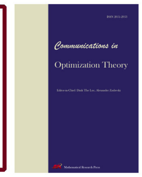
Commun. Optim. Theory 2026 (2026) 36 https://doi.org/10.23952/cot.2026.36

MATHRES

Communications in Optimization Theory

Available online at http://cot.mathres.org



ERGODIC STABILIZATION OF CONTROLLED STOCHASTIC NONLINEAR SYSTEMS UNDER INFORMATION CONSTRAINTS: GEOMETRIC ANALYSIS WITH STABLE AND UNSTABLE COORDINATE SPLITTING

NICOLÁS GARCÍA[1], CHRISTOPH KAWAN[2], SERDAR YÜKSEL[3,*]

[1]Department of Operations Research and Financial Engineering, Princeton University, USA
[2]InMach Intelligent Machines GmbH, Ulm, Germany
[3]Department of Mathematics & Statistics, Queen's University, Canada

Dedicated to the Memory of Professor Sanjoy Mitter

Abstract. This paper considers the problem of stabilizing a discrete-time non-linear stochastic dynamical system over a finite capacity noiseless channel. Asymptotic ergodicity of the state process is the stability notion considered. In this article, we first provide a review of recent results on the subject, and note that in the literature it has been established that under technical assumptions, the channel capacity must not be smaller than the logarithm of the determinant of the linearized system, averaged over the noise and ergodic state measures. In this paper, we establish that for systems with a stable subspace, it suffices to integrate over only the unstable dimensions, providing a refinement on the general data-rate bound for a large class of systems. The result is established using the notion of stabilization entropy, a notion adapted from invariance entropy, used in the study of noise-free systems under information constraints. This analysis and the associated refined bounds highlight the utility of a stochastic geometric approach when compared with information theoretic methods. A detailed comparison and reflection on these two approaches is also presented in the paper.

Keywords. Information and control; stochastic stabilization; control under communication constraints; ergodicity; invariance entropy

2020 Mathematics Subject Classification. 93E15, 93C10, 37A35.

Dedication

This paper is dedicated to the memory of Professor Sanjoy Kumar Mitter. We gratefully acknowledge the influence of Professor Mitter and his collaborators, whose insights have shaped

*Corresponding author.

E-mail address: ng6303@princeton.edu (N. García), christoph.kawan@gmx.de (C. Kawan), yuksel@queensu.ca (S. Yüksel).

Received November 30, 2024; Accepted July 29, 2025.

A preliminary conference version of this paper was presented at the 2021 IEEE Conf. Decision Control in Dec. 2021.

both the results presented here and our earlier work on control under information-theoretic constraints.

Professor Sanjoy Mitter and his collaborators were at the forefront of investigating the interaction between *information and control*. This was in fact a vibrant area of research during the mid-to-late 20th century, before information theory, control theory, and probability/dynamics diverged into essentially disjoint communities with distinct scholarly venues and discipline-specific research questions. Sanjoy Mitter was one of the few scholars who consistently studied all aspects of this interface, leaving a lasting legacy with fundamental contributions in each of these domains. The authors of this paper have been greatly influenced by this scholarship; the last author recalls that during the first days of his graduate studies he received copies of then-recent papers [1, 2, 3, 4] co-authored by Sanjoy Mitter, and Ph.D. theses of his students Anant Sahai [5] and Sekhar Tatikonda [6]. Together with [7] and [8], these works generated much enthusiasm and provided encouragement for further study. In part because of this intellectual legacy, many generations of scholars - including the authors of this paper - have studied the interaction between information, control, and probability.

1. Introduction

A commonly studied problem in the field of control under communication constraints is to characterize the minimum amount of information required by a controller to achieve a given task. In this paper, we study systems with dynamics of the form

$$x_{t+1} = f(x_t, w_t) + Bu_t, \tag{1.1}$$

and consider the control objective of rendering the $\mathbb{R}^N$-valued state process stochastically stable in the sense of (asymptotic) ergodicity. In the above display, x_t, w_t and u_t are the state, noise, and control at time t, respectively, and B is an appropriately sized matrix. Additionally, we impose that the state information travel through a finite capacity noiseless channel at each time step before reaching the controller, as depicted in Figure 1. We formalize the notion of a coding and control policy as follows. First, let $\mathcal{M} := \{1, \dots, M\}$ denote the alphabet of the channel, thus its capacity in bits is given by $C := \log_2 M$. At time t, the coder (also known as the encoder) generates a channel input q_t from past state realizations $x_0, \dots, x_t$. The channel input $q_t \in \mathcal{M}$ is therefore determined by a map $\gamma_t^e : (\mathbb{R}^N)^{t+1} \to \mathcal{M}$. The symbol q_t is transmitted over the channel, reaching the controller. The controller generates u_t based on channel outputs $q_0, \dots, q_t$ according to a map $\gamma_t^c : \mathcal{M}^{t+1} \to \mathbb{R}^N$. A coding and control policy is therefore a sequence of functions $(\gamma_t^e)_{t\in\mathbb{N}}$ and $(\gamma_t^c)_{t\in\mathbb{N}}$. Once we fix a coding and control policy, $(x_t)_{t\in\mathbb{N}}$ is a well defined autonomous stochastic process, with randomness coming from the possibly random initial state x_0, and the noise process $(w_t)_{t\in\mathbb{N}}$.

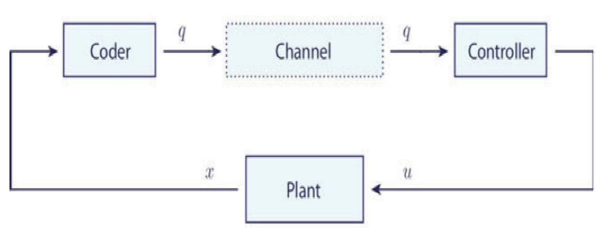


Figure 1. Control of a system over a finite capacity noiseless channel.

While our analysis will primarily focus on the case where the channel is a finite capacity noiseless channel, later on in the paper, and also during the literature review, we will additionally discuss the noisy channel case, as depicted in Figure 2. In this case, a channel input q_t leads to a channel output q'_t, and the receiver has access also to the noisy channel outputs $q'_{[0,t]}$ at time t, whereas the encoder has access to the state/action history up to time t (including x_t at time t), past channel inputs $q_{[0,t-1]}$, and past channel outputs $q'_{[0,t-1]}$, prior to the selection of the channel input q_t at time t under an encoder policy.

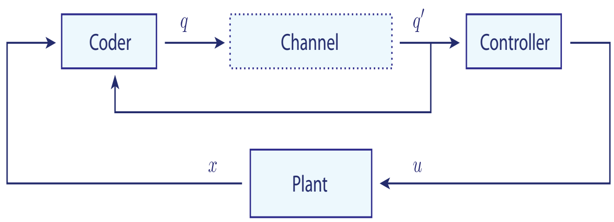


FIGURE 2. Control of a system over a noisy channel with feedback.

In this paper, we establish a necessary lower bound on the channel capacity (i.e. the data-rate) required for the existence of coding and control policies which result in asymptotic ergodicity of the state process $(x_t)_{t\in\mathbb{N}}$. The lower bound is obtained by integrating (over the unstable coordinates only) the base two logarithm of the determinant of the system linearization with respect to the noise law and (asymptotic) state law. This main result is Theorem 4.1, and requires some regularity assumptions on f and on the law of x_0 which are discussed in the theorem statement and proof.

Notation: Throughout this paper, $\mathbb{Z}$ denotes the integers, $\mathbb{R}$ the real numbers, and $\mathbb{N}$ the non-negative integers. The Lebesgue measure is denoted by m where the dimension will be clear from context. A discrete interval in the integers is denoted by $[a;b]$ (i.e., $[a;b] = \{a, a+1, \dots, b-1, b\}$ for $a \le b$ in $\mathbb{Z}$). Given a topological space $\mathcal{X}$, $\mathcal{B}(\mathcal{X})$ denotes its Borel σ-algebra. For a function $f : \mathbb{R}^n \to \mathbb{R}^m$, we denote the Jacobian (matrix of partial derivatives) by Df. For $n \in \mathbb{N}$, f^n denotes the $n-$fold composition of f with itself whenever the expression makes sense (we use the convention that f^0 is the identity). We use $\sqcup$ to emphasize that a union in question is disjoint. When applied to a set, $|\cdot|$ denotes cardinality. Given a sequence $x := (x_n)_{n\in\mathbb{N}}$, θ denotes the left shift map so that $(\theta x)_n = x_{n+1}$ for every $n \in \mathbb{N}$. Given a topological space $\mathcal{X}$, we let $\mathcal{X}^{\mathbb{N}}$ denote the set of sequences taking values in $\mathcal{X}$ and endow $\mathcal{X}^{\mathbb{N}}$ with the product topology. Given a map $f : \mathcal{X} \times \mathcal{W} \to \mathcal{X}$ we write $f_w(\cdot) := f(\cdot, w)$ to denote the resulting map from $\mathcal{X}$ to itself when $w \in \mathcal{W}$ is fixed. We write $\mu \ll \nu$ to denote that measure μ is absolutely continuous w.r.t. measure ν.

Asymptotic Ergodicity: A brief discussion on the stochastic stability notion considered throughout is in order. We begin by recalling some basic facts from ergodic theory: A measurable map $\mathcal{T} : \Omega \to \Omega$ on a probability space $(\Omega, \mathcal{F}, P)$ is called *measure-preserving* if $P(\mathcal{T}^{-1}(A)) = P(A)$ for all $A \in \mathcal{F}$. An event $A \in \mathcal{F}$ is $\mathcal{T}$*-invariant* if $A = \mathcal{T}^{-1}(A)$ (up to a set of measure zero). We denote by $\mathcal{F}_{\text{inv}(\mathcal{T})}$ the set of all $\mathcal{T}$-invariant measurable sets, which is a σ-algebra. A measure-preserving map $\mathcal{T}$ is called *ergodic* if $P(A) \in \{0, 1\}$ for all $A \in \mathcal{F}_{\text{inv}(\mathcal{T})}$. Note that ergodicity is a property of a system $(\Omega, \mathcal{F}, P, \mathcal{T})$, but sometimes we also say that "$\mathcal{T}$ is ergodic", or occasionally "P is ergodic", when the other components of the system are clear from the context. A fundamental result in ergodic theory is the following pointwise ergodic theorem.

Theorem 1.1. *(Pointwise Ergodic Theorem) Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\mathcal{T} : \Omega \to \Omega$ a measure-preserving map. Then for any $f \in L^1(\Omega, \mathcal{F}, P)$ we have*

$$\frac{1}{N}\sum_{k=0}^{N-1} f \circ \mathcal{T}^k \xrightarrow[N\to\infty]{a.s.} \varphi$$

for some $\varphi \in L^1(\Omega, \mathcal{F}_{\mathrm{inv}(\mathcal{T})}, P|_{\mathcal{F}_{\mathrm{inv}(\mathcal{T})}})$ satisfying $\int \varphi dP = \int f dP$. If, in addition, $\mathcal{T}$ is ergodic, then φ is almost everywhere constant and thus

$$\frac{1}{N}\sum_{k=0}^{N-1} f \circ \mathcal{T}^k \xrightarrow[N\to\infty]{a.s.} \int f dP.$$

We refer the reader to [9, Theorem 1.14] for a proof of this result and a detailed study of ergodic theory.

We now turn back to the setting of the paper at hand. In this context, the transformation of interest will be the left shift map, the underlying space will be the sequence space in which the controlled state process takes values in, and the measure will be the process measure on this sequence space. More explicitly, let $(\Omega, \mathcal{F}, P)$ denote the common probability space on which all random variables are defined. Fix a coding and control policy, specify an initial state law for x_0 in (1.1), and let $\mu \in \mathcal{B}((\mathbb{R}^N)^{\mathbb{N}})$ denote the resulting state process law.

Definition 1.2. We say that μ (or alternatively, the discrete-time stochastic process $(x_n)_{n\in\mathbb{N}}$ taking values in $\mathbb{R}^N$ with process law μ) is:

- *stationary* iff $\mu(\theta^{-1}(B)) = \mu(B)$ for all $B \in \mathcal{B}((\mathbb{R}^N)^{\mathbb{N}})$,
- *asymptotically mean stationary (AMS)* iff there exists a probability measure Q (called the asymptotic mean of the process) on $\mathcal{B}((\mathbb{R}^N)^{\mathbb{N}})$ such that
$$\lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1}\mu(\theta^{-k}(B)) = Q(B) \quad \text{for every } B \in \mathcal{B}((\mathbb{R}^N)^{\mathbb{N}}),$$
- *ergodic* iff it is stationary, and for $A \in \mathcal{B}((\mathbb{R}^N)^{\mathbb{N}})$ we have that $A = \theta^{-1}(A) \implies \mu(A) \in \{0,1\}$,
- *AMS ergodic* iff it is AMS, and the asymptotic mean is ergodic.

Note that if a process is *AMS*, then the asymptotic mean is a stationary measure on the sequence space. Note also that a stationary measure on $\mathcal{B}((\mathbb{R}^N)^{\mathbb{N}})$ can be unambiguously projected to a measure on $\mathcal{B}(\mathbb{R}^N)$. By slight abuse of notation, we use the same notation for a stationary measure on the sequence space and its projected coordinate measure. In this paper, the stability notion considered is *AMS ergodicity* (or informally, *asymptotic ergodicity*). Suppose that the stochastic process in Definition 1.2 is *AMS ergodic* with process measure μ and asymptotic mean Q. The application of the pointwise ergodic theorem (Theorem 1.1) to the L^1 map $\mathbb{1}_{x_0\in B} : (\mathbb{R}^N)^{\mathbb{N}} \to \{0,1\}$ for $B \in \mathcal{B}(\mathbb{R}^N)$ yields

$$\mu\Big(\Big\{x \in (\mathbb{R}^N)^{\mathbb{N}} : \lim_{T\to\infty}\frac{1}{T}\sum_{k=0}^{T-1}\mathbb{1}_B(x_k) = Q(B)\Big\}\Big) = 1. \tag{1.2}$$

In principle, Theorem 1.1 tells us that the set in (1.2) has measure one with respect to Q. From [10, Lem. 7.5 and Eq. (7.22)], it follows that μ and Q agree on all Q-trivial sets, justifying (1.2). This equation, which provides almost sure guarantees on asymptotic sample path behavior, is a crucial ingredient for the proofs in this paper. This concludes our discussion on ergodicity, and we now turn to a discussion on the motivation for the problems considered in this paper.

Problem Motivation: Suppose that system (1.1), controlled over a noiseless channel of finite capacity C, is rendered asymptotically ergodic with AMS mean Q. Under different sets of assumptions, [11] and [12] establish that

$$\int\int \log_2 |\det Df_w(x)| dQ(x) d\nu(w) \leq C \tag{1.3}$$

where ν is the distribution of the i.i.d. noise and f_w denotes the map $x \mapsto f(x,w)$ for a fixed $w \in \mathcal{W}$, where $\mathcal{W}$ is a standard Borel space. This bound, however, is in general not tight, as the following two examples illustrate.

Example 1.3. Consider the two dimensional linear system given by

$$\begin{bmatrix} x_{t+1} \\ y_{t+1} \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} x_t \\ y_t \end{bmatrix} + w_t + u_t \tag{1.4}$$

where u_t and w_t take values in $\mathbb{R}^2$ and the noise is i.i.d with zero mean. The LHS of (1.3) is easily seen to equal zero for this system, thus providing a vacuous bound on channel capacity. It is well known from the literature, however, that a tight bound on data-rate for linear system stabilization is the log-sum of the unstable eigenvalues. Note that by replacing $1/2$ in the above matrix with any number no smaller than one, the bound in (1.3) recovers the tight linear bound. As we will see, the refinement of the channel capacity bound in this paper will recover the tight bound in the general linear case (thus, also with stable eigenvalues).

Example 1.4. Consider the system $(x_t, y_t)_{t\in\mathbb{N}}$ in $\mathbb{R}^2$ evolving with scalar-valued i.i.d. noise according to

$$\begin{bmatrix} x_{t+1} \\ y_{t+1} \end{bmatrix} = \begin{bmatrix} (x_t^3 + x_t)(1 + y_t^2) \\ \frac{1}{2} y_t + w_t \end{bmatrix} + u_t \tag{1.5}$$

with x_0 and y_0 independent and admitting bounded densities. We note that the y-component of the above system is stochastically stable. Moreover, the presence of the y-term in the dynamics of the x-component cannot be modeled as noise, as the i.i.d. assumption required in data-rate theorems of non-linear systems is not satisfied. Suppose the above system is made asymptotically ergodic via a coding and control policy with AMS mean Q. We compute

$$Df_w(x,y) = \begin{bmatrix} (3x^2+1)(1+y^2) & (x^3+x)2y \\ 0 & 1/2 \end{bmatrix} \tag{1.6}$$

and apply (1.3) to obtain

$$\int \log_2 |\frac{(1+y^2)}{2}(3x^2+1)| dQ(x,y) \leq C. \tag{1.7}$$

Note that there is a factor of $1/2$ coming from the stable second component in the integrand. It seems sensible that the bound should hold without this factor, as the coding and control policy need not be concerned with the stochastically stable component. Indeed, the result in this paper establishes that the above bounds holds when removing the factor of $1/2$ and is therefore a strict refinement for certain systems. We now proceed with a literature review.

2. Literature Review and Information Requirements for Stochastic Stability

In the field of control under communication constraints, one wishes to study if and how it is possible to accomplish a control task under varying degrees of imperfect information. A ubiquitous problem in the field is to characterize minimum data rates required to stabilize a dynamical system. This problem has been considered extensively for linear deterministic and stochastic systems, for which one can usually characterize the minimum data rate required for closed-loop stability as the log-sum of the unstable open-loop eigenvalues.

Some related earlier papers considering the linear case include [2, 3, 7, 13, 14, 15, 16]. More recent contributions include [17, 18, 19, 20, 21, 22], and [23, 24, 25, 26, 27], with this latter group of resources presenting necessary and sufficient conditions for stability criteria such as existence of invariant measures, positive Harris recurrence and (asymptotic) ergodicity. There has been a separate line of work for the special Gaussian channel setup, which we do not review in this paper (see [28] for an extensive review).

For non-linear systems, the majority of papers have focused on deterministic systems. Some early works include [29], where it was established that global asymptotic stabilization of a non-linear continuous time system is feasible provided that data rates exceed a quantity related to system dimension and a Lipschitz constant, and [30] where non-linear feed-forward systems were considered. In [31] the authors presented the first systematic approach for determining minimal data rates for stabilization and introduced the notion of topological feedback entropy, a notion inspired by the classical open cover definition of topological entropy in dynamical systems due to Adler et al. [32]. It was established in [31] that a necessary and sufficient condition for stabilization to a compact set is that the data rate in the control loop exceeds the topological feedback entropy. For the same stabilization problem, invariance entropy was introduced in [33]. This notion serves as a way to quantify the difficulty of a control task through the minimum number of open loop control sequences required to achieve it. The monograph [34] provides a detailed account of the applications of invariance entropy in determining minimum data rates, particularly for continuous time (non-linear) systems. In [35], it was further established that under a strong invariance condition, the notions of topological feedback entropy and invariance entropy coincide in the discrete time case. A recent related development was the introduction of metric invariance entropy in [36]. Many more interesting results have been obtained under a wealth of settings, and we refer the reader to [37], [38], and [39] for a more detailed overview of the literature. The paper at hand builds on the techniques involving stabilization entropy, and provides a refinement for the lower bound in [11] and [12] for the stability notion of (asymptotic) ergodicity discussed in the previous section. Before proceeding with our main result, we outline specific data-rate theorems for non-linear systems in the next section.

3. Review of Information Theoretic and Stochastic Geometric Bounds

To the best of our knowledge, the first converse result on channel capacity for non-linear stochastic systems was established in [12] using information theoretic methods. The paper provided lower bounds on channel capacity necessary for stochastic stabilization of discrete time non-linear systems over both noisy and noiseless channels for stability notions of ergodicity and entropy growth conditions. With a fundamentally different approach via stochastic growth properties, a similar result was established in [11] for the ergodic case, relying on the notion of stabilization entropy. This notion, introduced in [40], was a modification of invariance entropy for the stochastic case and was first used to obtain lower bounds on channel capacity required for AMS stability. In the remainder of this section, we review both the information-theoretic and stabilization entropy results and techniques used respectively in [12] and [11].

3.1. **Information Theoretic Approach for Lower Bounds and Fundamental Limits.**

We start our review with several information theoretic bounds. In this subsection, we allow the channel to be noisy as depicted in Figure 2.

3.1.1. *Sublinear entropy growth for non-linear systems.* In this section, instead of a general $\mathbb{R}^N$-valued non-linear state model

$$x_{n+1} = f(x_n, u_n, w_n), \tag{3.1}$$

we will consider non-linear systems in either one of the following three forms:

$$x_{n+1} = f(x_n, w_n) + Bu_n, \tag{3.2}$$

$$x_{n+1} = f(x_n) + Bu_n + w_n, \tag{3.3}$$

$$x_{n+1} = f(x_n, u_n) + w_n. \tag{3.4}$$

We will also have an occasion to study non-linear systems of a fourth form:

$$x_{n+1} = f(x_n, w_n) + B(x_n)u_n. \tag{3.5}$$

In all of the models above, x_n is the $\mathbb{R}^N$-valued state, w_n is the $\mathbb{R}^N$-valued noise variable, u_n is $\mathbb{R}^s$-valued, B is appropriately sized, and w_n is assumed to be an i.i.d. process with $w_n \sim \nu$.

We assume throughout that f is measurable and continuously differentiable in the state variable. For a possibly non-linear differentiable function $f : \mathbb{R}^n \to \mathbb{R}^m$, the Jacobian matrix of f is an $m \times n$ matrix function consisting of partial derivatives of f such that

$$(Df(x))_{ij} = \frac{\partial f_i}{\partial x_j}(x), \quad 1 \le i \le m, 1 \le j \le n.$$

We have the following two assumptions.

Assumption 3.1. In the models considered above $f(\cdot, w) : \mathbb{R}^N \to \mathbb{R}^N$ is invertible for every realization of w.

In the second assumption below, $|D(f)|$ will denote the absolute value of the determinant of the Jacobian. Furthermore, with $f_w(x) = f(x, w)$, we define $D(f(x, w)) := D(f_w(x))$.

Assumption 3.2. There exist $M_1 \in \mathbb{R}$ and $L_1 \in \mathbb{R}$ so that for all x, w

$$L_1 \le \log_2(|D(f(x, w))|) \le M_1$$

The channels we consider satisfy the conditions given in the following definition.

Definition 3.3. [12, Definition 1.1] Channels are said to be of **Class A** type, if

- they satisfy the Markov chain condition:

$$q'_t \leftrightarrow q_t, q_{[0,t-1]}, q'_{[0,t-1]} \leftrightarrow \{x_0, w_t, t \ge 0\},$$

  for all $t \ge 0$, and

- their capacity with feedback is given by:

$$C = \lim_{T\to\infty} \max_{\{P(q_t|q_{[0,t-1]},q'_{[0,t-1]}),\quad 0\le t\le T-1\}} \frac{1}{T} I(q_{[0,T-1]} \to q'_{[0,T-1]}),$$

where the directed mutual information is defined by

$$I(q_{[0,T-1]} \to q'_{[0,T-1]}) = \sum_{t=1}^{T-1} I(q_{[0,t]}; q'_t | q'_{[0,t-1]}) + I(q_0; q'_0).$$

Discrete Memoryless Channels (DMCs), including noiseless channels, naturally belong to this class. We note that for DMCs, feedback from channel outputs does not increase the capacity [41]. Such a class also includes finite state stationary Markov channels which are indecomposable [42], and non-Markov channels which satisfy certain symmetry properties [43]. Further examples can be found in [44] and [45].

The result below provides conditions for sublinear entropy growth (in time) which implies quadratic stability. Let $\pi_t(B) = P(x_t \in B)$ for all Borel B.

Theorem 3.4. [12] *Let (i) f have the form in (3.2), (ii) Assumptions 3.1 and 3.2 hold, and (iii) x_0 have finite differential entropy. Then,*

a) If there is an admissible coding and control policy such that

$$\liminf_{t\to\infty} h(x_t)/t \le 0,$$

where $h(x_t)$ denotes the differential entropy of x_t, it must be that

$$C \ge \liminf_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T-1} \int \pi_t(dx) \Big(\int \nu(dw) \log_2(|D(f(x,w))|) \Big). \tag{3.6}$$

b) If $L := \inf_{x,w} \log_2 |D(f(x,w))|$, then $C \ge L$.

Remark 3.5. In the theorem, we would have arrived at the same results if we had replaced $\limsup_{t\to\infty} h(x_t)/t \le 0$ with $\limsup_{t\to\infty} \frac{1}{t} h(x_t | q'_{[0,t-1]}) \le 0$. This condition would be more relevant for state estimation problems, where the goal is not necessarily to make the state stable, but to make the estimation error stable (where u_t would be the state estimate and $x_t - u_t$ would be the estimation error). Since $h(x_t|q'_{[0,t-1]}) \le h(x_t)$, it is evident that the condition $h(x_t)/t \le 0$ implies that $h(x_t|q'_{[0,t-1]})/t \le 0$. ⋄

Remark 3.6. We note that if the system had been of the model (3.5), the expression involving $D(f(x,w))$ would explicitly depend on the control policy, which would in turn depend possibly on the entire past channel outputs, making the expression computationally more involved. ⋄

Consider now the system (3.3), under some admissible policy, controlled over a communication channel.

Assumption 3.7. Assume that

$$\begin{aligned} M &:= \sup_{x\in\mathbb{R}^N} \log_2 |D(f(x))| < \infty, \\ L &:= \inf_{x\in\mathbb{R}^N} \log_2 |D(f(x))| > -\infty. \end{aligned}$$

Lemma 3.8. [12] *Consider the system (3.3) with noisy channel with feedback and where* $h(x_0) < \infty$ *under Assumption 3.7. If* $C < L$*, under any admissible policy,*

$$\limsup_{T\to\infty} P(|x_T| \le b(T)) \le \frac{M-(L-C)}{M},$$

for all (sequences) $b(T) > 0$ *such that* $\lim_{T\to\infty} \frac{1}{T}\log_2(b(T)) = 0$.

Theorem 3.9. [12] *Consider the system (3.3) controlled over a Class A type noisy channel with feedback, and let Assumption 3.1 hold. If, under some causal encoding and controller policy, the state process is AMS, then the channel capacity* C *must satisfy* $C \ge L$.

3.1.2. *Stationarity and positive Harris recurrence under structured (stationary) policies.* In many applications, one uses a state-space formulation for coding and control policies. In the following, we will consider stationary update rules which have the form

$$\begin{aligned} q_t &= \gamma^e(x_t, m_t), \\ u_t &= \gamma^d(m_t, q'_t), \\ m_t &= \eta(m_{t-1}, q'_{t-1}), \end{aligned} \tag{3.7}$$

for functions γ^e, γ^d, and η. Here, m is an $\mathbb{S}$-valued memory or *quantizer state* variable. A large class of adaptive encoding policies have this form. This includes delta modulation, differential pulse coded modulation (DPCM), adaptive differential pulse coded modulation (ADPCM), Goodman-Gersho type adaptive quantizers (see, e.g., [46, 47]), as well as the coding schemes used for stabilization of networked control systems under fixed-rate codes [23]. Even further, jointly optimal source and channel codes for zero-delay coding schemes under infinite horizon optimization criteria also have the form above (where $\mathbb{S}$ is a space of probability measures [48]). We now present a necessary structural result on the encoders.

Now, instead of asymptotic mean stationarity, we will consider the more stringent condition of (asymptotic) stationarity of the controlled source process. For ease in presentation we will assume that m_t takes values in a countable set, even though the extension to more general spaces is possible, without substantial changes in the derivation of the results.

Lemma 3.10. *If the channel is memoryless, the process* (x_t, m_t) *is a Markov chain.*

In the following, we assume that the channel is memoryless. For the Markov chain (x_t, m_t), let $\pi_t(B) = E[1_{\{(x_t\in B)\}}]$ for all Borel B, that is, π_t is the marginal occupation probability for the state process x_t.

Theorem 3.11. [12] *Suppose that the encoding, control and the memory update laws are given by (3.7). Let (i)* f *have the form (3.2), (ii) Assumptions 3.1 and 3.2 hold, and (iii)* $h(x_0) < \infty$*. For the positive Harris recurrence of the process* (x_t, m_t) *(which implies the existence of a unique invariant measure* π *(and thus ergodicity)), it must be that*

$$C \ge \int \pi(dx)\left(\int \nu(dw)\log_2(|D(f(x,w))|)\right), \tag{3.8}$$

provided that $\limsup_{t\to\infty} \frac{1}{t} h(x_t) \le 0$.

3.2. **Stochastic and Geometric Volume Growth Approach.**

3.2.1. *Some intuition via a visit to the deterministic setup.* For deterministic nonlinear systems, *invariance entropy*, a concept introduced by Colonius and Kawan [33], measures the smallest average data rate of a noiseless channel above which a compact subset Q of the state space can be made invariant by a controller receiving its state information through this channel. The essence of the idea behind this concept is as follows: If the controller has n bits of information available, it can distinguish between at most 2^n different states, and hence generate at most 2^n different control inputs.

To explain the concept of *invariance entropy* [33] (which can be shown to be equivalent to the notion of topological feedback entropy [31]), let $\mathbb{X}$ be a topological space, $\mathbb{U}$ a non-empty set, and

$$x_{n+1} = f(x_n, u_n) \tag{3.9}$$

be a discrete-time control system on $\mathbb{X}$, where $f : \mathbb{X} \times \mathbb{U} \to \mathbb{X}$ is a map with the property that $f_u := f(\cdot, u) : \mathbb{X} \to \mathbb{X}$ is continuous for each $u \in \mathbb{U}$. We merge all solutions of (3.9) into one map $\varphi : \mathbb{N} \times \mathbb{X} \times \mathbb{U}^{\mathbb{N}} \to \mathbb{X}$,

$$\varphi(n, x, \underline{u}) := f_{u_{n-1}} \circ \cdots \circ f_{u_0}(x),$$

where $\mathbb{U}^{\mathbb{Z}_+}$ is the set of all sequences in $\mathbb{U}$ and $\underline{u} = (u_0, u_1, \ldots)$.

Let $Q \subset \mathbb{X}$ be a compact set with non-empty interior satisfying the following *strong invariance condition* (essentially, this is assumption SI in [31, p. 1586]): For every $x \in Q$ there is $u_x \in \mathbb{U}$ with $f(x, u_x) \in \mathrm{int}Q$.

For a number $\tau \in \mathbb{N}$, a set $\mathcal{S} \subset \mathbb{U}^\tau$ is called (τ, Q)-spanning if for every $x \in Q$ there is $\underline{u} \in \mathcal{S}$ such that $\varphi(j, x, \underline{u}) \in \mathrm{int}Q$ for $j = 1, \ldots, \tau$. The minimal cardinality of such a set is denoted by $r_{\mathrm{inv}}(\tau, Q)$ and the *invariance entropy* of Q is defined by

$$h_{\mathrm{inv}}(Q) := \lim_{\tau \to \infty} \frac{1}{\tau} \log r_{\mathrm{inv}}(\tau, Q), \tag{3.10}$$

where subadditivity guarantees the existence of the limit. We note also that in [35] it was proved that $h_{\mathrm{inv}}(Q) = h_{\mathrm{fb}}(Q)$, where $h_{\mathrm{fb}}(Q)$ denotes topological feedback entropy [31].

Theorem 3.12. [31] *(see also* [35]*) Consider system (3.9). Suppose that a sensor measures its states and is connected to a controller via a noiseless digital channel which carries one discrete-valued symbol per sampling interval, selected from a coding alphabet S_k of time-varying size. If the transmission data rate*

$$R = \liminf_{k \to \infty} \frac{1}{k} \sum_{j=0}^{k-1} \log |S_j|$$

of the channel satisfies $R > h_{\mathrm{inv}}(Q)$, then a coder-controller pair exists, which renders Q invariant. If $R < h_{\mathrm{inv}}(Q)$, then no such coder-controller pair exists.

An important result in the deterministic formulation is with regard to stabilization to a point: Under the assumptions that (i) f has the form in (3.9) with $\mathbb{X} = \mathbb{R}^N$, $\mathbb{U} = \mathbb{R}^m$ and has continuous partial derivatives, (ii) there exists an equilibrium pair (x^*, u^*), i.e., $x^* = f(x^*, u^*)$, (iii) a local strong invariability condition is satisfied which relates the size of an invariant set and the size of a control action set in the sense that for any $\epsilon > 0$, there exists $\rho > 0$ so that for all $\epsilon' \in (0, \rho]$, the set $\{x : |x - x^*| \leq \epsilon'\}$ is strongly invariant with the control action set $U = \{u : |u - u^*| \leq \epsilon\}$, and (iv) the pair (A, B) is controllable, where A, B are the Jacobians of f with respect to state and control at x^*, u^*, [31] has shown that for convergence to the equilibrium an average rate $R > \sum_{|\lambda_i|>1} \log(|\lambda_i|)$ is sufficient, where $\{\lambda_i\}$ are the eigenvalues of the Jacobian of f_{u^*} at x^*.

Consequently, the number of control inputs needed to achieve the control objective (on a finite time interval) is a measure for the necessary information, leading to the notion of invariance entropy (3.10) above, where in the definition $r_{\mathrm{inv}}(\tau, Q)$ is the minimal number of control inputs needed to achieve invariance of Q on the time interval $[0, \tau]$ for arbitrary initial states in Q. It is relatively immediate to observe that the growth rate of $r_{\mathrm{inv}}(\tau, Q)$ is directly related to the rate of volume expansion for subsets of Q under the evolution of the system. Indeed, the faster the volume is expanded, the more coding regions, and hence the more different control inputs, would be necessary to keep the whole volume inside Q. Since, for every reasonable stabilization objective, it is necessary to keep certain volumes bounded (or even shrink them to zero), the same ideas as used in the definition of invariance entropy should work universally for stabilization over discrete channels. This intuition was rigorously verified in [33, 35, 49, 50].

3.2.2. *The Stochastic Setup: Stability Under the AMS Criterion.* We next demonstrate that an approach similar to the one discussed above is also applicable to stochastic systems, stochastic channels, and to stochastic stability, and this can also be utilized to recover the deterministic theory as a special case.

As before, we will have two criteria: Asymptotic mean stationarity (AMS) and ergodicity. We will see that one can arrive at complementary conditions.

As an auxiliary quantity to derive lower bounds on the necessary channel capacity for generating an AMS state process, we discuss a new concept of stabilization entropy, introduced in [51] (and an ergodic theoretic generalization in Section 3.2.6 introduced in [52]), inspired by both invariance entropy and measure-theoretic entropy of dynamical systems, in particular by a characterization of the latter due to A. Katok [53] and a generalization thereof developed in [54]. Roughly speaking, stabilization entropy looks at the exponential growth rate of the number of length-n control sequences necessary to keep the state inside some set for a certain fraction of the number n of times with a certain positive probability. The corresponding set, the frequency of times, and the probability are parameters that can be adjusted, and the relation to channel capacity can only be established for certain choices of these parameters.

Recall that $X^{\mathbb{N}}$ denotes the set of all (one-sided) sequences with values in some set X and θ denotes the left-shift map on a sequence space. In this section, we will write $\bar{x} = (x_t)_{t\in\mathbb{N}}$ for elements of $X^{\mathbb{N}}$. Moreover, we write $\bar{x}_{[0,t]} = (x_0, x_1, \ldots, x_t)$ for $t \in \mathbb{N}$. We assume that all measurable spaces are standard Borel and all random variables associated with a given control system are modeled on a common (standard Borel) probability space $(\Omega, \mathcal{F}, P)$. The standard Borel space assumption leads to useful universal measurability properties which are utilized in this chapter: A measurable image of a Borel set is called an *analytic* set [55, App. 2]. We note here that this is equivalent to the seemingly more restrictive condition of being a *continuous* image of a Borel set. The following property will be utilized in our analysis: The image of a Borel set under a measurable map, and hence an analytic set, is universally measurable [55].

Consider

$$x_{t+1} = f(x_t, u_t, w_t), \quad t = 0, 1, 2, \ldots \tag{3.11}$$

This defines a measurable map $f : \mathbb{R}^N \times U \times W \to \mathbb{R}^N$, where $\mathbb{R}^N$ is endowed with the Borel σ-field $\mathcal{B}(\mathbb{R}^N)$, $(U, \mathcal{F}_U)$ is a measurable space and $(W, \mathcal{F}_W, \nu)$ a probability space. The noise is modeled by an i.i.d. sequence $(w_t)_{t\in\mathbb{N}}$ of random variables on $(W, \mathcal{F}_W)$ with associated probability measure ν. The initial state x_0 is modeled by another random variable with probability measure π_0 on $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$ and is assumed to be independent of $(w_t)_{t\in\mathbb{N}}$.

We write $\varphi(t, x_0, \bar{u}, \bar{w})$, $t \in \mathbb{N}$, for the unique trajectory with initial value $x_0 \in \mathbb{R}^N$ associated with the noise realization $\bar{w} \in W^{\mathbb{N}}$ and the control sequence $\bar{u} \in U^{\mathbb{N}}$.

We assume that an encoder, knowing the states $x_0, x_1, \dots, x_t$ at time $t \in \mathbb{N}$, transmits at time $t \in \mathbb{N}$ a symbol q_t through a noiseless discrete channel to a decoder/controller. We assume that the decoder receives the signals without delay. The finite coding alphabet is denoted by $\mathcal{M}$ and the capacity of the channel is

$$C = \log \#\mathcal{M} = \log M.$$

Thus, at time t, the controller has the symbol string $q_{[0,t]} = (q_0, q_1, \dots, q_t) \in \mathcal{M}^{t+1}$ available to generate the control input u_t. Any coding and control policy of this form is called a *causal coding and control policy*. The control objective considered is to render the state process $(x_t)_{t\in\mathbb{N}}$ asymptotically mean stationary (AMS).

3.2.3. *A New Concept: Stabilization Entropy.*

Definition 3.13. [51] For any Borel set $B \subset \mathbb{R}^N$, $T \in \mathbb{N}$ and $\rho, r \in (0,1)$, a set $S \subset U^T$ is called *(T, B, ρ, r)-spanning* if there exists a set $\tilde{\Omega} \in \mathcal{F}$ with $P(\tilde{\Omega}) \geq 1 - \rho$ so that for every $\omega \in \tilde{\Omega}$ there is $\bar{u} \in S$ with

$$\frac{1}{T}\#\{t \in [0; T-1] : \varphi(t, x_0(\omega), \bar{u}, \bar{w}(\omega)) \in B\} \geq 1 - r. \tag{3.12}$$

We write $s_B(T, \rho, r)$ to denote the smallest cardinality of a (T, B, ρ, r)-spanning set (where $s_B(T, \rho, r) = \infty$ if no finite (T, B, ρ, r)-spanning set exists) and define the *(B, ρ, r)-stabilization entropy of system* (3.11) *by*

$$h_B(\rho, r) := \limsup_{T\to\infty} \frac{1}{T} \log s_B(T, \rho, r).$$

Some remarks about this definition are now in order:

Remark 3.14. **(i):** As noted earlier, in (3.12), $\varphi(\cdot, x_0(\omega), \bar{u}, \bar{w}(\omega))$ denotes the solution of the recursion $x_{t+1} = f(x_t, u_t, w_t)$ with $x_0 = x_0(\omega)$, with u_t being the t-th element in the sequence $\bar{u}$ and $w_t = w_t(\omega)$, $\bar{w}(\omega) = (w_t(\omega))_{t\in\mathbb{N}}$. In particular, the control sequences $\bar{u}$ in the above definition are not generated by a coding and control policy. Indeed, $h_B(\rho, r)$ is an intrinsic quantity of the open-loop system.

(ii): The existence and finiteness of (T, B, ρ, r)-spanning sets are not immediately clear from the definition. However, as we will see below, in relevant cases this is guaranteed. In general, we always have $0 \leq h_B(\rho, r) \leq \infty$.

(iii): There are some immediate monotonicity properties of the function $h_B(\cdot, \cdot)$. Namely, if r or ρ become smaller, $h_B(\rho, r)$ increases. This in particular implies the existence of corresponding limits as $r \to 0$ and $\rho \to 0$ (which may be infinite).

We now present a key lemma which relates the channel capacity necessary for stabilization to the stabilization entropy. In particular, it shows that finite (T, B, ρ, r)-spanning sets exist for appropriate choices of B, ρ, r, provided that the AMS property can be achieved.

Lemma 3.15. [51] *Assume that the AMS property is achieved via a causal coding and control policy over a noiseless channel of capacity C. Then, for every Borel set $B \subset \mathbb{R}^N$ with $0 < Q(B) < 1$,* *and all sufficiently small $\varepsilon > 0$, we have*

$$C \geq h_B\left(\frac{1+\frac{\varepsilon}{2}}{1+\varepsilon}, (1+\varepsilon)Q(B^c)\right).$$

If $Q(B) = 1$, then for all $r \in (0,1)$ and $\varepsilon > 0$ sufficiently small, we have

$$C \geq h_B\left(\frac{1+\frac{\varepsilon}{2}}{1+\varepsilon}, (1+\varepsilon)r\right).$$

Lemma 3.15, while sounding technical, has significant consequences, since it allows for the application of volume-growth arguments that have been used in the literature for deterministic settings.

3.2.4. *Volume-expanding systems.* Here, we assume throughout that the measure π_0 of the random variable x_0 is absolutely continuous with respect to the Lebesgue measure m on $\mathbb{R}^N$ and that the associated density is essentially bounded.

Consider a system of the form

$$x_{t+1} = f(x_t) + u_t + w_t \tag{3.13}$$

with $U = W = \mathbb{R}^N$ and an injective C^1-map $f : \mathbb{R}^N \to \mathbb{R}^N$ satisfying

$$|\det D(f(x))| \geq 1 \quad \text{for all } x \in \mathbb{R}^N. \tag{3.14}$$

If μ, ν are two measures on the same measurable space, we write $\mu \ll_b \nu$ to denote that μ is absolutely continuous with respect to ν and its density is essentially bounded.

Theorem 3.16. [51] *Consider system (3.13) satisfying (3.14) and $\pi_0 \ll_b m$. Assume that the AMS property is achieved with an associated AMS measure Q via a causal coding and control policy over a noiseless channel of capacity C. Then for all Borel sets $B \subset \mathbb{R}^N$ with $0 < m(B) < \infty$, we have*

$$C \geq Q(B) \log \inf_{x \in B} |\det(Df(x))|. \tag{3.15}$$

Remark 3.17. The preceding theorem recovers, as a special case, Theorem 3.9, which shows that $C \geq \inf_{x \in \mathbb{R}^N} \log|\det D(f(x))|$. However, the result there is more general with regard to the allowed class of channels.

Remark 3.18. In the inequality (3.15), we see a trade-off between the Q-measure of the set B and the infimal volume growth on B. If some characteristics of the measure Q are known, one can try to optimize the lower bound by a careful choice of B. We also note that

$$\int Q(dx) \log|\det D(f(x))| \geq Q(B) \inf_{x \in B} \log|\det D(f(x))|$$

holds for all Borel sets B, where the left-hand side is the expected volume expansion with respect to the AMS measure Q. Hence, it is tempting to conjecture that also the integral above is a lower bound on the capacity. Under the stronger criterion of asymptotic ergodicity, we will show further below that this is indeed the case.

The next corollary shows that imposing further properties on the AMS measure Q can lead to more concrete bounds.

Corollary 3.19. *Consider system (3.13) satisfying (3.14) and $\pi_0 \ll_b m$. Assume that the AMS property is achieved via a noiseless channel of capacity C and the measure Q satisfies for some $M, p > 0$ the moment constraint*

$$\int Q(dx)|x|^p \leq M.$$

Then the channel capacity satisfies

$$C \geq \sup_{\kappa^p \geq M} \left(1 - \frac{M}{\kappa^p}\right) \min_{|x|\leq\kappa} \log|\det D(f(x))|. \tag{3.16}$$

The next example shows that for nonlinear systems the supremum in (3.16) is not necessarily attained as $\kappa \to \infty$, i.e., the lower bound (3.15) indeed expresses a trade-off between the measure of B and the minimal volume expansion on B.

Example 3.20. Consider a map $f : \mathbb{R} \to \mathbb{R}$ with derivative

$$f'(x) = \begin{cases} 2 & \text{if } |x| \leq 1 \\ 2^{\frac{1}{\sqrt{|x|}}} & \text{if } |x| > 1 \end{cases}$$

and note that $|f'(x)| = f'(x) > 1$ for all $x \in \mathbb{R}$. Since f' is symmetric and monotonically decreasing on $[0, \infty)$, we obtain

$$\min_{|x|\leq\kappa} \log|f'(x)| = \log|f'(\kappa)| \quad \text{for all } \kappa > 0.$$

Corollary 3.19, applied with $M = p = 1$ thus yields the capacity bound

$$C \geq \sup_{\kappa\geq 1} \left(1 - \frac{1}{\kappa}\right) \frac{1}{\sqrt{\kappa}}.$$

A straightforward analysis shows that this supremum is attained as a maximum at $\kappa = 3$, and hence $C \geq 2/(3\sqrt{3})$. $\diamond$

3.2.5. *The noisy channel case.* For discrete noiseless channels, the key idea combining the volume-growth based approaches for deterministic models with the stochastic system setup was the observation that the number of control sequences is bounded from above by the total number of received messages. This approach clearly does not directly apply to the noisy channel setup, for there can be an arbitrarily large number of possibly distinct received channel outputs, but these may not carry reliable information. In the following, we develop a new method to address this for a discrete memoryless channel (DMC).

The results in this section apply for channels with feedback for data transmission from the encoder to the controller, as depicted in [52, Fig 1]. The channel has a finite input alphabet $\mathcal{M}$ and a finite output alphabet $\mathcal{M}'$. The channel input q_t at time t is generated by a function γ_t^e so that $q_t = \gamma_t^e(x_{[0,t]}, q'_{[0,t-1]})$. The channel maps q_t to q'_t in a stochastic fashion so that $P(q'_t \in \cdot | q_t, q_{[0,t-1]}, q'_{[0,t-1]}) = P(q'_t \in \cdot | q_t)$ is a conditional probability measure on $\mathcal{M}'$ for all $t \in \mathbb{N}$, for every realization $q_t, q_{[0,t-1]}, q'_{[0,t-1]}$. The controller, upon receiving the information from the channel, generates its decision at time t, also causally: $u_t = \gamma_t^c(q'_{[0,t]})$.

Consider a DMC with channel capacity C. The following property, known as the *strong converse*, holds, see [56], [57, Problem 10.17]: For any $R > C$, under any coding policy:

$$\lim_{T\to\infty} p_e(T) = 1, \tag{3.17}$$

where $p_e(T)$ is the average probability of error among 2^{RT} equally likely messages after the channel is used T times under coding and decoding policies admissible according to the standard information-theoretic formulation of communication with noiseless feedback, cf. [58].

Now we consider a scalar system of the form

$$x_{t+1} = f(x_t) + u_t + w_t \tag{3.18}$$

with a C^1-function $f : \mathbb{R} \to \mathbb{R}$ satisfying

$$|f'(x)| \geq 1 \quad \text{for all } x \in \mathbb{R}. \tag{3.19}$$

Theorem 3.21. [51] *Consider system (3.18) satisfying (3.19). Assume that $\pi_0 \ll m$ with p denoting the density with respect to m, that $K := \mathrm{supp}(\pi_0)$ is a compact interval, and*

$$p_{\min} := \operatorname*{ess\,inf}_{x \in K} p(x) > 0, \qquad p_{\max} := \operatorname*{ess\,sup}_{x \in K} p(x) < \infty.$$

Then, if the AMS property is achieved via a causal coding and control strategy over a DMC of capacity C, we have

$$C \geq \inf_{x \in \mathbb{R}} \log |f'(x)|.$$

We next state the following variation where the initial measure may have non-compact support.

Theorem 3.22. [51] *Consider system (3.18) satisfying (3.19). Assume that $\pi_0 \ll m$ with p denoting the density with respect to m, and that for every $\epsilon > 0$, there exists a compact interval K_ϵ such that, $\pi_0(K_\epsilon) \geq 1 - \epsilon$, and with*

$$p^K_{\min} := \operatorname*{ess\,inf}_{x \in K} p(x) > 0, \qquad p^K_{\max} := \operatorname*{ess\,sup}_{x \in K} p(x) < \infty,$$

the following condition holds:

$$\lim_{\epsilon \to 0} \frac{\int_{\mathbb{R} \setminus K_\epsilon} p(x)dx}{p^{K_\epsilon}_{\min}} = 0. \tag{3.20}$$

Then, if the AMS property is achieved via a causal coding and control strategy over a DMC of capacity C, we have

$$C \geq \inf_{x \in \mathbb{R}} \log |f'(x)|.$$

Remark 3.23. A sufficient condition for (3.20) is that p is differentiable, positive everywhere, and monotone decreasing in either direction as $|x|$ increases for sufficiently large values of $|x|$, and $\lim_{|x| \to \infty} p'(x)/p(x) = \infty$. This follows from an application of L'Hospital's theorem to the expression

$$\lim_{x \to \infty} \frac{\int_{|s| > x} p(s)ds}{\min(p(x), p(-x))}.$$

Probability densities which decay faster than an exponential (such as the Gaussian) satisfy this condition. An exponential density (if one-sided, the denominator will just be $p(x)$) keeps this ratio a constant as $|x|$ increases and densities with a heavier tail than an exponential do not satisfy this condition.

3.2.6. *Stability under the Ergodicity Criterion via the Stochastic Geometric Method.* In this section, we consider the criterion that the closed-loop process is AMS Ergodic (see Definition 1.2).

Consider the system

$$x_{t+1} = f(x_t, w_t) + u_t \tag{3.21}$$

where x_t and u_t are $\mathbb{R}^N$-valued for some $N \in \mathbb{N}$ and w_t takes values in a standard probability space $\mathbb{W}$. Recall that for a fixed $w \in \mathbb{W}$, the map $x \mapsto f(x,w)$ is denoted by f_w. Suppose also that the following holds:

- (A1) The map $f : \mathbb{R}^N \times \mathbb{W} \to \mathbb{R}^N$ is Borel measurable.
- (A2) The noise process $(w_t)_{t\in\mathbb{N}}$ is i.i.d. By abuse of notation, ν denotes both the law of any individual w_t, as well as the process measure.
- (A3) The map $f_w : \mathbb{R}^N \to \mathbb{R}^N$ is C^1 and injective for any $w \in \mathbb{W}$.
- (A4) The initial state x_0 is random and independent of the noise process. We write π_0 for the associated probability measure.
- (A5) The measure π_0 is absolutely continuous with respect to the N-dimensional Lebesgue measure m, and its density (which exists by the Radon-Nikodym theorem) is bounded.
- (A6) There is a constant $c > 0$ with $|\det D(f_w(x))| > c$ for all $x \in \mathbb{R}^N$ and $w \in \mathbb{W}$.

We write $(\Omega, \mathcal{F}, P)$ for the probability space on which both x_0 and w_t are defined.

Theorem 3.24. [52] *Consider system (4.2) satisfying assumptions (A1)–(A6). Suppose the system is controlled over a discrete noiseless channel with capacity C and a coding and control policy achieves that the state process is AMS ergodic with asymptotic mean Q. Then, the capacity must satisfy*

$$\int\int \log|\det D(f_w(x))|\, Q(dx)\nu(dw) \le C.$$

Our second main theorem relaxes the condition of the channel being noiseless. On the other hand, the class of nonlinear systems considered is more restrictive.

Theorem 3.25. *Consider the scalar system*

$$x_{t+1} = f(x_t, w_t) + u_t$$

satisfying assumptions (A1)–(A5). Additionally, suppose that the following holds:

1. *$|f'_w(x)| \ge 1$ for every $x \in \mathbb{R}$.*
2. *The support of π_0 is a compact interval $K \subseteq \mathbb{R}$.*
3. *The essential infimum and supremum of the density of π_0, denoted by $\rho_{\min}$ and $\rho_{\max}$, respectively, satisfy $0 < \rho_{\min} \le \rho_{\max} < \infty$.*

Suppose that the system is controlled over a discrete memoryless channel with feedback of capacity C and a causal coding and control policy results in the state process being AMS ergodic with asymptotic mean Q. Then, the channel capacity must satisfy

$$\int\int \log|f'_w(x)| Q(dx)\nu(dw) \le C. \tag{3.22}$$

The methods involved in the proof of Theorem 3.25 rely on a multi-set stabilization entropy definition, the observation that ergodicity provides almost-surely exact asymptotic rates of visits to subsets of the state space, and volume-growth combinatorial arguments. These techniques and notions are used in the proof of Theorem 4.1 (the main result of this paper) and are discussed in detail in the following section.

4. Refined Information Rate Bounds via Coordinate Splitting: A Utility of Geometric Analysis

4.1. **Main Result.** Consider a subset $p \subseteq \{1, \dots, N\}$ of indices listed in increasing order as $p_1 < p_2 < \cdots < p_{|p|}$. Let $z_1 < \cdots < z_{N-|p|}$ denote the elements in $\{1, .., N\} \setminus p$. We define the permutation $\psi_p : \mathbb{R}^N \to \mathbb{R}^N$ by

$$\psi_p(x_1, ..., x_N)_i = \begin{cases} x_{p_i} & i \le |p| \\ x_{z_{i-|p|}} & i > |p| \end{cases}$$

for $i \in \{1, .., N\}$. Also, let $\pi_p : \mathbb{R}^N \to \mathbb{R}^{|p|}$ denote the natural projection of coordinates $p_1, .., p_{|p|}$. For a map $f : \mathbb{R}^N \to \mathbb{R}^N$, a set p as above, and a fixed vector $(y_1, .., y_{N-|p|})$ we define the map $f^p(\cdot, y_1, .., y_{N-|p|}) : \mathbb{R}^{|p|} \to \mathbb{R}^{|p|}$ by

$$f^p(x, y_1, .., y_{N-|p|}) := \pi_p(f(\psi_p^{-1}(x, y_1, .., y_{N-|p|}))) \tag{4.1}$$

where $x \in \mathbb{R}^{|p|}$. As an example, consider $n = 4, p = \{2, 4\}$, a fixed vector (y_1, y_2), and a function $f : \mathbb{R}^4 \to \mathbb{R}^4$ written as $f = (f_1, f_2, f_3, f_4)$ for maps $f_i : \mathbb{R}^4 \to R$. Then

$$f^p(x_1, x_2, y_1, y_2) = (f_2(y_1, x_1, y_2, x_2), f_4(y_1, x_1, y_2, x_2)).$$

This notation allows us to precisely state our main result. Consider the system

$$x_{t+1} = f(x_t, w_t) + Bu_t \tag{4.2}$$

where x_t is $\mathbb{R}^N$-valued for some $N \in \mathbb{N}$, $B \in \mathbb{R}^{N \times N'}$, u_t is $\mathbb{R}^{N'}$-valued, and w_t takes values in a standard probability space $\mathcal{W}$. For a fixed $w \in \mathcal{W}$, let us denote the map $x \mapsto f(x, w)$ by f_w. Suppose that the following holds:

(i) The state evolution map f is Borel measurable.
(ii) The noise process $(w_t)_{t \in \mathbb{N}}$ is i.i.d. By abuse of notation, ν denotes both the i.i.d. measure on $\mathcal{B}(\mathcal{W})$ and the noise process measure on $\mathcal{B}(\mathcal{W}^{\mathbb{N}})$.
(iii) The map $f_w(\cdot) : \mathbb{R}^N \to \mathbb{R}^N$ is C^1 and injective for any $w \in \mathcal{W}$.
(iv) The initial state $x_0 \in \mathbb{R}^N$ is random and independent of the noise process, and its law π_0 admits a bounded density.
(v) The set $\Gamma = \Big\{ p \subseteq \{1, \dots, N\} : \exists c_p > 0$ such that $|\det D f_w^p(x_{p_1}, \dots, x_{p_{|p|}}, x_{z_1}, \dots, x_{z_{N-|p|}})| > c_p, \forall x \in \mathbb{R}^N, w \in \mathcal{W} \Big\}$ is non-empty.
(vi) For $\{t_1, \dots, t_s\} \in \Gamma$ and writing the random initial state as $x_0 = (x_0^1, \dots, x_0^N)$, the law of $(x_0^{t_1}, x_0^{t_2}, \dots, x_0^{t_s})$ admits a bounded density when conditioned on some possible realization of the remaining initial state components.

The following is our main theorem. The proof is given in the following section.

Theorem 4.1. *Consider system (4.2) satisfying assumptions (i)–(vi), controlled over a noiseless channel with finite alphabet $\mathcal{M}$ and capacity $C := \log_2 |\mathcal{M}|$. if there exists a coding and* *control policy which renders the state process $(x_t)_{t\in\mathbb{N}}$ AMS ergodic (asymptotically ergodic) with asymptotic mean Q, then we must have that*

$$\max_{p\in\Gamma}\int\int \log|\det Df_w^p(x_{p_1},..,x_{p_{|p|}},x_{z_1},..,x_{z_{N-|p|}})| \\ dQ(x_1,..,x_N)d\nu(w)\le C, \tag{4.3}$$

where the Jacobian above is the $|p|\times|p|$ matrix of partial derivatives of $f_w^p(\cdot,x_{z_1},..,x_{z_{N-|p|}})$ evaluated at $(x_{p_1},..,x_{p_{|p|}})\in\mathbb{R}^{|p|}$.

Remark 4.2. Observe that by taking $p=\{1,..,M\}$ (if $\{1,..,M\}\in\Gamma$), we recover the bound (1.3) established previously in [11] and [12]. For a large class of systems however, it is clear that Theorem 4.1 is a strict refinement, as can be seen by noting that in Example 1.4, taking $p=\{1\}$ recovers the sharper bound

$$\int \log_2|(1+y^2)(3x^2+1)|dQ(x,y)\le C. \tag{4.4}$$

It is clear that for linear system, the new bound recovers the tight linear bound such as in Example 1.3.

Remark 4.3. In one dimension, (v) is the requirement that the absolute value of the derivative be bounded away from zero. This implies that the Lebesgue measure $m(f(B))$ of the image under f of a set $B\subseteq\mathbb{R}$ can be lower bounded in terms of its own Lebesgue measure, as $m(f(B))\ge m(B)c$ for some $c>0$. In higher dimensions, (v) amounts to the requirement that, along some subset of the coordinates (and uniformly over the remaining inputs) the volume of the image of a set under the restriction of f to the coordinate subset can be lower bounded proportionality to the volume of the set itself.

Remark 4.4. Note that the technical assumption (vi) is satisfied if the initial state has independent components each admitting a bounded density. Suppose now that for a given system, the assumptions (i-v) of Theorem 4.1 are satisfied, but (vi) only holds for certain subsets of Γ. Then the theorem will still hold, however the max in (4.3) should be taken only over subsets of Γ for which the assumption (vi) holds. This last observation will become clear from the proof.

Remark 4.5. Noting that the Jacobian determinant is invariant under a linear change of coordinates, we note that the bound in (4.1) is invariant under a linear change of coordinates. Note however that assumption (v) is not coordinate independent; it is not hard to see that for certain systems, the choice of coordinates may result in a different (or empty) set Γ. Under a non-linear coordinate change, it is not clear if the above bound is invariant (or if control even remains additive), thus a possible future research direction is to consider the problem of optimizing the coordinate system chosen in order to maximize the bound.

4.2. **Proof of Theorem 4.1.** We first fix an integer $m\le M$ and view the map f_w as a function of two vectors, i.e. we decompose the state into a pair (x,y) where $x\in\mathbb{R}^m$ and $y\in\mathbb{R}^{N-m}$. Consider the control system in Theorem 4.1 and note that for a fixed sequence of controls $u:=(u_t)_{t\in\mathbb{N}}$, a fixed sequence of noise symbols $w:=(w_t)_{t\in\mathbb{N}}$, and a fixed initial state $x_0\in\mathbb{R}^N$, the state process $(x_t)_{t\in\mathbb{N}}$ is deterministic. Let us introduce the notation $\varphi(t,x_0,u,w):=x_t$ for every $t\in\mathbb{N}$. Letting π_m and π_{-m} denote the natural projection of $\mathbb{R}^N$ on to the first m and last $N-m$ coordinates respectively, we further define $\varphi^m(t,x_0,u,w) := \pi_m(x_t)$ and $\varphi^{-m}(t,x_0,u,w) := \pi_{-m}(x_t)$ so that

$$\varphi(t,x_0,u,w) = (\varphi^m(t,x_0,u,w), \varphi^{-m}(t,x_0,u,w)) \text{ for every } t \in \mathbb{N}.$$

Stabilization Entropy: The proof relies on the notion of stabilization entropy and an associated lemma relating it to channel capacity. Compared to [11], we consider a version of stabilization entropy with an additional collection of sets since we are decomposing the state space into two components. The definition follows:

Definition 4.6. (Spanning Sets) Let $(D_j)_{j=1}^{j=d} \subseteq \mathcal{B}(\mathbb{R}^m)$, $(E_k)_{k=1}^{k=e} \in \mathcal{B}(\mathbb{R}^{N-m})$ and $(F_l)_{l=1}^{f} \in \mathcal{B}(\mathcal{W})$ be finite disjoint unions of Borel sets and define

$$D := \bigsqcup_{j=1}^{d} D_j \qquad E := \bigsqcup_{k=1}^{e} E_k \qquad F := \bigsqcup_{l=1}^{f} F_l.$$

Let also R denote a collection of numbers $r_{j,k,l} \in [0,1]$ for $j \in \{1,\dots,d\}$, $k \in \{1,\dots,e\}$ and $l \in \{1,\dots,f\}$ satisfying

$$1 - r := \sum_{j=1}^{d}\sum_{k=1}^{e}\sum_{l=1}^{f}(1 - r_{j,k,l}) \in [0,1]$$

and fix $T \in \mathbb{N}$ and $\rho \in (0,1)$. A set $S \subseteq (\mathbb{R}^N)^T$ of control sequences of length T is called (T,D,E,F,ρ,R)-spanning iff there exists $\tilde{\Omega} \in \mathcal{F}$ such that the following conditions both hold:

- $P(\tilde{\Omega}) \geq 1 - \rho$.
- For each $\omega \in \tilde{\Omega}$, there exists a control sequence $u \in S$ such that the following both hold:

$$\frac{1}{T}|\{t \in [0;T-1] : (\varphi^m(t,x_0(\omega),u,w(\omega)), \varphi^{-m}(t,x_0(\omega),u, w(\omega)), w_t(\omega)) \in D_j \times E_k \times F_l\}| \geq 1 - r_{j,k,l}$$

  for all j, k and l.

We slightly abuse notation writing (T,D,E,F,ρ,R)-spanning instead of $(T,(D_j)_{j=1}^{d},(E_k)_{k=1}^{e},(F_l)_{l=1}^{f},\rho,R)$-spanning. Whenever we do this however, the specific sequences of sets making up the disjoint unions will be clear from context. We will use the size of spanning sets to quantify the difficulty of a control task. This leads to:

Definition 4.7. (Stabilization Entropy) For the system (4.2), and sequences of sets as in Definition 4.7, we define the (D,E,F,ρ,R)-stabilization entropy by

$$h(D,E,F,\rho,R) := \limsup_{T\to\infty} \frac{1}{T}\log s(T,D,E,F,\rho,R),$$

where $s(T,D,E,F,\rho,R)$ denotes the smallest cardinality of a (T,D,E,F,ρ,R)-spanning set. We define this quantity to be ∞ if no finite spanning set exists.

Finite (T,D,E,F,ρ,R)-spanning sets need not exist in general but as we will shortly see, they exist in desired scenarios. The following lemma relates stabilization entropy with channel capacity.

Lemma 4.8. *Consider system (4.2) with the assumptions of Theorem 4.1 (i.e., a coding and control policy exists over a noiseless channel of capacity $C = \log_2|\mathcal{M}|$ which makes the state process AMS ergodic with asymptotic mean Q). Let D, E and F be as in Definition 4.6 and let $\rho \in$*

$(0,1)$ *be arbitrary. Let* $\epsilon > 0$ *and define the collection of numbers* $R_\epsilon := (r_{j,k,l})_{1\le j\le d, 1\le k\le e, 1\le l\le f}$, *where*

$$r_{j,k,l} := \begin{cases} (1+\epsilon)(1-Q(D_j\times E_k)\nu(F_l)) & \text{if } \kappa_{j,k,l}\in(0,1) \\ 1 & \text{if } \kappa_{j,k,l}=0 \\ \epsilon & \text{if } \kappa_{j,k,l}=1 \end{cases}$$

where we use the shorthand $\kappa_{j,k,l} := Q(D_j\times E_k)\nu(F_l)$. *Although the* $r_{j,k,l}$*'s are* ϵ*-dependent, we suppress this from the notation. The claim of the lemma is that for all sufficiently small* $\epsilon > 0$, *the stabilization entropy is well defined and satisfies*

$$h(D,E,F,\rho,R_\epsilon) \le C. \tag{4.5}$$

Proof. We note that for $\epsilon > 0$ sufficiently small enough, the conditions

(i) $1-r := \sum_{j,k,l}(1-r_{j,k,l}) \in [0,1]$,
(ii) $1-(1+\epsilon)(1-Q(D_j,E_k)\nu(F_l)) \in (0,1)$ for all j,k,l with $Q(D_j,E_k)\nu(F_l)\in(0,1)$,

are both satisfied, thus ensuring that for such an ϵ the stabilization entropy $h(D,E,F,\rho,R_\epsilon)$ is well defined. Consider system (4.2) evolving according to the fixed coding and control policy which renders the state process $(x_t)_{t\in\mathbb{N}}$ AMS ergodic with AMS mean Q. To prove that inequality we consider three cases:

Case 1: We first consider the case where $Q(D_j\times E_k)\nu(F_l)\in(0,1)$ for all j,k,l. Let $\epsilon > 0$ be small enough such that $\epsilon < \rho$ as well as conditions (i) and (ii) are satisfied. We will show that for any such ϵ the claim holds. Let us denote the process measure by μ, which is AMS by assumption. Now, for any $V\in\mathcal{B}(\mathcal{W})$, it is clear by the i.i.d. property that

$$P\Big(\Big\{\omega\in\Omega : \lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{1}_V(w_t(\omega)) = \nu(V)\Big\}\Big) = 1.$$

Noting that x_t and w_t are independent at each time step t, $(w_t)_{t\in N}$ is i.i.d, and recalling equation (1.2), it follows that $P(\hat{\Omega}) = 1$ where

$$\hat{\Omega} := \{\omega\in\Omega : \lim_{T\to\infty}\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{1}_{D_j}(\pi_m(x_t(\omega)))\cdot \tag{4.6}$$
$$\mathbb{1}_{E_k}(\pi_{-m}(x_t(\omega)))\cdot\mathbb{1}_{F_l}(w_t(\omega)) = Q(D_j\times E_k)\nu(F_l),\ \forall j,k,l\Big\}$$

where we note that the above set can be written as the intersection of a finite number of full measure sets. We continue by defining the events

$$E_I^J := \Big\{\omega\in\Omega : \Big|\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{1}_{D_j}(\pi_m(x_t(\omega)))\mathbb{1}_{E_k}(\pi_{-m}(x_t(\omega)))$$
$$\mathbb{1}_{F_l}(w_t(\omega)) - Q(D_j\times E_k)\nu(F_l)\Big| < \frac{1}{I}\forall j,k,l \text{ whenever } T\ge J\Big\}$$

and note that for any $I\in\mathbb{N}$, it is clear that $\hat{\Omega}\subseteq\bigcup_{J=1}^\infty E_I^J$ therefore $P\Big(\bigcup_{J=1}^\infty E_I^J\Big) = 1$. Let now I_0 be large enough such that

$$\frac{1}{I_0} \le \epsilon(1-Q(D_j\times E_k)\nu(F_l)) \quad \text{for all } j,k,l$$

and observe that $E^1_{I_0} \subseteq E^2_{I_0} \subseteq E^3_{I_0} \subseteq \cdots$. By continuity of probability, we have

$$\lim_{J\to\infty} P(E^J_{I_0}) = P\left(\bigcup_{J=1}^{\infty} E^J_{I_0}\right) = 1,$$

and thus there exists $J_0 \in \mathbb{N}$ such that $P(E^J_{I_0}) \geq 1-\epsilon$ for all $J \geq J_0$. For an arbitrary $T \geq J_0$, we define the set of control sequences

$$S_T := \{u_{[0;T-1]}(\omega) : \omega \in E^T_{I_0}\}.$$

We claim that this set is $(T, D, E, F, \rho, R_\epsilon)$-spanning. We use the set $\tilde{\Omega}_T := E^T_{I_0} \in \mathcal{F}$ to show this, where we note that $P(\tilde{\Omega}_T) \geq 1-\epsilon > 1-\rho$, satisfying the first requirement of the spanning set definition (Definition 4.6). To check the second condition, observe that for every $\omega \in \tilde{\Omega}_T$ and every triple j, k, l, the control sequence $u_{[0;T-1]}(\omega) \in S_T$ results in the joint state-noise process satisfying

$$\Big|\frac{1}{T}\sum_{t=0}^{T-1} \mathbb{1}_{D_j}(\pi_m(x_t(\omega)))\mathbb{1}_{E_k}((\pi_{-m}(x_t(\omega)))\mathbb{1}_{F_l}(w_t(\omega)) - Q(D_j \times E_k)\nu(F_l)\Big| < \frac{1}{I_0} \leq \epsilon(1 - Q(D_j \times E_k)\nu(F_l))$$

which implies that

$$\frac{1}{T}|\{t \in [0;T-1] : (\varphi(t, x_0(\omega), u_{[0;T-1]}(\omega), w(\omega)), w_t(\omega)) \in D_j \times E_k \times F_l\}| \geq 1 - (1+\epsilon)(1 - Q(D_j, E_k)\nu(F_l)) = 1 - r_{j,k,l}$$

which establishes the second condition, since the triple j, k, l was arbitrary. We have thus established that S_T is $(T, D, E, F, \rho, R_\epsilon)$-spanning. Since the fixed causal coding and control policy can generate at most $|\mathcal{M}|^T$ distinct control sequences by time T, it follows that $|S_T| \leq |\mathcal{M}|^T$, therefore $s(T, D, E, F, \rho, R_\epsilon) \leq |\mathcal{M}|^T$. Recalling that $T \geq J_0$ was arbitrary, we find that

$$\log s(T, B, D, \rho, R_\epsilon) \leq T \log_2 |\mathcal{M}| = TC \quad \text{for all } T \geq J_0,$$

and therefore dividing by T and letting $T \to \infty$ yields the desired capacity bound (4.5), completing the proof for Case 1.

Case 2: We now consider the case where every triple of sets (D_j, E_k, F_l) satisfies $Q(D_j \times E_k)\nu(F_l) \in [0,1)$. Suppose that j, k, l, is such that $Q(D_j \times E_k)\nu(F_l) = 0$. Then $1 - r_{j,k,l} = 0$ and the second condition in Definition 4.6 is vacuously satisfied. Combining this with Case 1, the result follows.

Case 3: Finally, we consider the case where for some indices j, k, l, $Q(D_j \times E_k)\nu(F_l) = 1$. Because each collection of sets is disjoint, $Q(D_{j'} \times E_{k'})\nu(F_{l'}) = 0$ whenever $(j, k, l) \neq (j', k', l')$. The analysis reduces to establishing the second condition in Definition 4.6 for the single set $D_j \times E_k \times F_l$ with $Q(D_j \times E_k)\nu(F_l) = 1$. Using an almost identical argument as in Case 1, the result follows. Alternatively, the analysis of a single set can be found in [40], where AMS was considered instead of AMS ergodicity as the control objective. Since AMS ergodicity implies AMS, and $h(D, E, F, \rho, R)$ reduces to the stabilization entropy notion used in [40] in case of a single set, the desired inequality follows. □

Proof of Theorem 4.1 To prove Theorem 4.1, we will approximate the integral in equation (4.3) from below using simple functions. We will prove that each of these approximations is upper bounded by the stabilization entropy, which in turn is no larger than the channel capacity. Taking a limit will yield the result.

Proof. Recalling that we fixed an integer $m \le N$, we define $p := \{1,..,m\}$. WLOG, it suffices to establish

$$\int\int \log_2 |\det Df_w^p(x_{p_1},\dots,x_{p_m},x_{z_1},\dots,x_{z_{N-m}})| dQ(x_1,\dots,x_N) d\nu(w) \le C \tag{4.7}$$

since by a relabeling of coordinates, any other set $p' \in \Gamma$ can be written in the form $\{1,2,3,..,|p'|\}$.

Recall now that by assumption (v) in Theorem 4.1, there exists a realization $\hat{x} \in \mathbb{R}^{N-m}$ such that when conditioned on the event $\{\omega \in \Omega : (x_0^{m+1},..,x_0^{N-m})(\omega) = \hat{x}\}$, the law of the random vector $(x_0^1,..,x_0^m)$ admits a bounded density. Let π_0' denote this conditional law. We now establish inequality (4.7) under slightly different assumptions than those of Theorem 4.1. More specifically, we impose that

- The last $N-m$ components of the initial state are deterministic, taking the value $\hat{x}$.
- The initial m components of x_0 are distributed according to the law π_0'.

We claim that if we can establish (4.7) under these modified assumptions, the inequality will also hold under the assumptions of Theorem 4.1. To see this, suppose otherwise. Then a coding and control policy exists which stabilizes the system in Theorem 4.1 over a channel of capacity strictly less than the LHS of (4.7). Since the stabilizing scheme works for all possible values of the initial state, this coding and control policy would also stabilize the system under the modified assumptions above, resulting in a contradiction since we are assuming that (4.7) holds for the modified system. We now proceed under the modified assumptions, and redefine $\pi_0 := \pi_0'$ to refer to the conditional law of $\pi_m(x_0)$.

Let $c \in (0,1)$ be such that $c < |\det Df_w^p(x,y)|$ for all $(x,y) \in \mathbb{R}^m \times \mathbb{R}^{N-m}$ and $w \in \mathcal{W}$. Let also $\delta > 0$ (think of this as small) and $\rho \in (0,1)$ (think of this as close to 1) be arbitrary. Next, fix Borel sets $D \subset \mathbb{R}^m$ and $E \subset \mathbb{R}^{N-m}$ satisfying that $D \times E$ have finite N-dimensional Lebesgue measure and that

$$Q(D \times E) > 1 - \frac{\delta}{2|\log c|}$$

holds (such sets can easily be found due to continuity of probability). Put also $F = \mathcal{W}$ and let $(D_j)_{j=1}^d$, $(E_k)_{k=1}^e$ and $(F_l)_{l=1}^f$ be (disjoint) partitions of D, E and F respectively. Let now $\epsilon > 0$ be small enough so that Lemma 4.8 holds, resulting in

$$h(D,E,F,\rho,R_\epsilon) \le C,$$

where R_ϵ is the associated collection of $r_{j,k,l}$'s as defined in Lemma 4.8. Let also $1-r := \sum(1-r_{j,k,l})$. Expanding out, it is easy to see (recalling that $\nu(F)=1$) that $r = 1-(1+\epsilon)Q(D\times E)+def\epsilon$ (or $r=\epsilon$ if one of the $D_j \times E_k \times F_l$'s has full $Q\times\nu$-measure) thus we see that for every sufficiently small ϵ,

$$2r < \frac{\delta}{|\log c|}. \tag{4.8}$$

Now fix a sufficiently large $T \in \mathbb{N}$ and let S_T be a finite $(T,D,E,F,\rho,R_\epsilon)$-spanning set (whose existence is guaranteed by the proof of Lemma 4.8) with $\tilde{\Omega}_T \in \mathcal{F}$, $P(\tilde{\Omega}) \ge 1-\rho$, the associated subset of Ω. Letting x_0^m denote the vector consisting of the first m components of x_0, we proceed by defining

$$\begin{aligned}
A &:= \{(w(\omega), x_0^m(\omega)) : \omega \in \tilde{\Omega}\},\\
A(u) &:= \{(w,x) \in \mathcal{W}^{\mathbb{N}} \times \mathbb{R}^m : \forall j,k,l,\\
&\quad \frac{1}{T}\sum_{t=0}^{T-1} \mathbb{1}_{D_j \times E_k \times F_l}(\varphi(t,(x,\hat{x}),u,w), w_t) \geq 1 - r_{j,k,l}\}.\\
A(u,w) &:= \{x \in \mathbb{R}^m : (w,x) \in A(u)\}.
\end{aligned}$$

Letting m denote the m-dimensional Lebesgue measure, we see that

$$A \subseteq \bigcup_{u \in S_T} A(u), \quad (\nu \times m)(A(u)) = \int m(A(u,w)) d\nu(w), \tag{4.9}$$

where the equality follows from the Fubini-Tonelli theorem (and the containment by definition of the sets). Letting $M > 0$ be an upper bound for the density of π_0, we have that

$$1 - \rho \leq (\nu \times \pi_0)(A) \leq M \cdot (\nu \times m)(A). \tag{4.10}$$

Combining (4.9) and (4.10), we obtain the key inequality:

$$\frac{1}{M}(1-\rho) \leq (\nu \times m)(A) \leq |S_T| \max_{u \in S_T} (\nu \times m)(A(u)) = |S_T| \max_{u \in S_T} \int m(A(u,w)) d\nu(w). \tag{4.11}$$

The next step in the proof is to obtain upper bounds for the volume $m(A(u,w))$. We proceed by defining a set consisting of disjoint collections of subsets of $\{0,\dots,T-1\}$:

$$\begin{aligned}
\mathbb{A} := \{\Lambda = \{\Lambda_{j,k,l}\}_{j,k,l} &: \bigsqcup_{j=1}^{d}\bigsqcup_{k=1}^{e}\bigsqcup_{l=1}^{f} \Lambda_{j,k,l} \subseteq \{0,\dots,T-1\},\\
&|\Lambda_{j,k,l}| \geq (1 - r_{j,k,l})T, \forall j = 1,..,d,\ k = 1,..,e,\ l = 1,..,f\}
\end{aligned}$$

and note that as a consequence of the definition, $|\bigsqcup_{j=1}^{d}\bigsqcup_{k=1}^{e}\bigsqcup_{l=1}^{f}\Lambda_{j,k,l}| \geq (1-r)T$ for all $\Lambda \in \mathbb{A}$. We note that such sets can only be found for T sufficiently large, however as we will be taking a limit as $T \to \infty$, this is not a problem. For $\Lambda \in \mathbb{A}$, define the set

$$A(u,w,\Lambda) := \{x \in \mathbb{R}^m : (\varphi(t,(x,\hat{x}),u,w), w_t) \in D_j \times E_k \times F_l \Leftrightarrow t \in \Lambda_{j,k,l} \text{ for all } j,k,l\}.$$

It is not hard to see that $A(u,w) = \bigsqcup_{\Lambda \in \mathbb{A}} A(u,w,\Lambda)$ is a disjoint union, thus (4.11) becomes

$$\frac{1}{M}(1-\rho) \leq |S_T| \max_{u \in S_T} \int \sum_{\Lambda \in \mathbb{A}} m(A(u,w,\Lambda)) d\nu(w). \tag{4.12}$$

Our next step is to bound the volumes of the form $m(A(u,w,\Lambda))$. Writing $\varphi_{t,u,w}(\cdot) := \varphi(t,(\cdot,\hat{x}),u,w)$ we define

$$A_t(u,w,\Lambda) := \varphi_{t,u,w}(A(u,w,\Lambda)), \quad t = 0,1,\dots,T-1,$$

and observe that

$$A_t(u,w,\Lambda) \subseteq D_j \quad \text{whenever } t \in \Lambda_{j,k,l}\ \forall\ j,k,l.$$

Next, we define the following numbers:

$$c_{j,k,l} := \inf_{(x,y,w) \in D_j \times E_k \times F_l} |\det Df_w^p(x,y)|.$$

Recalling that by assumption $f_w^p(\cdot, y)$ is injective and C^1, it follows that for all (j,k,l) we have

$$m(A_{t+1}(u,w,\Lambda)) \geq c_{j,k,l} \cdot m(A_t(u,w,\Lambda)) \text{ whenever } t \in \Lambda_{j,k,l},$$
$$m(A_{t+1}(u,w,\Lambda)) \geq c \cdot m(A_t(u,w,\Lambda)) \text{ whenever } t \notin \bigsqcup \Lambda_{j,k,l}.$$

Letting $t^*(\Lambda_{j,k,l}) := \max \Lambda_{j,k,l}$, $t^*(\Lambda) := \max_{j,k,l} t^*(\Lambda_{j,k,l})$, applying the above inequalities repeatedly, and recalling that $c \leq c_{j,k,l}$, it is not hard to see that

$$m(A(u,w,\Lambda))\Big(\prod_{j=1}^{d}\prod_{k=1}^{e}\prod_{l=1}^{f} c_{j,k,l}^{|\Lambda_{j,k,l}|-1}\Big)c^{rT+def} \leq m(A_{t^*(\Lambda)}(u,w,\Lambda)).$$

where in principle, all the exponents of the $c_{j,k,l}$'s should be $|\Lambda_{j,k,l}|$, except for possibly one which should be $|\Lambda_{j,k,l}|-1$. We do not know which one though, so we write the weaker inequality as above. Combining this with (4.12), we obtain

$$\frac{1}{M}(1-\rho) \leq |S| \max_{u \in S_T} \sum_{\Lambda \in \mathbb{A}} \int m(A_{t^*(\Lambda)}(u,w,\Lambda))c^{-(rT+def)}$$
$$\prod_{j=1}^{d}\prod_{k=1}^{e}\prod_{l=1}^{f} c_{j,k,l}^{-(|\Lambda_{j,k,l}|-1)} d\nu(w),$$

and note that the right hand side of the above can be written as

$$|S| \cdot c^{-(rT+def)} \max_{u \in S_T} \sum_{t_{1,1,1}=(1-r_{1,1,1})T}^{T} \cdots \sum_{t_{d,e,f}=(1-r_{d,e,f})T}^{T}$$
$$\int \sum_{\Lambda \in \mathbb{A}:\ t^*(\Lambda_{j,k,l})=t_{j,k,l} \forall j,k,l} m(A_{t^*(\Lambda)}(u,w,\Lambda))$$
$$\prod_{j=1}^{d}\prod_{k=1}^{e}\prod_{l=1}^{f} c_{j,k,l}^{-(|\Lambda_{j,k,l}|-1)} d\nu(w)$$
$$\leq |S| \cdot c^{-(2rT+def)} \max_{u \in S_T} \sum_{t_{1,1,1}=(1-r_{1,1,1})T}^{T} \cdots \sum_{t_{d,e,f}=(1-r_{d,e,f})T}^{T}$$
$$\int \sum_{\Lambda \in \mathbb{A}:\ t^*(\Lambda_{j,k,l})=t_{j,k,l} \forall j,k,l} m(A_{t^*(\Lambda)}(u,w,\Lambda))$$
$$\prod_{j=1}^{d}\prod_{k=1}^{e}\prod_{l=1}^{f} c_{j,k,l}^{-((1-r_{j,k,l})T-1)} d\nu(w).$$

where the last inequality follows by noting that

$$\begin{aligned} c^{rT+def}\prod_{j,k,l}c_{j,k,l}^{|\Lambda_{j,k,l}|-1} &= c^{rT+\sum_{j,k,l}|\Lambda_{j,k,l}|}\prod_{j,k,l}\Big(\frac{c_{j,k,l}}{c}\Big)^{|\Lambda_{j,k,l}|-1} \\ &\geq c^{rT+\sum_{j,k,l}|\Lambda_{j,k,l}|}\prod_{j,k,l}\Big(\frac{c_{j,k,l}}{c}\Big)^{(1-r_{j,k,l})T-1} \\ &= c^{rT+\sum_{j,k,l}|\Lambda_{j,k,l}|-(1-r)T+def}\prod_{j,k,l}c_{j,k,l}^{(1-r_{j,k,l})T-1} \\ &\geq c^{2rT+def}\prod_{j,k,l}c_{j,k,l}^{(1-r_{j,k,l})T-1}. \end{aligned}$$

Observe that the sets $A_{t^*(\Lambda)}(u,w,\Lambda)$ with $\Lambda \in \mathbb{A}$, $t^*(\Lambda)$ fixed, are pairwise disjoint, since they are the images of the corresponding sets $A(u,w,\Lambda)$ under the injective map $\varphi_{t^*(\Lambda),u,w}$. Moreover, all of these sets are contained in D, hence

$$\sum_{\Lambda\in\mathbb{A}:t^*(\Lambda_{j,k,l})=t_{j,k,l}\forall j,k,l} m(A_{t^*(\Lambda)}(u,w,\Lambda)) \leq m(D).$$

Together with the above chain of inequalities, this implies

$$\frac{1}{M}(1-\rho) \leq |S_T|\cdot m(D)\cdot c^{-(2rT+def)}.$$

$$\prod_{j=1}^{d}\prod_{k=1}^{e}\prod_{l=1}^{f}c_{j,k,l}^{-((1-r_{j,k,l})T-1)}\prod_{j=1}^{d}\prod_{k=1}^{e}\prod_{l=1}^{f}(r_{j,k,l}T+1).$$

Since this inequality holds for every T sufficiently large, we can take logarithms on both sides, divide by T and let $T\to\infty$. This results in

$$0 \leq h(D,E,F,\rho,R_\epsilon) - 2r\log c - \sum_{j=1}^{d}\sum_{k=1}^{e}\sum_{l=1}^{f}(1-r_{j,k,l})\log c_{j,k,l}.$$

Recalling the definition of $r_{j,k,l}$, the fact that ϵ can be chosen arbitrarily small and (4.8), this leads to the estimate

$$C+\delta \geq \sum_{j=1}^{d}\sum_{k=1}^{e}\sum_{l=1}^{f}Q(D_j\times E_k)\nu(F_l)\inf_{(x,y,w)\in D_j\times E_k\times F_l}\log|\det Df_w^p(x,y)|.$$

Considering the supremum of the right-hand side over all finite measurable partitions of D,E and $F=\mathcal{W}$ leads to

$$C+\delta \geq \int\int \mathbb{1}_{D\times E}(x_1,..,x_N)\log|\det Df_w^p(x_1,..,x_N)|dQ(x_1,..,x_N)d\nu(w),$$

where we use that the integrand is uniformly bounded below by $\log c$ (and hence, we can assume that it is non-negative). Considering now an increasing sequence of sets $D_k\times E_k\subset\mathbb{R}^N$ whose union is $\mathbb{R}^N$, we can invoke the theorem of monotone convergence to obtain the desired estimate, observing that δ can be made arbitrarily small as $D_k\times E_k$ becomes arbitrarily large. This completes the proof. □

5. Reflection: Comparison of the information theoretic and stochastic growth (dynamical systems) approaches

In view of our main theorem above and the review presented earlier, we present in this section a comparison between the information theoretic and the stochastic geometric approaches.

For the AMS criterion, Theorem 3.9 was arrived at via an information theoretic method. Theorem 3.16 (and the noisy channel generalization Theorem 3.21) was arrived at via the geometric method. Theorem 3.9 allowed for more general channels, but Theorem 3.16 is more relaxed.

For the ergodicity criterion, the information theoretic method culminated in Theorem 3.11. Theorem 3.24 and Theorem 3.25 established ergodicity via the stochastic geometric method. In particular, Theorem 4.1 allowed for splitting the dynamical system into subsystems that are stable and unstable via a novel approach building on a stochastic geometric argument.

We note that obtaining the sharper bound established in the main theorem, Theorem 4.1, using only information theoretic methods does not appear to be possible, with the main impediment being the fact that when splitting the state and conditioning on a past state realization (or some other sufficient information), the unstable and stable state components may not be independent random variables, which was required for the information theoretic methods to be applicable in this context. The geometric analysis proved to be versatile in allowing us to enhance the results.

In summary, below are some of the benefits of the stochastic geometric approach:

(i) Refined stochastic stability results applicable to a more general class of system models (Theorem 3.16) and more refined stability criteria such as the AMS property in combination with moment conditions (see Corollary 3.19);
(ii) A direct derivation, building on volume growth arguments, applicable to a plethora of criteria;
(iii) More refined bounds for a large class of systems through trading-off growth rates with the measures of sets under the coordinate projection of a stationary measure (see Theorem 3.16);
(iv) The unification of the theory developed for deterministic systems controlled over noise-free communication channels with their stochastic counterparts, involving both stochastic nonlinear dynamical systems and noisy communication channels (see Theorem 3.21);
(v) For the ergodicity criterion, when we compare Theorem 3.24 and Theorem 3.25 with Theorem 3.11; Theorem 3.24 is more general in the sense that it applies to arbitrary causal coding and control policies, not just Markov ones. Moreover, it does not require the assumption of sublinear growth of the differential entropy of the state process. Theorem 3.11 assumes that the state process is positive Harris recurrent, which implies unique ergodicity, while Theorem 3.24 only assumes ergodicity of the AMS measure. On the other hand, compared with Theorem 3.25, Theorem 3.11 considers a more general class of channels (involving memory) as well as systems taking values in higher dimensions.

Theorem 4.1 allowed for splitting the dynamical system into subsystems that are stable and unstable via a novel approach. This is a consequential result in that the entropy growth of a system may be negative but this may be due to the presence of a stable mode compensating the growth in the direction of an unstable direction. This was already observed for the linear systems case by the sum of the unstable eigenvalues analysis [3]. A non-linear generalization of such a splitting turned out to be challenging via information theoretic methods, but the stochastic geometric method allowed for such a generalization.

(vi) On the other hand, the information theoretic method allows for the well-developed machinery of Shannon theory. A stochastic dynamical system generalization of some of the concepts, such as conditional entropy, requires rather indirect arguments.

6. Concluding Remarks

In this paper, after a general review on the problem of stabilizing a discrete-time non-linear stochastic dynamical system over a finite capacity channel via information theoretic as well as geometric methods, we have - for a certain class of non-linear systems - established a sharper bound on channel capacity required for ergodic stabilization. The techniques involved in our proof were stabilization entropy, a volume growth argument, and the property that almost surely, system sample paths visit regions of the state space at a frequency given by an ergodic measure.

There are two possible avenues of further investigation. First, it would be interesting to enlarge the class of noise processes for which the bounds in this paper hold. We have considered only i.i.d. noise, however it is possible that an ergodic-like property (i.e. that equation (4.6) holds) for the joint state-noise process will hold for less restrictive classes of noise. Secondly, it seems interesting to attempt the generalization of this result for the noisy channel case. Using stabilization entropy techniques, [11] established the bound (1.3) for scalar systems controlled over Discrete Memoryless Channels. Given that sharper bounds can be established for multi-dimensional systems, it seems worthwhile to attempt to generalize the one dimensional result to many dimensions, and combine it with the arguments in this paper to sharpen the bound.

Acknowledgements

The first and third authors were supported by the Natural Sciences and Engineering Council of Canada (NSERC).

References

[1] V. S. Borkar, S. K. Mitter, and S. Tatikonda, Optimal sequential vector quantization of Markov sources, SIAM J. Control and Optimization, vol. 40, pp. 135–148, 2001.

[2] N. Elia and S. K. Mitter, Stabilization of linear systems with limited information, IEEE Transactions on Automatic Control, vol. 46, no. 9, pp. 1384–1400, 2001.

[3] S. Tatikonda and S. Mitter, Control under communication constraints, IEEE Transactions on Automatic Control, vol. 49, no. 7, pp. 1056–1068, 2004.

[4] S. Tatikonda, A. Sahai, and S. Mitter, Stochastic linear control over a communication channels, IEEE Transactions on Automatic Control, vol. 49, pp. 1549–1561, 2004.
[5] A. Sahai, Anytime Information Theory, Massachusetts Institute of Technology, Cambridge, MA: Ph.D. dissertation, 2001.
[6] S. Tatikonda, Control under Communication Constraints, Massachusetts Institute of Technology, Cambridge, MA: PhD dissertation, 2000.
[7] R. Bansal and T. Başar, Simultaneous design of measurement and control strategies in stochastic systems with feedback, Automatica, vol. 45, pp. 679–694, 1989.
[8] V. S. Borkar and S. K. Mitter, LQG Control with Communication Constraints, In: Kailath Festschrift, Kluwer Academic Publishers, Boston, 1997.
[9] P. Walters, An Introduction to Ergodic Theory, Springer, vol. 79, 2000.
[10] R. M. Gray, Probability, Random Processes, and Ergodic Properties, Springer, 2009.
[11] N. Garcia, C. Kawan, and S. Yüksel, Ergodicity conditions for controlled stochastic nonlinear systems under information constraints: A volume growth approach, SIAM Journal on Control and Optimization, vol. 59, no. 1, pp. 534–560, 2021.
[12] S. Yüksel, Stationary and ergodic properties of stochastic nonlinear systems controlled over communication channels, SIAM Journal on Control and Optimization, vol. 54, no. 5, pp. 2844–2871, 2016.
[13] J. Hespanha, A. Ortega, and L. Vasudevan, Towards the control of linear systems with minimum bit-rate, In: Proc. of the Int. Symp. on the Mathematical Theory of Networks and Syst. 2002.
[14] D. F. Delchamps, Stabilizing a linear system with quantized state feedback, IEEE transactions on automatic control, vol. 35, no. 8, pp. 916–924, 1990.
[15] G. N. Nair and R. J. Evans, Stabilizability of stochastic linear systems with finite feedback data rates, SIAM Journal on Control and Optimization, vol. 43, no. 2, pp. 413–436, 2004.
[16] W. S. Wong and R. W. Brockett, Systems with finite communication bandwidth constraints. ii. stabilization with limited information feedback, IEEE Transactions on Automatic Control, vol. 44, no. 5, pp. 1049–1053, 1999.
[17] A. S. Matveev and A. V. Savkin, Estimation and Control over Communication Networks, Birkhäuser, Boston, 2008.
[18] A. Sahai and S. Mitter, The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link part I: Scalar systems, IEEE Transactions on Information Theory, vol. 52, no. 8, pp. 3369–3395, 2006.
[19] P. Minero, L. Coviello, and M. Franceschetti, Stabilization over markov feedback channels: the general case, IEEE Transactions on Automatic Control, vol. 58, no. 2, pp. 349–362, 2012.
[20] V. Kostina, Y. Peres, G. Ranade, and M. Sellke, Exact minimum number of bits to stabilize a linear system, In: 2018 IEEE Conference on Decision and Control (CDC), pp. 453–458, IEEE, 2018.
[21] N. C. Martins and M. A. Dahleh, Feedback control in the presence of noisy channels: 'Bode-like' fundamental limitations of performance, IEEE Transactions on Automatic Control, vol. 53, pp. 1604–1615, 2008.
[22] G. N. Nair, F. Fagnani, S. Zampieri, and J. R. Evans, Feedback control under data constraints: an overview, Proceedings of the IEEE, pp. 108–137, 2007.
[23] S. Yüksel, Stochastic stabilization of noisy linear systems with fixed-rate limited feedback, IEEE Transactions on Automatic Control, vol. 55, pp. 2847–2853, 2010.
[24] S. Yüksel and S. P. Meyn, Random-time, state-dependent stochastic drift for Markov chains and application to stochastic stabilization over erasure channels, IEEE Transactions on Automatic Control, vol. 58, pp. 47–59, 2013.

[25] A. P. Johnston and S. Yüksel, Stochastic stabilization of partially observed and multi-sensor systems driven by unbounded noise under fixed-rate information constraints, IEEE Transactions on Automatic Control, vol. 59, no. 3, pp. 792–798, 2013.

[26] S. Yüksel, Characterization of information channels for asymptotic mean stationarity and stochastic stability of non-stationary/unstable linear systems, IEEE Transactions on Information Theory, vol. 58, pp. 6332–6354, 2012.

[27] S. Yüksel and T. Başar, Stochastic Networked Control Systems: Stabilization and Optimization under Information Constraints, Springer, New York, 2013.

[28] S. Yüksel and T. Başar, Stochastic Teams, Games, and Control under Information Constraints, Springer, Cham, 2024.

[29] D. Liberzon and J. P. Hespanha, Stabilization of nonlinear systems with limited information feedback, IEEE Transactions on Automatic Control, vol. 50, no. 6, pp. 910–915, 2005.

[30] C. De Persis, n-bit stabilization of n-dimensional nonlinear systems in feedforward form, IEEE Transactions on Automatic Control, vol. 50, no. 3, pp. 299–311, 2005.

[31] G. N. Nair, R. J. Evans, I. M. Mareels, and W. Moran, Topological feedback entropy and nonlinear stabilization, IEEE Transactions on Automatic Control, vol. 49, no. 9, pp. 1585–1597, 2004.

[32] R. L. Adler, A. G. Konheim, and M. H. McAndrew, Topological entropy, Transactions of the American Mathematical Society, vol. 114, no. 2, pp. 309–319, 1965.

[33] F. Colonius and C. Kawan, Invariance entropy for control systems, SIAM Journal on Control and Optimization, vol. 48, no. 3, pp. 1701–1721, 2009.

[34] C. Kawan, Invariance entropy for deterministic control systems, Lecture notes in mathematics, vol. 2089, 2013.

[35] F. Colonius, C. Kawan, and G. Nair, A note on topological feedback entropy and invariance entropy, Systems & Control Letters, vol. 62, no. 5, pp. 377–381, 2013.

[36] F. Colonius, Metric invariance entropy and conditionally invariant measures, Ergodic Theory and Dynamical Systems, vol. 38, no. 3, pp. 921–939, 2018.

[37] M. Franceschetti, M. J. Khojasteh, and M. Z. Win, The many facets of information in networked estimation and control, Annual Review of Control, Robotics, and Autonomous Systems, vol. 6, no. 1, pp. 233–259, 2023.

[38] G. N. Nair, F. Fagnani, S. Zampieri, and R. J. Evans, Feedback control under data rate constraints: An overview, Proceedings of the IEEE, vol. 95, no. 1, pp. 108–137, 2007.

[39] A. S. Matveev and A. V. Savkin, Estimation and control over communication networks, Springer, 2009.

[40] C. Kawan and S. Yüksel, Invariance properties of nonlinear stochastic dynamical systems under information constraints, IEEE Transactions on Automatic Control, to appear (arXiv: 1901.02825), 2020.

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

[42] H. H. Permuter, T. Weissman, and A. J. Goldsmith, Finite state channels with time-invariant deterministic feedback, IEEE Transactions on Information Theory, vol. 55, no. 2, pp. 644–662, 2009.

[43] N. Şen, F. Alajaji, and S. Yüksel, Feedback capacity of a class of symmetric finite-state Markov channels, IEEE Transactions on Information Theory, vol. 56, pp. 4110 – 4122, 2011.

[44] S. Tatikonda and S. Mitter, The capacity of channels with feedback, IEEE Transactions on Information Theory, vol. 55, no. 1, pp. 323–349, 2009.

[45] R. Dabora and A. Goldsmith, On the capacity of indecomposable finite-state channels with feedback, Proceedings of the Allerton Conf Commun Control Comput, pp. 1045–1052, 2008.

[46] J. C. Kieffer, Stochastic stability for feedback quantization schemes, IEEE Transactions on Information Theory, vol. 28, pp. 248–254, 1982.

[47] J. C. Kieffer and J. G. Dunham, On a type of stochastic stability for a class of encoding schemes, IEEE Transactions on Information Theory, vol. 29, pp. 793–797, 1983.

[48] T. Linder and S. Yüksel, On optimal zero-delay quantization of vector Markov sources, IEEE Transactions on Information Theory, vol. 60, pp. 2975–5991, 2014.

[49] A. Da Silva and C. Kawan, Robustness of critical bit rates for practical stabilization of networked control systems, Automatica, vol. 93, pp. 397–406, 2018.

[50] C. Kawan and J.-C. Delvenne, Network entropy and data rates required for networked control, IEEE Transactions on Control of Network Systems, vol. 3, no. 1, pp. 57–66, 2015.

[51] C. Kawan and S. Yüksel, Invariance properties of nonlinear stochastic dynamical systems under information constraints, IEEE Transactions on Automatic Control, vol. 66, pp. 4514–4529, 2021.

[52] N. Garcia, C. Kawan, and S. Yüksel, Ergodicity conditions for controlled stochastic non-linear systems under information constraints, SIAM Journal on Control and Optimization, vol. 59, no. 1, pp. 534–560, 2021.

[53] A. Katok, Lyapunov exponents, entropy and periodic orbits for diffeomorphisms, Publications Mathématiques de l'IHÉS, vol. 51, pp. 137–173, 1980.

[54] Y. Ren, L. He, J. Lü, and G. Zheng, Topological r-entropy and measure-theoretic r-entropy of a continuous map, Science China Mathematics, vol. 54, no. 6, pp. 1197–1205, 2011.

[55] E. B. Dynkin and A. A. Yushkevich, Controlled Markov processes, vol. 235, Springer, 1979.

[56] J. H. Kemperman, Strong converses for a general memoryless channel with feedback, In: Proceedings 6th Prague Conference on Information Theory, Statistical Decision Functions, and Random Processes, pp. 375–409, 1971.

[57] I. Csiszár and J. Korner, Information Theory: Coding Theorems for Discrete Memoryless Channels, Budapest, Akademiai Kiado, 1981.

[58] C. E. Shannon, The zero error capacity of a noisy channel, IEEE Transactions on Information Theory, vol. 2, pp. 8–19, 1956.