\alpha+\delta)\ell}\rceil\right)$, where we note that the first two terms on the right-hand side vanish when T goes to infinity. Thus, we have that $h_{\text{est}}(\alpha,K) \leq \frac{1}{\ell}\sum_{i=1}^d \log\left(\lceil e^{(\lambda_i+\alpha+\delta)\ell}\rceil\right)$. Since $\delta > 0$ can be arbitrarily small, this shows that $h_{\text{est}}(\alpha,K) \leq \frac{1}{\ell}\sum_{i=1}^d \log\left(\lceil e^{(\lambda_i+\alpha)\ell}\rceil\right)$. Finally, because ℓ can be made arbitrarily large, we get inequality (3.2). Here, we used the fact that $\lim_{\ell\to\infty}\frac{1}{\ell}\log(\lceil e^{y\ell}\rceil) = \max\{y,0\}$ for $y \in \mathbb{R}$. To see this, note that we have $\lceil e^{y\ell}\rceil = 1$ for $y \leq 0$, so $\log(\lceil e^{y\ell}\rceil) = 0$, and we have that $y \leq \frac{1}{\ell}\log(\lceil e^{y\ell}\rceil) \leq \frac{1}{\ell}\log(e^{y\ell}(1+e^{-y\ell})) = y+\frac{1}{\ell}\log(1+e^{-y\ell})$ for $y > 0$, from which we conclude that the limit equals $\max\{y,0\}$.

For the lower bound, assume that $(A_n)_{n\in\mathbb{N}}$ is regular. Let $\{v_1,\ldots,v_d\}$ be a normal basis for the Oseledets' filtration associated with $(A_n)_{n\in\mathbb{N}}$. Further, fix an arbitrary $\delta > 0$ and pick an $\epsilon > 0$. Define $\mathcal{I} := \{i \in \{1,\ldots,d\} : \lambda_i+\alpha+\delta > 0\}$, a set of indices, and $U := \{\sum_{i\in\mathcal{I}}\gamma_i v_i : \underline{\kappa}_i \leq \gamma_i \leq \overline{\kappa}_i\}$, with $\underline{\kappa}_i$ and $\overline{\kappa}_i$ such that $U \subset K$, which is always possible because K has a nonempty interior. For simplicity, assume that $\underline{\kappa}_i = 0$ and $\overline{\kappa}_i = \overline{\kappa}$ for all $i \in \mathcal{I}$. If this is not the case, translate the set K so that the origin will be in its interior, a transformation that does not change the $\#\mathcal{I}$th volume. Therefore, U is the parallelepiped $\{\kappa_i v_i : \kappa_i \in [0,\overline{\kappa}]\}$ with the $\#\mathcal{I}$th volume given by $\text{vol}(U) = (\overline{\kappa})^{\#\mathcal{I}}\text{vol}(\{v_i : i \in \mathcal{I}\})$.

Now, from the regularity hypothesis and the second bullet in Lemma 3.13, for our $\delta > 0$, $\exists N \in \mathbb{N}$ such that for all $j > N$ we have that $\left|\frac{1}{j}\log\text{vol}(\{\Phi_j v_i : i \in \mathcal{I}\}) - \sum_{i\in\mathcal{I}}\lambda_i\right| \leq \delta\#\mathcal{I}$, which implies that $\text{vol}(\{\Phi_j v_i : i \in \mathcal{I}\}) \geq e^{\sum_{i\in\mathcal{I}}(\lambda_i-\delta)j}$.

[8]This is an immediate consequence of equation (3.3).

[9]We need a minor change for the final step, i.e., $j = \lceil (T-1)/\ell\rceil$, in the induction process. Because of the domain of $\hat{x}_q(\cdot)$, we have to consider $(\lceil (T-1)/\ell\rceil - 1)\ell \leq t \leq T$ instead of $(\lceil (T-1)/\ell\rceil - 1)\ell \leq t \leq \lceil (T-1)/\ell\rceil\ell$. This is the only change needed to prove the induction.

Notice that the parallelepiped $\Phi_j U = \{\sum_{i\in\mathcal{I}} \gamma_i \Phi_j v_i : 0 \le \gamma_i \le \overline{\kappa}\}$ has the $\#\mathcal{I}$th volume equal to $\mathrm{vol}(\Phi_j U) = (\overline{\kappa})^{\#\mathcal{I}} \mathrm{vol}(\{\Phi_j v_i : i \in \mathcal{I}\})$. Now, let $C = \{x_1, \ldots, x_N\}$ be a (T, ϵ, α, U)-spanning set. We show how the cardinality of C compares with the minimum cardinality, s^*_{est}, of a (T, ϵ, α, U)-spanning set.

First, recall that $\mathbb{B}(x, r)$ is the infinity-norm ball (hypercube) centered at x with radius r. Define $\mathbb{B}^{(j,\mathcal{I})}(x, r) := \mathbb{B}(x, r) \cap \{\sum_{i\in\mathcal{I}} \gamma_i \Phi_j v_i : \gamma_i \in \mathbb{R}\}$, i.e., the intersection of the ball with the subspace spanned by the vectors $\Phi_j v_i$ for $i \in \mathcal{I}$. Now, since C is (T, ϵ, α, U)-spanning, we cover $\Phi_T U$ with balls of radius $\epsilon e^{-\alpha T}$ centered at $\Phi_T x_i$ for $x_i \in C$. Because the balls $\mathbb{B}^{(j,\mathcal{I})}(\Phi_T x_i, \epsilon e^{-\alpha T})$ cover $\Phi_T U$, the sum of their $\#\mathcal{I}$th volumes, i.e., the cardinality of C times the $\#\mathcal{I}$th volume of a single ball, is greater than or equal to the $\#\mathcal{I}$th volume of $\Phi_T U$. From this, we conclude that $\#C \ge \frac{\mathrm{vol}(\Phi_T U)}{\mathrm{vol}(\mathbb{B}(\Phi_T x_i, \epsilon e^{-T\alpha}))}$. Lastly, because $s^*_{\mathrm{est}}(T, \epsilon, \alpha, U)$ is the lowest value for the cardinality of any (T, ϵ, α, U)-spanning set, we conclude that $s^*_{\mathrm{est}}(T, \epsilon, \alpha, U) \ge \frac{\mathrm{vol}(\Phi_T U)}{\mathrm{vol}(\mathbb{B}(\Phi_T x_i, \epsilon e^{-T\alpha}))} = \left(\frac{\overline{\kappa}}{2\epsilon e^{-T\alpha}}\right)^{\#\mathcal{I}} \mathrm{vol}\left(\{\Phi_T v_i : i \in \mathcal{I}\}\right)$.

It is straightforward to see that $s^*_{\mathrm{est}}(T, \epsilon, \alpha, K) \ge s^*_{\mathrm{est}}(T, \epsilon, \alpha, U)$ by the fact that any (T, ϵ, α, K)-spanning set is also a (T, ϵ, α, U)-spanning set. Thus, we arrive at $s^*_{\mathrm{est}}(T, \epsilon, \alpha, K) \ge \left(\frac{\overline{\kappa}}{2\epsilon e^{-T\alpha}}\right)^{\#\mathcal{I}} \mathrm{vol}\left(\{\Phi_T v_i : i \in \mathcal{I}\}\right)$. Furthermore, we have that $h_{\mathrm{est}}(K, \alpha) \ge \lim_{\epsilon\to 0} \limsup_{T\to\infty} \frac{1}{T}\Big(\log\left(\frac{\overline{\kappa}}{2\epsilon}\right)^{\#\mathcal{I}} + \log \mathrm{vol}\left(\{\Phi_T v_i : i \in \mathcal{I}\}\right)\Big) + \alpha\#\mathcal{I}$, and, since T can be taken to be larger than N, we derive that $h_{\mathrm{est}}(K, \alpha) \ge \lim_{\epsilon\to 0} \limsup_{T\to\infty} \frac{1}{T}\Big(\log\left(\frac{\overline{\kappa}}{2\epsilon}\right)^{\#\mathcal{I}}\Big) + \sum_{i\in\mathcal{I}}(\lambda_i + \alpha - \delta)$, and we conclude that $h_{\mathrm{est}}(\alpha, K) \ge \sum_{i\in\mathcal{I}}(\lambda_i + \alpha - \delta) = \sum_{i=1}^{d} \max\{\lambda_i + \alpha - \delta, 0\}$, where the last equality comes from the definition of $\mathcal{I}$. Finally, by the fact that $\delta > 0$ was arbitrary, we have that $h_{\mathrm{est}}(\alpha, K) \ge \sum_{i\in\mathcal{I}} \max\{\lambda_i + \alpha, 0\}$. ∎

Example 3.15 (Example 2.1 revisited). Now, we can analyze Example 2.1 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 $h_{\mathrm{est}}(\alpha, K) \le \max\left\{\frac{1}{2}\log(0.99) + \alpha, 0\right\} + \max\{\alpha, 0\}$ with probability 1.

3.1. Connection with previous results. Some important remarks about the connections between this work and [33] must be made. The bounds obtained in that paper rely on the individual modes and their active rates. We will discuss why the above result is not a direct consequence of that work and why the geometry of the Oseledets' filtration is essential to get equality for the regular case. It is important to mention that those results concern the continuous-time model, i.e., described by (2.1). Thus, for consistency, all the examples in this remark will be in continuous-time as well.

First, we need to state some definitions to discuss those results. Denote by $\mu(A) = \lim_{t\downarrow 0} \frac{\|I+tA\|-1}{t}$ the *matrix measure* of the matrix A, and by $\mathbb{1}_p(\sigma)$ the *indicator function* of mode p, i.e., $\mathbb{1}_p(q) = 1$ if $p = q$ and $\mathbb{1}_p(q) = 0$, otherwise. Finally, define the *active rate* of mode p as $\rho_p(t) = \frac{1}{t}\int_0^t \mathbb{1}_p(\sigma(\tau))d\tau$. Next, for the sake of the discussion, we transcribe here the upper bound for the topological entropy from [33], i.e., $h_{\mathrm{est}}(0, K) \le \max\{\limsup_{t\to\infty} \sum_{p\in\Sigma} \mu(\mathcal{A}_{\sigma(t)})\rho_p(t)d, 0\}$. Second, to obtain that upper bound, the authors utilized a consequence of Coppel's inequality (see, e.g., [27]) to say that $\xi(x, t)$ is upper bounded by an exponential term that depends on the averaged sum of matrix measures. Nonetheless, it is easy to see why that bound is conservative. Take, for instance, a system with one mode, whose matrix has only one unstable eigenvalue and all others are stable, e.g., $A = \left(\begin{smallmatrix} 2 & 0 \\ 0 & -2 \end{smallmatrix}\right)$. It is easy to see that $\mu(A) \geq 2$, the unstable eigenvalue, but the entropy is 2 and the upper bound is greater than or equal to 4. With this example, we show that if we do not take into consideration the stable subspaces, we are going to overestimate the entropy value. For linear time-varying systems, the Oseledets' filtration plays a role similar to that of the eigenspaces. When we remove the vectors of the normal basis of the Oseledets' filtration that correspond to negative Lyapunov exponents, we are only looking at the directions where the system expands. In this way, we avoid the above conservative bound.

Now, for the lower bound, the authors of [33] obtain the inequality $h_{\text{est}}(0, K) \geq \{\limsup_{t\to 0} \sum_{p\in\Sigma} \operatorname{tr}(\mathcal{A}_p)\rho_p(t)\}$. For its proof, the authors used a volume counting argument just as in our Theorem 3.14. However, similar geometric reasons, as in the case of the upper bound, prevent that bound from being tight. First, notice that there are continuous-time-invariant linear systems with negative trace and positive entropy, such as $\binom{2\ \ 0}{0\ -3}$. Second, the volume form those authors take is d-dimensional. In that way, if there is a direction in the state space such that the state is contracting, the volume will shrink. In the above Theorem 3.14, we deal with that issue by removing those shrinking directions and only looking at the expansive ones. Once again, the restrictive setting of only knowing the active rate prevents the bound from being tight.

We need to stress that the knowledge of such directions, originating from the Oseledets' filtration, requires the knowledge of the entire switching signal. Therefore, the result of Theorem 3.14 requires more information about the switching signal than just the active rates used in [33]. Also, the switching signals are not restricted to being regular for the equality presented in Theorem 2 from [33], where the modes are assumed to be commuting.

Finally, regularity should be understood as the compatibility between the geometric notion of expanding directions and the notion of an expanding volume form. This is what allows us to get the identity in Theorem 3.14.

4. Quantization algorithm. In this section, we describe the quantization algorithm. This algorithm's goal is to estimate the state of system (2.2), with a desired exponential decay rate for the estimation error, using quantized measurements. The algorithm works by giving an overapproximation 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 $\mathcal{V}_j = \{v_1^j, \dots, v_d^j\}$, $j \in \mathbb{Z}_{\geq 0}$, for $\mathbb{R}^d$. Using this family, the proposed algorithm generates an overapproximation for the reachable set. Then, we show that by using a proper choice of family $(\mathcal{V}_j)_{j\in\mathbb{Z}_{\geq 0}}$ 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 $(\mathcal{V}_j)_{j\in\mathbb{Z}_{\geq 0}}$ that makes the algorithm achieve an average data-rate arbitrarily close to the estimation entropy online, assuming that the switching signal is known. Also, as in section 3, we will let $T_p > 0$ be a sampling time, and the sequence $(A_n)_{n\in\mathbb{N}}$ corresponds to[10] the exact discrete-time model of some continuous-time model described by (2.1), i.e., $A_n = \Phi(T_p n, T_p(n-1))$.

4.1. 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 $t \in \mathbb{R}_{\geq 0}$. Under the hypothesis that model (2.2) holds, the previous assumption becomes the hypothesis of knowing the sequence $(A_n)_{n\in\mathbb{N}}$. We also assume that we are given

[10] As described after (3.1), i.e., $A_n = \tilde{A}_k$ for $n = k + 1$ and $k \in \mathbb{Z}_{\geq 0}$.

an arbitrary family of orthonormal[11] 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 3.14, 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 3.14—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 $\{v_i^j\}_{i=1}^d$ from the family $\mathcal{V}^j$, build a new parallelepiped $\tilde{B}^j$ with sides parallel to the v_i^j's that contain $\bar{B}^{j-1}$. Now, we flow $\tilde{B}^j$ forward using $\Phi_{j\ell+1}$ and denote it by B^j. More precisely, 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, which corresponds to the symbol q^j, backward by $\Phi_{j\ell+1}$ and define the result to be $\bar{B}^j$. Finally, we repeat the procedure.

We emphasize that the bases $\{v_i^j\}_{i=1}^d$ with $j \in \mathbb{Z}_{\geq 0}$ 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 4.2 and Theorem 4.4 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 $j \in \mathbb{Z}_{\geq 0}$ and at the initial time $t = 0$. The idea of using the block length was borrowed from the block coding approach,[12] 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, $\epsilon > 0$ be a prescribed precision, $T_p > 0$ be the sampling time, and $\ell \in \mathbb{N}$ be the block length. Also, consider the sequence $(A_n)_{n\in\mathbb{N}}$, where[13] $A_n = \Phi(T_p n, T_p(n-1))$ and $\Phi_n = A_n \dots A_1$. Further, let $\mathcal{V}_j = \{v_1^j, \dots, v_d^j\}$, $j \in \mathbb{Z}_{\geq 0}$, be a family of orthonormal bases for $\mathbb{R}^d$. We define $\Gamma_i^0 = 1$ for all $i \in \{1, \dots, d\}$. If the system is known to be regular, we set

$$\Gamma_i^j := \max_{k\in\{0,\dots,\ell-1\}} \left\| \Phi_{j\ell-k} v_i^j \right\|; \tag{4.1}$$

otherwise, we set

$$\Gamma_i^j := \max \left\{ \max_{k\in\{0,\dots,\ell-1\}} \left\| \Phi_{j\ell-k} v_i^j \right\|, e^{T_p(\lambda_i+\delta)j\ell}, e^{T_p(\lambda_i+\delta)((j-1)\ell+1)} \right\} \tag{4.2}$$

[11] We omit the word "orthonormal" from this point onward, as the orthonormality is implied.

[12] See, e.g., Chapter 5 of [9].

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

for a prescribed $\delta > 0$ and[14] $\lambda_i := \limsup_{j\to\infty} \frac{1}{j} \log \left(||\Phi_j v_i^j|| \right)$. Also, let $\alpha \geq 0$ be the prescribed exponential decay rate for the estimation error.

Step 0:

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

- Define

$$B^0 := \left\{ \sum_{i=1}^{d} \gamma_i v_i^0 : \underline{\kappa}_i^0 \leq \gamma_i < \overline{\kappa}_i^0 \right\}, \tag{4.3}$$

  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 $q^0 = (q_1^0, \ldots, q_d^0)$, constructed as follows: Define $\mathcal{C}_i^0 := \left\{1, \ldots, \lceil d \frac{\overline{\kappa}_i^0 - \underline{\kappa}_i^0}{\epsilon} \rceil \right\}$. We define q_i^0, for every $i \in \{1, \ldots, d\}$, as the $k \in \mathcal{C}_i^0$ such that

$$\beta_i^0 \in \left[\underline{\kappa}_i^0 + \frac{\epsilon}{d}(k-1), \underline{\kappa}_i^0 + \frac{\epsilon}{d} k \right) \tag{4.4}$$

  holds true.
- Denote

$$\hat{\beta}_i^0 := \underline{\kappa}_i^0 + \frac{\epsilon}{d} \left(q_i^0 - 1/2 \right). \tag{4.5}$$

  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 ith 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 $1 \leq t \leq \ell$. Notice that we generated a box

$$\bar{B}^0 := \left\{ \sum_{k=1}^{d} \mu_k v_k^0 : \underline{\kappa}_k^0 + \frac{\epsilon}{d}(q_k^0 - 1) \leq \mu_k < \underline{\kappa}_k^0 + \frac{\epsilon}{d} q_k^0 \right\} \tag{4.6}$$

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

$$\tilde{B}^1 := \left\{ \sum_{i=1}^{d} \gamma_i v_i^1 : \underline{\kappa}_i^1 \leq \gamma_i < \overline{\kappa}_i^1 \right\}.$$

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 $\gamma_i = \langle \sum_{k=1}^d \mu_k v_k^0, v_i^1 \rangle = \sum_{k=1}^d \mu_k \langle v_k^0, v_i^1 \rangle$.

[14]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.

Hence, to find the smallest such box, we need to take

$$\underline{\kappa}_i^1 := \min\Big\{ \sum_{k=1}^{d} \mu_k \langle v_k^0, v_i^1\rangle : \underline{\kappa}_k^0 + \frac{\epsilon}{d}\big(q_k^0 - 1\big) \le \mu_k \le \underline{\kappa}_k^0 + \frac{\epsilon}{d} q_k^0, \quad k = 1, \dots, d\Big\}, \tag{4.7}$$

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

$$B^1 := \left\{ \sum_{i=1}^{d} \gamma_i \Phi_1 v_i^1 : \underline{\kappa}_i^1 \le \gamma_i < \overline{\kappa}_i^1 \right\} \tag{4.8}$$

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 $B^1 := \{\sum_{i=1}^d \gamma_i \Phi_1 v_i^1 : \underline{\kappa}_i^1 \le \gamma_i < \overline{\kappa}_i^1\}$, 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 $q^1 = (q_1^1, \dots, q_d^1)$. Define $\mathcal{C}_i^1 := \Big\{1, \dots, \lceil d\Gamma_i^1 e^{T_p \alpha \ell} \frac{\overline{\kappa}_i^1 - \underline{\kappa}_i^1}{\epsilon} \rceil\Big\}$. We define q_i^1, for every $i \in \{1, \dots, d\}$, as the $k \in \mathcal{C}_i^1$ such that

$$\beta_i^1 \in \left[\underline{\kappa}_i^1 + \frac{\epsilon}{d} \frac{e^{-T_p\alpha\ell}}{\Gamma_i^1}(k-1), \underline{\kappa}_i^1 + \frac{\epsilon}{d}\frac{e^{-T_p\alpha\ell}}{\Gamma_i^1} k \right) \tag{4.9}$$

  holds true.
- Denote by

$$\hat{\beta}_i^1 := \underline{\kappa}_i^1 + \frac{\epsilon}{d}\frac{e^{-T_p\alpha\ell}}{\Gamma_i^1}(q_i^1 - 1/2). \tag{4.10}$$

  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

$$\bar{B}^j := \Big\{ \sum_{k=1}^{d} \mu_k v_k^j : \underline{\kappa}_k^j + \frac{\epsilon}{d}\frac{e^{-T_p\alpha j\ell}}{\Gamma_k^j}(q_k^j - 1) \le \quad \mu_k < \underline{\kappa}_k^j + \frac{\epsilon}{d}\frac{e^{-T_p\alpha j\ell}}{\Gamma_k^j} q_k^j \Big\} \tag{4.11}$$

at the end of the previous step, Step j, and that $x \in \bar{B}^j$. Now, in this step we generate the smallest box aligned with the new basis $\{v_i^{j+1}\}_{i=1}^d$ 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 a manner analogous to how we obtained $\underline{\kappa}_i q$ and $\overline{\kappa}_i^1$ in Step 1. 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

$$B^{j+1} := \left\{ \sum_{i=1}^{d} \gamma_i \Phi_{j\ell+1} v_i^{j+1} : \underline{\kappa}_i^{j+1} \leq \gamma_i < \overline{\kappa}_i^{j+1} \right\}, \tag{4.12}$$

where

$$\underline{\kappa}_i^{j+1} := \min \left\{ \sum_{k=1}^{d} \mu_k \langle v_k^j, v_i^{j+1} \rangle : \underline{\kappa}_k^j + \frac{\epsilon}{d} \frac{e^{-T_p \alpha j \ell}}{\Gamma_k^j} \left(q_k^j - 1 \right) \leq \mu_k \leq \underline{\kappa}_k^j + \frac{\epsilon}{d} \frac{e^{-T_p \alpha j \ell}}{\Gamma_k^j} q_k^j, k = 1, \ldots, d \right\}, \tag{4.13}$$

and $\overline{\kappa}_i^{j+1}$ is obtained in an analogous fashion by changing min to 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 $q^{j+1} = (q_1^{j+1}, \ldots, q_d^{j+1})$. Let

$$\mathcal{C}_i^{j+1} := \left\{ 1, \ldots, \left\lceil d e^{T_p \alpha (j+1) \ell} \Gamma_i^{j+1} \frac{\overline{\kappa}_i^{j+1} - \underline{\kappa}_i^{j+1}}{\epsilon} \right\rceil \right\}.$$

We define q_i^{j+1} as the $k \in \mathcal{C}_i^{j+1}$ such that

$$\beta_i^{j+1} \in \left[\underline{\kappa}_i^{j+1} + \frac{\epsilon}{d} \frac{e^{-T_p \alpha (j+1) \ell}}{\Gamma_i^{j+1}} (k-1), \underline{\kappa}_i^{j+1} + \frac{\epsilon}{d} \frac{e^{-T_p \alpha (j+1) \ell}}{\Gamma_i^{j+1}} k \right) \tag{4.14}$$

holds true.

- Denote

$$\hat{\beta}_i^{j+1} := \underline{\kappa}_i^{j+1} + \frac{\epsilon}{d} \frac{e^{-T_p \alpha (j+1) \ell}}{\Gamma_i^{j+1}} (q_i^{j+1} - 1/2). \tag{4.15}$$

Then, our state estimate for the time instants $j\ell + 1 \leq t \leq (j+1)\ell$ is

$$\hat{x}(t) := \sum_{i=1}^{d} \hat{\beta}_i^{j+1} \Phi_t v_i^{j+1}.$$

Theorem 4.1 shows that our previous algorithm 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 4.1. *Let* $(A_n)_{n \in \mathbb{N}}$ *be a sequence of matrices that comes from the exact discretization of the system* (2.1) *with sampling time* $T_p > 0$. *Then, the algorithm from section* 4.1 *gives a sequence of estimates* $(\hat{x}(t))_{t \in \mathbb{Z}_{\geq 0}}$ *such that* $||\hat{x}(t) - \xi(x,t)|| \leq \frac{\epsilon}{2} e^{-T_p \alpha t}$. *Further, the average data-rate of the algorithm from section* 4.1 *is given by* $b = \limsup_{j \to \infty} \frac{1}{T_p t \ell} \sum_{j=0}^{t} \log \left(\# \mathcal{C}^j \right)$, *with* $\mathcal{C}^j := \prod_{i=1}^{d} \mathcal{C}_i^j$ *and* $\# \mathcal{C}^j := \prod_{i=1}^{d} \# \mathcal{C}_i^j$, *where* $\# \mathcal{C}_i^{j+1} \leq \lceil e^{T_p \alpha \ell} \frac{\Gamma_i^{j+1}}{\Gamma_i^j} \sum_{k=1}^{d} \left| \langle v_k^j, v_i^{j+1} \rangle \right| \rceil$ *for* $j \in \mathbb{Z}_{\geq 0}$ *and* $\# \mathcal{C}_i^0 \leq \lceil d \frac{\operatorname{diam}(B^0)}{\epsilon} \rceil$.

Proof.

Step 0: Recall that $|\hat{\beta}_i^0 - \beta_i^0| \leq \epsilon/2d$ by (4.4) and (4.5). Then, $||\hat{x}(0) - \xi(x,0)|| = \left\| \sum_{i=1}^d \left(\hat{\beta}_i^0 - \beta_i^0\right) v_i^0 \right\| \leq \frac{\epsilon}{2}$ and $\#\mathcal{C}_i^0 = \lceil d\frac{\overline{\kappa}_i^0 - \underline{\kappa}_i^0}{\epsilon}\rceil \leq \lceil d\frac{\operatorname{diam}(B^0)}{\epsilon}\rceil$. Finally, notice that $x \in \bar{B}^0$ by (4.6) and (4.5).

Step 1: We need to show that $\Phi_1\left(\bar{B}^0\right) = \left\{ \sum_{i=1}^d \gamma_i \Phi_1 v_i^0 : \underline{\kappa}_i^0 + \frac{\epsilon}{d}(q_i^0 - 1) \leq \gamma_i < \underline{\kappa}_i^0 + \frac{\epsilon}{d} q_i^0 \right\} \subset B^1$. Take $y \in \bar{B}^0$ and write it as $y = \sum_{k=1}^d y_k v_k^0$ and recall that $\underline{\kappa}_k^0 + \frac{\epsilon}{d}(q_k^0 - 1) \leq y_k \leq \underline{\kappa}_k^0 + \frac{\epsilon}{d} q_k^0$ for $k \in \{1, \dots, d\}$ by (4.6). Now, rewriting $y = \sum_{i=1}^d (\sum_{k=1}^d y_k \langle v_k^0, v_i^1\rangle) v_i^1$, we can check that $\underline{\kappa}_i^1 \leq (\sum_{k=1}^d y_k \langle v_k^0, v_i^1\rangle) \leq \overline{\kappa}_i^1$ by the definitions of $\underline{\kappa}_i^1$ and $\overline{\kappa}_i^1$. Thus, $\Phi_1 y \in B^1$ by (4.8). Since $y \in \bar{B}^0$ was arbitrary, we have that $\Phi_1(\bar{B}^0) \subset B^1$.

Now, we need to find an estimate for $\#\mathcal{C}_i^1$. First, let $(\underline{\gamma}_1^1, \dots, \underline{\gamma}_d^1)$ be any argument of the minimum corresponding to the minimization used to define $\underline{\kappa}_i^1$, and let $(\overline{\gamma}_1^1, \dots, \overline{\gamma}_d^1)$ be any argument of the maximum corresponding to the maximization used to define $\overline{\kappa}_i^1$. Next, notice that $\left|\overline{\kappa}_i^1 - \underline{\kappa}_i^1\right| = \left|\sum_{k=1}^d \left(\overline{\gamma}_k^1 - \underline{\gamma}_k^1\right) \langle v_k^0, v_i^1\rangle\right| \leq \frac{\epsilon}{d} \sum_{k=1}^d \left|\langle v_k^0, v_i^1\rangle\right|$ because $|\overline{\gamma}_k^1 - \underline{\gamma}_k^1| \leq \epsilon/d$ by the fact that[15] $\underline{\kappa}_i^0 + \frac{\epsilon}{d}(q_i^0 - 1) \leq \gamma_i < \underline{\kappa}_i^0 + \frac{\epsilon}{d}$ for every $i \in \{1, \dots, d\}$. Thus, we get the upper bound $\#\mathcal{C}_i^1 \leq \lceil \Gamma_i^1 e^{T_p \alpha \ell} \sum_{k=1}^d |\langle v_k^0, v_i^1\rangle| \rceil$.

Further, by (4.9) and (4.10), we have that $\left|\hat{\beta}_i^1 - \beta_i^1\right| \leq \frac{\epsilon}{2d} \frac{e^{-T_p \alpha \ell}}{\Gamma_i^1}$. Then, for $1 \leq t \leq \ell$ we have that $||\hat{x}(t) - \xi(x,t)|| = \left\|\sum_{i=1}^d \left(\hat{\beta}_i^1 - \beta_i^1\right) \Phi_t v_i^1 \right\| \leq \frac{\epsilon}{2d} e^{-T_p \alpha \ell} \left\| \sum_{i=1}^d \frac{\Phi_t v_i^1}{\Gamma_i^1} \right\| \leq \frac{\epsilon}{2} e^{-T_p \alpha t}$, where the last inequality comes from the facts that $\left\| \frac{\Phi_t v_i^1}{\Gamma_i^1} \right\| \leq 1$ and $1 \leq t \leq \ell$. Finally, notice that $x \in \bar{B}^1$ because $\sum_{i=1}^d \beta_i^1 v_i^1 \in \bar{B}^1$ by the fact that[16] $\Phi_1 \bar{B}^1 \subset B^1$ and by (4.8).

Step $\boldsymbol{j+1}$: By our induction hypothesis, we have that $x \in \bar{B}^j$. We need to show that $\Phi_{j\ell+1}\left(\bar{B}^j\right) = \left\{ \sum_{i=1}^d \gamma_i \Phi_{j\ell+1} v_i^j : \underline{\kappa}_i^j + \frac{\epsilon}{d}\frac{e^{-T_p \alpha j \ell}}{\Gamma_i^j}(q_i^j - 1) \leq \gamma_i < \underline{\kappa}_i^j + \frac{\epsilon}{d}\frac{e^{-T_p \alpha j \ell}}{\Gamma_i^j} q_i^j \right\} \subset B^{j+1}$. Take $y \in \bar{B}^j$ and write it as $y = \sum_{k=1}^d y_k v_k^j$ and recall that $\underline{\kappa}_k^j + \frac{\epsilon}{d}\frac{e^{-T_p \alpha j \ell}}{\Gamma_i^j}(q_k^j - 1) \leq y_k \leq \underline{\kappa}_k^j + \frac{\epsilon}{d}\frac{e^{-T_p \alpha j \ell}}{\Gamma_i^j} q_k^j$ for $k \in \{1, \dots, d\}$ by equation (4.11). Now, rewriting $y = \sum_{i=1}^d (\sum_{k=1}^d y_k \langle v_k^j, v_i^{j+1}\rangle) v_i^{j+1}$, we can check that $\underline{\kappa}_i^{j+1} \leq (\sum_{k=1}^d y_k \langle v_k^j, v_i^{j+1}\rangle) \leq \overline{\kappa}_i^{j+1}$ by the definitions of $\underline{\kappa}_i^{j+1}$ and $\overline{\kappa}_i^{j+1}$. Thus, $\Phi_{j\ell+1} y \in B^{j+1}$ by equation (4.12). Since $y \in \bar{B}^j$ was arbitrary, we have that $\Phi_{j\ell+1}(\bar{B}^j) \subset B^{j+1}$.

Now, we need to find an estimate for $\#\mathcal{C}_i^{j+1}$. First, let $(\underline{\gamma}_1^{j+1}, \dots, \underline{\gamma}_d^{j+1})$ be any argument of the minimum corresponding to the minimization used to define $\underline{\kappa}_i^{j+1}$, and let $(\overline{\gamma}_1^{j+1}, \dots, \overline{\gamma}_d^{j+1})$ be any argument of the maximum corresponding to the maximization used to define $\overline{\kappa}_i^{j+1}$. Next, notice that $\left|\overline{\kappa}_i^{j+1} - \underline{\kappa}_i^{j+1}\right| = \left|\sum_{k=1}^d \left(\overline{\gamma}_k^{j+1} - \underline{\gamma}_k^{j+1}\right) \langle v_k^j, v_i^{j+1}\rangle\right| \leq \frac{\epsilon}{d}\frac{e^{-T_p \alpha \ell}}{\Gamma_i^j} \sum_{k=1}^d \left|\langle v_k^j, v_i^{j+1}\rangle\right|$ because $\left|\overline{\gamma}_k^{j+1} - \underline{\gamma}_k^{j+1}\right| \leq \frac{\epsilon}{d}\frac{e^{-T_p \alpha j \ell}}{\Gamma_i^j}$ by the fact that[17] $\underline{\kappa}_i^j + \frac{\epsilon}{d}\frac{e^{-T_p \alpha j \ell}}{\Gamma_i^j}(q_i^j - 1) \leq \gamma_i < \underline{\kappa}_i^j + \frac{\epsilon}{d}\frac{e^{-T_p \alpha j \ell}}{\Gamma_i^j} q_i^j$. Thus, we arrive at the bound $\#\mathcal{C}_i^{j+1} \leq \lceil e^{T_p \alpha \ell} \frac{\Gamma_i^{j+1}}{\Gamma_i^j} \sum_{k=1}^d \left|\langle v_k^j, v_i^{j+1}\rangle\right| \rceil$.

[15]See (4.7) and the discussion below.

[16]To see this, look at compare (4.8) to (4.11) with $j = 1$.

[17]See (4.13) and the discussion below.

Further, by (4.14) and (4.15), we have the inequality $\left|\hat{\beta}_i^{j+1} - \beta_i^{j+1}\right| \leq \frac{\epsilon}{2d}\frac{e^{-T_p\alpha(j+1)\ell}}{\Gamma_i^{j+1}}$. Then, for $j\ell+1 \leq t \leq (j+1)\ell$ we have that $||\hat{x}(t) - \xi(x,t)|| = \left|\left|\sum_{i=1}^d \left(\hat{\beta}_i^{j+1} - \beta_i^{j+1}\right)\Phi_t v_i^{j+1}\right|\right| \leq \frac{\epsilon}{2d}e^{-T_p\alpha(j+1)\ell}\left|\left|\sum_{i=1}^d \frac{\Phi_t v_i^{j+1}}{\Gamma_i^{j+1}}\right|\right| \leq \frac{\epsilon}{2}e^{-T_p\alpha t}$, where the last inequality comes from the facts that[18] $\left|\left|\frac{\Phi_t v_i^{j+1}}{\Gamma_i^{j+1}}\right|\right| \leq 1$ 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 the fact that $\Phi_{j\ell+1}\bar{B}^{j+1} \subset B^{j+1}$ and by (4.12). ∎

It is important to note that if $\mathcal{V} = \{v_1, \ldots, v_d\}$ is a normal basis for the Oseledets' 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 $\sum_{k=1}^d |\langle v_k^j, v_i^{j+1}\rangle| = 1$ and $\lambda_i = \limsup_{j\to\infty}\frac{1}{j}\log\left(\left\|\Phi_j v_i^j\right\|\right) = \limsup_{j\to\infty}\frac{1}{j}\log(\|\Phi_j v_i\|)$, i.e., the 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 for all $j \geq \lceil\frac{N-1}{\ell} + 1\rceil$ 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 4.1. 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 $j \geq \lceil\frac{N-1}{\ell} + 1\rceil$ we have that $\max_{\{0,\ldots,\ell-1\}}\{\|\Phi_{j\ell-k}v_i\|\} \leq \max\{e^{T_p(\lambda_i+\delta+\eta)j\ell}, e^{T_p(\lambda_i+\delta+\eta)((j-1)\ell+1)}\}$.

Hence, as a consequence of our previous discussion and (4.2), if $\lambda_i + \delta < 0$, then we have that $\Gamma_i^j = e^{T_p(\lambda_i+\delta)((j-1)\ell+1)}$ for all $j \geq \lceil\frac{N-1}{\ell} + 1\rceil$ and all $i \in \{1, \ldots, d\}$; otherwise, we have that $\Gamma_i^j = e^{T_p(\lambda_i+\delta)j\ell}$ for all $j \geq \lceil\frac{N-1}{\ell} + 1\rceil$ and all $i \in \{1, \ldots, d\}$. Note that for $\lambda_i + \delta \geq 0$, we have that $e^{T_p(\lambda_i+\delta-\eta)j\ell} \leq \Gamma_i^j \leq e^{T_p(\lambda_i+\delta+\eta)j\ell}$ 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 4.1, we have that $\#\mathcal{C}_i^{j+1} \leq \lceil e^{T_p(\lambda_i+\alpha+\delta+2\eta)\ell}\rceil$ for all $j \geq \lceil\frac{N-1}{\ell} + 1\rceil$ and every $i \in \{1, \ldots, d\}$. We conclude, by showing that the first $\lceil\frac{N-1}{\ell} + 1\rceil + 1$ terms of the sum in the definition of b go to zero, that $\#\mathcal{C}^j \leq \prod_{i=1}^d \lceil e^{T_p(\lambda_i+\alpha+\delta+2\eta)\ell}\rceil$ for all $j \geq \lceil\frac{N-1}{\ell} + 1\rceil$, and that[19] $b \leq \frac{1}{T_p\ell}\sum_{i=1}^d \log\lceil e^{T_p(\lambda_i+\alpha+\delta+2\eta)\ell}\rceil$. Also, because η can be arbitrarily small, we have that $b \leq \frac{1}{T_p\ell}\sum_{i=1}^d \log\lceil e^{T_p(\lambda_i+\alpha+\delta)\ell}\rceil$. Finally, by choosing ℓ large enough, b can get as close to $\sum_{i=1}^d \max\{\lambda_i + \alpha + \delta, 0\}$ 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 $b \leq \frac{1}{T_p\ell}\sum_{i=1}^d \log\lceil e^{T_p(\lambda_i+\alpha)\ell}\rceil$, 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 Corollary 4.2.

COROLLARY 4.2. *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* $b \leq \frac{1}{T_p\ell}\sum_{i=1}^d \log\lceil e^{T_p(\lambda_i+\alpha)\ell}\rceil$

[18]This is implied by the defining equations (4.1) and (4.2).

[19]These steps are similar to those used in the proof of the entropy's upper bound in Theorem 3.14.

if the system is known to be regular and $b \leq \frac{1}{T_p \ell} \sum_{i=1}^{d} \log \lceil e^{T_p(\lambda_i+\alpha+\delta)\ell} \rceil$ *otherwise. Furthermore, b can be placed as close as desired to* $h_{\text{est}}(\alpha, K)$ *by choosing* ℓ *large enough in the case when the system is known to be regular; otherwise, b can be placed as close as desired to* $\sum_{i=1}^{d} \max\{\lambda_i + \alpha + \delta, 0\}$.

Remark 4.3. The algorithm in section 4.1 reconstructs the state at the end of the interval $j\ell + 1 \leq t \leq (j+1)\ell$ for $j \in \mathbb{Z}_{\geq 0}$, i.e., we need to wait until time $(j+1)\ell$ to build our estimate for a time inside this interval. If we want to build an estimate at the beginning of the interval, all we have to do is choose an arbitrary $\bar{\delta} > 0$ and redefine, for all $i \in \{1, \ldots, d\}$ and all $j \in \mathbb{Z}_{\geq 0}$, Γ_i^j as $\Gamma_i^{'j} := \|\Phi_{j\ell} v_i^j\| e^{T_p \bar{\delta}(j+1)\ell}$, if the system is known to be regular, or $\Gamma_i^{'j} := \max\{\|\Phi_{j\ell} v_i^j\|, e^{T_p(\lambda_i+\delta)j\ell}, e^{T_p(\lambda_i+\delta)((j-1)\ell+1)} e^{T_p\bar{\delta}(j+1)\ell}\}$, otherwise. The reason why this works is because, by temperedness, we have that there exists some $N \in \mathbb{N}$ such that we have $\|\Phi_t v_i^t\| \leq e^{T_p \bar{\delta}(t-j\ell)} \|\Phi_{j\ell} v_i^j\| \leq e^{T_p \bar{\delta}(j+1)\ell} \|\Phi_{j\ell} v_i^j\|$ for all $t \geq N$. This latter fact implies that $\frac{\|\Phi_t v_i^t\|}{\Gamma_i^{'j}} \leq 1$, which is all that is needed for the proof of Theorem 4.1 to hold. Further, notice that the only other place where Γ_i^j appears is in the fraction $\frac{\Gamma_i^{j+1}}{\Gamma_i^j}$ that gives our data-rate estimate. Note, however, that the data-rate analysis presented in the proof of Corollary 4.2 holds with the minor change of $\frac{\Gamma_i^{'j+1}}{\Gamma_i^{'j}} \leq e^{T_p(\lambda_i+2\eta+\bar{\delta})\ell}$ for all $i \in \{1, \ldots, d\}$ and for all $j \geq N$, where $N \in \mathbb{N}$. Since $\bar{\delta} > 0$ is arbitrary, our claim in Corollary 4.2 remains unchanged.

4.2. 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 4.2, is unrealistic. Recall that, because of the limit superior in Definition 3.4 of the 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 3.7 and 3.11 should help to make 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 of time 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.

THEOREM 4.4. *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* $\mathcal{V}_j = \{v_1^j, \ldots, v_d^j\}$ *be an orthonormal basis that diagonalizes* Q_j, *with an order on the elements of the basis induced by the order on their corresponding eigenvalues* $e^{\rho_i(j)}$. *Then the average data-rate of the algorithm from section* 4.1 *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 $\#\mathcal{C}_i^j$ for j large enough. For that purpose, we will use the upper bound obtained in Theorem 4.1. So, we need to find upper bounds or expressions for $\sum_{k=1}^{d} |\langle v_k^j v_i^{j+1} \rangle|$ 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 4.1 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 $\|Q_j v_i^j\| = e^{\rho_i(j)}$ and that $\lambda_i = \limsup_{j\to\infty} \frac{1}{j}\log\left\|\Phi_j v_i^j\right\| = \limsup_{j\to\infty} \frac{1}{j}\log\left((v_i^j)^\top \Phi_j^\top \Phi_j v_i^j\right)^{1/2} = \limsup_{j\to\infty} \frac{1}{j}\log((v_i^j)^\top Q_j^{2j} v_i^j)^{1/2} = \limsup_{j\to\infty} \rho_i(j)$, 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 3.13, 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, denote the limit of Q_j 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 $\sum_{k=1}^d |\langle v_k^j, v_i^{j+1}\rangle| \leq 1 + \eta_1$ 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 $\{v_1^j, \ldots, v_d^j\}$ 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 $\sum_{k=1}^d |\langle v_k^j, v_i^{j+1}\rangle| \leq \sqrt{d}$ for every $i \in \{1, \ldots, d\}$ always hold, even without simplicity.

Third, again because of regularity, for an arbitrary choice of $\eta_2 > 0$ such that $\lambda_i + \eta_2 < 0$ for all $\lambda_i < 0$[20] 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, $\Gamma_i^j := \max_{k\in\{0,\ldots,\ell-1\}} \|\Phi_{j\ell-k} v_i^j\| = \max_{k\in\{0,\ldots,\ell-1\}} \|e^{\rho_i(j\ell-k)}\|$. 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 $b = \limsup_{t\to\infty} \frac{1}{T_p t\ell} \sum_{j=0}^t \sum_{i=1}^d \log(\#\mathcal{C}_i^j)$. Denote $N := \max\{N_1, N_2\}$. So, for $j \geq N$ we have that $\#\mathcal{C}_i^j \leq \lceil e^{T_p(\alpha+\lambda_i+2\eta_2)\ell}(1+\eta_1)\rceil$. Further, define $M = \sum_{j=0}^{N-1}\sum_{i=1}^d \log(\#\mathcal{C}_j^i)$. We can upper-bound the average data-rate by $b \leq \limsup_{t\to\infty} \frac{1}{T_p t\ell}\left(M + \sum_{k=N}^t \sum_{i=1}^d \log(g_i)\right)$, where $g_i = \lceil e^{T_p(\alpha+\lambda_i+2\eta_2)\ell}(1+\eta_1)\rceil$.

Notice that $\log(\lceil x\rceil) \leq \max\{\log(x)+1, 0\}$. To see this, 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 $\log(\lceil e^{T_p(\alpha+\lambda_i+2\eta_2)\ell}(1+\eta_1)\rceil) \leq \max\{T_p(\alpha+\lambda_i+2\eta_2)\ell(1+\eta_1)+1, 0\}$. Thus, $b \leq \limsup_{t\to\infty} \frac{1}{T_p t\ell}\left(M + (t-N)\sum_{i=1}^d \max\left\{T_p(\alpha+\lambda_i+2\eta_2)\ell + \log(1+\eta_1)+1, 0\right\}\right)$ 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 can replace $\log(1+\eta_1)$ by $\log(\sqrt{d})$ and obtain $b \leq \sum_{i=1}^d \max\left\{\alpha+\lambda_i+\frac{\log(\sqrt{d})+1}{T_p\ell}, 0\right\}$, and therefore in both cases, by

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

choosing ℓ sufficiently large, the upper bound on b can be placed arbitrarily close to the estimation entropy $h_{\text{est}}(\alpha, K)$ as given by the last statement of Theorem 3.14. $\square$

Remark 4.5. It is important to mention what still holds without regularity and simplicity. First, it is always true that $\sum_{k=1}^{d} |\langle v_k^j, v_i^{j+1}\rangle| \le \sqrt{d}$ 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 $\frac{\Gamma_i^{j+1}}{\Gamma_i^j} \le e^{T_p(\lambda_i+\delta+2\eta_2)\ell}$ for all $j \ge N$, where $\delta > 0$ is the same that appears in the definition of Γ_i^j in algorithm 12. Furthermore, from these inequalities, we conclude that $\#\mathcal{C}_i^j \le \lceil e^{T_p(\alpha+\lambda_i+\delta+2\eta_2)\ell}\sqrt{d}\rceil$ for $j \ge N$ and $i \in \{1, \ldots, d\}$. Using this upper bound for $\#\mathcal{C}_i^j$ and following the steps of the proof above, we conclude that $b \le \sum_{i=1}^{d} \max\left\{(\alpha + \lambda_i + \delta) + \frac{\log(\sqrt{d})+1}{T_p\ell}, 0\right\}$. 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 wellknown that these λ_i's are less 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 3.14.

Furthermore, note that, without the regularity assumption, we need to have a priori knowledge of either the λ_i's or an upper bound to them. Both hypotheses 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 [28].

5. Sufficient conditions for regularity.

In this section, we show that regularity is a condition that arises naturally in many practical examples. We start by claiming that continuous-time regular systems give rise to discrete-time regular systems after sampling. Afterward, we conclude with a class of systems that preserve a given probability measure, such as ergodic Markov jump linear systems (MJLS). Some of the results in this section were stated without proof in [29]. Also, we refer the reeader to that work for results concerning the regularity and entropy of upper-triangular systems.

5.1. Sampled continuous-time regular systems.

For continuous-time systems, i.e., described by (2.1), regularity is defined in the following analogous way. Let $\lambda^c(v) := \limsup_{t\to\infty} \frac{1}{t} \log ||\Phi(t, 0)v||$, where $\Phi(t, 0)$ is the state transition matrix of system (2.1), be the Lyapunov exponent of system (2.1); see, e.g., Chapter 3 of [2]. Furthermore, let the Oseledets' filtration and Lyapunov exponents with multiplicities, λ_i^c with $i \in \{1, \ldots, d\}$, be defined analogously to Definition 3.6 by changing the definition of the Lyapunov exponent used. Then system (2.1) is regular if $\lim_{t\to\infty} \frac{1}{t} \int_0^t \operatorname{tr}(\mathcal{A}_{\sigma(\tau)})d\tau = \sum_{i=1}^{d} \lambda_i^c$. Finally, we define temperedness as $\lim_{t\to\infty} \frac{1}{t} \int_t^{t+1} \left\|\mathcal{A}_{\sigma(\tau)}\right\| d\tau = 0$. Further, an analogue of Lemma 3.13 holds for the continuous-time case; see, e.g., Chapter 4 of [2]. More specifically, in this subsection we use a consequence of the analogue of the second bullet in Lemma 3.13, i.e., for tempered and regular systems it holds that $\lambda_i^c = \lim_{t\to\infty} \frac{\log(\|\Phi(t,0)v_i\|)}{t}$, where $\{v_1, \ldots, v_d\}$ is a normal basis for the Oseledets' filtration.

Proposition 5.1 shows that the discrete-time system that originates from sampling a regular and tempered continuous-time system is regular and tempered. Therefore, regularity and temperedness are preserved after sampling.

PROPOSITION 5.1. *Consider a continuous-time switched linear system as in* (2.1). *Define* $x_n := x(T_p n)$ *and* $A_n := \Phi(nT_p, (n-1)T_p)$ *for* $n \in \mathbb{N}$, *where* $\Phi(t, 0)$ *is the fundamental matrix of* (2.1), *and* T_p *is the sampling time. If the continuous-time system is tempered and regular, then the sequence* $(A_n)_{n\in\mathbb{N}}$ *is tempered and regular.*

Proof. First, note that since $\lambda^c(v) = \lim_{t\to\infty} \frac{1}{t} \log(\|\Phi(t,0)v\|)$, we can take a subsequence $t_j = T_p j$ and conclude that $\lambda^c(v) = \lim_{j\to\infty} \frac{1}{T_p j} \log(\|\Phi_j v\|) = \frac{\lambda(v)}{T_p}$. Thus, $\lambda_i^c = \frac{\lambda_i}{T_p}$. Notice that, by Liouville's formula, we have that $\log(|\det(\Phi(t,0))|) = \int_0^t \operatorname{tr}(\mathcal{A}_{\sigma(\tau)})d\tau$. Finally, we conclude that $\sum_{i=1}^d \lambda_i = T_p \sum_{i=1}^d \lambda_i^c = T_p \lim_{t\to\infty} \frac{\log\left(|\det(\Phi(t,0))|\right)}{t} = \lim_{j\to\infty} \frac{\log\left(|\det(\Phi(T_p j,0))|\right)}{j}$. Therefore, the sampled system is regular. Now, for temperedness, notice that $\|A_n\| \le e^{\int_{(n-1)T_p}^{nT_p} \|\mathcal{A}_{\sigma(\tau)}\| d\tau}$ by the Bellman–Grönwall lemma.[21] Taking the logarithm on both sides, we get that $\frac{\log(\|A_n\|)}{nT_p} \le \frac{1}{nT_p} \int_{(n-1)T_p}^{nT_p} \|\mathcal{A}_{\sigma(\tau)}\| d\tau$, and after a change of variables and using the fact that temperedness implies $\lim_{n\to\infty} \frac{1}{n} \int_{n-1}^{n} \|\mathcal{A}_{\sigma(\tau)}\| d\tau = 0$, we get that $\limsup_{n\to\infty} \frac{\log(\|A_n\|)}{n} \le 0$. For the lower bound, note that we can apply the Bellman–Grönwall lemma to conclude that $\|A_n\| \le e^{\int_{nT_p}^{(n-1)T_p} \|\mathcal{A}_{\sigma(\tau)}\| d\tau}$ and get that $\limsup_{n\to\infty} \frac{\log(\|A_n^{-1}\|)}{n} \le 0$. Finally, we recall that $\|A_n^{-1}\| \ge \|A_n\|^{-1}$, which implies that $\liminf_{n\to\infty} \frac{\log(\|A_n\|)}{n} \ge 0$, and conclude that $\lim_{n\to\infty} \frac{\log(\|A_n\|)}{n} = 0$. ∎

5.2. Randomly switched systems. In this subsection, we study conditions under which a system with random switches will give rise to a regular sequence $(A_n)_{n\in\mathbb{N}}$. It might be helpful to keep the following example in mind. Consider the modes $\{B_1, \dots, B_m\}$ with B_i an invertible $d \times d$ matrix for $i \in \{1, \dots, m\}$. Assume that the probability of choosing mode B_i is given by p_i, i.e., $p = (p_1, \dots, p_m)$ is a probability mass function on the modes $\{B_1, \dots, B_m\}$. Repeat this process indefinitely to get a sequence $(B_{i_n})_{n\in\mathbb{N}}$. The following natural question arises: What is the probability that this sequence is regular? To answer this question, we first need some definitions.

DEFINITION 5.2 (linear cocycle [28]). *Let* $(M, \mathscr{B}, \mu)$ *be a probability space, and let* $f : M \to M$ *be a measure-preserving map. Let*[22] $L : M \to \mathrm{GL}(\mathbb{R}, d)$. *The linear cocycle defined by* L *over* f *is the transformation* $F : M \times \mathbb{R}^d \to M \times \mathbb{R}^d$ *with* $F(x, v) = (f(x), L(x)v)$. *It follows that* $F^n(x, v) = (f^n(x), L(f^n(x)) \cdots L(x)v)$ *for every* $n \ge 1$. *Moreover, if* f *is invertible, then so is* F, *with inverse* $F^{-1}(x, v) = (f^{-1}(x), (L(f^{-1}(x)))^{-1}v)$.

In this subsection, all the linear cocycles will have the following structure. Let $\mathbf{B} := \{B_1, \dots, B_m\} \subset \mathrm{GL}(d, \mathbb{R})$ be the set of modes. Further, let $M := \mathbf{B}^{\mathbb{N}}$ be the set of sequences over the modes in $\mathbf{B}$. Define $f : M \to M$ to be the shift, i.e., $f((A_n)_{n\in\mathbb{N}}) = (A_{n+1})_{n\in\mathbb{N}}$, and let $L : M \to \mathrm{GL}(d, \mathbb{R})$ be the projection to the first coordinate, i.e., $L((A_n)_{n\in\mathbb{N}}) = A_1$. We define a *cylinder of rank* k as a set of the form $[(A_n)_{n\in\mathbb{N}} : A_1 = B_{i_1}, \dots, A_k = B_{i_k}]$, where $i_j \in \{1, \dots, m\}$ and $j \in \{1, \dots, k\}$. We define $\mathscr{B}$ as the smallest σ-algebra that contains the cylinder sets of all ranks; see, e.g., section 2 of [6].

THEOREM 5.3 (Oseledets [1, 2, 28]). *Let* $(M, \mathscr{B}, \mu)$ *be a probability space, and let* $f : M \to M$ *be a measure-preserving map. Let* $L : M \to \mathrm{GL}(\mathbb{R}, d)$ *be such that*[23]

[21] See, e.g., Chapter 2 of [7].

[22] Recall that $\mathrm{GL}(\mathbb{R}, d)$ is the set of $d \times d$ invertible matrices.

[23] Here we use the notation $\log^+(x) = \max\{\log(x), 0\}$.

$\log^+ ||L|| \in L^1(\mu)$. *Also consider the linear cocycle defined by L over f. Further, denote $\Phi_n(x) = L(f^n(x)) \cdots L(x)$.*

Then, for μ-almost every $x \in M$, there are $k(x)$ positive integers, $\lambda_k(x) > \cdots > \lambda_1(x)$, and a filtration $\{0\} = E_x^1 \subsetneq \cdots \subsetneq E_x^k = \mathbb{R}^d$ such that for all $i = 1, \ldots, k(x)$,

- *$k(f(x)) = k(x)$ and $\lambda_i(f(x)) = \lambda_i(x)$ and $L(x)(E_x^i) = E_{f(x)}^i$;*
- *$\lim_{n\to\infty} \frac{1}{n} \log ||\Phi_n(x)v|| = \lambda_i(x)$ for all $v \in E_x^{i+1} \setminus E_x^i$, with $E_x^1 = \{0\}$;*
- *the $\lim_{n\to\infty} \left(\Phi_n^\top(x)\Phi_n(x)\right)^{\frac{1}{2n}}$ exists.*

Furthermore, if f is ergodic, the multiplicities $k(x)$ of the Lyapunov exponents $\lambda_i(x)$ are constant and, consequently, so are the dimensions of the subspaces E_x^i. Also, in the ergodic case, $\lambda_i(x) = \lambda_i$ is constant almost everywhere.

We denote by C the measurable set with $\mu(C) = 1$, on which this theorem is true, and call it the set of *regular realizations* with respect to μ.

Note that by the third bullet of Theorem 5.3 and the third equivalent condition in the characterization of regularity given by Lemma 3.13, the elements of the set of regular realizations are regular sequences in the previously defined sense. Therefore, the set of regular sequences has full measure for any shift-invariant probability measure over $\mathscr{B}$.

COROLLARY 5.4 (Periodically Switched Systems). *Let $(A_n)_{n\in\mathbb{N}} \subset \mathbf{B}^{\mathbb{N}}$ be such that $A_{n+T} = A_n$ for some $T \in \mathbb{N}$ and every $n \in \mathbb{N}$. Then, this sequence is regular.*

Proof. Let $\mathcal{N} \in \mathscr{B}$, $x = (A_n)_{n\in\mathbb{N}}$, and $f(x) = (A_{n+1})_{n\in\mathbb{N}}$. Define the measure $\mu(\mathcal{N}) = \frac{1}{T}\sum_{i=0}^{T-1} \delta_{f^i(x)}(\mathcal{N})$, where δ_x is a Dirac measure, i.e., $\delta_x(\mathcal{N}) = 1$ if $x \in \mathcal{N}$ and $\delta_x(\mathcal{N}) = 0$ otherwise. This measure is trivially forward invariant under the shift, and, because $\|A_n\| < \infty$, we have that $\log^+ \|L\| \in L^1(\mu)$. Therefore, we can apply Oseledets' theorem and conclude that there exists $C \in \mathscr{B}$ with $\mu(C) = 1$ such that all of its realizations are regular. Notice that $K := y \cup_{i\ge 0} f^i(x) = \cup_{i=0}^{T-1} f^i(x) \in \mathscr{B}$ and that $\mu(K) = 1$ by construction. Finally, notice that $C \cap K = K$. To see this, note that K is a finite set, and μ gives the same measure for each point of K; more specifically, $\mu(f^i(x)) = \frac{1}{n}$ for $i \in \{0, \ldots, T-1\}$. Hence, if $\#C \cap K < \#K$, we would have that $1 = \mu(C \cap K) \le \mu(K) - \frac{1}{n}$, which is a contradiction. Therefore, the sequence $(A_n)_{n\in\mathbb{N}}$ is regular. Also, notice that the Lyapunov exponents with multiplicity are constant on K. ∎

Some definitions about Markov chains are necessary for stating Corollary 5.2 next. Let $P = (p_{ij})$ be the $m \times m$ transition probability matrix of a Markov chain. Then, a *stationary distribution* of the chain $\pi^* = (\pi_1, \ldots, \pi_m)$ is defined as a solution of $\pi^{*\top} = P\pi^{*\top}$, where $\sum_{i=1}^m \pi_i^* = 1$ and $\pi_i^* \ge 0$ for all $i \in \{1, \ldots, m\}$. It is a well-known result that if the chain is irreducible and aperiodic, then the stationary distribution is unique. Further, we can use any $\pi^0 = (\pi_1^0, \ldots, \pi_m^0)$ with $\pi_i^0 \ge 0$ for $i \in \{1, \ldots, m\}$ and $\sum_{i=1}^m \pi_i^0 = 1$ and P to define a measure on the cylinder sets. So, we define the value of the measure μ on a cylinder of rank k, namely $\mathcal{N}_k = [(A_n)_{n\in\mathbb{N}} \in \mathbf{B}^{\mathbb{N}} : A_1 = B_{i_1}, \ldots, A_k = B_{i_k}]$, as

$$\mu(\mathcal{N}_k) = \pi_{i_1}^0 p_{i_1 i_2} p_{i_2 i_3} \cdots p_{i_{k-1} i_k}$$

for every cylinder of rank k and for every $k \in \mathbb{N}$. As proved by Kolmogorov, defining a measure in the cylinders is equivalent to defining a measure; see, e.g., section 24 of [6] or section 4 of chapter III of [14]. We call such a measure μ the *probability measure induced by π^0 and P.* Notice that the measure evaluated on the cylinder $\mathcal{N}_k$ is the same as the probability of the event $(B_{i_1}, \ldots, B_{i_k})$ given an initial distribution π^0 on the modes. Thus, we can think of the measure on a cylinder as being the same as the probability of seeing a sequence given an initial distribution. Finally, note that we can choose π^0 to be the stationary distribution of the chain, i.e., π^*.

COROLLARY 5.5 (Markov jump linear systems). *Let $P = (p_{ij})$ be the $m \times m$ transition matrix of an irreducible and aperiodic discrete-time, discrete-state Markov chain that represents the switching of the modes $B_i \in \mathbf{B}$. Let $\pi^{*\top} \in \mathbb{R}^d$ be the Markov chain's stationary distribution. Let $\mu^* : \mathscr{B} \to [0,1]$ be the probability measure induced by π^* and P. Then, the set of regular realizations with respect to μ^* has full probability.*

Proof. Let $\mathcal{N}_k = \{(A_{n+1})_{n\in\mathbb{N}} \in \mathbf{B}^{\mathbb{N}} : A_1 = B_{i_1}, \dots, A_k = B_{i_k}\}$ be a cylinder of rank k, and let $f((A_n)_{n\in\mathbb{N}}) = (A_{n+1})_{n\in\mathbb{N}}$ be the shift. Notice that the probability measure induced by π^* is ergodic under the shift f; see, e.g., Chapter 1 of [30] or section 24 of [6]. Because $\#\{B_1, \dots, B_m\} = m$, we have that $\log^+ \|L\| \in L^1(\mu^*)$. Therefore, we can apply Oseledets' theorem and get that the set of regular realizations C of a Markov jump linear system with an irreducible and aperiodic probability transition matrix has probability 1 under μ. Furthermore, because of the ergodicity, the Lyapunov exponents with multiplicity are constant, i.e., have the same value for any realization, in the set C. ∎

Remark 5.6. One may wonder if, given a Markov chain with an arbitrary initial distribution π^0 on the modes, the result still holds. Indeed, the result does not change. Intuitively, that is true because the distribution $\pi^0 P^n$ converges to π^*. First, notice that the Lyapunov exponents and the determinant $\det(\Phi_n)$ only depend on the tail of the sequence; i.e., if $(A_n)_{n\in\mathbb{N}}$ is regular, then $(A_{n+T})_{n\in\mathbb{N}}$ for any $T \in \mathbb{N}$ is also regular. Second, we want to prove that the measure μ_n induced by $\pi^n = \pi^0 P^n$ and P converges to the measure μ^* induced by π^* and P on the set of regular realizations C. To prove this, we show that μ^n converges to μ^* in the total variation distance, i.e., in the distance defined by $\|\mu_n - \mu^*\| = \sup_{\mathcal{B}\in\mathscr{B}} \|\mu_n(\mathcal{B}) - \mu^*(\mathcal{B})\|$. Now, notice that for a given cylinder $\mathcal{N}_k = [(A_n)_{n\in\mathbb{N}} \in \mathbf{B}^{\mathbb{N}} : A_1 = B_{i_1}, \dots, A_k = B_{i_k}]$ of rank k, we have that $\|\mu_n(\mathcal{N}_k) - \mu^*(\mathcal{N}_k)\| = \|\pi^n_{i_1} - \pi^*_{i_1}\| p_{i_1 i_2} \cdots p_{i_{k-1} i_k} \le \|\pi^n_{i_1} - \pi^*_{i_1}\|$. Therefore, the measure μ_n converges to μ^* on the cylinders and, consequently, for any measurable set. Hence, we have that $\lim_{n\to\infty} \|\mu_n - \mu^*\| = 0$. In particular, $\mu^n(C) \to 1$ from which we conclude that, with probability 1, our realizations will be regular.

Notice that Corollary 5.2 answers the question we posed at the beginning of this subsection by saying that the sequence $(B_{i_n})_{n\in\mathbb{N}}$ is regular with probability 1. In addition, now we can revisit Example 2.1 and determine the average data-rate needed for the algorithm in section 4.1 to work.

Example 5.7 (Example 2.1 revisited). Notice that by the Corollary 5.2, the realizations of the system presented in Example 2.1 are regular with probability 1. Therefore, the upper bound found in Example 3.15 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

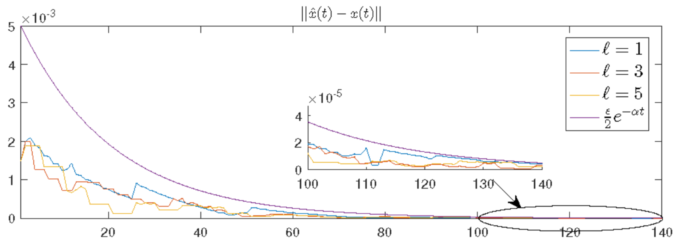

FIG. 1. *Evolution of error for several block lengths.*

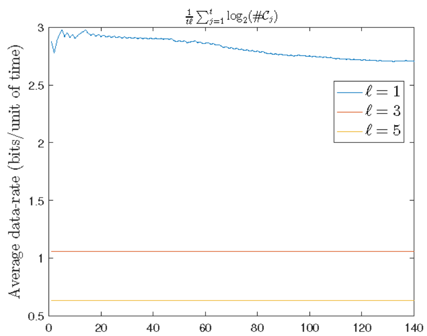

FIG. 2. *Evolution of the empirical average data-rate for several block lengths.*

upper bounded by the purple curve $\epsilon e^{-\alpha t}/2$ for all values of ℓ. Further, the empirical average data-rate, i.e., $\frac{1}{t\ell}\sum_{j=1}^{t}\log\left(\mathcal{C}_i^j\right)$, 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 4.4. That happens because the result in Theorem 4.4 is only asymptotic.

6. Conclusion and future work. In this paper, we studied how the concepts of Lyapunov exponents relate to the estimation entropy of a switched linear system. Also, we discussed how the geometric concept of Oseledets' filtration is associated with those notions. Further, we addressed the problem of finding a quantization scheme that operates close to the minimum average data-rate for regular switched linear systems. Furthermore, we showed how to adapt our algorithm to work close to the optimal data-rate, even if the underlying system is not regular. Additionally, we showed that regular switches occur in several practical conditions, including periodic switching and almost all switches that come from Markov jump linear systems. Finally, we presented simulation results.

As a future research direction, 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.

REFERENCES

[1] L. ARNOLD, *Random Dynamical Systems*, Springer, Berlin, Heidelberg, 1998.
[2] L. BARREIRA, *Lyapunov Exponents*, Birkhäuser/Springer, Cham, 2017.
[3] G. O. BERGER AND R. M. JUNGERS, *Topological entropy and minimal data rate for state observation of ltv systems*, in Proceedings of the 2020 21st IFAC World Congress, 2020.

[4] G. O. Berger and R. M. Jungers, *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, ACM, New York, 2020.

[5] G. O. Berger and R. M. Jungers, *Quantized stabilization of continuous-time switched linear systems*, IEEE Control Syst. Lett., 5 (2021), pp. 319–324.

[6] P. Billingsley, *Probability and Measure*, Wiley Ser. Probab. Math. Statist., John Wiley & Sons, New York, 1986.

[7] R. Brockett, *Finite Dimensional Linear Systems*, Classics Appl. Math. 74, SIAM, Philadelphia, 2015, https://doi.org/10.1137/1.9781611973884.

[8] F. Colonius, *Minimal bit rates and entropy for exponential stabilization*, SIAM J. Control Optim., 50 (2012), pp. 2988–3010, https://doi.org/10.1137/110829271.

[9] T. Cover and J. Thomas, *Elements of Information Theory*, Wiley Ser. Telecomm. Signal Process., John Wiley & Sons, New York, 1991.

[10] G. Yang, J. P. Hespanha, and D. Liberzon, *On topological entropy and stability of switched linear systems,* in Proceedings of the 22nd ACM International Conference on Hybrid Systems: Computation and Control, ACM, New York, 2019.

[11] J. P. Hespanha, A. Ortega, and L. Vasudevan, *Towards the control of linear systems with minimum bit-rate*, in Proceedings of the International Symposium on the Mathematical Theory of Networks and Systems, 2002.

[12] C. Kawan, *Invariance Entropy for Deterministic Control Systems*, Lecture Notes in Math. 2089, Springer, Cham, 2013.

[13] C. Kawan, *Exponential state estimation, entropy and Lyapunov exponents*, Syst. Control Lett., 113 (2018), pp. 78–85.

[14] A. Kolmogorov and N. Morrison, *Foundations of the Theory of Probability*, 2nd Eng. Ed., Dover Books Math., Dover Publications, New York, 2018.

[15] D. Liberzon and S. Mitra, *Entropy notions for state estimation and model detection with finite-data-rate measurements*, in Proceedings of the 2016 IEEE 55th Conference on Decision and Control (CDC), IEEE, Piscataway, NJ, 2016.

[16] D. Liberzon and S. Mitra, *Entropy and minimal bit rates for state estimation and model detection*, IEEE Trans. Automat. Control, 63 (2018), pp. 3330–3344.

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

[18] A. S. Matveev and A. V. Savkin, *Estimation and Control over Communication Networks*, Control Engineering, Birkhäuser Boston, Boston, MA, 2009.

[19] R. Mañé, *A proof of Pesin's formula*, Ergod. Theor. Dyn. Syst., 1 (1981), pp. 95–102.

[20] G. Nair, S. Dey, and R. J. Evans, *Communication-limited stabilisability of jump Markov linear systems,* in Proceedings of the 15th International Symposium on the Mathematical Theory of Networks and Systems, 2002.

[21] G. Nair, R. Evans, I. Mareels, and W. Moran, *Topological feedback entropy and nonlinear stabilization*, IEEE Trans. Automat. Control, 49 (2004), pp. 1585–1597.

[22] Y. Pesin, *Dimension Theory in Dynamical Systems: Contemporary Views and Applications,* Chicago Lectures in Math., University of Chicago Press, Chicago, IL, 1997.

[23] M. Rungger and M. Zamani, *Invariance feedback entropy of nondeterministic control systems*, in Proceedings of the 20th International Conference on Hybrid Systems: Computation and Control, ACM, 2017.

[24] H. Sibai and S. Mitra, *Optimal data rate for state estimation of switched nonlinear systems*, in Proceedings of the 20th International Conference on Hybrid Systems: Computation and Control, ACM, 2017.

[25] H. Sibai and S. Mitra, *State estimation of dynamical systems with unknown inputs*, in Proceedings of the 21st International Conference on Hybrid Systems: Computation and Control, ACM, 2018.

[26] A. D. Silva and C. Kawan, *Lyapunov exponents and partial hyperbolicity of chain control sets on flag manifolds*, Israel J. Math., 232 (2019), pp. 947–1000.

[27] E. D. Sontag, *Contractive systems with inputs*, in Perspectives in Mathematical System Theory, Control, and Signal Processing, Springer, Berlin, Heidelberg, 2010, pp. 217–228.

[28] M. Viana, *Lectures on Lyapunov Exponents*, Cambridge Stud. Adv. Math., Cambridge University Press, Cambridge, UK, 2014.

[29] G. S. Vicinansa and D. Liberzon, *Estimation entropy for regular linear switched systems*, in Proceedings of the 2019 IEEE 58th Conference on Decision and Control (CDC), IEEE, Piscataway, NJ, 2019.

[30] P. Walters, *An Introduction to Ergodic Theory*, Grad. Texts Math., Springer, New York, 2000.

[31] G. Yang, *Switched and Hybrid Systems with Inputs: Small-Gain Theorems, Control with Limited Information, and Topological Entropy*, Ph.D. thesis, University of Illinois at Urbana-Champaign, 2017.

[32] G. Yang, A. J. Schmidt, and D. Liberzon, *On topological entropy of switched linear systems with diagonal, triangular, and general matrices*, in Proceedings of the 2018 IEEE Conference on Decision and Control (CDC), IEEE, Piscataway, NJ, 2018.

[33] G. Yang, A. J. Schmidt, D. Liberzon, and J. P. Hespanha, *Topological entropy of switched linear systems: General matrices and matrices with commutation relations*, Math. Control Signals Systems, 32 (2020), pp. 411–453.