On the Threshold Effect in the Estimation of Chaotic Sequences

Ilan Hen* and Neri Merhav
Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, ISRAEL
{ilanh@tx, merhav@ee}.technion.ac.il

Abstract - Chaotic sequences and chaotic dynamical systems are attractive candidates for use in signal synthesis and analysis as well as in communications applications. In previous works, various methods for the estimation of chaotic sequences under noise were developed. However, although the methods were different, their qualitative performance was the same: for high SNR the performance was good, but below some threshold SNR, a sharp degradation in performance occurred. We quantify this threshold effect and derive lower bounds on the value of the threshold SNR. Using information-theoretic tools, we prove that for any ergodic chaotic system, there is a certain threshold SNR level, below which the ratio between the mean square error obtained by any estimator of the system's initial state at the output of AWGN channel and the Bayesian Cramér-Rao bound increases exponentially fast as the number of observations, N, grows without bound. We derive lower bounds on SNR_{th}, the value of the threshold SNR, as a function of the system's Lyapunov exponent. Our bounds have two versions, one for a finite number of observations, and one for the asymptotic regime as $N \to \infty$. We explain the connection between the existence of threshold effect in the estimation process of chaotic sequences and the converse to the joint source-channel coding theorem. We demonstrate our results on the chaotic system governed by the r-diadic map.

Index Terms - Chaotic systems, Kolmogorov-Sinai entropy, Lyapunov exponent, symbols of dynamical systems, Bayesian Cramér-Rao bound, joint source-channel coding theorem, channel capacity.

*This paper summarizes part of the M.Sc. thesis of the first author.

1 Introduction

Chaotic discrete-time dynamical systems are used in various applications due to there ability to generate highly complicated signals by a simple recursive procedure. Chaotic sequences are attractive candidates for use in signal analysis, signal synthesis, practical engineering and communications applications. Chaotic systems are used as models for wide range of signal processing applications [18] as well as for practical engineering systems like analog-to-digital converters [26] and power converters [6]. Chaotic systems have the potential to give rise to good joint source-channel codes [3] due to there ability to separate orbits of nearby initial states while maintaining global boundness, thus conforming with energy and peak amplitude constraints. Chaotic signals are also used in spread spectrum applications [8], where the power of the transmitted signal is spread across a broad range of frequencies.

In most of the above applications, there is a frequent need to estimate the chaotic sequence from noisy observations. Farmer and Sidorowich [10], based on the results of Hammel [11], proposed to estimate the chaotic sequence as the sequence whose Euclidean distance to the noisy sequence is minimum, under the constraint of obeying the dynamics of the chaotic system. As this extremum problem turned out to be rather difficult, they have reformulated the problem by using a linearization of the chaotic system, which leads to the so-called manifold decomposition approach for estimating chaotic sequences under noise. This approach is based on the observation that chaotic systems work in two kinds of directions, stable and unstable, and it deals with errors along these directions accordingly. Cuomo and Oppenheim [5] made use of the self-synchronization property of some class of chaotic systems to recover chaotic sequences from noisy observations. The property of self-synchronization [22] refers to the ability of some chaotic systems to synchronize independently. Cuomo and Oppenheim focused on the Lorenz system and compared their self-synchronization approach to extended Kalman filtering. Cazelles and Boudjema [2], So, Ott and Dayawansa [29], and Singer, Wornell and Oppenheim [28] proposed methods for recovering chaotic sequences from noise-corrupted observations which are based on different techniques from control theory. Myers, Kay and Richard [17] considered the maximum likelihood estimator of chaotic sequences. They reported that, due to the chaotic nature of the sequences, the corresponding log-likelihood function is highly irregular, containing multiple local maxima which are very narrow, and so their detection is difficult. They also derived the parametric Cramér-Rao bound associated with the estimation of the initial state of a multidimensional chaotic system. Papadopoulos and Wornell [20] derived the maximum likelihood estimator of chaotic sequences generated by the so called "tent" map. They showed that the maximum likelihood estimator is implementable as a certain extended Kalman filter, whose computational complexity grows linearly with the number of observations . However, as shown by Chen [3] later on, not only does not the maximum likelihood estimator achieve the Cramér-Rao bound, but neither it is a consistent estimator for every value of the SNR. Kay and Nagesha [15] developed a dynamical programing (DP) algorithm for the realization of the maximum likelihood estimation of chaotic sequences. This algorithm is similar to the Viterbi algorithm. However, the analytical derivation of the DP algorithm is very hard due to the nonlinear nature of chaotic systems. In fact, the maximum likelihood estimator of the tent map sequences is the only one that could have been obtained analytically using Kay's DP algorithm. Kay [14] investigated the asymptotic performance of the maximum likelihood estimator of chaotic sequences. He showed, under some regulation conditions, that when $SNR \to \infty$ the estimation error is asymptotically normally distributed with zero mean and variance given by the Cramér-Rao bound, i.e., the maximum likelihood estimator is asymptotically unbiased and efficient. However, when $SNR \to 0$, violating a necessary condition [27], [30] for consistency of least squares (LS) estimator, the maximum likelihood estimator becomes inconsistent.

Although the above mentioned approaches for estimation of chaotic sequences under noise are different, their qualitative performance remains the same: good performance for high enough SNR but sharp degradation in performance below a certain level of the SNR. In other words, in the estimation process of the chaotic sequences, there is a threshold effect. In this paper, we prove the existence of the threshold effect in the estimation of chaotic. We derive lower bound on the threshold SNR as a function of the system's Lyapunov exponent and the mutual information between the chaotic sequence and the noisy observations. This bound is further simplified to a bound which depends on the system's Lyapunov exponent and the power spectrum of the chaotic sequence. We explain how the existence of the threshold effect actually derives from the converse to the joint source-channel coding theorem. We obtain the bound on the threshold SNR by comparing the Bayesian Cramér-Rao bound to a bound on the estimation error mean square error which is derived using the data processing inequality. Although the data processing inequality does not yield a tight bound (due to the suboptimality of the chaotic system for coding purposes), the interesting point here is that it is still powerful enough to yield a better bound than the Cramér-Rao bound below the threshold SNR.

The outline of the paper is as follows. In Section 2, we review the basics of dynamical systems, focusing on issues that will be needed in the next sections. Readers that are familiar with these basics may skip to Section 3 where we quantify the threshold effect for the asymptotic regime. In Section 4, we derive a lower bound on the threshold SNR. In Section 5, we consider the threshold effect in the case where the number of observations is finite. In Section 6, we demonstrate our results on the chaotic system governed by the r-diadic map. Section 7 contains some concluding remarks and directions for future research.

2 Preliminaries

In this section, we review basic terminology, properties and features of dynamical systems. The review brought here is based on [1], [13], [16], [19], [23], [24]. Throughout this paper we denote random variables with upper case letters and their realization with lower case letters. In general, there are two kinds of dynamical systems: continuous-time dynamical systems and discrete-time dynamical systems. The continuous-time systems are described by partial differential equations, whereas the discrete time systems are described by difference equations. In our work, we focus on discrete-time dynamical systems only. A discrete-time dynamical system is defined on a probability space $(\mathcal{X}, \mathcal{A}, \mu)$ and described by the following difference equation

$$\begin{aligned} &x[0] = x_0 \in \mathcal{X} \\ &x[n] = F(x[n-1]) \qquad n = 1, 2, 3, \ldots \end{aligned} \tag{1}$$

the transformation $F : \mathcal{X} \to \mathcal{X}$ is the system's *map*, $\mathcal{X}$ is the *state space* and x_0 is the system's *initial state*. The probability space $(\mathcal{X}, \mathcal{A}, \mu)$ together with the system map $F : \mathcal{X} \to \mathcal{X}$, constitutes the dynamical system, denoted here by $\{(\mathcal{X}, \mathcal{A}, \mu), F\}$. If we denote the n-th order composition of the map F by F^n then $x[n] = F^n(x_0)$. The sequence of states $O_F^+(x_0) \triangleq \{F^n(x_0)\}_{n=0}^{\infty}$ is an *orbit* of the system. In this paper, we focus on the dynamical systems which are defined on the *Borel* probability space, $\{([a,b], \mathcal{B}, \mu_B), F\}$. One class of such systems, which is interesting and mathematically amenable, is the class of systems which are governed by the *eventually expanding piecewise linear Markov* maps. A map is eventually expanding, piecewise linear Markov map if the following conditions are satisfied:

- There exists a set of points, called the *partition points*, $a = \alpha_0 < \alpha_1 < \ldots < \alpha_L = b$, such that in each interval $\Delta_i = [\alpha_{i-1}, \alpha_i)$ the map F is affine.
- Partition points are mapped to partition points, i.e., for all α_i, $i = 0, 1, \ldots, L$, $F(\alpha_i) = \alpha_j$ for some $j = 0, 1, \ldots, L$.
- The map is eventually expanding map,i.e., $\exists n \in \mathcal{N}, \quad \inf_{x \in [a,b]} \left| \frac{dF^n(x)}{dx} \right| > 1.$

An example of such system is the one governed by the r-diadic map, $\{([0,1), \mathcal{B}, \mu_B), F_r\}$,

$$F_r(x) = (rx) \text{mod} 1 \tag{2}$$

where r is an integer greater than one.

A. *Invariant Density*

When the system's initial state is a random variable the sequence of states is a stochastic process, whose stochastic properties are derived exclusively from the distribution of initial state. The following questions naturally arise: What is the probability law of this stochastic process? Is it stationary?

Is it ergodic? The answers to these questions are based on understanding how does the density function of the state $X[n]$ evolve as the system iterates, i.e., given the density function of $X[n-1]$, denoted here by $f_{n-1}(x)$, what is the density function of $X[n]$, $f_n(x)$? The density function $f_n(x)$ is obtained recursively as a functional of $f_{n-1}(x)$ by using the *Frobenius-Perron (FP)* operator [16]. The FP operator associated with the system $\{([a,b],\mathcal{B},\mu_B),F\}$ is given by

$$FPf(x)=\frac{d}{dx}\int_{F^{-1}([a,x])} f(\alpha)d\alpha, \quad x\in[a,b] \tag{3}$$

where $f\in L_1(a,b)$. If the system $\{([a,b],\mathcal{B},\mu_B),F\}$ initial state x_0 is a random variable with density f_0 then it can be shown [16] that

$$f_n(x)=FPf_{n-1}(x) \tag{4}$$

The stochastic process $\{X[n]\}$ is, in general, not stationary. A density f_0 for which $f_0=f_1=f_2=\dots$ is called an *invariant density*. In other words, the invariant density is a fixed point of the Frobenius-Perron operator

$$FPf=f. \tag{5}$$

A fixed point does not necessarily exist and if exists, may not be unique (the conditions required for the existence and uniqueness of a fixed point are given by the *fixed point theorem*). When f_0 is an invariant density, it is straightforward to show that the stochastic process generated by the dynamical system is wide sense stationary. Furthermore, if the invariant density is unique, it is also *ergodic* [16] i.e., time averages equal ensemble averages with probability one. The map F is then called ergodic. For example, in the case of the r-diadic map, the uniform density is easily shown to be invariant. This density is unique [16] and therefore the stochastic process generated by the system is wide sense stationary ergodic process. From now on, throughout this paper, we consider the dynamical system $\{([a,b],\mathcal{B},\mu_f),F\}$ where F is ergodic, f is the invariant density and

$$\mu_f(A)\triangleq\int_A f(x)dx \quad A\in\mathcal{B}. \tag{6}$$

B. *Kolmogorov-Sinai Entropy*

The *Kolmogorov-Sinai (KS)* entropy is the greatest average amount of information that the dynamical system produces about its initial state per iteration and therefore the additional minimum amount of information, required on the average, in each iteration to maintain an arbitrarily fine localization of the initial state. Consider a partition $\xi=\{A_k\}_{k=0}^{K-1}$ of $[a,b]$. The *entropy of the partition* ξ is given by

$$H_{\mu_f}(\xi)\triangleq-\sum_{k=0}^{K-1}\mu_f(A_k)\ln\mu_f(A_k). \tag{7}$$

When no iterations of F on the system initial state x_0 are considered, we need, on the average, $H_{\mu_f}(\xi)$ nats in order to determine in which set A_k the initial state X_0 is. When $N-1$ iterations are considered

$$H_{\mu_f}\left(\bigvee_{n=0}^{N-1} F^{-n}(\xi)\right) \tag{8}$$

nats are needed to determine which set of

$$\bigvee_{n=0}^{N-1} F^{-n}(\xi) \triangleq \left\{A_{k_1} \bigcap F^{-1}(A_{k_2}) \bigcap \ldots \bigcap F^{-(N-1)}(A_{k_N})\right\}_{k_1,k_2,\ldots,k_N=0}^{K-1}, \tag{9}$$

where

$$F^{-n}(A) \triangleq \left\{x \in [a,b] \middle| F^n(x) \in A\right\}, \tag{10}$$

includes X_0. The KS entropy, $H_{\mu_f}(F)$, is given [13] by

$$H_{\mu_f}(F) \triangleq \sup_{\xi: H_{\mu_f}(\xi)<\infty} H_{\mu_f}(F,\xi) \tag{11}$$

where

$$H_{\mu_f}(F,\xi) \triangleq \lim_{N\to\infty} \frac{1}{N} H_{\mu_f}\left(\bigvee_{n=0}^{N-1} F^{-n}(\xi)\right) \tag{12}$$

where the limit always exists [13]. A partition ξ achieving the supremum is called a *generating partition.* For example, in the case of the r-diadic map the partition

$$\xi = \left\{\left[\frac{i-1}{r}, \frac{i}{r}\right)\right\}_{i=1}^{r} \tag{13}$$

is a generating partition [13] and

$$H_{\mu_f}(F) = \ln r \quad \text{nats.} \tag{14}$$

C. *Chaotic Systems*

Chaotic dynamical systems are characterized by high sensitivity to the initial state. In chaotic systems, the orbits $O_F^+(x_0)$ and $O_F^+(x_0+\varepsilon)$ originating from nearby initial states, x_0 and $x_0+\varepsilon$, $\varepsilon << 1$, respectively, diverge exponentially fast as N grows large. Chaotic systems exhibit high local instability while maintaining global boundness. A commonly used definition of chaotic system is that based on the statistical quantities called the *Lyapunov exponents* [19]. The *Lyapunov exponent*, $\lambda(x)$, associated with the dynamical system $\{([a,b],\mathcal{B},\mu_f),F\}$ is defined as

$$\lambda(x) \triangleq \lim_{N\to\infty} \frac{1}{N} \ln\left|\frac{dF^{N-1}(x)}{dx}\right|, \qquad x \in [a,b] \tag{15}$$

provided that the limit exists. In general, $\lambda(x)$ depends on x. However, when F is ergodic, *Oseledec's* Theorem [19] tells us that

$$\lambda(x) = \lambda \triangleq E \ln\left|\dot{F}(X)\right| \quad a.e. \tag{16}$$

$$\dot{F}(x) \triangleq \frac{dF(x)}{dx}. \tag{17}$$

where the expectation is with respect to the invariant density f, and is assumed to exist. If $\lambda > 0$, the system is chaotic, otherwise the orbits eventually converge to a fixed point or behave periodically. For example, in the case of the r-diadic map we have

$$\lambda = E \ln\left|\dot{F}_r(X)\right| = \ln r. \tag{18}$$

The r-diadic map is defined for $r > 1$, so $\lambda > 0$, and therefore the system is chaotic.

D. *Symbols of Dynamical Systems*

Orbits of dynamical systems can be represented by using the so called *symbols* of the system, denoted here by $\{S[n]\}_{n=0}^{N-1}$. A symbol $S[n] \in \mathcal{S} = \{s_0, \dots, s_{K-1}\}$ indicates to which subset of the generating partition $\xi = \{A_0, \dots, A_{K-1}\}$, the state $X[n]$ belongs, i.e.,

$$S[n] = s_k \quad \text{if} \quad X[n] \in A_k. \quad k = 0, 1, \dots, K-1. \tag{19}$$

In the case of a non-invertible map F, knowing $X[n]$ and $S[n-1] \in \mathcal{S}$ is sufficient to determine $X[n-1]$ and then it can be written that

$$X[n-1] = \phi(S[n-1], X[n]). \tag{20}$$

where the map ϕ is referred to as the inverse of F given symbol. Using (20), an orbit of a dynamical system, $\{X[n]\}_{n=0}^{N-1}$, can be represented in the symbol-based form, $(\{S[n]\}_{n=0}^{N-2}, X[N-1])$. Sometimes there is an advantage in using the symbol-based representation like in [20] where it was used to derive the maximum likelihood estimator of the chaotic "tent" map sequences. The symbols $\{S[n]\}_{n=0}^{N-1}$ are identically distributed, discrete-valued random variables with the probability law

$$\mathcal{P}_r(S[n] = s_k) = \int_{A_k} f(x)dx, \quad k = 0, \dots, K-1. \tag{21}$$

For example, in the case of the r-diadic map we have that

$$S[n] = k \quad \text{if} \quad X[n] \in \left[\frac{k-1}{r}, \frac{k}{r}\right). \quad k = 0, 1, \dots, r-1 \tag{22}$$

$$\phi(s, x) = \frac{1}{r}x + \frac{1}{r}s, \tag{23}$$

$$X[n] = \sum_{i=n}^{N-2} S[i] r^{-(i-n+1)} + \frac{1}{r^{N-n-1}} X[N-1], \quad n = 0, 1, \dots, N-1 \tag{24}$$

and the symbols are uniformly distributed across $\mathcal{S} = \{0, 1, \dots, r-1\}$.

For chaotic systems, we can expect that as N increases, the subsets visited during the system evolution become more and more different for different initial states and in the limit as $N \to \infty$, they determine exclusively the initial state $X[0]$. This intuition is true. There is a one-to-one map a.e. between the initial state $X[0]$ and the infinite sequence of symbols $\{S[n]\}_{n=0}^{\infty}$ [13]. For example, in the case of the r-diadic map (assigning $n = 0$ and taking $N \to \infty$ in (24)) the infinite sequence of symbols is the infinite-length radix-r representation of $X[0]$, which is known to constitutes one-to-one map a.e.

The symbols $\{S[n]\}_{n=0}^{N-1}$ completely determine the subset to which the initial state belongs, therefore $H_{\mu_f}\left(\bigvee_{n=0}^{N-1} F^{-n}(\xi)\right) = H\left(\{S[n]\}_{n=0}^{N-1}\right)$. Thus

$$H_{\mu_f}(F) = \lim_{N \to \infty} \frac{1}{N} H\left(\{S[n]\}_{n=0}^{N-1}\right) = H(S). \tag{25}$$

where $H(S)$ is the entropy rate of $S[n]$. Combining (25) with the *Pesin* relation for chaotic systems [24], stating that $H_{\mu_f}(F) \leq \lambda$, we have the relation

$$H(S) = H_{\mu_f}(F) \leq \lambda. \tag{26}$$

3 The Threshold Effect

In this section, we prove the existence of a threshold effect in the estimation of chaotic sequences as the number of observations grows without bound. Using the data processing theorem, we derive a lower bound on the mean square error obtained by any estimator of the initial state. We compare that bound with the Bayesian Cramér-Rao bound and show that below certain SNR level, the Bayesian Cramér-Rao bound is not tight, thus the existence of the threshold effect is proved.

We consider the problem of estimating the initial state of the chaotic system $\{([a,b], \mathcal{B}, \mu_f), F\}$ given the noisy observations $Y[0], \ldots, Y[N-1]$,

$$Y[n] = X[n] + W[n] \quad n = 0, 1, , \ldots, N-1, \tag{27}$$

where $\{X[n]\}_{n=0}^{N-1}$ is the chaotic sequence and $\{W[n]\}_{n=0}^{N-1}$ is zero-mean white Gaussian noise with variance σ^2. Let $\underline{X} = [X[0] \ldots X[N-1]]^T, \underline{Y} = [Y[0] \ldots Y[N-1]]^T$. The average power of $\{X[n]\}_{n=0}^{N-1}$ is

$$P = \frac{1}{N} \sum_{n=0}^{N-1} E\{X^2[n]\} = E\{X_0^2\}, \tag{28}$$

and the channel SNR is then

$$\mathrm{SNR} = \frac{P}{\sigma^2} = \frac{E\{X_0^2\}}{\sigma^2}. \tag{29}$$

The Bayesian Cramér-Rao bound, associated with the estimation of the initial state of the dynamical system $\{([a,b], \mathcal{B}, \mu_f), F\}$, is given by

$$CRB = \left[E(\varphi(X)) + \frac{1}{\sigma^2} \sum_{n=0}^{N-1} E\left(\dot{F}^n(X)\right)^2 \right]^{-1}, \tag{30}$$

where

$$\dot{F}^n(x) \triangleq \frac{dF^n(x)}{dx}, \quad n \in \mathcal{N} \tag{31}$$

$$\varphi(x) \triangleq \left(\frac{d \ln f(x)}{dx} \right)^2, \tag{32}$$

and the expectation is with respect to the invariant density f. This bound can readily be derived by extending the previously obtained parametric Cramér-Rao bound [17] into the Bayesian case.

Theorem 3.1. *Let $\{([a,b], \mathcal{B}, \mu_f), F\}$ be a dynamical system with an ergodic mapping F, let f be the invariant density assumed to satisfy $f \geq \varepsilon > 0$ a.e., and let λ be the system's Lyapunov exponent.*

Let CRB denote the Bayesian Cramér-Rao bound associated with the estimation of initial state. Then,

$$\liminf_{N\to\infty} \frac{1}{N} \ln CRB \leq -2\lambda. \tag{33}$$

The proof appears in Appendix A.

We now use the data processing theorem to derive a lower bound on the mean square error obtained by any estimator of the initial state.

Theorem 3.2. *For a given dynamical system $\{([a,b],\mathcal{B},\mu_f),F\}$, and an arbitrary estimator $\hat{X}[0]$ of the initial state, let*

$$D \triangleq E\{(\hat{X}[0]-X[0])^2\}. \tag{34}$$

Under the assumptions of Theorem 3.1, we have

$$\liminf_{N\to\infty} \frac{1}{N} \ln D \geq -2\overline{I}(\underline{X};\underline{Y}) \tag{35}$$

where

$$\overline{I}(\underline{X};\underline{Y}) \triangleq \limsup_{N\to\infty} \frac{I(\underline{X};\underline{Y})}{N}, \tag{36}$$

and $I(\underline{X};\underline{Y})$ is the mutual information between $\underline{X}$ and $\underline{Y}$.

Proof. The random variables $(X[0],\underline{X},\underline{Y},\hat{X}[0])$ constitute a Markov chain

$$X[0] \to \underline{X} \to \underline{Y} \to \hat{X}[0]. \tag{37}$$

Therefore, according to the data processing theorem [4] we have

$$R(D) \leq I(X[0];\hat{X}[0]) \leq I(\underline{X};\underline{Y}), \tag{38}$$

where $R(D)$ is the rate-distortion function of a random variable distributed according to f. Using *Shannon's* lower bound [4], we have

$$R(D) \geq \frac{1}{2}\ln\left(\frac{e^{2h(X)}}{2\pi e D}\right), \tag{39}$$

where $h(X)$ is the differential entropy induced by f. Incorporating (39) into (38,) we have

$$I(\underline{X};\underline{Y}) \geq \frac{1}{2}\ln\left(\frac{e^{2h(X)}}{2\pi e D}\right). \tag{40}$$

Dividing both sides of by N and taking $N\to\infty$, we have

$$\liminf_{N\to\infty} \frac{1}{N} \ln D \geq -2\overline{I}(\underline{X};\underline{Y}), \tag{41}$$

which is the desired result ■.

Lemma 3.3. *For a given channel input distribution, the mutual information $I(\underline{X};\underline{Y})$ is a strictly increasing function of the SNR.*

Proof. Follows immediately from the data processing theorem [4] ■.

Motivated by Lemma 3.3, we now introduce the notation:

$$I(SNR, N) \triangleq I(\underline{X};\underline{Y}), \quad I(SNR) \triangleq \overline{I}(\underline{X};\underline{Y}). \tag{42}$$

As $I(SNR)$ is a strictly increasing function of the SNR, it has an inverse function, denoted here as $I^{-1}(\cdot)$, which is also an increasing function.

Corollary 3.4. *Under the assumptions of Theorem 3.1,*

$$\liminf_{N\to\infty} \frac{1}{N} \ln\left(\frac{D}{CRB}\right) \geq 2\left(\lambda - I(SNR)\right). \tag{43}$$

Proof. Combining (35) and (33) immediately yields the desired result ■.

Corollary 3.5. *Under the assumptions of Theorem 3.1, we have that for*

$$SNR < I^{-1}(\lambda) \triangleq SNR^* \tag{44}$$

no estimator of the initial state can be asymptotically efficient. Moreover,

$$\lim_{N\to\infty} \frac{D}{CRB} = \infty. \tag{45}$$

Proof. Follows directly from Corollary 3.4 ■.

If an estimator of the initial state asymptotically achieves the Bayesian Cramér-Rao bound it may do so only beyond a threshold SNR, denoted here by SNR_{th}, which is lower bounded by

$$SNR_{th} \geq SNR^*. \tag{46}$$

Since the system is chaotic ($\lambda > 0$) then:

$$SNR^* = I^{-1}(\lambda) > I^{-1}(0) = 0, \tag{47}$$

ensuring that the bound on the threshold SNR is nontrivial. The above analysis explains why there will always be a threshold effect: below some threshold SNR, a sharp degradation in performance, of exponential order, as Corollary 3.4 indicates, will occur. Corollary 3.5 provides us with a lower bound on the threshold SNR. We now show how the existence of the threshold effect is actually tied to the *converse to the joint source-channel coding theorem* [4]. As mentioned above, sharp degradation, of exponential order, in performance starts when

$$I(SNR) < \lambda. \tag{48}$$

Suppose that (26) is met with equality, we then deduce that whenever

$$I(SNR) < H(S), \tag{49}$$

degradation in the performance (in comparison with the lowest possible distortion, the Cramér-Rao bound) of every estimator occurs. According to the converse to the joint source-channel coding theorem, a finite alphabet stationary source $V[n]$ with entropy rate $H(V)$ cannot be sent through a channel whose capacity is C, with arbitrary low probability of error if $C < H(V)$. There is a one-to-one map between the initial state $X[0]$ and the infinite sequence of symbols $\{S[n]\}_{n=0}^{\infty}$ a.e., therefore, as $N \to \infty$, we can consider the problem of chaotic sequence estimation as a joint source-channel coding scheme with stationary source $S[n]$ having entropy rate of $H(S)$, a joint source-channel coder which is the chaotic system and channel with capacity $I(SNR)$. This scheme is drawn in Fig. 1. Note that the maximum number of nats that the channel can convey with arbitrary low probability of error is $I(SNR)$, not the channel capacity C, since the density of $\underline{X}$ is predetermined according to the invariant density and the map F, and does not achieve the capacity. When (26) is not met with equality, degradation in performance occurs already in $SNR^* = I^{-1}(\lambda)$ which is larger than $I^{-1}(H(S))$. This is because the chaotic system is not an optimal joint source-channel coder.

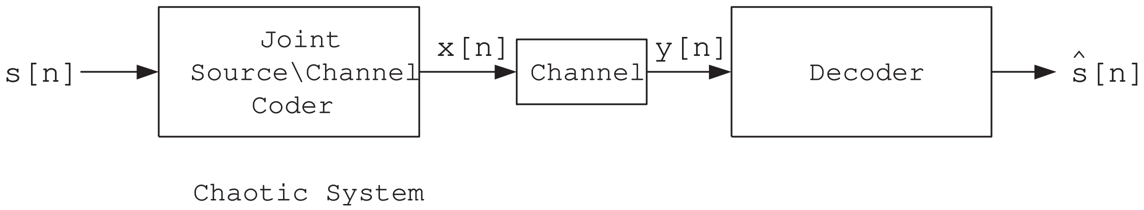

Figure 1: The estimation scheme of chaotic sequences as a joint source-channel coding scheme.

4 Lower Bound on the Threshold SNR

The bound on the threshold SNR is a function of the mutual information between the chaotic sequence and the noisy observations. However, due to the nonlinearity of the map F, calculating this mutual information is very difficult. Therefore, we derive an upper bound on the mutual information and use it to obtain a lower bound on the threshold SNR. If

$$I(SNR) < I_B(SNR) \quad \forall SNR \in \mathcal{R}^+ \tag{50}$$

where $I_B(SNR)$ is some increasing function of SNR then

$$SNR_{th} > I_B^{-1}(\lambda) \triangleq SNR_B^*, \tag{51}$$

where I_B^{-1} is the inverse function of I_B. However, the above bound is less tight than the one in (46).

The mutual information $I(\underline{X};\underline{Y})$ is given by

$$\begin{aligned} I(\underline{X};\underline{Y}) &= h(\underline{Y}) - h(\underline{Y}|\underline{X}) \\ &= h(\underline{Y}) - h(\underline{X}+\underline{W}|\underline{X}) \\ &= h(\underline{Y}) - \frac{N}{2}\ln(2\pi e\sigma^2). \end{aligned} \tag{52}$$

The last equality is due to the fact that the channel noise $\underline{W}$ is white Gaussian noise, independent of $\underline{X}$. Let us denote the zero mean random vector $\underline{\widetilde{Y}}$ as

$$\underline{\widetilde{Y}} = \underline{Y} - E\underline{Y} = \underline{Y} - E\underline{X}. \tag{53}$$

Since the differential entropy is invariant to translation,

$$\begin{aligned} h(\underline{Y}) &= h(\underline{\widetilde{Y}}) && (54)\\ &= \sum_{n=0}^{N-1} h\left(\widetilde{Y}[n]\middle|\{\widetilde{Y}[i]\}_{i=0}^{n-1}\right) && (55)\\ &= \sum_{n=0}^{N-1} h\left(\widetilde{Y}[n] - g_n\left(\{\widetilde{Y}[i]\}_{i=0}^{n-1}\right)\middle|\{\widetilde{Y}[i]\}_{i=0}^{n-1}\right) && (56)\\ &\le \sum_{n=0}^{N-1} h\left(\widetilde{Y}[n] - g_n\left(\{\widetilde{Y}[i]\}_{i=0}^{n-1}\right)\right) && (57)\\ &\le \frac{1}{2}\sum_{n=0}^{N-1} \ln\left(2\pi e E\left(\widetilde{Y}[n] - g_n\left(\{\widetilde{Y}[i]\}_{i=0}^{n-1}\right)\right)^2\right) && (58) \end{aligned}$$

where in (55) we made use of the entropy chain rule, (56) holds for arbitrary functions $g_n(\cdot)$, (57) follows from the fact that conditioning reduces entropy, and (58) follows from the fact that the Gaussian density maximizes the differential entropy for a given second moment. To obtain the tightest bound, we select $g_n(\cdot)$ to be the optimal linear predictor of $\widetilde{Y}[n]$ given $\{\widetilde{Y}[i]\}_{i=0}^{n-1}$. Let us denote the mean square error obtained by the optimal linear predictor by σ_n^2, which depends on the autocorrelation function of $\widetilde{Y}[n]$, $R_{\widetilde{Y}\widetilde{Y}}[k]$. Let

$$\widetilde{X}[n] \triangleq X[n] - E\{X_0\}, \tag{59}$$

then,

$$R_{\widetilde{Y}\widetilde{Y}}[k] = \sigma^2\delta[k] + R_{\widetilde{X}\widetilde{X}}[k], \quad R_{\widetilde{X}\widetilde{X}}[k] = R_{XX}[k] - (E\{X_0\})^2, \tag{60}$$

$$R_{XX}[k] = E(X[n]F^k(X[n])) = \int_a^b xF^k(x)f(x)dx. \tag{61}$$

Substituting σ_n^2 into (58), we obtain

$$h(\underline{Y}) \le \frac{1}{2}\sum_{n=0}^{N-1} \ln\left(2\pi e\sigma_n^2\right), \tag{62}$$

which in turn yields (by (52))

$$I(\underline{X};\underline{Y}) \leq \frac{1}{2}\sum_{n=0}^{N-1} \ln\left(\frac{\sigma_n^2}{\sigma^2}\right). \tag{63}$$

Dividing (63) by N and taking $N \to \infty$, we have that

$$\limsup_{N\to\infty} \frac{I(\underline{X};\underline{Y})}{N} \leq \frac{1}{2}\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} \ln\left(\frac{\sigma_n^2}{\sigma^2}\right) \tag{64}$$

Invoking *Cèsaro's theorem* [25], we have that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} \ln\left(\frac{\sigma_n^2}{\sigma^2}\right) = \lim_{n\to\infty} \ln\left(\frac{\sigma_n^2}{\sigma^2}\right) \tag{65}$$

Since the process $\widetilde{Y}[n]$ is stationary and ergodic, then according to the *Kolmogorov formula* [25], we have

$$\lim_{n\to\infty} \sigma_n^2 = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \ln S_{\widetilde{Y}\widetilde{Y}}(\omega)d\omega\right\}. \tag{66}$$

Using (60), the power spectrum of $\widetilde{Y}[n]$ is given by

$$S_{\widetilde{Y}\widetilde{Y}}(\omega) = \sigma^2 + S_{\widetilde{X}\widetilde{X}}(\omega). \tag{67}$$

Substituting (67) into (66) and then the result into (65), we obtain

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1} \ln\left(\frac{\sigma_n^2}{\sigma^2}\right) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \ln\left(1+\frac{S_{\widetilde{X}\widetilde{X}}(\omega)}{\sigma^2}\right) d\omega. \tag{68}$$

Substituting (68) into (64) we have

$$\limsup_{N\to\infty} \frac{I(\underline{X};\underline{Y})}{N} \leq \frac{1}{4\pi}\int_{-\pi}^{\pi} \ln\left(1+\frac{S_{\widetilde{X}\widetilde{X}}(\omega)}{\sigma^2}\right) d\omega. \tag{69}$$

Let

$$I_L(SNR) \triangleq \frac{1}{4\pi}\int_{-\pi}^{\pi} \ln\left(1+\frac{S_{\widetilde{X}\widetilde{X}}(\omega)SNR}{E\{X_0^2\}}\right) d\omega. \tag{70}$$

Note that (70) is the capacity of the Gaussian channel where the channel input is forced to have a given power spectrum $S_{\widetilde{X}\widetilde{X}}(\omega)$. The channel capacity is then achieved with a Gaussian input having this power spectrum. Using (51), a lower bound on the threshold SNR, is the value of the SNR satisfying

$$\lambda = \frac{1}{4\pi}\int_{-\pi}^{\pi} \ln\left(1+\frac{S_{\widetilde{X}\widetilde{X}}(\omega)SNR}{E\{X_0^2\}}\right) d\omega. \tag{71}$$

Isabelle and Wornell [12] used the matrix representation of the Frobenius-Perron operator [9] to obtain an expression for the autocorrelation function and power spectrum corresponding to the eventually expanding piecewise linear Markov maps. Using their results, (71) and (75) could be obtained for this class of maps.

5 The Threshold Effect - Finite N

In this section, we consider the case where the number of observations is finite. Rearranging (40), we have

$$D \geq \frac{1}{2\pi e} \exp\left\{-2(I(SNR,N) - h(X))\right\}, \tag{72}$$

Comparing the above bound on D to the Cramér-Rao bound bound $D \geq CRB \triangleq CRB(SNR,N)$, a lower bound on the threshold SNR, is the highest value of the SNR satisfying

$$I(SNR,N) = \frac{1}{2}\ln\left(\frac{e^{2h(X)}}{2\pi e CRB(SNR,N)}\right) \triangleq R(SNR,N). \tag{73}$$

If the mutual information $I(SNR,N)$ is upper bounded by a increasing function of SNR, $I_B(SNR,N)$, the threshold SNR is also lower bounded by the highest value of the SNR satisfying

$$I_B(SNR,N) = \frac{1}{2}\ln\left(\frac{e^{2h(X)}}{2\pi e CRB(SNR,N)}\right). \tag{74}$$

Using the upper bound we obtained on the mutual information (63), a lower bound on the threshold SNR, is the highest value of the SNR satisfying

$$I_L(SNR,N) = \frac{1}{2}\ln\left(\frac{e^{2h(X)}}{2\pi e CRB(SNR,N)}\right) \tag{75}$$

where

$$I_L(SNR,N) \triangleq \frac{1}{2}\sum_{n=0}^{N-1}\ln\left(\frac{\sigma_n^2 SNR}{E\{X_0^2\}}\right). \tag{76}$$

Next we demonstrate our results on the dynamical system governed by the r-diadic map.

6 Results For The r-diadic Map

The r-diadic map is eventually expanding piecewise linear Markov map. However, it is easier to calculate the autocorrelation function directly, as we do next. It is easy to verify that for every nonnegative integer k,

$$F_r^k(x) = (r^k x)\mathrm{mod}1 = r^k x - i + 1, \quad x \in \left[\frac{i-1}{r^k}, \frac{i}{r^k}\right), \quad i = 1, 2, \ldots, r^k. \tag{77}$$

Using (77), the autocorrelation function is given by

$$\begin{aligned} R_{XX}[k] &= \int_0^1 x F_r^{|k|}(x)dx \\ &= \sum_{i=0}^{r^{|k|}-1}\int_{\frac{i}{r^{|k|}}}^{\frac{i+1}{r^{|k|}}} x(r^{|k|}x - i)dx \\ &= \frac{1}{r^{2|k|}}\sum_{i=0}^{r^{|k|}-1}\frac{3i+2}{6} \\ &= \frac{r^{-|k|}}{12} + \frac{1}{4}. \end{aligned} \tag{78}$$

For the r-diadic map we have

$$f \equiv 1. \quad \frac{dF^n(x)}{dx} = r^n \quad a.e. \tag{79}$$

Using the above we have

$$CRB(SNR, N) = \frac{r^2 - 1}{3(r^{2N} - 1)SNR}. \tag{80}$$

$$h(X) = 0. \tag{81}$$

Substituting (80) into (75), a lower bound on the threshold SNR for finite N, is the highest value of the SNR satisfying

$$I_L(SNR, N) = \frac{1}{2} \ln \left(\frac{3(r^{2N} - 1)SNR}{2\pi e(r^2 - 1)} \right). \tag{82}$$

Now we deal with the asymptotic result. Substituting (78) into (60) we have

$$R_{\tilde{Y}\tilde{Y}}[k] = \frac{r^{-|k|}}{12} + \sigma^2 \delta[k], \qquad k \in \mathcal{Z} \tag{83}$$

and

$$R_{\tilde{Y}\tilde{Y}}(z) = \mathcal{Z}\left\{R_{\tilde{Y}\tilde{Y}}[k]\right\} = \frac{1}{12} \frac{1 - r^{-2}}{(1 - r^{-1}z^{-1})(1 - r^{-1}z)} + \sigma^2. \tag{84}$$

After some algebraic manipulation, (84) is rewritten in the *canonical form* [25] :

$$R_{\tilde{Y}\tilde{Y}}(z) = \sigma_u^2 B(z) B(z^{-1}), \tag{85}$$

where the minimum phase filter $B(z)$ is given by

$$B(z) = \frac{1 - \beta^{-1} z^{-1}}{1 - r^{-1} z^{-1}} \tag{86}$$

with

$$\beta \equiv \beta(\sigma^2, r) = \frac{\alpha(\sigma^2, r) + \sqrt{\alpha^2(\sigma^2, r) - 576 r^{-2} \sigma^4}}{24 r^{-1} \sigma^2} \tag{87}$$

and

$$\alpha(\sigma^2, r) \triangleq 1 - r^{-2} + 12\sigma^2(1 + r^{-2}) \tag{88}$$

and the variance of the *innovation process* is

$$\sigma_u^2 = \frac{\beta(\sigma^2, r)\sigma^2}{r}. \tag{89}$$

According to Kolmogorov's theorem [25], we have

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \ln S_{\tilde{Y}\tilde{Y}}(\omega) d\omega = \ln \sigma_u^2. \tag{90}$$

Substituting (90) and $E\{X_0^2\} = 1/3$ into (70), we obtain

$$I_L(SNR) = \frac{1}{2} \ln \left(\frac{\beta\left(\frac{1}{3SNR}, r\right)}{r} \right). \tag{91}$$

Substituting (91) and $\lambda = \ln r$ into (71), a lower bound on the threshold SNR is the value of the SNR that solves the equation

$$r^3 = \beta\left(\frac{1}{3SNR}, r\right). \tag{92}$$

In [12], it was shown that the spectra associated with the eventually expanding piecewise linear Markov maps are rational (thus the processes generated by those maps are referred to as chaotic ARMA processes). Therefore, for this class of chaotic maps, the representation of the spectrum in the canonical form, in order to obtain the innovation process variance, is relatively easy.

In Fig. 2, $I_L(SNR)$ (91), for $r = 2$, together with the $AWGN$ channel capacity, $C_{AWGN} = 0.5\ln(1 + SNR)$, are drawn. It is clearly seen that there is considerable difference between the two. We obtained for $N \to \infty$ that $SNR_{th} \geq 11.8dB$. In Fig. 3, $I_L(SNR, N)$ and $R(SNR, N)$ (82) are drawn for $N = 8$ and $r = 2$. Here we obtained that $SNR_{th} \geq 11.5dB$. The values of the lower bound on the threshold SNR for $N = 5, 6, 7, 8$ are summarized in Table 1. In [7] the optimal (MMSE) estimator of the initial state for the 2-diadic map was obtained. We performed Monte-Carlo simulations to obtain the mean square error of the MMSE estimator of the initial state as function of the SNR. The result for $N = 8$ is drawn in Fig. 4. The form of representation is MSE^{-1} and CRB^{-1} in dB. It is seen that the Bayesian Cramér-Rao bound is achieved beyond some threshold SNR, $SNR_{th} = 20dB$ (up to the SNR resolution we used in the simulation, $0.5dB$). The bound we obtained is therefore not tight enough.

There are three possible reasons for the fact that the bound on the threshold SNR is not tight: the first reason is that the bound we used on the mutual information $I(\underline{X};\underline{Y})$ is not tight, the second reason is that the data processing inequality is not tight, and last reason is that Shannon's lower bound on the rate- distortion function is not tight. We have carefully examined the above three factors and reached the conclusion that the main reason for the bound on the threshold SNR not being tight is that the data processing inequality is not tight.

7 Concluding Remarks

In this paper, we quantified the threshold effect that exists in the estimation process of chaotic sequences. We derived a lower bound on the threshold SNR, showing it depends on the system's Lyapunov exponent and the mutual information between the chaotic sequence and the noisy observations. This bound was further simplified to a bound depending on the system's Lyapunov exponent and the power spectrum of the chaotic sequence. For the wide and important class of eventually expanding piecewise linear Markov maps this bound can be easily calculated. We explained our results using the converse to the joint-source channel coding theorem. Essentially, for SNR's below the threshold, the amount of information that the chaotic system produces about its initial state is larger than the maximum information that the channel can convey with low probability of error.

Thus, degradation in distortion performance is unavoidable. Although the bound on the estimation error mean square error which derives from the data processing inequality is not tight, it is tighter than the Bayesian Cramér-Rao bound below the threshold SNR. An interesting future research direction would be to consider the data processing theorem suggested by Ziv and Zakai [31], which for a finite number N of channel uses, provides better bounds on distortion than the classical data processing theorem.

N	5	6	7	8	∞
$SNR_{th} \geq$	$11.3dB$	$11.4dB$	$11.4dB$	$11.5dB$	$11.8dB$

Table 1: Lower bound on the threshold SNR for $r = 2$, finite valued N's and $N \rightarrow \infty$.

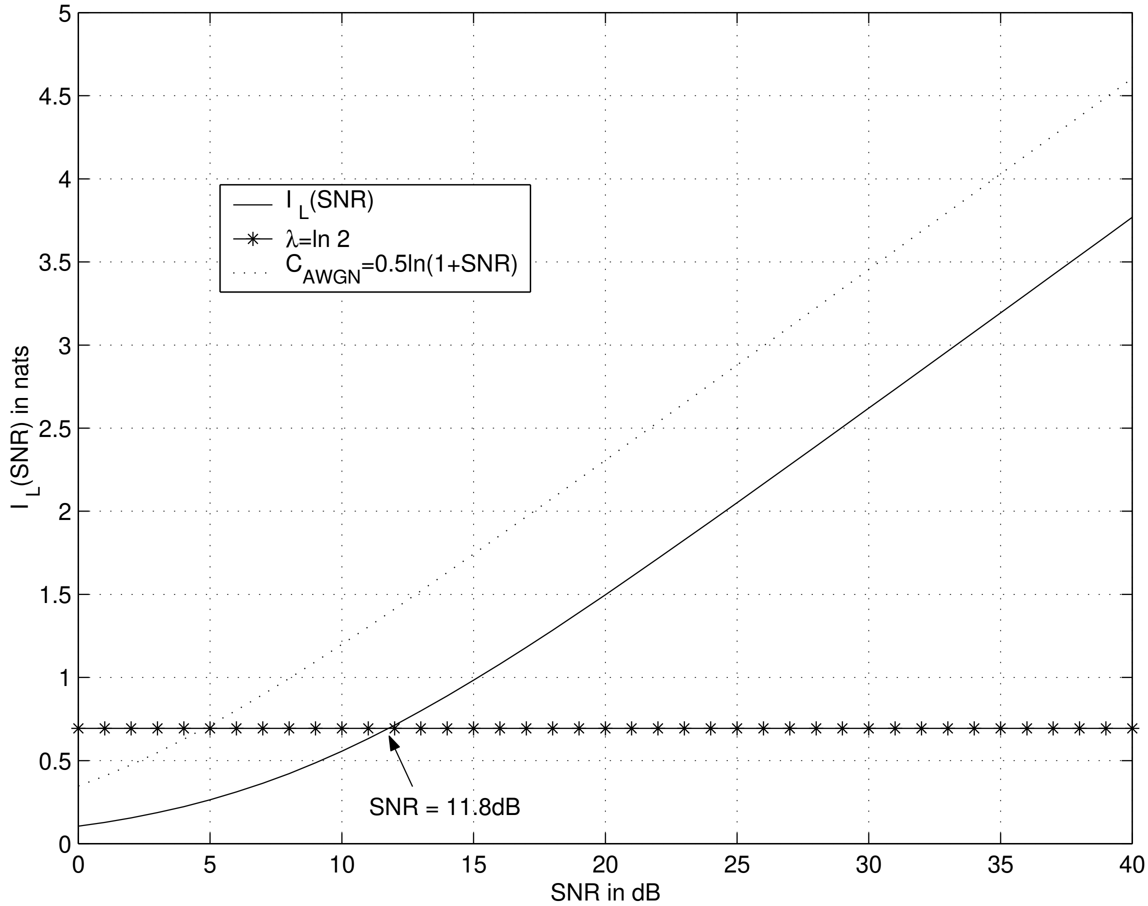

Figure 2: The asymptotic mutual information upper bound, $I_L(SNR)$, for $r = 2$.

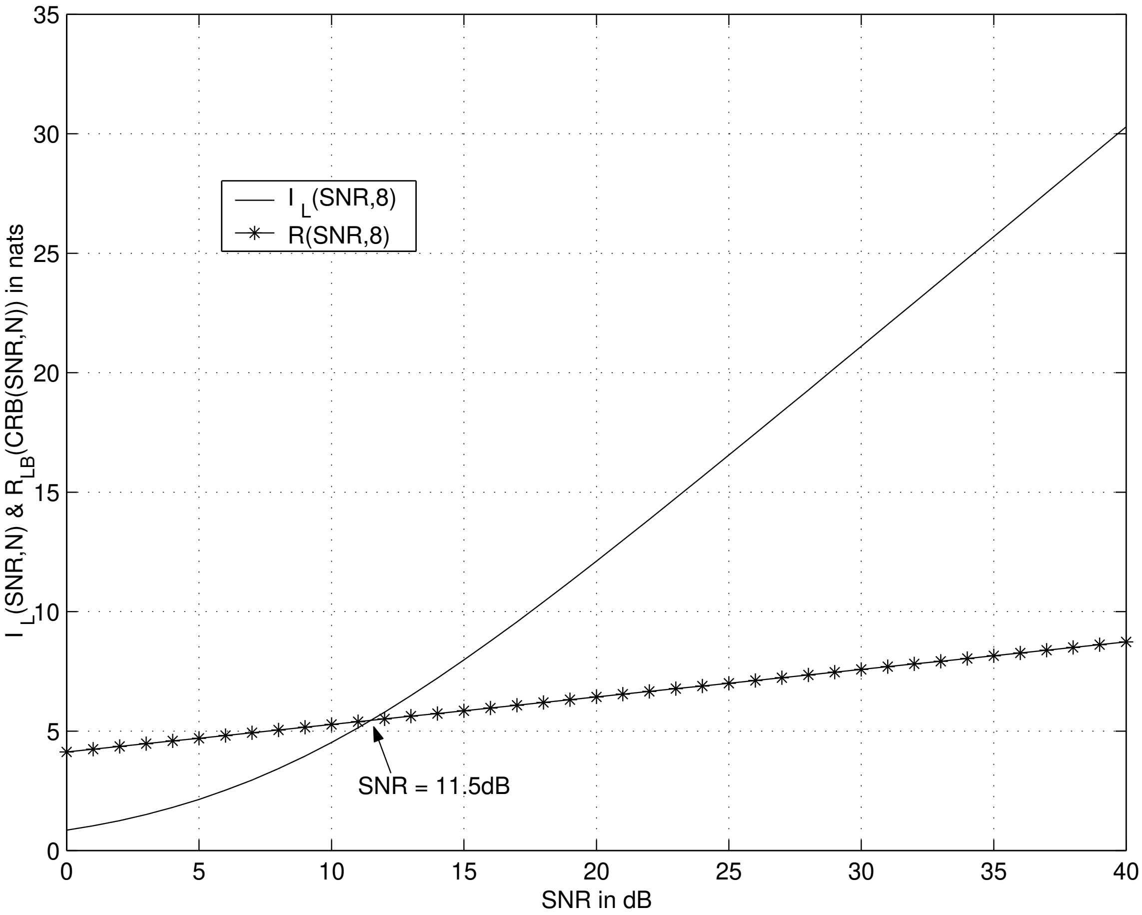

Figure 3: The mutual information upper bound, $I_L(SNR, N)$, and $R(SNR, N)$ for $r = 2$ and $N = 8$.

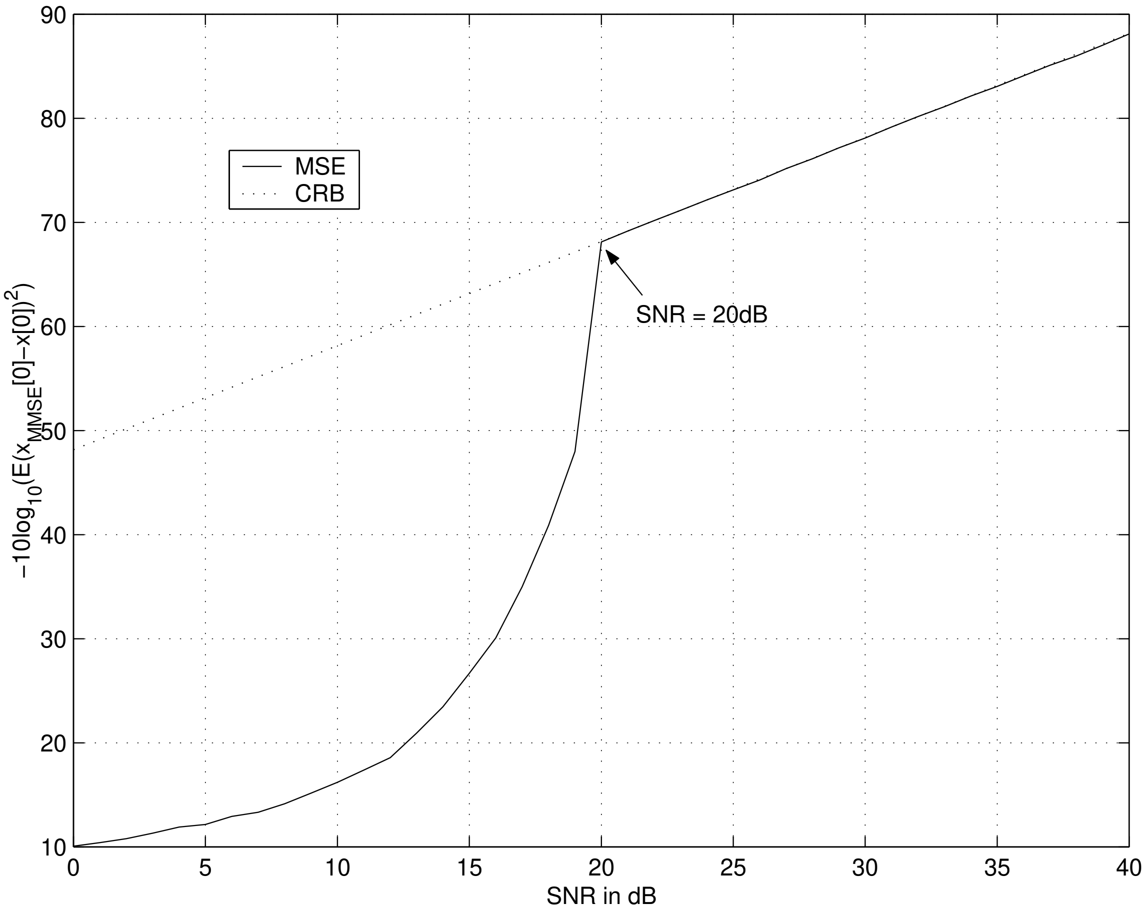

Figure 4: The MMSE and CRB for $r = 2$ and $N = 8$.

A Proof of Theorem 3.1

We first prove the following lemma:

Lemma A.1. *Let $\{([a,b],\mathcal{B},\mu_f),F\}$ be a dynamical system. Under the assumptions of Theorem 3.1, we have that for all $N \in \mathcal{N}$ the function*

$$\frac{1}{N}\ln\left|\frac{dF^N(x)}{dx}\right| \tag{93}$$

is bounded almost everywhere.

Proof. According to Oseledec's theorem [19],

$$E\ln\left|\dot{F}(X)\right| = \int_a^b f(x)\ln\left|\frac{dF(x)}{dx}\right|dx = \lambda < \infty. \tag{94}$$

Thus, the above integrand is bounded almost everywhere,

$$|f(x)|\left|\ln\left|\frac{dF(x)}{dx}\right|\right| \leq B \quad a.e. \tag{95}$$

Since $f \geq \varepsilon > 0$ a.e. we have,

$$\left|\ln\left|\frac{dF(x)}{dx}\right|\right| \leq \frac{B}{\varepsilon} \quad a.e. \tag{96}$$

yielding that

$$m \leq \left|\frac{dF(x)}{dx}\right| \leq M \quad a.e, \tag{97}$$

where $0 < m \leq M$. Using the derivative chain rule, we have

$$\frac{dF^N(x)}{dx} = \prod_{n=1}^{N}\frac{dF(t)}{dt}\Big|_{t=F^{n-1}(x)}. \tag{98}$$

Using (98) and (97), we have that

$$m^N \leq \left|\frac{dF^N(x)}{dx}\right| \leq M^N \quad a.e \tag{99}$$

and that

$$\left|\frac{1}{N}\ln\left|\frac{dF^N(x)}{dx}\right|\right| \leq \max\{|\ln m|, |\ln M|\}, \quad a.e. \tag{100}$$

for all $N \in \mathcal{N}$ ■.

We now prove Theorem 3.1.

Using we (30) have

$$\begin{aligned}\frac{1}{CRB} &= E(\varphi(X)) + \frac{1}{\sigma^2}\sum_{n=0}^{N-1}E\left(\dot{F}^n(X)\right)^2 \\ &> \frac{1}{\sigma^2}E\left(\dot{F}^{N-1}(X)\right)^2 \\ &\geq \frac{1}{\sigma^2}\left(E\left|\dot{F}^{N-1}(X)\right|\right)^2\end{aligned} \tag{101}$$

where the in last line, we used Jensen's inequality and the convexity of the function $g(u) = u^2$. Taking the logarithm and dividing by N, we get:

$$\begin{aligned}\frac{1}{N}\ln\left(\frac{1}{CRB}\right) &> \frac{1}{N}\ln\left(\frac{1}{\sigma^2}\left(E\left|\dot{F}^{N-1}(X)\right|\right)^2\right)\\ &= \frac{2}{N}\ln\left(E\left|\dot{F}^{N-1}(X)\right|\right) + \frac{1}{N}\ln\left(\frac{1}{\sigma^2}\right)\\ &\geq \frac{2}{N}E\ln\left|\dot{F}^{N-1}(X)\right| + \frac{1}{N}\ln\left(\frac{1}{\sigma^2}\right),\end{aligned} \tag{102}$$

where in the last line we used again Jensen's inequality. Taking $N \to \infty$ at both sides of (102), we obtain

$$\begin{aligned}\liminf_{N\to\infty}\frac{1}{N}\ln\left(\frac{1}{CRB}\right) &\geq 2\lim_{N\to\infty}E\frac{1}{N}\ln\left|\dot{F}^{N-1}(X)\right|\\ &= 2\lim_{N\to\infty}\int_a^b f(x)\frac{1}{N}\ln\left|\frac{dF^{N-1}(x)}{dx}\right|dx.\end{aligned} \tag{103}$$

Using Oseledec's Theorem [19], we have

$$\lim_{N\to\infty}\frac{1}{N}\ln\left|\frac{dF^{N-1}(x)}{dx}\right| = \lambda \quad a.e, \tag{104}$$

and from Lemma A.1, we have for all $N \in \mathcal{N}$, that the expression in the above limit is bounded almost everywhere. Therefore, according to the Lebesgue dominated convergence theorem [16], we have

$$\begin{aligned}\lim_{N\to\infty}\int_a^b f(x)\frac{1}{N}\ln\left|\frac{dF^{N-1}(x)}{dx}\right|dx &= \int_a^b f(x)\lim_{N\to\infty}\frac{1}{N}\ln\left|\frac{dF^{N-1}(x)}{dx}\right|dx\\ &= \int_a^b f(x)\lambda dx = \lambda.\end{aligned} \tag{105}$$

Substituting (105) into (103,) we obtain that

$$\liminf_{N\to\infty}\frac{1}{N}\ln\left(\frac{1}{CRB}\right) \geq 2\lambda, \tag{106}$$

which completes the proof.

References

[1] A. Boyarsky and M. Scarowsky, " On a class of transformations which have unique absolutely continious invariant measures," *Trans. Amer. Math. Soc.*, Vol. 255, pp. 243-262, 1979.

[2] B. Cazelles, G. Boudjema and N.P. Chau, "Adaptive control of chaotic systems in noisy environment," *Physics Letters A*, Vol. 196, pp. 326-330, Jan. 1995.

[3] B. Chen and G.W. Wornell, "Analog error-correcting codes based on chaotic dynamical systems," *IEEE Trans. Comm.* Vol. 46, No. 7, pp. 881-890, Jul. 1998.

[4] T.M. Cover and J.A. Thomas, *Elements of Information Theory.* John Wiley & Sons, Inc, 1991.

[5] K.M. Cuomo and A.V. Oppenheim, "Synchronization of Lorenz based chaotic circuits with applications to communications," *IEEE Trans. Circuits and Systems II*, Vol. 40, No. 10, pp. 626-633, Oct. 1993.

[6] J. H. Deane and D.C. Hamill, "Chaotic behavior in a current-mode controlled DC-DC converter," *Electron. Lett.*, Vol. 27, pp. 1172-1173, 1991.

[7] D.F. Drake, *Information's role in the estimation of chaotic signals.* Ph.D. Thesis, Georgia Ins. of Tech. Aug. 1998.

[8] D.F. Drake and D.B. Williams, "Spread spectrum communications using chaotic systems," In *IEEE Dual-Use Thech. and Appl. Conf.* Vol. 2, pp. 385-391, 1994.

[9] N. Friedman and A. Boyarsky, "Matrices and eigienfunctions induced by markov maps," *Lin. Alg. App.*, Vol. 39, pp. 141-147, 1981.

[10] J.D. Farmer and J. Sidorowich, "Optimal shadowing and noise reduction," *Physica D*, Vol. 47, pp. 373-392, 1991.

[11] S. Hammel, "A noise reduction method for chaotic systems," *Physics Letters A*, Vol. 148, No. 9, pp. 421-428, Sep. 1990.

[12] S. H. Isabelle and G.W. Wornell, "Statistical analysis and spectral estimation techniques for one-dimensional chaotic signals," *IEEE Trans. Signal Processing*, Vol. 45. No. 6, pp. 1495-1506, Jun. 1997.

[13] A. Katoke and B. Hasselblatt, *Introduction to the Modern Theory of Dynamical Systems.* Cambridge University Press 1998.

[14] S. Kay, "Asymptotic maximum likelihood estimator performance for chaotic signals in noise," *IEEE Trans. Signal Processing* Vol. 43 , pp. 1009-1012, Apr. 1995.

[15] S. Kay and V. Nagesha, "Methods for chaotic signal estimation," *IEEE Trans. Signal Processing*, Vol. 43, No. 8, pp. 2013-2016, Aug. 1995.

[16] A. Lasota and M.C. Mackey, *Chaos Fractals and Noise.* Springer-Verlag New York 1994.

[17] C. Myers, S. Kay and M. Richard, "Signal separation for nonlinear dynamical systems," in *Proc. Int. Conf. on Acoust., Speech and Signal Processing*, Vol. IV, pp. 129-132, 1992.

[18] A.V. Oppenheim, G.W. Wornell, S. H. Isabelle and K.M. Cuomo, "Signal processing in the context of chaotic signals," in *Proc. Int. Conf. Acoust., Speech and Signal Processing*, Vol. IV, pp. 117-120, 1992.

[19] V.I. Oseledec, "A multiplicative ergodic theorem: Lyapunov characteristic numbers for dynamical systems," *Trans. Mosc. Math. Soc.* Vol. 19, p. 197, 1968.

[20] H.C. Papadopoulos and G.W. Wornell, "Maximum likelihood estimation of a class of cahotic signals," *IEEE Trans. Inform. Theory*, Vol. 41, No. 1, pp. 312-317, Jan. 1995.

[21] T.S. Parker and L.O. Chua, "Chaos: A Tutorial for Engineers," *Proc. of the IEEE.*, Vol. 75, No. 8, pp. 982-1008, Aug. 1987.

[22] L.M. Pecora and T.L. Carroll, "Synchronization in chaotic systems," *Physical Review Letters*, Vol. 64, No. 8, pp. 821-824, Feb. 1990.

[23] H. Peitgen, H. Jurgens and D. Saupe, *Chaos and Fractals: New Frontiers of Science.* Springer Verlag, 1992.

[24] Y.B. Pesin, "Lyapunov characteristic exponent and ergodic properties of smooth dynamical systems with an invariant measure," *Sov. Math. Dokl.* 17 p. 196, 1976.

[25] B. Porat, *Digital Signal Processing of Random Signals.* Prentice Hall, 1994.

[26] L. Risbo, "On the design of tone-free sigma-delta modulators," *IEEE Trans. Circuits Sys. II*, Vol. 42, pp. 52-55, Jan. 1995

[27] G. A. Seber and C.J. Wild, *Nonlinear Regression.* John Wiley & Sons, Inc, New York, 1989.

[28] A.C. Singer, G.W. Wornell and A.V. Oppenheim, "Nonlinear autoregresive modeling and estimation in the presence of noise," *Digital Signal Processing*, Vol. 4, pp. 207-221, Oct. 1994.

[29] P. So, E. Ott and W.P. Dayawansa, "Observing chaos: Deducing and tracking the state of chaotic system from limited observation," *Physical Review E*, Vol. 49, No. 4, pp. 2650-2660, Apr. 1994.

[30] C. Wu, "Asymptotic theory of nonlinear least squares estimation," *Ann. Stat.*, Vol. 9, pp. 501-513, 1981.

[31] J. Ziv and M. Zakai, "On functionals satisfying a data-processing theorem," *IEEE Trans. Inform. Theory*, Vol. IT-19, No. 3, pp. 275-283, May 1973.