Metric and topological entropy bounds on state estimation for stochastic non-linear systems

Christoph Kawan and Serdar Yüksel

***Abstract*— This paper studies state estimation over noisy channels for stochastic non-linear systems. We consider three estimation objectives, a strong and a weak form of almost sure stability of the estimation error as well as quadratic stability in expectation. For all three objectives, we derive lower bounds on the smallest channel capacity C_0 above which the objective can be achieved with an arbitrarily small error. Lower bounds are obtained via a dynamical systems (through a novel construction of a dynamical system), an information-theoretic and a random dynamical systems approach. The first two approaches show that for a large class of systems, such as additive noise systems, $C_0 = \infty$, i.e., the estimation objectives cannot be achieved via channels of finite capacity. The random dynamical systems approach is shown to be operationally non-adequate for the problem, since it yields finite lower bounds C_0 under mild assumptions. Finally, we prove that a memoryless noisy channel in general constitutes no obstruction to asymptotic almost sure state estimation with arbitrarily small errors, when there is no noise in the system.**

I. Introduction

State estimation over noisy channels is a first step towards a complete theory of control of non-linear systems over noisy channels. The results in this area have either almost exclusively considered linear systems, or in the non-linear case only been on deterministic systems over deterministic channels, with few exceptions. State estimation over digital channels was studied in [11], [14] for linear discrete-time systems in a stochastic framework with the objective to bound the estimation error in probability. In these works, the inequality

$$C \geq H(A) := \sum_{\lambda \in \sigma(A)} \max\{0, n_\lambda \log|\lambda|\} \qquad (1)$$

for the channel capacity C was obtained as a necessary and almost sufficient condition. Here A is the dynamical matrix of the system and the summation is over its eigenvalues λ with multiplicities n_λ.

Some relevant studies that have considered non-linear systems are the following. The papers [9], [15] and [13] studied state estimation for non-linear deterministic systems and noise-free channels. In [9], Liberzon and Mitra characterized the critical bit rate C_0 for exponential state estimation with a given exponent $\alpha \geq 0$ for a continuous-time system on a compact subset K of its state space. As a measure for C_0, they introduced a quantity named estimation entropy $h_{\mathrm{est}}(\alpha, K)$, which coincides with the topological entropy on K when $\alpha = 0$, but for $\alpha > 0$ is no longer a purely topological quantity. Furthermore, they derived an upper bound R of $h_{\mathrm{est}}(\alpha, K)$ in terms of α, the dimension of the state space and a Lipschitz constant of the dynamical system. The paper [6] provided a lower bound on $h_{\mathrm{est}}(\alpha, K)$ in terms of Lyapunov exponents under the assumption that the system preserves a smooth measure. In [13], Matveev and Pogromsky studied three estimation objectives of increasing strength for discrete-time non-linear systems. For the weakest one, the smallest bit rate was again shown to be equal to the topological entropy. For the other ones, general upper and lower bounds were obtained which can be computed directly in terms of the linearized right-hand side of the equation generating the system. A further related paper is [12], which studies state estimation for a class of non-linear systems over noise-free digital channels and where connections with topological entropy are established.

The rate-distortion function $R(D)$ for a stochastic process is a related information-theoretic performance metric. In the information theory community, the study of problems on state estimation has almost exclusively focused on stationary processes. For a large class of such processes, the rate-distortion function often admits a simple formula, referred to as a *a single-letter characterization*. We also note that if the source process is a finite-valued stationary and ergodic process, the rate-distortion function with $D = 0$ reduces to the metric entropy of the source. There have been few contributions in the information theory literature on non-causal coding of non-stationary/unstable sources. These have only, to our knowledge, focused on Gaussian linear models: Consider the Gaussian auto-regressive (AR) process $x_t = -\sum_{k=1}^m a_k x_{t-k} + w_t$, where $\{w_t\}$ is an i.i.d. zero-mean, Gaussian random sequence with variance $E[w_t^2] = \sigma^2$. The rate-distortion function (with the distortion being the expected, normalized Euclidean error) is given parametrically by the following [3]:

$$\begin{aligned} D_\theta &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \min\Big(\theta, \frac{1}{g(w)}\Big)\mathrm{d}w, \\ R(D_\theta) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \max\Big(\frac{1}{2}\log\frac{1}{\theta g(w)}, 0\Big)\mathrm{d}w \\ &\quad + \sum_{k=1}^{m} \frac{1}{2}\max\big(0, \log(|\rho_k|^2)\big), \end{aligned}$$

where $g(w) = \frac{1}{\sigma^2}|1 + \sum_{k=1}^m a_k \mathrm{e}^{-ikw}|^2$ and $\{\rho_k\}$ are the roots of the polynomial H. We may emphasize that the second term on the right-hand side of the equation is exactly the topological entropy for a linear system with eigenvalues ρ_k.

We also note that the above formulation is what is known

C. Kawan is with the Faculty of Computer Science and Mathematics, University of Passau, 94032 Passau, Germany (e-mail: christoph.kawan@uni-passau.de). S. Yüksel is with the Department of Mathematics and Statistics, Queen's University, Kingston, Ontario, Canada, K7L 3N6 (e-mail: yuksel@mast.queensu.ca)

978-1-5090-4096-4/17/$31.00 ©2017 IEEE

as a non-causal construction. When one also inserts causality constraints, the problem typically becomes far more challenging, even when the source process is an i.i.d. process, because in this case the optimization problem becomes non-convex (see e.g. [10],[19, Ch. 5]).

A related problem is the control of stochastic non-linear systems over communication channels. This problem has been studied in few publications, and mainly only for deterministic systems or deterministic channels. Recently, [18] studied stochastic stability properties for a more general class of stochastic non-linear systems building on information-theoretic bounds and Markov-chain-theoretic constructions, however these bounds do not distinguish between the unstable and stable components of the tangent space associated with a dynamical non-linear system, except for the linear system case. Our paper here provides such a refinement, but only for estimation problems and in the low-distortion regime.

The proofs of the results presented in this paper can be accessed at [7].

II. Preliminaries

a) Notation and definitions: All logarithms in this paper are taken to the base 2. An important concept used in the paper is the topological entropy of a dynamical system (see, e.g., [5]). The topological entropy quantifies the average exponential divergence rate of nearby trajectories, and as such is a measure for the total exponential orbit complexity. Its precise definition is as follows.

If $f : X \to X$ is a continuous map on a metric space (X, d), and $K \subset X$ is compact, we say that $E \subset K$ is $(n, \epsilon; f)$-separated for some $n \in \mathbb{N}$ and $\epsilon > 0$ if for all $x, y \in E$ with $x \neq y$, $d(f^i(x), f^i(y)) > \epsilon$ for some $i \in \{0, 1, \ldots, n-1\}$. We write $r_{\mathrm{sep}}(n, \epsilon, K; f)$ for the maximal cardinality of an $(n, \epsilon; f)$-separated subset of K and define the topological entropy $h_{\mathrm{top}}(f, K)$ of f on K by

$$h_{\mathrm{sep}}(f, \epsilon; K) := \limsup_{n\to\infty} \frac{1}{n} \log r_{\mathrm{sep}}(n, \epsilon, K; f),$$
$$h_{\mathrm{top}}(f, K) := \lim_{\epsilon \downarrow 0} h_{\mathrm{sep}}(f, \epsilon; K).$$

If X is compact and $K = X$, we also omit the argument K and call $h_{\mathrm{top}}(f)$ the topological entropy of f. Alternatively, one can define $h_{\mathrm{top}}(f, K)$ using (n, ϵ)-spanning sets. A set $F \subset X$ (n, ϵ)-spans another set $K \subset X$ if for each $x \in K$ there is $y \in F$ with $d(f^i(x), f^i(y)) \leq \epsilon$ for $i = 0, 1, \ldots, n-1$. Letting $r_{\mathrm{span}}(n, \epsilon, K; f)$ (or $r_{\mathrm{span}}(n, \epsilon, K)$ if the map f is clear from the context) denote the minimal cardinality of a set which (n, ϵ)-spans K, the topological entropy of f on K satisfies $h_{\mathrm{top}}(f, K) = \lim_{\epsilon\downarrow 0} \limsup_{n\to\infty} \frac{1}{n} \log r_{\mathrm{span}}(n, \epsilon, K; f)$.

If $f : X \to X$ is a measure-preserving map on a probability space $(\Omega, \mathcal{F}, \mu)$, i.e., $\mu(f^{-1}(F)) = \mu(F)$ for all $F \in \mathcal{F}$, its metric entropy $h_\mu(f)$ is defined as follows. Let $\mathcal{A}$ be a finite measurable partition of X. Then the entropy of f w.r.t. $\mathcal{A}$ is defined by $h_\mu(f; \mathcal{A}) := \lim_{n\to\infty} \frac{1}{n} H_\mu\left(\bigvee_{i=0}^{n-1} f^{-i}\mathcal{A}\right)$. Here $\bigvee$ denotes the join operation, i.e., $\bigvee_{i=0}^{n-1} f^{-i}\mathcal{A}$ is the partition of X consisting of all intersections of the form $A_0 \cap f^{-1}(A_1) \cap \ldots \cap f^{-n+1}(A_{n-1})$ with $A_i \in \mathcal{A}$. For any partition $\mathcal{B}$ of X, $H_\mu(\mathcal{B}) = -\sum_{B\in\mathcal{B}} \mu(B) \log \mu(B)$. The metric entropy of f is then defined by $h_\mu(f) := \sup_{\mathcal{A}} h_\mu(f; \mathcal{A})$, where the supremum is taken over all finite measurable partitions of X.

Topological and metric entropy are related to each other via the variational principle: For a continuous map f on a compact metric space X, $h_{\mathrm{top}}(f) = \sup_\mu h_\mu(f)$, where the supremum is taken over all f-invariant Borel probability measures.

Throughout the paper, we assume that all random variables are modeled on a common probability space $(\Omega, \mathcal{F}, P)$. We will also use several concepts from information theory, which will not be defined here but are available at [7].

b) Stochastic networked systems and estimation objectives: In this paper, we consider non-linear noisy systems given by an equation of the form

$$x_{t+1} = f(x_t, w_t). \tag{2}$$

Here x_t is the state at time t and $(w_t)_{t\in\mathbb{Z}_+}$ is an i.i.d. sequence of random variables with common distribution $w_t \sim \nu$, modeling the noise. In general, we assume that $f : X \times W \to X$ is a Borel measurable map, where X and W are Polish spaces, so that for any $w \in W$ the map $f(\cdot, w)$ is a homeomorphism of X. We further assume that x_0 is a random variable on X with an associated probability measure π_0, independent of $(w_t)_{t\in\mathbb{Z}_+}$. We use the notations $f_w : X \to X$, $f_w(x) = f(x, w)$ and $f^x : W \to X$, $f^x(w) = f(x, w)$.

This system is connected over a noisy channel of finite capacity to an estimator, as shown in Fig. 1. The estimator has access to the information it has received through the channel. A source coder maps the source symbols (i.e., state values) to corresponding channel inputs. The channel inputs are transmitted through the channel. We assume that the channel is a discrete channel with input alphabet $\mathcal{M}$ and output alphabet $\mathcal{M}'$.

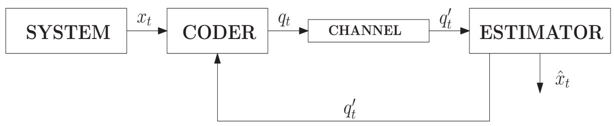

Fig. 1: Estimation over a noisy channel with feedback

We refer by a *Coding Policy* Π, to a sequence of functions $(\gamma^e_t)_{t\in\mathbb{Z}_+}$ which are causal such that the channel input at time t, $q_t \in \mathcal{M}$, under Π is generated by a function of its local information, i.e. $q_t = \gamma^e_t(\mathcal{I}^e_t)$, where $\mathcal{I}^e_t = \{x_{[0,t]}, q'_{[0,t-1]}\}$ and $q_t \in \mathcal{M}$, the channel input alphabet given by $\mathcal{M} = \{1, 2, \ldots, M\}$, for $0 \leq t \leq T-1$. Here, we use the notation $x_{[0,t-1]} = \{x_s, 0 \leq s \leq t-1\}$ for $t \geq 1$.

The channel maps q_t to q'_t in a stochastic fashion so that $P(q'_t|q_t, q_{[0,t-1]}, q'_{[0,t-1]})$ is a conditional probability measure on $\mathcal{M}'$ for all $t \in \mathbb{Z}_+$. If this expression equals $P(q'_t|q_t)$, the channel is said to be memoryless, i.e., the past variables do not affect the channel output q'_t given the current input q_t.

The receiver, upon receiving the information from the channel, generates an estimate $\hat{x}_t$ at time t, also causally: An admissible causal estimation policy is a sequence of functions $(\gamma_t^d)_{t\in\mathbb{Z}_+}$ such that $\hat{x}_t = \gamma_t^d(q'_{[0,t]})$ with

$$\gamma_t^d : (\mathcal{M}')^{t+1} \to X, \quad t \geq 0.$$

For a given $\epsilon > 0$, we denote by C_ϵ the smallest channel capacity above which there exist an encoder and an estimator so that one of the following estimation objectives is achieved:

(E1) Eventual almost sure stability of the estimation error: There exists $T(\epsilon) \geq 0$ so that

$$\sup_{t\geq T(\epsilon)} d(x_t, \hat{x}_t) \leq \epsilon \quad \text{a.s.} \tag{3}$$

(E2) Asymptotic almost sure stability of the estimation error:

$$P\big(\limsup_{t\to\infty} d(x_t, \hat{x}_t) \leq \epsilon\big) = 1. \tag{4}$$

(E3) Asymptotic quadratic stability of the estimation error in expectation:

$$\limsup_{t\to\infty} E[d(x_t, \hat{x}_t)^2] \leq \epsilon. \tag{5}$$

It is easy to see that (3) implies (4) and (5). On the other hand, (4) and (5) do not imply one another in general.

The primary goal of the paper is to find C_ϵ and in particular

$$C_0 := \lim_{\epsilon\downarrow 0} C_\epsilon. \tag{6}$$

Observe that this limit exists as a number in $[0,\infty]$, since C_ϵ is non-decreasing as $\epsilon \downarrow 0$.

III. Bounds through a dynamical systems approach

In this section, we assume that the channel is noiseless. Under this assumption, we provide lower bounds of C_0 for the estimation objectives (3) and (4).

We consider the space $X^{\mathbb{Z}_+}$ of all sequences in X, equipped with the product topology. We write $\bar{x} = (x_0, x_1, x_2, \ldots)$ for the elements of $X^{\mathbb{Z}_+}$. A natural dynamical system on $X^{\mathbb{Z}_+}$ is the shift map $\theta : X^{\mathbb{Z}_+} \to X^{\mathbb{Z}_+}$, $(\theta\bar{x})_t \equiv x_{t+1}$, which is continuous with respect to the product topology. An analogous shift map is defined on $W^{\mathbb{Z}_+}$ and denoted by ϑ.

Observing that the sequence of random variables $(x_t)_{t\in\mathbb{Z}_+}$ forms a Markov chain, the following lemma shows how a stationary measure of this Markov chain defines an invariant measure for θ.

Lemma 3.1: Let π be a stationary measure of the Markov chain $(x_t)_{t\in\mathbb{Z}_+}$. Then an invariant Borel probability measure μ for θ is defined on cylinder sets by

$$\begin{aligned}&\mu(B_0 \times B_1 \times \cdots \times B_n \times X^{[n+1,\infty)})\\ &:= \int_{B_0\times\cdots\times B_n} \pi(\mathrm{d}x_0)P(\mathrm{d}x_1|x_0)P(\mathrm{d}x_2|x_1)\cdots P(\mathrm{d}x_n|x_{n-1}),\end{aligned}$$

where $B_0, B_1, \ldots, B_n$ are arbitrary Borel sets in X. Here

$$P(x_{n+1} \in B|x_n = x) = P(f(x_n, w) \in B|x_n = x).$$

The support of μ is contained in the closure of the set of all trajectories, i.e., $\operatorname{supp}\mu \subset \mathrm{cl}\mathcal{T}$ with

$$\mathcal{T} := \big\{\bar{x} \in X^{\mathbb{Z}_+} \,:\, \exists w_t \in W \text{ s.t. } x_{t+1} \equiv f(x_t, w_t),\ t \in \mathbb{Z}_+\big\}.$$

Theorem 3.1: Consider the estimation objective (3) for an initial measure π_0 which is stationary under the Markov chain $(x_t)_{t\in\mathbb{Z}_+}$. If $\operatorname{supp}\mu$ is not compact, we have $C_0 = \infty$. Otherwise,

$$C_0 \geq h_{\mathrm{top}}(\theta|_{\operatorname{supp}\mu}).$$

As a consequence, the metric entropy of θ with respect to μ is also a lower bound of C_0.

Remark 3.1: To make the statement of the theorem clearer, let us consider the two extreme cases when there is no noise and when there is only noise: **(i)** If the system is deterministic, i.e., $x_{t+1} = f(x_t)$ for a homeomorphism $f : X \to X$ of a compact metric space X, then π_0 is an invariant measure of f. Moreover, $P(x_t \in B|x_{t-1} = x) = 1$ if $f(x) \in B$ and 0 otherwise, implying

$$\begin{aligned}&\mu(B_0 \times B_1 \times \cdots \times B_n \times X^{[n+1,\infty)})\\ &= \pi_0\big(B_0 \cap f^{-1}(B_1) \cap f^{-2}(B_2) \cap \ldots \cap f^{-n}(B_n)\big).\end{aligned}$$

From this expression, we see that the support of μ is contained in the set $\mathcal{T}$ of all trajectories of f (which in this case coincides with its closure). The map $h : \mathcal{T} \to X$ defined by $h(\bar{x}) := x_0$, is easily seen to be a homeomorphism, which conjugates $\theta|_{\mathcal{T}}$ and f. That is, the following diagram commutes:

$$\begin{array}{ccc} \mathcal{T} & \xrightarrow{\theta} & \mathcal{T} \\ {\scriptstyle h}\big\downarrow & & \big\downarrow{\scriptstyle h} \\ X & \xrightarrow[f]{} & X \end{array}$$

Since $\mu(h^{-1}(\cdot)) = \pi_0(\cdot)$ and conjugate systems have the same entropy, our theorem implies

$$C_0 \geq h_{\mathrm{top}}(f; \operatorname{supp}\pi_0).$$

The right-hand side of this inequality is finite under mild assumptions. These conditions are in particular satisfied if f is a diffeomorphism on a finite-dimensional manifold. However, one should be aware that even on a compact interval there exist continuous maps with infinite topological entropy on the support of an invariant measure.

(ii) Assume that $X = W$ is compact and the system is given by $x_{t+1} = w_t$, i.e., the trajectories are only determined by the noise. In this case, with $\pi_0 := \nu$, the measure μ is the product measure $\nu^{\mathbb{Z}_+}$. Hence, C_0 is bounded below by the topological entropy of the shift on $W^{\mathbb{Z}_+}$ restricted to $\operatorname{supp}\nu^{\mathbb{Z}_+} = (\operatorname{supp}\nu)^{\mathbb{Z}_+}$. This number is finite if and only if $\operatorname{supp}\nu$ is finite and in this case is given by $\log|\operatorname{supp}\nu|$.

If the system is not deterministic, then usually $C_0 = \infty$. In fact, this is always the case when the noise can be recovered from the state trajectory [7] to a sufficiently large extent. The following corollary treats the case, when the noise can be recovered completely from the state trajectory.

Corollary 3.1: Additionally to the assumptions in Theorem 3.1, suppose that W and X are compact and $f^x : W \to X$ is invertible for every $x \in X$ so that $(x, y) \mapsto (f^x)^{-1}(y)$ is continuous. Then

$$C_0 \geq h_{\mathrm{top}}(\Phi|_{\operatorname{supp}(\pi_0\times\nu^{\mathbb{Z}_+})}) \geq h_{\mathrm{top}}(\vartheta|_{\operatorname{supp}\nu^{\mathbb{Z}_+}}), \tag{7}$$

where $\Phi : X \times W^{\mathbb{Z}_+} \to X \times W^{\mathbb{Z}_+}$ is the skew-product map $(x, \bar{w}) \mapsto (f_{w_0}(x), \vartheta\bar{w})$. As a consequence, $C_0 = \infty$ whenever $\operatorname{supp} \nu$ contains infinitely many elements.

Remark 3.2: One might wonder why it is not immediate that $C_0 = \infty$, since we want to achieve a state estimate of arbitrarily small error from information sent through a fixed channel. Indeed, it would be obvious that $C_0 = \infty$ (if X is infinite) if instead of (3) we wanted to achieve $d(x_t, \hat{x}_t) \leq \epsilon$ almost surely for $t \geq 0$. Since we are allowing a waiting time $T(\epsilon)$, however, it becomes possible to achieve the estimation objective with a channel of finite capacity for most deterministic systems. This is shown in [15] and Theorem 6.1 below.

For the asymptotic estimation objective (4) we can prove the following result.

Theorem 3.2: Consider the estimation objective (4) for an initial measure π_0 which is stationary and ergodic under the Markov chain $(x_t)_{t\in\mathbb{Z}_+}$. Then, if $\operatorname{supp}\mu$ is compact,

$$C_0 \geq h_\mu(\theta).$$

IV. An information-theoretic view

In this section, we again allow noise in the channel and assume that $X = \mathbb{R}^N$. We provide further impossibility results via information-theoretic methods. These results will shed more light on the precise conditions that force C_0 to be infinite.

Theorem 4.1: Consider system (2) with state space $X = \mathbb{R}^N$. Suppose that

$$\limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-1} h(x_t|x_{t-1}) > 0$$

and $h(x_t) < \infty$ for all $t \in \mathbb{Z}_+$. Then, under (5) (and thus under (3)),

$$C_\epsilon \geq \left(\liminf_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T-1} h(x_t|x_{t-1}) \right) - \frac{N}{2} \log(2\pi e \epsilon).$$

In particular, $C_0 = \infty$.

The proof of the above theorem, given in [7], in part builds on what is known as Shannon's Lower Bounding technique (see e.g. [10]).

For the asymptotic almost sure criterion (4) for additive noise systems on $\mathbb{R}^N$ we have the following theorem.

Theorem 4.2: Suppose that $X = W = \mathbb{R}^N$ and the system is given by $x_{t+1} = f(x_t) + w_t$, where the noise measure ν admits a bounded density which is positive everwhere. Then, under (4), $C_0 = \infty$.

V. A random dynamical systems view

In this section, the system (2) is viewed as a random dynamical system (briefly an RDS). We present a result showing that the metric entropy of the RDS is a lower bound on C_0 for the objective (4), and hence also for (3), under the assumption of a compact state space and a noiseless channel.

To this end, we first build a random dynamical system associated to the equation (2). The base space is defined as $B := W^{\mathbb{Z}_+}$; for its elements we use the notation $\bar{w} = (w_t)_{t\geq 0}$. The σ-field on B is the product field of the Borel σ-field on W (generated by cylinder sets), and the measure on this space is $\nu^{\mathbb{Z}_+}$, the corresponding product measure of ν. The dynamics on B is given by the left shift operator $(\theta\bar{w})_t = w_{t+1}$, which preserves the measure $\nu^{\mathbb{Z}_+}$ and is easily seen to be ergodic. A random dynamical system over θ is given by

$$\begin{aligned} \varphi(t, \bar{w})x &:= f_{w_{t-1}} \circ \cdots \circ f_{w_1} \circ f_{w_0}(x), \\ \varphi &: \mathbb{Z}_+ \times B \to \operatorname{Homeo}(X). \end{aligned}$$

The associated skew-product transformation is given by

$$\begin{aligned} &\Phi : \mathbb{Z}_+ \times B \times X \to B \times X, \\ &\quad (t, (\bar{w}, x)) \mapsto \Phi_t(\bar{w}, x) = (\theta^t\bar{w}, \varphi(t, \bar{w})x). \end{aligned}$$

We actually work with a time-invertible extension of φ, which replaces B with $B^* := W^{\mathbb{Z}}$ (two-sided sequences) endowed with the measure $\nu^{\mathbb{Z}}$ and the shift operator θ^* on two-sided sequences. Since f_w is invertible by assumption for every $w \in W$, it is evident that also the cocycle φ over (B, θ) can be extended to a cocycle $\varphi^* : \mathbb{Z} \times B^* \times X \to X$ over (B^*, θ^*). For simplicity, we drop the superscript $*$ again and just write (θ, φ) for the time-invertible extension.

Consider a probability measure μ on $B \times M$, invariant under the time-1-map $\Phi_1 : B \times M \to B \times M$, with marginal $\nu^{\mathbb{Z}}$ on B. Then μ is called a *φ-invariant measure*.

The general definition of the metric entropy $h_\mu(\varphi)$ of an RDS φ with respect to an invariant measure μ can be found in [2]. In the proof of the following theorem we use a characterization of $h_\mu(f)$ for ergodic measures μ due to Katok [4], which was generalized to RDS by Zhu [20].

Theorem 5.1: Consider the objective (4) to be achieved for an initial measure π_0 which is equivalent to an ergodic measure π of the Markov chain $\{x_t\}_{t\geq 0}$. Then there exists an ergodic measure μ for the RDS φ such that $h_\mu(\varphi) < \infty$ implies

$$C_0 \geq h_\mu(\varphi).$$

It should be mentioned that the metric entropy $h_\mu(\varphi)$ can be expressed in terms of the positive Lyapunov exponents of the RDS φ, given that the state space is a smooth manifold and the system satisfies some regularity assumptions. In particular, this includes that f is twice differentiable w.r.t. x and the invariant measure μ satisfies the so-called SRB property, cf. [8].

VI. Almost sure estimation for noise-free systems

In this section, we consider the estimation objective (4) in the case when there is only noise in the channel, but not in the source.

A. A general result for non-linear systems

The following result shows that in case of a noise-free system, the topological entropy provides an upper bound on C_0 for the objective (4). The proof uses similar arguments as employed in [13] for the case of a noiseless channel.

Theorem 6.1: Consider a non-linear deterministic system $x_{t+1} = f(x_t)$ on a compact state space X, estimated via a discrete memoryless channel (DMC). Then, for the asymptotic estimation objective (4), $C_0 \leq h_{\text{top}}(f)$.

The proof of the result crucially depends on the fact that the system is deterministic and is admittedly impractical to implement. For noise-free linear systems there exist stationary policies which guarantee the almost sure estimation criterion, studied in the following.

B. Refinement for linear systems

We consider a noiseless linear system

$$x_{t+1} = Ax_t \tag{8}$$

with $x_t \in \mathbb{R}^N$. The following result gives a positive answer to the question whether the estimation objective (4) can be achieved, when no noise is present in the system, via an erasure channel of finite capacity. Recall that such a channel has $\mathcal{M} = \{0, 1\}$ and $\mathcal{M}' = \{0, 1, e\}$, with

$$P(q'|q) = (1-p)1_{\{q'=q\}} + p1_{\{q'=e\}},$$

where e denotes the erasure symbol and p is the erasure probability.

Theorem 6.2: Consider system (8) estimated over a memoryless erasure channel with finite capacity. Then, for (4), we have

$$C_0 \le \sum_{|\lambda_i|>1} \log(\lceil|\lambda_i|\rceil). \tag{9}$$

Sahai and Mitter [17, Corollary 5.3, Theorem 4.3] show that for a discrete memoryless channel indeed it suffices that $C > \sum_{|\lambda_i|\ge 1} \log(|\lambda_i|)$ for the existence of encoder and controller policies leading to stochastic stability. This result is in agreement with Theorems 6.1 and 6.2.

VII. An Example

In this section, we illustrate the concepts and results of the paper with an example.

Example 7.1: Consider the diffeomorphism $f_A : \mathbb{T}^2 \to \mathbb{T}^2$ on the 2-torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$, induced by the linear map

$$A = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \tag{10}$$

i.e., $f_A(x + \mathbb{Z}^2) = Ax + \mathbb{Z}^2$. Note that the inverse of f_A is given by $f_{A^{-1}}$, which is well-defined, since $\det A = 1$. The map f_A is known as *Arnold's Cat Map*, and is probably the simplest example of an Anosov diffeomorphism.

Since $\det \mathrm{D}f_A(x) \equiv \det A \equiv 1$, the map f_A is area-preserving. The eigenvalues of the matrix A are given by

$$\gamma_1 = -\frac{3}{2} - \frac{1}{2}\sqrt{5} \quad \text{and} \quad \gamma_2 = -\frac{3}{2} + \frac{1}{2}\sqrt{5}$$

and satisfy $|\gamma_1| > 1$, $|\gamma_2| < 1$. It is well-known that both the topological entropy and the metric entropy of f_A with respect to Lebesgue measure are given by $\log|\gamma_1| > 0$. Hence, Theorem 6.1 yields

$$C_0 \le \log\left|-\frac{3}{2} - \frac{1}{2}\sqrt{5}\right| \approx 1.3885$$

for the objective (4) to be achieved over a DMC.

Now, suppose we have additive noise for the cat map, so that $f_A(x+\mathbb{Z}^2) = Ax+w+\mathbb{Z}^2$, with $w \sim \nu$ which admits a density supported on $\mathbb{T}^2$. In this case, the map $f^x : \mathbb{T}^2 \to \mathbb{T}^2, w \mapsto Ax + w$ is invertible and $(x, y) \mapsto (f^x)^{-1}(y) = y - Ax$ is continuous. By Corollary 3.1, $C_0 = \infty$ for the estimation objective (3), under a stationary initial measure. For the objective (4) it can be shown that, under corresponding initial measure conditions, Theorem 3.2 leads to $C_0 = \infty$, whereas Theorem 5.1 leads to a lower bound of $\log\left|-\frac{3}{2} - \frac{1}{2}\sqrt{5}\right|$.

VIII. Conclusion

In this paper, we considered three estimation objectives for stochastic non-linear systems $x_{t+1} = f(x_t, w_t)$ with i.i.d. noise (w_t), assuming that the estimator receives state information via a noisy channel of finite capacity.

References

[1] T. Berger. Information rates of Wiener processes. *IEEE Transactions on Information Theory*, 16:134–139, 1970.

[2] T. Bogenschütz. Entropy, pressure, and a variational principle for random dynamical systems. *Random Comput. Dynam*, 1(1):99–116, 1992.

[3] R. M. Gray and T. Hashimoto. A note on rate-distortion functions for nonstationary Gaussian autoregressive processes. *IEEE Transactions on Information Theory*, 54:1319–1322, March 2008.

[4] A. Katok. Lyapunov exponents, entropy and periodic orbits for diffeomorphisms. *Inst. Hautes Études Sci. Publ. Math.*, (51):137–173, 1980.

[5] A. Katok, B. Hasselblatt. Introduction to the modern theory of dynamical systems. Vol. 54. Cambridge university press, 1997.

[6] C. Kawan. State estimation under communication constraints. *arXiv preprint arXiv:1605.03210*, 2016.

[7] C. Kawan and S. Yüksel. Entropy bounds on state estimation for stochastic non-linear systems under information constraints. *arXiv preprint arXiv:1612.00564*, 2016.

[8] F. Ledrappier and L.-S.Young. Entropy formula for random transformations. *Probability theory and related fields*, 80(2):217–240, 1988.

[9] D. Liberzon and S. Mitra. Entropy and minimal data rates for state estimation and model detection. In *Proceedings of the 19th International Conference on Hybrid Systems: Computation and Control*, pp. 247–256. ACM, 2016.

[10] T. Linder and R. Zamir. Causal coding of stationary sources and individual sequences with high resolution. *IEEE Transactions on Information Theory*, 52:662–680, February 2006.

[11] A. S. Matveev. State estimation via limited capacity noisy communication channels. *Mathematics of Control, Signals, and Systems*, 20:1–35, 2008.

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

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

[14] A. S. Matveev and A. V. Savkin. *Estimation and Control over Communication Networks*. Birkhäuser, Boston, 2008.

[15] A. Yu. Pogromsky and A. S. Matveev. A topological entropy approach for observation via channels with limited data rate. *IFAC Proceedings Volumes*, 44(1):14416–14421, 2011.

[16] A. Sahai. Coding unstable scalar Markov processes into two streams. In *Proceedings of the IEEE International Symposium on Information Theory*, page 462, 2004.

[17] 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*, 52(8):3369–3395, 2006.

[18] S. Yüksel. Stationary and ergodic properties of stochastic non-linear systems controlled over communication channels. *SIAM Journal on Control and Optimization*, pp. 2844–2871, 2016.

[19] S. Yüksel and T. Başar. *Stochastic Networked Control Systems: Stabilization and Optimization under Information Constraints*. Birkhäuser, New York, NY, 2013.

[20] J. J. Zhu. Two notes on measure-theoretic entropy of random dynamical systems. *Acta Mathematica Sinica*, 25(6):961–970, 2009.