MSC

# Information Theory and Chaotic Systems

conducted at the
Signal Processing and Speech Communications Laboratory
Graz University of Technology, Austria

by
Gebhard Josef Wallinger, 0831047

Supervisors:
Univ.-Prof. Dipl.-Ing. Dr.techn. Gernot Kubin
Dipl.-Ing. Bernhard Geiger

Assessors/Examiners:
Univ.-Prof. Dipl.-Ing. Dr.techn. Gernot Kubin

Graz, October 25, 2013

# Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

| | |
|---|---|
| date | (signature) |

# Abstract

In this thesis a link between the concepts of entropy in statistical mechanics and the concept of information loss from a system-theoretic point-of-view is given. An inequality between the Kolmogorov-Sinai entropy, a measure of information generation, and the information loss for autonomous dynamical systems is discovered. Equality can be achieved, e.g., for absolutely continuous invariant measures. Based on this inequality, univariate dynamical systems are analyzed. The information loss is upper semicontinuous in the probability measure of the dynamical system; the latter converges weakly to the invariant measure. This thesis thus utilizes the sequence of information loss values obtained by iterating the systems to get an upper bound on the Kolmogorov-Sinai entropy.

# Zusammenfassung

In dieser Master-Arbeit wird eine Verbindung zwischen den Konzepten der Entropie in der statistischen Mechanik und dem Informationsverlust im Sinne der Systemtheorie hergestellt. Eine Ungleichung zwischen der Kolmogorov-Sinai Entropie, einem Maß für die Informationgenerierung, und dem Informationsverlust für autonome dynamische System wurde gezeigt. Gleichheit wird dabei für absolut stetige invariante Maße erreicht. Basierend auf dieser Ungleichung werden eindimensionale dynamische Systeme betrachtet und analysiert. Der Informationsverlust ist halbstetig in Bezug auf das Wahrscheinlichkeitsmaß des dynamischen Systems, welches schwach zum invarianten Maß konvergiert. Die Folge der Werte des Informationsverlusts über mehrere Iterationen des Systems konvergiert daher zu einer oberen Schranke fuer die Kolmogorov-Sinai Entropie.

# Contents


**1 Introduction** **4**

**2 Information-Theoretic Preliminaries** **7**
- 2.1 Entropy . . . 8
- 2.2 Conditional Entropy . . . 9
- 2.3 Differential Entropy . . . 9
- 2.4 Information Loss . . . 9
- 2.5 Upper Semicontinuity of Information Loss . . . 11

**3 Preliminaries for Dynamical Systems** **12**
- 3.1 Lyapunov Exponent . . . 13
- 3.2 Kolmogorov-Sinai Entropy (Metric Entropy) . . . 15
- 3.3 Connection of K-S entropy and Lyapunov exponent (Margulis-Ruelle inequality) . . . 16
- 3.4 Sinai, Ruelle, and Bowen (SRB)-Measures . . . 18
- 3.5 Pesin Identity . . . 20

**4 Motivating Example: Tent map** **22**

**5 Main Result** **24**

**6 Examples** **29**
- 6.1 Asymmetric Tent Map . . . 29
  - 6.1.1 Lyapunov exponent $\lambda$ and K-S entropy $H_{KS}$ . . . 30
  - 6.1.2 Information Loss (Analytic) . . . 31
  - 6.1.3 Information Loss (Numeric Upper Bound) . . . 32
- 6.2 Gaussian Map . . . 35
  - 6.2.1 Lyapunov exponent $\lambda$ and K-S entropy $H_{KS}$ . . . 36
  - 6.2.2 Information Loss (Analytic) . . . 36
  - 6.2.3 Information Loss (Numeric Upper Bound) . . . 38
- 6.3 Sine-Circle Map . . . 42
  - 6.3.1 Lyapunov exponent $\lambda$ and K-S entropy $H_{KS}$ . . . 43
  - 6.3.2 Information Loss (Analytic) . . . 44
  - 6.3.3 Information Loss (Numeric Upper Bound) . . . 46
- 6.4 Generalized $2x$ *mod* 1 map . . . 52
  - 6.4.1 Lyapunov exponent $\lambda$ and K-S entropy $H_{KS}$ . . . 54
  - 6.4.2 Information Loss (Analytic) . . . 56
  - 6.4.3 Information Loss (Numeric Upper Bound) . . . 56

**7 Conclusion and Outlook** **61**

# 1 Introduction

In the pioneering work of *Eckmann and Ruelle* [1] the complexity of statistical mechanical entropy, associated with different kinds of statistical mechanical systems is investigated. Specifically, in systems which are characterized by strange attractors. The statistical mechanical entropy is a concept which measures the amount of uncertainty an observer has about an atom or particle. It is a measure of the disorder of a dynamical system and measures the unpredictability of the observed value. Therefore, it is also related to the degree of the chaotic behaviour of a dynamical system.
*Ruelle* [2] investigated the heating of an idealized thermostat, and it became clear that entropy changes as the dynamical system evolves in time. One is speaking of entropy production or of the entropy which gets pumped out of the system in each iteration. This quantity is the difference between the statistical mechanical entropy of the current state and the previous state. The difference is always positive and will not exceed the sum of Lyapunov exponents for autonomous dynamical diffeomorphisms. Chaotic behaviour is characterized by the rate of exponential divergence of neighboring values, quantified by the Lyapunov exponents of the dynamical system.
Another characterization of chaotic behaviour is given by the Kolmogorov-Sinai entropy of a dynamical system, also known as metric entropy. It measures the amount of information generated in each iteration of the dynamical system and is related to exponential stretching. The Margulis-Ruelle inequality states that the Kolmogorov-Sinai entropy cannot exceed the sum of positive Lyapunov exponents.
Through the invention of the Axiom-A attractor and the Anosov diffeomorphism *Ruelle* [3], a certain invariant measure - the Sinai, Ruelle, and Bowen (SRB) measure [4] is found. The SRB-measure assumes that the dynamical system admits absolutely continuous conditional measures. For those systems admitting the SRB-measure property, the inequality mentioned above becomes an equality. The entropy production does not exceed the sum of the Lyapunov exponents and is always positive.

The Lyapunov exponents and the Kolmogorov-Sinai entropy are powerful concepts for describing the chaotic behaviour of dynamical systems. Since these quantities are tough to compute and since there is the need of an infinite amount of measurements, just an estimation of these quantities can be made.

For this reason, there is the need to find other quantities to gain knowledge about the chaotic behaviour of such dynamical systems and to make proper estimations of such quantities.

Changing the discipline from statistical mechanics to information theory [5], a few more concepts of entropy can be found. Note, the principle concept of entropy is equal in both disciplines. In information-theoretic terms, the context of atoms or particles is replaced with a random event, denoted by a random variable or random vector in multivariate cases, respectively.

Thinking about the concept of conditional entropy, the uncertainty about a random variable given another random variable can be determined. Imagine a dynamical system evolves in time and let the input be a random variable. This introduces the concept of information loss: In the work of *Geiger et al.* [6],[7] it is defined as the conditional entropy of the input given the output.

The main result of [7], where the information loss is induced by a piecewise bijective function, shows an interesting relation. It became clear that this result provides a link between the information loss and differential entropies of the input and output. This result assumes the existence of the probability density function of the input as well as the probability density function of the output. An similar expression can be also found in *Papoulis and Pillai* [8, pp. 660].

Of particular interest is the last term of the formula of the main result presented in [7]. To be precise, the expectation of the logarithm of magnitude of the derivative of the given function, which is related to the Kolmogorov-Sinai entropy. However, a very similar expression of this main result can be found also in [2], denoted as folding entropy.

This work tries to close the gap between the main result of [7] and the entropy production mentioned in [2], as well as to show the interplay between information theory in system-theoretic terms and statistical mechanics. For the univariate case, the main result can be shown and is explored for some examples. The main result provides a general inequality between the Kolmogorov-Sinai entropy and the information loss. Since the information loss is upper semicontinuous, the information loss can be taken as a quantitative measure for the chaotic behaviour of a autonomous dynamical system.

This work is organized as follows: In section 2 a general description of related information-theoretic basics is given as well as a the basic concepts of the information theory are described in the following subsections. Section 3 introduces the basic concepts of dynamical systems. This sections is followed by the motivating example of the symmetric tent map in section 4. The main result and its proof is provided in section 5. This work will be closed by the examples (section 6) of the asymmetric tent map, Gaussian map, sine-circle map, and the generalized $2x$ *mod* 1 map, providing analytical and numerical results.

# 2 Information-Theoretic Preliminaries

For the rest of this thesis the common basics are defined as follows:

The finite, countable, or uncountable set of all possible outcomes of a random variable $X$ will be denoted by $\mathcal{X}$. For each element of a *countable* set $x \in \mathcal{X}$, a certain probability $p(x)$ can be defined that satisfies

$$p(x) \in [0, 1] \tag{1}$$
$$\sum_{x \in \mathcal{X}} p(x) = 1 \tag{2}$$

Thus, the probability of an certain event $A$ is given as

$$P(A) = \sum_{x \in A} p(x). \tag{3}$$

The function $p(x)$ is called the probability mass function (PMF), which is mapping a probability to each sample $x$ out of the countable set $\mathcal{X}$. Note, it represents a collection of positive discrete masses.

If a continuous random variable with distribution $F_X(x)$ is considered, the probability density function (PDF) can be defined as

$$f_X(x) = \frac{dF_X(x)}{dx} \tag{4}$$

and satisfies

$$\int_{\mathcal{X}} f_x(x) dx = 1 \tag{5}$$

provided the PDF exists.

Finally the expectation value should be defined w.r.t. the random variable $X$ in the discrete case as

$$E\{X\} = \sum_{x \in \mathcal{X}} x p(x) \tag{6}$$

and for the continuous case as

$$E\{X\} = \int_{\mathcal{X}} x f(x) dx. \tag{7}$$

## 2.1 Entropy

In this section the concept of entropy (see [5, pp. 13]) will be introduced. The entropy in information theory can be seen as the uncertainty of a random variable and gives the upper limit of possible lossless data compression.

Let $X$ be a discrete random variable with alphabet $\mathcal{X}$ and with probability mass function $p(x)$. The Shannon entropy is

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x) \tag{8}$$

where $\log p(x)$ is taken to base 2 (thus the entropy $H(X)$ is measured in *bits*). Note, unless otherwise specified, the logarithms of form $\log(x)$ will be taken to base 2 for the whole thesis.

The entropy is a functional of the distribution of random variable $X$ and it is independent from the actual values $x \in \mathcal{X}$ taken by the random variable $X$ (see [5]).

**Properties:**

- $H(X) \geq 0$, since the probability mass function $0 \leq p(x) \leq 1$ thus $\log(\frac{1}{p(x)}) \geq 0$.

- $H_b(x) = (\log_b a) H_a(X)$, since the base of the logarithm can be changed by multiplying with a certain factor e.g $\log_b p(x) = \log_b a \log_a p(x)$. Thus, the measure of entropy can be changed just by multiplying with the appropriate factor. Note, if the base of logarithm is taken to 2, the entropy is measured in *bits*. The entropy is measured in *nats* by taking the natural logarithm.

**Example:**

Consider an experiment with two outcomes $x_1$, $x_2$ and the probabilities $p(x_1) = p$ and $p(x_2) = 1 - p$. Let $X$ be the discrete random variable with $\mathcal{X} = [x_1, x_2]$. Thus the entropy computes to

$$H(X) = -p \log p - (1-p) \log(1-p). \tag{9}$$

Let the experiment be a fair coin toss with probability $p = 1/2$, the entropy calculates to $H(X) = 1\,\text{bit}$. This is the maximum, since above quantity is a concave function.

## 2.2 Conditional Entropy

In general the conditional entropy $H(Y|X)$ of a pair of discrete random variables $(X, Y) \sim p(x, y)$ is used to measure the information needed to describe the outcome of $Y$ given the outcome of $X$.

Let $(X, Y)$ be a pair of discrete random variables defined on $x \in \mathcal{X}$, $y \in \mathcal{Y}$ and with joint probability mass function $p(x, y)$. The conditional entropy $H(Y|X)$ is defined as follows (see [5, pp. 17])

$$\begin{aligned} H(Y|X) &= \sum_{x\in\mathcal{X}} p(x)H(Y|X=x) \\ &= -\sum_{x\in\mathcal{X}} p(x) \sum_{y\in\mathcal{Y}} p(y|x)\log p(y|x) \\ &= -\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y)\log p(y|x). \end{aligned} \tag{10}$$

Note that $H(Y|X) \neq H(X|Y)$, since $H(X)-H(X|Y) = H(Y)-H(Y|X)$. This relation is better known in the context of the mutual information $I(X;Y) = H(X) - H(X|Y)$ or $I(X;Y) = H(Y) - H(Y|X)$, respectively (see [5, pp. 21]).

## 2.3 Differential Entropy

The concept of the differential entropy describes the entropy of a continuous random variable.

**Definition 1.** *Let $X$ be a continuous random variable with the probability density function (PDF) $f_X(x)$ on the space $\mathcal{X}$ where $f_X(x) > 0$. The differential entropy is defined as (see [5, pp. 243])*

$$h(X) = -\int_{\mathcal{X}} f_X(x)\log f_X(x)dx \tag{11}$$

It can be seen that the differential entropy is, as the Shannon entropy, a functional of the distribution. Note, the differential entropy can be negative.

## 2.4 Information Loss

Consider a nonlinear input-output system $X_{n+1} = g(X_n)$ with $g$ mapping the output onto itself. According to [5, pp.38], the data processing inequality states that information can only be lost by passing a random variable through a nonlinear system. This information loss depends on the input density as well as on the nonlinearity of the system and is strongly related to the non-injectivity of the system. By the work of [6] and [7] the information loss can be defined as follows:

**Definition 2.** *Let $X_n$ be a random variable with alphabet $\mathcal{X} \subseteq \mathbb{R}$ and let a second random variable defined as $X_{n+1} = g(X_n)$. Then, the information loss by passing $X_n$ through the nonlinear system $g$ is*

$$L(X_n \to X_{n+1}) = \lim_{\hat{X}_n \to X_n} (I(\hat{X}_n; X_n) - I(\hat{X}_n; X_{n+1})) \tag{12}$$

*where $\hat{X}_n$ is the quantized version of the input $X_n$.*

By refining the quantization such that $\hat{X}_n \to X_n$, above quantity reduces to the conditional entropy of the input given the output:

$$\begin{aligned} L(X_n \to X_{n+1}) &= \lim_{\hat{X}_n \to X_n} (I(\hat{X}_n; X_n) - I(\hat{X}_n; X_{n+1})) \\ &= H(X_n | X_{n+1}) \end{aligned} \tag{13}$$

Above equality can be proved through the definitions of the entropy and mutual information. Roughly speaking, the mutual information of a random variable $X$ with itself is equal to its entropy $I(X; X) = H(X)$.

According to Theorem 2 and Corollary 1 of [7] for the univariate case with a piecewise bijective map $g : \mathcal{X} \to \mathcal{X}$, $\mathcal{X} \subseteq \mathbb{R}$, the information loss is given as

$$H(X_n | X_{n+1}) = h(X_n) - h(X_{n+1}) + E\{\log |g^{'}(X_n)|\} \tag{14}$$

where $h(.)$ denotes the differential entropy, assuming the probability density functions of corresponding random variables exist.

The quantity in (14) is of particular interest, since it provides a link between the differential entropy and information loss, as well as to the Kolmogorov-Sinai entropy (discussed later) through the last term of the formula.

## 2.5 Upper Semicontinuity of Information Loss

In section 2.4, the information loss $H(X_n|X_{n+1})$ was defined by equation (14). A very similar expression can be found in [2, pp. 11], where is the speaking of the folding entropy $F(\mu)$.

It turns out that the information loss is identical to the folding entropy, thus,

$$H(X_n|X_{n+1}) = F(P_{X_n}) \tag{15}$$

where $P_{X_n}$ is the probability measure of $X_n$ on the compact manifold $\mathcal{X}$. Note, the probability measure $P_{X_n}$ tends vaguely to the invariant measure $\mu$ (see [2, pp. 15])

$$P_{X_n} \xrightarrow{\text{vague}} \mu \tag{16}$$

**Proposition 1** (Proposition 2.1 [2, pp. 11])**.** *Let $I$ be the set of invariant measures and let $P$ be the set of probability measures on $\mathcal{X}$ with the vague topology and*

$$I = \{\mu \in P : \mu \textit{ is } g\textit{-invariant}\} \tag{17}$$

$$P_{\setminus\sigma} = \{\mu \in P : \mu(\sigma) = 0\} \tag{18}$$

$$I_{\setminus\sigma} = I \cap P_{\setminus\sigma} \tag{19}$$

*(a) The function $F : P_{\setminus\sigma} \to \mathbb{R}$ is concave upper semicontinuous.*
*(b) The restriction of $F$ to $I_{\setminus\sum}$ is affine concave upper semicontinuous.*

Since the probability measure $P_{X_n}$ tend vaguely to $\mu$ it can be said that

$$F(\mu) \leq \lim_{n\to\infty} F(P_{X_n}) \tag{20}$$

Thus,

$$\lim_{n\to\infty} H(X_n|X_{n+1}) \geq L_g := \text{“information loss in the invariant case”} \tag{21}$$

# 3 Preliminaries for Dynamical Systems

A dynamical system describes the time dependence of a point by a fixed rule. The motion of such point is usually given by differential equations. According to the order $r$ of a dynamical system, $r$ differential equations are given.

In the discrete-time case the dynamical system is called map and is usually defined as,

$$x_{n+1} = Ax_n + b \tag{22}$$

$$x_{n+1} = g(x_n) \tag{23}$$

where $A$ is a matrix and $b$ a vector in the multivariate case. For the univariate case $A$ and $b$ are scalars. Equation (22) is the usual form for linear discrete-time maps, whereas notation (23) can be used for nonlinear discrete-time maps. Note, for the rest of this thesis, univariate maps will be denoted as $g$, and $\mathbf{G}$ defines the multivariate case.

In general the motion of a point can be seen by the trajectory in the phase space. The phase space is characterized by the eigenvalues of the map $g(x)$ and defines the space of all possible states which the map can apply. Note, the phase space is also often called manifold and will be denoted as $\mathcal{X}$. This expression implies the same meaning as the uncountable set of outcomes in section 2.

Finally the term *chaos* should be defined. The chaotic systems are highly sensitive to initial conditions, which means that slight differences in the initial conditions yield completely different outcomes. Even though these dynamical systems are deterministic, their outcomes are not predictable. This behaviour is called *deterministic chaos* or just simply *chaos*.

## 3.1 Lyapunov Exponent

The Lyapunov exponent is a common quantity to characterize dynamical systems. In particular, it describes the chaotic motion of a nonlinear dynamical systems. It measures the exponential stretching of adjacent points, which become separated under the forward iteration of a map $x_{n+1} = g(x_n)$, $g : \mathcal{X} \to \mathcal{X}$, $\mathcal{X} \subseteq \mathbb{R}$.
The Lyapunov exponent is also often used to gain knowledge about the stability of the trajectory. Generally speaking, if the local Lyapunov exponent $\lambda(x_0) < 0$, the trajectory is stable and will reach the attractor. Otherwise, if the local Lyapunov exponent $\lambda(x_0) > 0$, the trajectory is not stable.

According to the lines of [9, pp. 21] and [10, pp. 64], the Lyapunov exponent for a one-dimensional discrete-time map can be defined as follows. Figure 1 illustrates the exponential stretching of an interval of adjacent points $[x_0, x_0 + \epsilon_0]$ in the symmetric tent map.

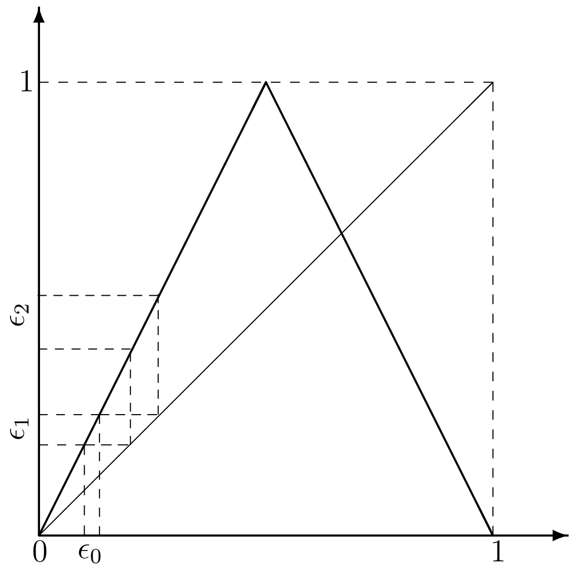


Figure 1: Exponential separation of an interval of adjacent points through the action of the symmetric tent map.

In this simple example, the length of the interval after $N$ iterations is

$$\epsilon_0 \, e^{N\lambda(x_0)} = |g^N(x_0 + \epsilon_0) - g^N(x_0)| \tag{24}$$

Taking the limits $\epsilon_0 \to 0$ and $N \to \infty$, the difference quotient can be obtained and the local Lyapunov exponent $\lambda(x_0)$ can be formally expressed as (see [9, pp. 22])

$$\begin{aligned} \lambda(x_0) &= \lim_{N\to\infty} \lim_{\epsilon_0\to 0} \frac{1}{N} \log \frac{g^N(x_0+\epsilon_0) - g^N(x_0)}{\epsilon_0} \\ &= \lim_{N\to\infty} \frac{1}{N} \log \frac{dg^N(x_0)}{dx_0}. \end{aligned} \tag{25}$$

As it can be seen from equation (24), (25), the interval $\epsilon_0$ is stretched by the factor $e^{\lambda(x_0)}$ at each iteration. This can be interpreted as some information generation about the position of some point $x \in [x_0, x_0 + \epsilon_0]$. Using the chain rule for derivatives of compositions of functions, equation (25) can be rewritten as

$$\lambda(x_0) = \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} \log |g^{'}(x_i)| \tag{26}$$

where $x_i = g(x_{i-1})$ and $|g^{'}(x_i)| = \frac{dg(x)}{dx}|_{x=x_i}$. Note, above quantity needs just the first derivative $g^{'}(x_i)$ at point $x_i$. If the map $g$ is ergodic, then the Lyapunov exponent is the average over all locations $x_i$ and will not change in the basin of attracton and can be written as (see [11, pp. 56])

$$\lambda = \int_{\mathcal{X}} \log |g^{'}(x_0)| d\mu(x_0) \tag{27}$$

In the multivariate case, the dynamical system $\mathbf{G} : \mathcal{X} \to \mathcal{X}$, $\mathcal{X} \subseteq \mathbb{R}^r$, has $r$ Lyapunov exponents $\lambda_1,\ \lambda_2, ...\lambda_r$. For calculating the Lyapunov exponents, the exponential expansion approach can be used (see [9, pp. 98]):

$$(e^{\lambda_1},\, e^{\lambda_2}, \cdots, e^{\lambda_r}) = \lim_{N\to\infty} \left( magnitude\ of\ the\ eigenvalues\ of\ \left\{ \prod_{k=0}^{N-1} J_{\mathbf{G}}(\mathbf{x}_k) \right\} \right)^{1/N} \tag{28}$$

where $J_{\mathbf{G}} = \partial \mathbf{G} / \partial x_j$ is the Jacobian matrix of the dynamical system $\mathbf{G}$.

Thus, by taking the logarithm of above equation and with $\log \prod_{k=0}^{N-1} x_k = \sum_{k=0}^{N-1} \log x_k$, the sum of the Lyapunov exponents can be expressed as

$$\sum_{i=1}^{r} \lambda_i = \lim_{N\to\infty} \frac{1}{N} \sum_{k=0}^{N-1} \log\left(\left|\det\left(J_{\mathbf{G}}(\mathbf{x}_k)\right)\right|\right). \tag{29}$$

Concerning the stability consideration, the sum of all Lyapunov exponents $\sum_i \lambda_i$ are taken into account. If the $\sum_i \lambda_i < 0$, then the trajectory is stable and an existing attractor of the map will be reached. Otherwise, if $\sum_i \lambda_i > 0$, the trajectory is unstable.

Note, unlike the continuous-time case of dynamical systems where chaotic behavior can only appear at $r \geq 3$ dimensions, chaotic behavior of discrete-time systems can appear with just a single dimension. For a more detailed definition of the Lyapunov exponent the interested reader is referred to [11, pp. 129].

## 3.2 Kolmogorov-Sinai Entropy (Metric Entropy)

The metric entropy also known as the Kolmogorov-Sinai entropy (K-S entropy), is one of the most important measures of chaotic motion of an arbitrary dimensional phase space (see [11, pp. 138] and [9, pp. 96]).

The K-S entropy can be seen as a value measuring the creation of information at each iteration under the action of a chaotic map. Generally speaking, the K-S entropy is positive for chaotic systems and zero for nonchaotic systems. Before the K-S entropy can be defined, the Shannon definition of entropy, which describes the uncertainty of predicting the outcome of an probabilistic event, should be recalled (see section 2.1).

The following definition of the K-S entropy is along the lines of [11, pp. 140]. Assuming an invariant probability measure $\mu$ of some map $g(x)$ on the defined space $\mathcal{X}$, the K-S entropy can be denoted as $H_{KS}(\mu)$. Consider a bounded region $M \subseteq \mathcal{X}$ such that $\mu(M) = 1$ and be invariant under the transformation of the map $g(x)$. Let $M$ consist of $k$ disjoint partitions such that

$$M = M_1 \cup M_2 \cup \cdots \cup M_k. \tag{30}$$

Thus, the entropy of the partition $\{M_i\}$ can be written as

$$H(\mu, \{M_i\}) = -\sum_{i=1}^{k} \mu(M_i) \log\left(\mu(M_i)\right). \tag{31}$$

Since the map $g(x)$ evolves in time, it produces a series of intersections $\{M_j^{(n)}\}$ of the form $M_j \cap g^{-1}(M_i)$, $j, i = 1, 2, ..., k$ such that for $n$ iterations the refined partitions $\{M_j^{(n)}\}$ are given as

$$M_{j_1} \cap g^{-1}(M_{j_2}) \cap g^{-2}(M_{j_3}) \cap \cdots \cap g^{-(n-1)}(M_{j_n}) \tag{32}$$

with $j_1, j_2, ..., j_n = 1, 2, ..., k$.

Thus the entropy for the partitions $\{M_i\}$ can be written as

$$\bar{H}(\mu, \{M_i\}) = \lim_{n\to\infty} \frac{1}{n} H(\mu, \{M_i^{(n)}\}) \tag{33}$$

From equation (33) can be seen, that the entropy depends on the original partition $\{M_i\}$. Now, the K-S entropy can be formally defined as the supremum of equation (33) over all initial partitions $\{M_i\}$.

$$H_{KS}(\mu) = \sup_{\{M_i\}} \bar{H}(\mu, \{M_i\}) \tag{34}$$

In other words, the K-S entropy is the remaining uncertainty of the next outcome $x_{n+1}$, if all the past outcomes $x_n, x_{n-1}, ..., x_0$ with a certain *uncertainty* are known[1].

## 3.3 Connection of K-S entropy and Lyapunov exponent (Margulis-Ruelle inequality)

In this section the Margulis-Ruelle inequality will be introduced. This theorem was first proved by Margulis [12] for diffeomorphisms preserving a smooth measure. A more general proof is given by Ruelle [13], which is valid for $C^1$ maps and non-invertible $C^1$ maps (see [14]). Roughly speaking, the Margulis-Ruelle inequality gives an upper bound on the metric entropy (K-S entropy). It states that the K-S entropy does not exceed the sum of positive Lyapunov exponents. By Ruelle [13], the definition of the Margulis-Ruelle inequality is given as follows.

[1]Note that $H(X_{n+1}|X_n) = 0$ by definition, it maybe follows from the discontinuity of entropy (see section 2.2)

**Theorem 1** (Margulis-Ruelle Inequality. [13]). *Let $\mathcal{X}$ be a $C^\infty$ compact manifold and $g : \mathcal{X} \to \mathcal{X}$ a $C^1$ map. Let $I$ be the set of $g$-invariant probability measures on $\mathcal{X}$. Assuming that there exists the Jacobian matrix $J_g(\mathcal{X})$ consisting of a increasing succession of subspaces $V_x^i$, $i = 1, ..., s(x)$ such that the Lyapunov exponents can be written as*

$$\lambda_i(x) = \lim_{n\to\infty} \frac{1}{n} \log \left|\det\{J_g^n(u)\}\right| \text{ if } u \in V_x^i \setminus V_x^{i-1} \tag{35}$$

*Therefore the positive Lyapunov exponent can be calculated as*

$$\lambda_+(x) = \sum_{i:\lambda_i(x)>0} m_i(x)\lambda_i(x) \tag{36}$$

*where $m_i(x)$ is the multiplicity of the Lyapunov exponent $\lambda_i(x)$.*

*Then for every $\mu \in I$ the K-S entropy has upper bound*

$$H_{KS}(\mu) \leq \int \lambda_+(x) d\mu(x). \tag{37}$$

*Proof.* The following proof is along the lines of [13]. There it is said, that a compact Riemannian manifold is partitioned into cubic pieces. Consider now a local simplex that is subjacent of each partition, and consider a set of points (given through the simplex plane) out of those partitions, which gets transformed under the map. The number of intersections between the cubic partition and this transformed set is bounded. Therefore an upper bound on the metric entropy through the entropy of these partitions can be found, since each point of the compact Riemannian manifold can be found in one of these sets.

Let $\mu \in I$ be fixed and the positive Lyapunov exponent defined as by equation (35), (36).

Obtain a partition $\delta_N$ of the compact manifold $\mathcal{X}$ decomposed by a m-dimensional simplex. Let $\Delta^m$ be a $m$-dimensional simplex of planes $t_i$ such that

$$\Delta^m = \left\{ t_i = k_i/N,\ k_i = 1,\ \dots\ ,\ N,\ i = 1,\ ,\dots\ ,m \mid \sum_{i=1}^{m} t_i \leq 1 \text{ and } t_i \geq 0,\ \forall i \right\} \tag{38}$$

where $N > 0$ is a given integer.

Assume that $\mu(t_i) = 0, \forall N$ and $\mathcal{X}$ be a Riemannian manifold. There exists a $C > 0$, such that the number of intersections of a partion $\delta_N$ and $g^n(S)$, $S \in \delta_N$ is the set of points of the simplex $t_i$

$$N(n) < N < C \left|\det\{J_{g^n}(x)^{\Lambda}\}\right| \text{ for any } x \in S. \tag{39}$$

where $\Lambda$ is the spectrum of the Lyapunov exponents at point $x$: $\lambda_x^1 < \lambda_x^2 <$ , ... , $< \lambda_x^{s(x)}$ with related multiplicities $m_x^1$, ... ,$m_x^{s(x)}$.

In fact the number of intersections between $\delta_N$ and $g^n(S)$ is bounded, since they are decomposed by the set of simplex planes $\Delta^m$.

The K-S entropy with respect to $g^n(S)$ and the partition $\delta_N$ satisfies

$$H(\mu, \delta_N) \leq \log N(n) < \log C + \int \log\left(\left|\det\{J_{g^n}(x)^{\Lambda}\}\right|\right) d\mu(x) \tag{40}$$

The proof can be completed by dividing (40) by $n$ and $n \to \infty$, which yields to

$$\bar{H}(\mu, \delta_N) \leq \lim_{n\to\infty} \frac{1}{n} \log C + \lim_{n\to\infty} \int \frac{1}{n} \log\left(\left|\det\{J_{g^n}(x)^{\Lambda}\}\right|\right) d\mu(x) \tag{41}$$

where $\frac{1}{n}\log\left(\left|\det\{J_{g^n}(x)^{\Lambda}\}\right|\right)$ is positive and bounded. If $N$ tends to $N \to \infty$, it can be written for this case

$$H_{KS}(\mu) := \sup_{\delta_N} \bar{H}(\mu, \delta_N). \tag{42}$$

Finally it can be concluded that

$$H_{KS}(\mu) \leq \int \lambda_+(x) d\mu(x). \tag{43}$$

$\square$

## 3.4 Sinai, Ruelle, and Bowen (SRB)-Measures

The SRB-measures play an important role in the ergodic theory of dissipative dynamical systems with chaotic behavior and have their origin in statistical mechanics. They were originally introduced with the theory about the Anosov-diffeomorphisms and Axiom A attractors [3],[1].
Roughly speaking, the SRB measure is an invariant measure under some map

$g$ which has absolutely continuous conditional measures on the unstable direction and it is most compatible with volume when volume is not preserved (see [4]).

The following definition of Axiom A attractors and Anosov-diffeomorphism is taken from [4] and [1, pp. 636].

**Definition 3.** *Let $g : \mathcal{X} \to \mathcal{X}$ be a diffeomorphism on the Riemannian manifold $\mathcal{X}$, i.e. a differentiable map with differentiable inverse $g^{-1}$. The map $g$ is called an Anosov-diffeomorphism, if the tangent space $T_x g$ at every $x \in \mathcal{X}$ is split into the linear unstable and stable eigenspaces $E^u(x)$ and $E^s(x)$. These subspaces are $\partial g/\partial x$ (usually written as $Dg$)-invariant where the unstable part $Dg_{|E^u}$ is uniformly expanding and the stable part $Dg_{|E^s}$ is uniformly contracting.*

**Definition 4.** *A $g$-invariant set $\Lambda$ is called an attractor, and has a basin of attraction $U$ in its vicinity, if every $x \in U$ reaches the attractor as the map evolves in time, $g^n x \to \Lambda$. The attractor $\Lambda$ is called an Axiom-A attractor, if the tangent bundle $T_x\Lambda$ is split into $E^u\Lambda$ and $E^s\Lambda$ as by definition above.*

Theorem 1 of [4] states the original definition of the SRB-measure by Sinai, Ruelle, and Bowen, and can be summarized as follows:

**Theorem 2.** *Let $g$ be a $C^2$ diffeomorphism with an Axiom-A attractor $\Lambda$. Then there is a unique $g$-invariant Borel probability measure $\mu$ on $\Lambda$ that is characterized by each of the following (equivalent) conditions:*

(i) $\mu$ has absolutely continuous conditional measures on unstable manifolds, i.e. every Axiom-A attractor $\Lambda$ has an invariant measure with density $\rho$ on the unstable manifold

(ii)

$$H_{KS}(\mu) = \int \lambda_+(x) d\mu(x) \tag{44}$$

where $\lambda_+(x) = \sum_{i:\lambda_i(x)>0} m_i(x)\lambda_i(x)$. Note this quantity is known as Pesin identity (discussed in the next subsection).

(iii) there is a set $V \subset U$ having full Lebesgue measure such that for every continuous observable $\phi : U \to \mathbb{R}$, $\forall x \in V$,

$$\frac{1}{n}\sum_{i=0}^{n-1} \phi(g^i x) \to \int \phi d\mu \tag{45}$$

This means that $\mu$ is observable, since it hast positive Lebesgue measure and is therefore called a physical measure.

(iv) $\mu$ is the zero-noise limit of small random perturbations of $g$

.

Note that in general property (ii) is an inequality, if the continuity assumption is not made (see section 3.3).

**Remark:** A stable manifold is the set of points $x$, such that the trajectory starting from a point $x_0$ approaches a stationary point or a closed curve (limit cycles) as some arbitrary map evolves in time. Otherwise, the unstable manifold is said to be the set of points $x$, such that the trajectory going backward in time starting from some $x_n$ approaches the stationary point or the limit cycle. This assumes invertibility of the map.

Assume that an invariant measure $\mu$ has a probability density function (PDF) $f_X$ w.r.t Lebesgue measure. By the Corollary 1.1.1 of [15] if follows that $\mu$ is an SRB measure:

**Corollary 1.** *Let $g$ be a $C^2$ endomorphism on $\mathcal{X}$ with a $g$-invariant Borel probability measure $\mu$. If $\mu$ is absolutely continuous with respect to the Lebesgue measure on $\mathcal{X}$, then $\mu$ has the SRB property.*

## 3.5 Pesin Identity

The Pesin Identity is a special case of the Margulis-Ruelle inequality (see subsection 3.3) which states that the K-S entropy is always bounded by the sum of positive Lyapunov exponents. Through the properties of the Pesin Identity, equality between the K-S entropy and the sum of positive Lyapunov exponents is given.

Let $\mathcal{X}$ be a $C^\infty$ compact manifold and $g : \mathcal{X} \to \mathcal{X}$ a $C^1$ map. Let $I$ be the set of $g$-invariant probability measures on $\mathcal{X}$. Assume a tangent bundle $T_x\mathcal{X}$ of the compact manifold consisting of a increasing succession of subspaces $V_x^i$, $i = 1, ..., s(x)$. Let $\mu \in I$ be fixed, the positive Lyapunov exponent is

$$\lambda_+(x) = \sum_{i:\lambda_i(x)>0} m_i(x)\lambda_i(x) \tag{46}$$

where $m_i(x)$ is the multiplicity of the corresponding Lyapunov exponent. Therefore, the Margulis-Ruelle inequality is given as (see [13])

$$H_{KS}(\mu) \leq \int \sum_{i:\lambda_i(x)>0} m_i(x)\lambda_i(x)d\mu(x) \ . \tag{47}$$

The standing hypotheses for the definition of the Pesin Identity are the following (see [16]):

(i) $\mathcal{X}$ is a $C^\infty$ compact Riemannian manifold without boundary,

(ii) $g$ is a $C^2$ diffeomorphism of $\mathcal{X}$ onto itself and is measure preserving,

(iii) $\mu$ is a $g$-invariant Borel probability measure on $\mathcal{X}$.

Through these hypotheses we are able to define the Pesin Identiy:

**Theorem 3.** *Let $g : \mathcal{X} \to \mathcal{X}$ be a $C^2$ diffeomorphism of a compact Riemannian manifold $\mathcal{X}$ preserving the SRB measure $\mu$. The K-S entropy is given as*

$$H_{KS}(\mu) = \int \sum_{i:\lambda_i(x)>0} m_i(x)\lambda_i(x)d\mu(x) \ . \tag{48}$$

Note, this result follows by property (ii) of the SRB measure in Definition 2. Not only that an SRB measure $\mu$ implies equality between the K-S entropy and the Lyapunov exponents, the reversed implication is also valid (see [16] for a precise statement).

An extension to a greater class of maps is given by in [4]. There it is said that equation (48) also holds for $g$ being an abritrary diffeomorphism and $\mu$ being an SRB measure. Due to Ruelle [13], this result also holds for non-invertible maps.

# 4 Motivating Example: Tent map

As a motivating example consider the symmetric tent map. The tent map is piecewise linear on its corresponding subspaces as well as it is piecewise bijective.

Let the tent map $g(x)$ be defined on the interval $\mathcal{X} = [0, 1]$ as follows,

$$g(x) = \begin{cases} 2x, & \text{if } 0 \leq x < 1/2 \\ 2(1-x), & \text{if } 1/2 \leq x \leq 1 \end{cases} . \tag{49}$$

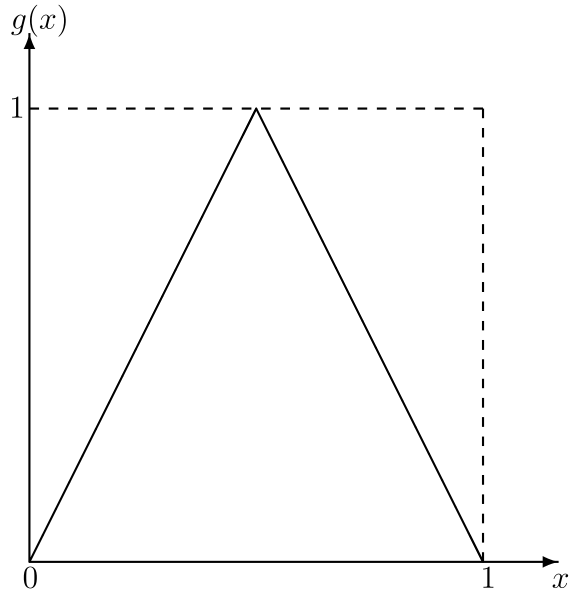


Figure 2: Piecewise bijective symmetric tent map.

Consider now a real-valued random variable $X_n$ with PDF $f_{X_n} = 1$, $\forall x \in \mathcal{X}$ and define a second random variable as $X_{n+1} = g(X_n)$. It can be shown that the PDF is invariant under the action of the map $g(x)$. From Corollary 1 follows that the absolute continuous conditional measure is given (SRB measure property), which means that Pesin's identity holds for this map. Thus the K-S entropy is equal to the sum of Lyapunov exponents. Since the map is univariate, there exists just one Lyapunov exponent. As it can be seen from Figure 4 the map is non-injective on $\mathcal{X}$, therefore the map has some information loss.

From the following can be obtained that, the K-S entropy evaluates to $H_{KS} = 1$ Bit and the information loss also evaluates to $H(X_n|X_{n+1}) = 1\,\mathrm{Bit}$.

By using the method of transformation [8, pp. 130], the marginal PDF of the second random variable is given as

$$f_{X_{n+1}}(x_{n+1}) = \sum_{x_i \in g^{-1}(x_{n+1})} \frac{f_{X_n}(x_i)}{|g^{'}(x_i)|} = \frac{1}{2} + \frac{1}{2} = 1 \tag{50}$$

where the derivative of the map is constant with $|g^{'}(x)| = 2$.

Since the logarithm of the PDF computes to zero, the differential entropies of the input and output compute also to zero. Thus, these results yield,

$$\begin{aligned} &H(X_n|X_{n+1}) = h(X_n) - h(X_{n+1}) + E_\mu\{\log|g^{'}(X_n)|\} = \\ &= 0 - 0 + 1 = 1\,\mathrm{Bit} \end{aligned} \tag{51}$$

The Lyapunov exponent and the Kolmogorov-Sinai entropy computes to,

$$H_{KS} = \lambda = \int_{\mathcal{X}} f_X(x) \log\left(\frac{dg(x)}{dx}\right) dx = \frac{1}{2}\log 2 + \left(1 - \frac{1}{2}\right)\log 2 = 1\,\mathrm{Bit}. \tag{52}$$

By (51), (52) can be seen, that the information loss is equal to the Kolmogorov-Sinai entropy. The question what arises, does equality between K-S entropy and information loss hold every time?

$$H_{KS} \stackrel{?}{=} H(X_n|X_{n+1}) \tag{53}$$

# 5 Main Result

The main result is connected to the formalism about the entropy production $e_g(\mu)$, and takes us to the mechanisms of statistical mechanics (see [2]). In this work of Ruelle the entropy production of an idealized thermostat is investigated. Generally speaking, the entropy production is the amount of entropy pumped out under the action of some map $g$ at each iteration.

**Definition 5.** *Let $\mathcal{X}$ be a compact Riemannian manifold, and $g : \mathcal{X} \to \mathcal{X}$ a $C^1$ diffeomorphism. Consider a real valued random variable $X_n$ with some density $f_{X_n}$ and define a second random variable by $X_{n+1} = g(X_n)$. Thus the entropy production of the map $g$ at each iteration is*

$$e_g = h(X_n) - h(X_{n+1}). \tag{54}$$

**Corollary 2.** *If $f_{X_n}$ is invariant under the transformation of $g$, then the entropy production is zero $e_g = 0$.*

In Corollary 2, the differential entropies get equal to each other, $h(X_n) = h(X_{n+1})$.

Let $\mu$ be a vague limit of the density $f_{X_n}(x) = (1/n)\sum_{k=0}^{n-1} g^k(f_X(x))$ which gets transformed under the action of some map $g$, and let $g$ be as stated above by Definition 5, it can be generally said that the entropy production is always positive [2, pp. 16].

$$e_g(\mu) \geq 0 \tag{55}$$

Consider a dynamical system $g : \mathcal{X} \setminus \Psi \to \mathcal{X}$, where $\Psi \subset \mathcal{X}$ is a closed subset of $\mathcal{X}$. Let $\mu$ be a positive measure on $\mathcal{X} \setminus \Psi$, such that $\mu(\Psi) = 0$. The entropy production $e_g(\mu)$ for this non-invertible maps is originally given as ([2, pp. 12])

$$e_g(\mu) = F(\mu) - \mu(\log J) \tag{56}$$

where $F(\mu)$ denotes the folding entropy and $J$ is the Jacobian matrix of $g$.

Let $P_{X_n}$ be the probability measure of $X_n$ on the compact manifold $\mathcal{X}$. If the probability measure $P_{X_n}$ tends vaguely to $\mu$, then the information loss is upper semicontinuous,i.e., $\lim_{n\to\infty} H(X_n|X_{n+1}) \geq F(\mu)$ (see section 2.5).

Assume that $\mu$ is a $g$-ergodic probability measure, then quantity $\mu(\log J)$ satisfies (see (26), (27) in section 3.1),

$$\mu(\log J) = \sum \text{pos. Lyapunov exp.} + \sum \text{neg. Lyapunov exp.}. \tag{57}$$

Note, the Kolmogorov-Sinai entropy $H_{KS} \leq \sum$ pos. Lyapunov exp (see section 3.3).

**Theorem 4** (Main Result)**.** *Let $\mathcal{X}$ be a compact Riemannian manifold, and $g : \mathcal{X} \to \mathcal{X}$ a $C^1$ map or a non-invertible map and $P_{X_n}$ be a g-ergodic probability measure. Let $X_n$ be a real-valued random variable such that $X_n \sim P_{X_n}$, and define a second random variable by $X_{n+1} = g(X_n)$. Then,*

$$H_{KS} \leq \lim_{n\to\infty} H(X_n|X_{n+1}) + \left|\sum \text{neg. Ljapunov exponents}\right|. \tag{58}$$

This inequality is through the conjecture made by Ruelle [2, pp.16], and is valid for $C^1$ maps and non-invertible maps.

**Corollary 3.** *If $P_{X_n}$ invariant and has a PDF $f_{X_n}$, then one has equality in* (58)*.*

One inequality behind equation (58) is due to the positivity of entropy production $e_g(\mu) \geq 0$. The other inequality is the Margulis-Ruelle inequality (see section 3.3). The third inequality is due to the upper semicontinuity of the information loss (see section 2.5) The following proof is along the lines of [2, pp. 16].

*Proof of positivity of entropy production:*
The quantity $e_g(\mu) \geq 0$ for a physically reasonable $\mu$ is close to the result obtained when $g$ is a diffeomorphism.

**Theorem 5** (Theorem 2.4 [2, pp. 16] )**.** *Let $X$ be a random variable with alphabet $\mathcal{X}$ and with the probability density function $f_X(x)$ such that the obtained differential entropy $h(X)$ is finite. If $\mu$ is a vague limit of the density $f_{X_n}(x) = (1/n)\sum_{k=0}^{n-1} g^k(f_X(x))$ which gets transformed under the action of some map $g$ and if time $n$ tend towards $\infty$, then $e_g(\mu) \geq 0$.*

Through the proposition 2.2(c) and 2.2(b) of [2, pp. 13] can be obtained,

$$e_g(\mu) \geq \limsup_{n\to\infty} e_g(f_X^n(x)) \tag{59}$$

$$e_g(\mu) \geq (1/n)\sum_{k=0}^{n-1} e_g(g^k(f_X(x))) \tag{60}$$

where $\sum_{k=0}^{n-1} e_g(g^k(f_X(x))) = -h(X_n) + h(X)$. Therefore we can rewrite (60) to,

$$e_g(\mu) \geq \limsup_{n\to\infty} \frac{1}{n}[-h(X_n) + h(X)] \tag{61}$$

Since $h(X) \leq \log vol\mathcal{X}$ it can be obtained $e_g(\mu) \geq 0$. □

With this proof and the proof of the Margulis-Ruelle inequality (section 3.3) as well as taking the upper semicontinuity of the information loss into account (see section 2.5), the main result can be proved as follows:

*Proof of main result* (58).
Let $\mu$ be a vague limit of the probability measure $P_{X_n}$ of $X_n$ on the compact manifold $\mathcal{X}$ and let $\mu$ be a $g$-ergodic measure. Thus, the following can be stated:

$$\begin{aligned} 0 &\leq e_g(\mu) \\ &= F(\mu) - \mu(\log J_g) \\ &\leq \lim_{n\to\infty} H(X_n|X_{n+1}) - \\ &- \sum pos.\ Lyapunov\ exp. + |\sum neg.\ Lyapunov\ exp.| \end{aligned} \tag{62}$$

$$\begin{aligned} H_{KS} &\leq \sum pos.\ Lyapunov\ exp. \leq \\ &\leq \lim_{n\to\infty} H(X_n|X_{n+1}) + |\sum neg.\ Lyapunov\ exp.| \end{aligned} \tag{63}$$

The Kolmogorov Sinai entropy $H_{KS}$ is bounded by the sum of positive Lyapunov exponents due to the Margulis-Ruelle inequality

$$H_{KS} \leq \sum pos.\ Lyapunov\ exp. \tag{64}$$

which completes the proof of the main result

$$H_{KS} \leq \lim_{n\to\infty} H(X_n|X_{n+1}) + |\sum neg.\ Lyapunov\ exp.|. \tag{65}$$

□

The entropy production $e_g(\mu)$ can fall into three cases, and further subdivided into bijective and non-injective maps. Note, bijectivity assumes surjectivity and injectivity of the map. Figure 5 illustrates such distinction.

The table below gives a general overview of certain maps admitting such required properties. For the following examples, only univariate maps will be considered.

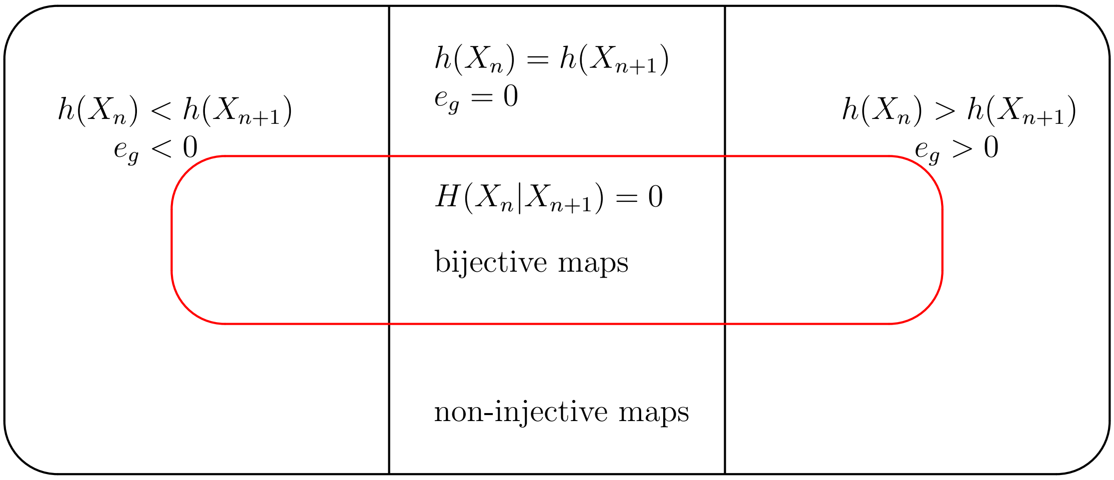


Figure 3: Classification of discrete-time maps, according to the entropy production and the bijectivity of the maps. The inner circle encloses the bijective maps, whereas the outer circle encloses the non-injective maps.

| | injectivity | maps | section |
|---|---|---|---|
| $e_g(\mu) = 0$: | bijective: | identity map | |
| | non-injective: | asymmetric tent map<br>... | sec. 6.1 |
| $e_g(\mu) > 0$: | bijective:<br>(multivariate) | e.g. Baker's Map<br>...maps that are measure preserving and admitting a cantor set | |
| | non-injective: | Gauss iterated map<br>Sine-circle map<br>... | sec. 6.2<br>sec. 6.3 |
| $e_g(\mu) < 0$: | bijective,<br>non-injective | Generalized $2x$ *mod* 1 map<br>...maps that are not measure preserving, open systems | sec. 6.4 |

Since in the univariate case just one Lyapunov exponent exists, there exists no negative Lyapunov exponent, otherwise map $g$ would be not chaotic. Note, chaotic behaviour of discrete type maps is possible with a single positive Lyapunov exponent. In fact the main result reduces to following,

$$H_{KS} \leq H(X_n|X_{n+1}) + \underbrace{|\sum \text{neg. Ljapunov exponents}|}_{=0} \tag{66}$$

$$H_{KS} \leq H(X_n|X_{n+1}) \tag{67}$$

assuming that $\mu$ is a $g$-ergodic probability measure.

Equation (67) is of particular interest, implying that there exists no chaotic univariate map which is bijective.[2]

[2]For bijective maps the information loss vanishes, while chaotic behaviour is characterized by a positve K-S entropy.

# 6 Examples

In this section some examples already mentioned in section 5 before, will be presented in the following subsections. As it was said, only univariate maps will be considered.

The presented numerical results for the information loss $H(X_n|X_{n+1})$, were generated via MATLAB/Simulink 2010b. The starting distribution was chosen to be the uniform distribution. To get representative results, a starting vector of $10^6$ values was generated and passed through the map which evolves in time. In general the iteration time was chosen to be $n = 20$, except it was necessary to iterate longer.

## 6.1 Asymmetric Tent Map

The asymmetric tent map is a non-injective map $g : \mathcal{X} \to \mathcal{X}$ and is defined with symmetry parameter $a$ by equation (68). This map admits $e_g(\mu) = 0$, as can be seen later.

Let the map (shown in Figure 4) be given as:

$$g(x) = \begin{cases} \frac{x}{a}, & \text{if } 0 \leq x < a \\ \frac{1-x}{1-a}, & \text{if } a \leq x \leq 1 \end{cases} \tag{68}$$

with $a \in (0, 1)$.

Assume that at time $n$ the state $g(x_n)$ is uniformly distributed, i.e., $f_{X_n}(x) = 1,\ \forall\ x \in [0, 1]$. Now, think of a second RV $X_{n+1} = g(X_n)$. Using the method of transformation [8, pp. 130], the marginal PDF is given as

$$f_{X_{n+1}}(x_{n+1}) = \sum_{x_i \in g^{-1}(x_{n+1})} \frac{f_{X_n}(x_i)}{|g'(x_i)|} \tag{69}$$

where $g^{-1}(x_{n+1})$ is the preimage of $x_{n+1}$.

The derivative of $g(x_n)$ is given by equation (70), the elements of the preimage by equation (71) and equation (72).

$$|g'(x_n)| = \begin{cases} \frac{1}{a}, & \text{if } 0 \leq x_n < a \\ \frac{1}{1-a}, & \text{if } a \leq x_n \leq 1 \end{cases} \tag{70}$$

$$x_{1,n} = a \cdot g(x_n) = \begin{cases} x_n, & \text{if } 0 \leq x_n < a \\ \frac{a}{1-a}(1 - x_n), & \text{if } a \leq x_n \leq 1 \end{cases} \tag{71}$$

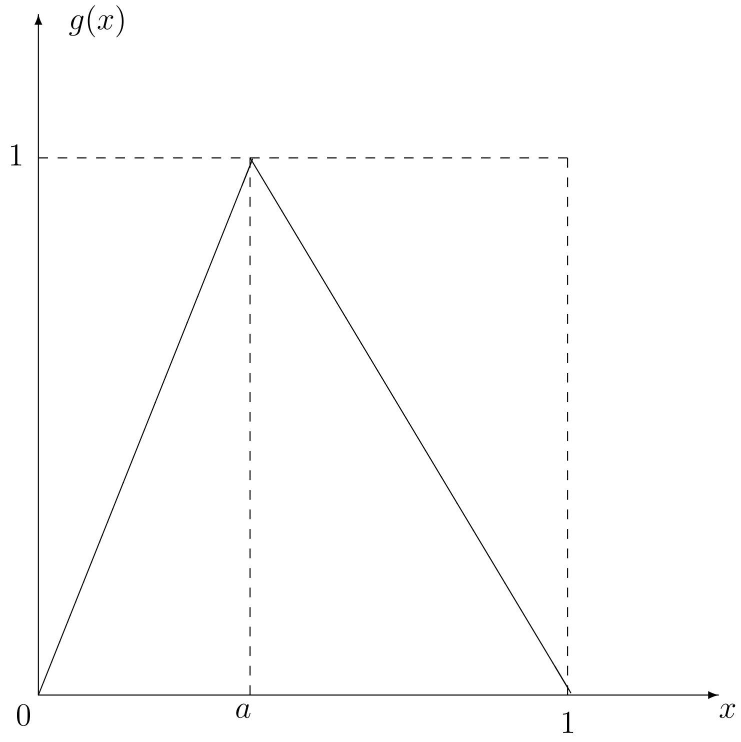


Figure 4: Asymmetric tent map defined via equation (68)

$$x_{2,n} = 1 - (1-a) \cdot g(x_n) = \begin{cases} 1 - (1-a)\frac{x_n}{a}, & \text{if } 0 \leq x_n < a \\ x_n, & \text{if } a \leq x_n \leq 1 \end{cases} \tag{72}$$

Since $f_{X_n}$ is constant and we always have an element of the preimage in $[0, a)$ and $[a, 1]$, the marginal PDF of $X_{n+1}$ is equivalent to the one of $X_n$, $f_{X_{n+1}} = f_{X_n}$. In other words, $X_{n+1}$ is a uniformly distributed RV on $[0, 1]$. Therefore, the asymmetric tent map has an invariant PDF and the differential entropy $h(X_n)$ and $h(X_{n+1})$ must be equivalent.

### 6.1.1 Lyapunov exponent $\lambda$ and K-S entropy $H_{KS}$

The KS-entropy will be calculated along the lines of [9]. With $p_1(x) = P(0 \leq x < a)$, the KS-entropy is [9, pp. 116]

$$H_{KS} = p_1(x) \log\left(\frac{1}{p_1(x)}\right) + (1 - p_1(x)) \log\left(\frac{1}{1 - p_1(x)}\right). \tag{73}$$

Finally we obtain from equation (73) and the probability $p_1(x) = a$ the KS-entropy,

$$H_{KS} = a \cdot \log\left(\frac{1}{a}\right) + (1-a) \cdot \log\left(\frac{1}{1-a}\right) \tag{74}$$

which is equal to the conditional entropy (77).
It is expected that the KS-Entropy is equal to the positive Lyapunov exponent $\lambda_p$ (see section 3.5) which can be easily proved.

Since the asymmetric tent map is a one dimensional problem, there is just one Lyapunov exponent. Therefore, by assuming to be $g$-ergodic, the Lyapunov exponent tends to the expectation value,

$$\lambda = \int_{\mathcal{X}} f_X(x) \log |g^{'}(x)| dx. \tag{75}$$

Thus, the Lyapunov exponent calculates according to the symmetry parameter $a$ to,

$$\lambda = a \cdot \log\left(\frac{1}{a}\right) + (1-a) \cdot \log\left(\frac{1}{1-a}\right). \tag{76}$$

#### 6.1.2 Information Loss (Analytic)

In this subsection equality between the conditional entropy $H(X_n|X_{n+1})$ and $E\{\log|g^{'}(X_n)|\}$ as well as the equality between the conditional entropy $H(X_n|X_{n+1})$ and the Kolmogorov-Sinai entropy $H_{KS}$ will be shown.

Taking the equation of the information loss (14) from section 2.4 and the knowledge about the differential entropy $h(X_n)$ and $h(X_{n+1})$, the equality $H(X_n|X_{n+1}) = E\{\log|g^{'}(X_n)|\}$ will be obtained. The expression $E\{\log|g^{'}(X_n)|\}$ can be calculated as,

$$\begin{aligned} H(X_n|X_{n+1}) = E\{\log|g^{'}(X_n)|\} &= \int_{\mathcal{X}} f_{X_n}(x) \log(|g^{'}(x)|) dx \\ &= \int_0^a \log\left(\frac{1}{a}\right) dx + \int_a^1 \log\left(\frac{1}{1-a}\right) dx \\ &= a \cdot \log\frac{1}{a} + (1-a) \cdot \log\frac{1}{1-a} =: H_S(a, 1-a) \end{aligned} \tag{77}$$

Finally, these results can be combined and we get

$$H_{KS} = H(X_n|X_{n+1}). \tag{78}$$

This result means, the information gain denoted by the Kolmogorov-Sinai entropy rate, which is a result of the exponential stretching, will be abolished by the information lost due to the non-injectivity of the map.

### 6.1.3 Information Loss (Numeric Upper Bound)

This map is defined by (68) on $\mathcal{X} = [0, 1]$ and is controlled by parameter $a$. The analytical results show, that $H_{KS} = H(X_n|X_{n+1})$, and the uniform distribution is invariant.

Let $a = 1/3$. The estimated information loss is plotted in Figure 5. It can be seen, that the value is constant over the time and is in between 0.91 Bit and 0.92 Bit. This is something what has been expected, since the analytical result show for the information loss and K-S entropy $H(X_n|X_{n+1}) = H_{KS} =$ 0.918 Bit. By Figure 6 can be clearly seen, that $X_n$ has a invariant PDF under the action of the map.

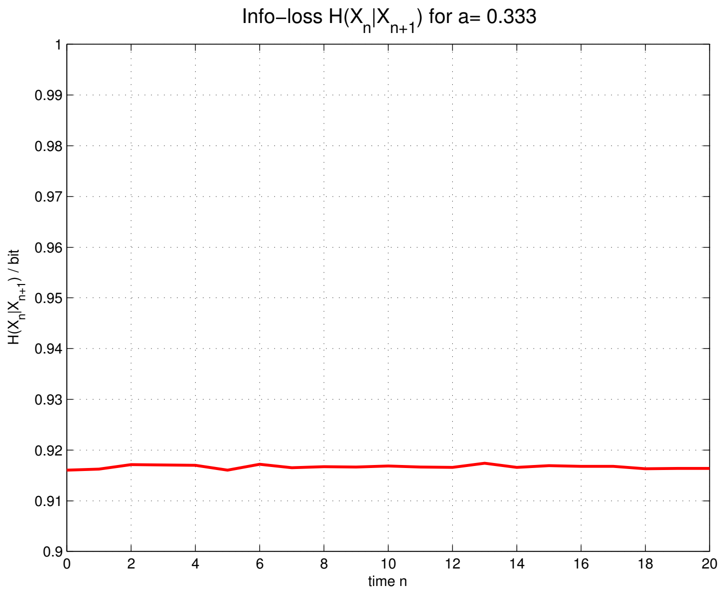


Figure 5: Information Loss over time n for parameter $a = 1/3$.

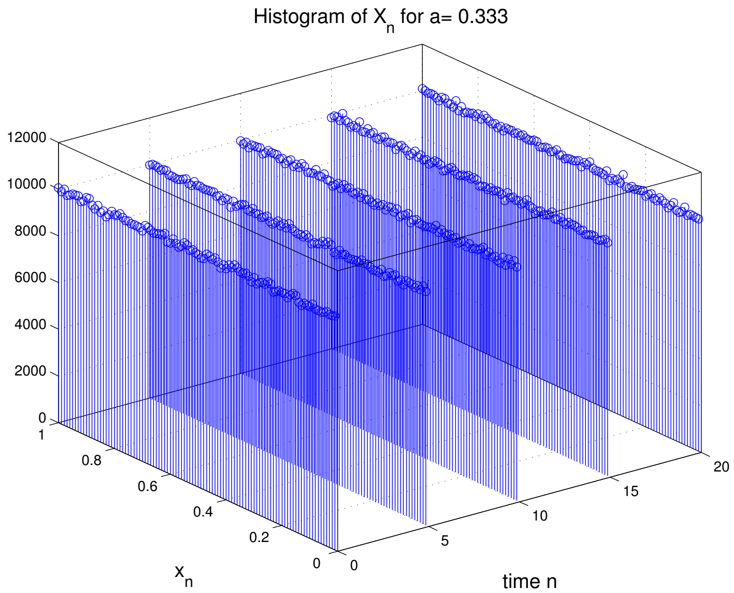


Figure 6: Histogram of $X_n$ under the action of the asymmetric tent map ($a = 1/3$) as time evolves.

Consider now the symmetric case of the map with $a = 1/2$. As Figure 7 shows, the numerical result shows numbers close to the analytical value of $H(X_n|X_{n+1}) = H_{KS} = 1\,\text{Bit}$. At time $n = 12$ a jump of the graph can be seen. By looking at the histogram (Figure 8) there can be no remarkable difference seen. So this jump will be due to nummerical issues. A third experiment with $a = 2/3$ was computed. But as it is identical to the first case with $a = 1/3$ the figures have been omitted.

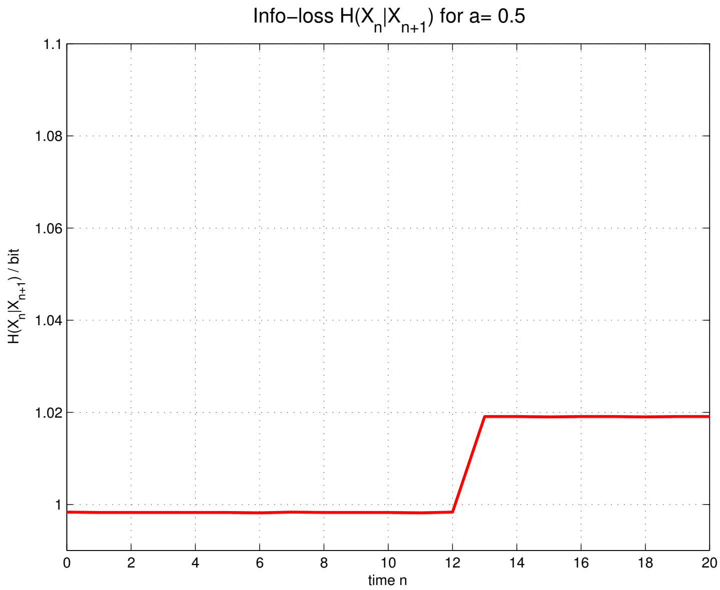


Figure 7: Information Loss over time n for parameter $a = 1/2$.

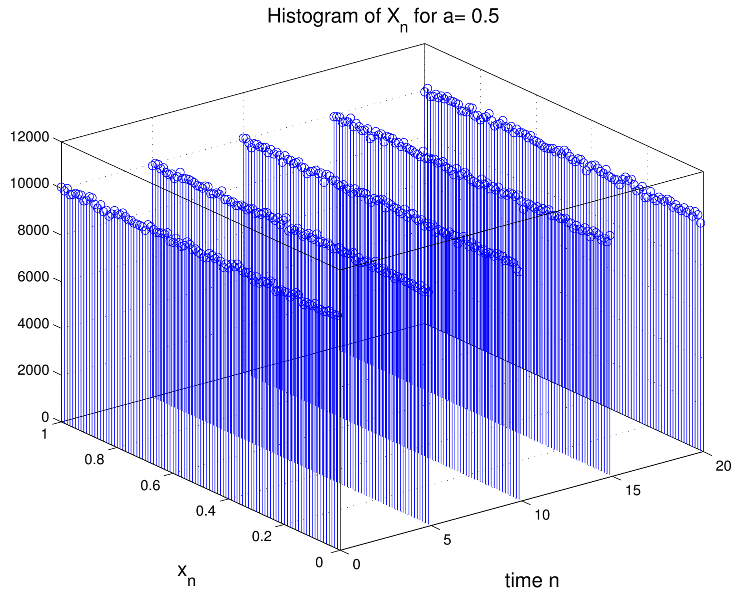


Figure 8: Histogram of $X_n$ for the symmetric case ($a = 1/2$).

## 6.2 Gaussian Map

The Gaussian map is characterized by its exponential function and by two control parameters. The map is defined as

$$g(x) = e^{-\alpha x^2} + \beta \tag{79}$$

where $\alpha > 0$ controls the width and $-1 < \beta < 1$ gives the offset. As it can be seen from Figure 9, the map is non-injective on its domain $\mathcal{X} \in [-1, 1]$, and is mapping the input onto itself.

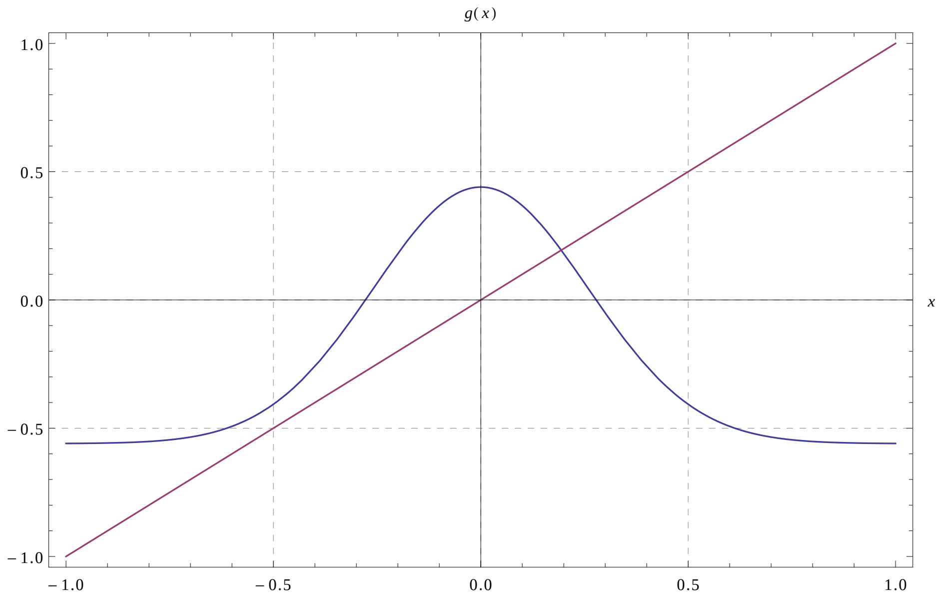


Figure 9: Gaussian map plotted with $\alpha = 7.5$ and $\beta = -0.56$.

As it can be seen in Figure 9, the Gaussian map has a similar shape as the logistic map, but it behaves completely different since it has two parameters. By a closer look at the Gaussian map, someone may see that the map can exhibit up to three fixed points, stable and unstable ones, respectively. Looking at the bifurcation diagram [17, pp. 196] for the Gaussian map, chaotic behaviour appears only at certain parameter constellations [18], where at least one unstable fixed point exists.

The derivative of the map (79) is given as,

$$|g'(x_n)| = -2\alpha x_n e^{-\alpha x_n^2} \tag{80}$$

and the preimage $g^{-1}(x_n)$ of $x_n$ has the elements

$$g^{-1}(x_n)_{(1)} = +\sqrt{\frac{\ln(g(x_n)) - \beta}{-\alpha}} \tag{81}$$

$$g^{-1}(x_n)_{(2)} = -\sqrt{\frac{\ln(g(x_n)) - \beta}{-\alpha}}. \tag{82}$$

Since $|g'(x_n)|$ is not constant, the uniform distribution will be not invariant under the action of the map. Note, not necessarily every PDF is noninvariant, there might be some certain PDF which is invariant under the action of the map. Here, however, the invariant measure does not posses a density. Thus, the differential entropy of the input and output are supposed to be $h(X_n) > h(X_{n+1})$ and the entropy production $e_g(\mu) > 0$.

### 6.2.1 Lyapunov exponent $\lambda$ and K-S entropy $H_{KS}$

The Lyapunov exponent could not be theoretically determined, since it is a tough job to find an analytical expression. Thus a numerical estimation of the Lyapunov exponent as a function of $\beta$ is plotted at Figure 10 for a given $\alpha = 7.5$ (figure taken from [18]). It can be clearly seen that chaotic behaviour can only appear at a small region of $\beta$.

The invariant measure of the map is an SRB-measure that satisfies the Pesin identity (see section 3.5). Thus, the Kolmogorov-Sinai entropy can be estimated through the Lyapunov exponent and is given as,

$$H_{KS} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \log\left|2\alpha x_n e^{-\alpha x_n^2}\right|. \tag{83}$$

According to the plot of the Lyapunov exponent, the K-S entropy will not exceed 1/2 Bit for given $\alpha = 7.5$.

### 6.2.2 Information Loss (Analytic)

Since, we do not have a good knowledge about the invariant measure $\mu$, the information loss can not be computed by the usual way given in section 2.4.

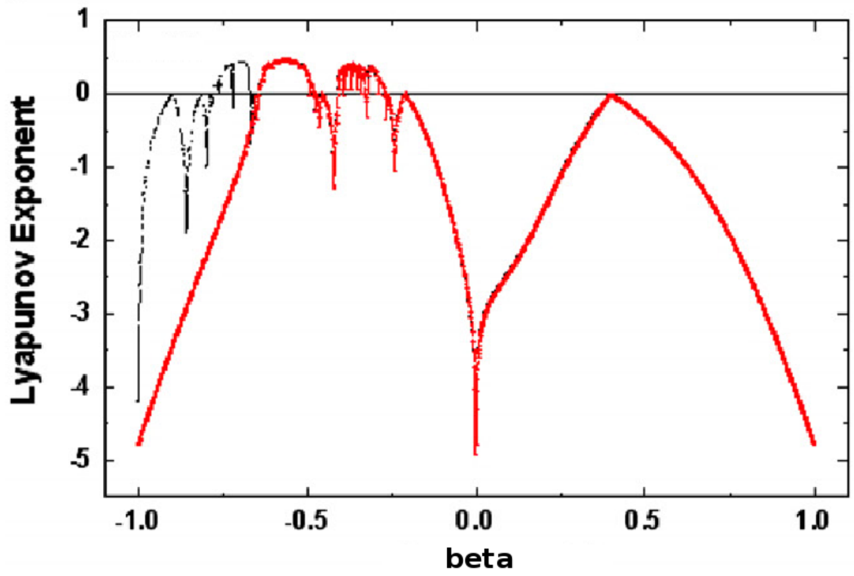


Figure 10: Estimation of the Lyapunov exponent as a function of $\beta$ for the Gaussian map with $\alpha = 7.5$. Figure taken from [18]

An estimation of the upper limit to the information loss can be given as follows:

$$H(X_n|X_{n+1}) = \int_{\mathcal{X}} H(X_n|X_{n+1} = x) dP_{X_{n+1}}(x) \tag{84}$$

$$\leq \int_{\mathcal{X}} \log card(g^{-1}(x)) dP_{X_{n+1}}(x) \tag{85}$$

where $card(g^{-1}(x))$ is the cardinality of the preimage. But $card(g^{-1}(x)) = 2$ for all $x$, thus,

$$H(X_n|X_{n+1}) \leq \log 2 = 1\,\mathrm{Bit}. \tag{86}$$

To summarize, assuming $\mu$ being a $g$-ergodic SRB-measure and the entropy production being strictly positive, it can be said for a given parameter $\alpha = 7.5$

$$\underbrace{H_{KS}}_{\leq 1/2\,\mathrm{Bit}} \leq \underbrace{H(X_n|X_{n+1})}_{\leq 1\,\mathrm{Bit}} \tag{87}$$

### 6.2.3 Information Loss (Numeric Upper Bound)

The Gaussian map or the so-called mouse map is defined by (79) on $\mathcal{X} = [-1, 1]$. This map will be controlled by two parameters $\alpha > 0$ and $-1 < \beta < 1$. Since the map has two branches, the information loss will not exceed one bit.

By the first experiment the parameters are defined as $(\alpha, \beta) = (7.5, -0.6)$. For this configuration the information loss tends to be approximately $H(X_n|X_{n+1}) \approx 0.6\,\text{Bit}$ after the transient has gone. As the Lyapunov exponent tends to $1/2$ for this setting, the K-S entropy can be computed as $H_{KS} \approx 0.5\,\text{Bit}$. The information loss is plotted in Figure 11. The histogram in Figure 12 shows that the uniform distribution is not invariant.

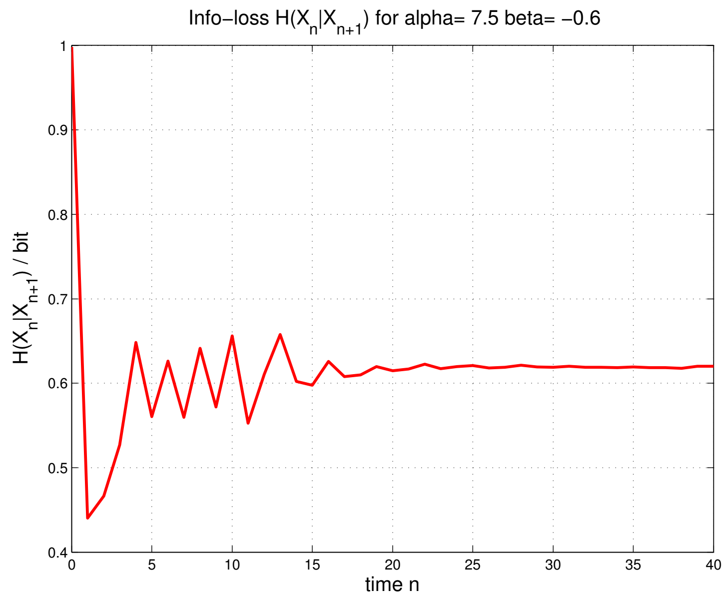


Figure 11: Information Loss over time $n$ for parameter set $(\alpha, \beta) = (7.5, -0.6)$. After the transient has gone, $H(X_n|X_{n+1})$ tends to 0.6 Bit.

The next two experiments with the setting $(\alpha, \beta) = (7.5, -0.55)$ and $(\alpha, \beta) = (7.5, -0.3)$ show that the information loss still converges to a certain value. Figure 13 and Figure 15 illustrate the information loss $H(X_n|X_{n+1})$ as a function of time. It can be clearly seen by Figure 13 that the information loss slightly greater than by the first experiment and by Figure 15 can be obtained that the information loss is converging slower than by the other two

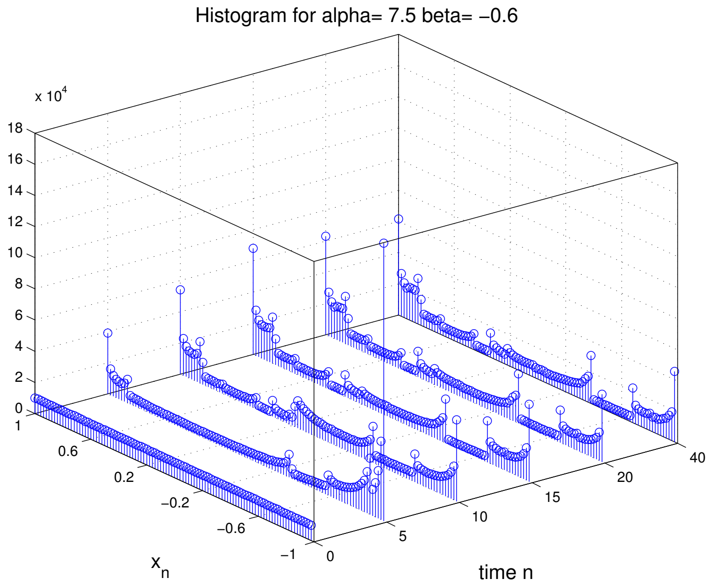


Figure 12: Histogram for the Gaussian map with $(\alpha, \beta) = (7.5, -0.6)$. The uniform distribution is not invariant under the action of the map.

experiments. It is not obvious, by the plot of the Lyapunov exponent shown at Figure 10 at section 6.2 which value the Lyapunov exponent takes, it is supposed that it is still $\lambda \approx 1/2$. The histograms of the experiments with the parameter sets $(\alpha, \beta) = (7.5, -0.55)$ and $(\alpha, \beta) = (7.5, -0.3)$ are plotted by Figure 14 and Figure 16, respectively.

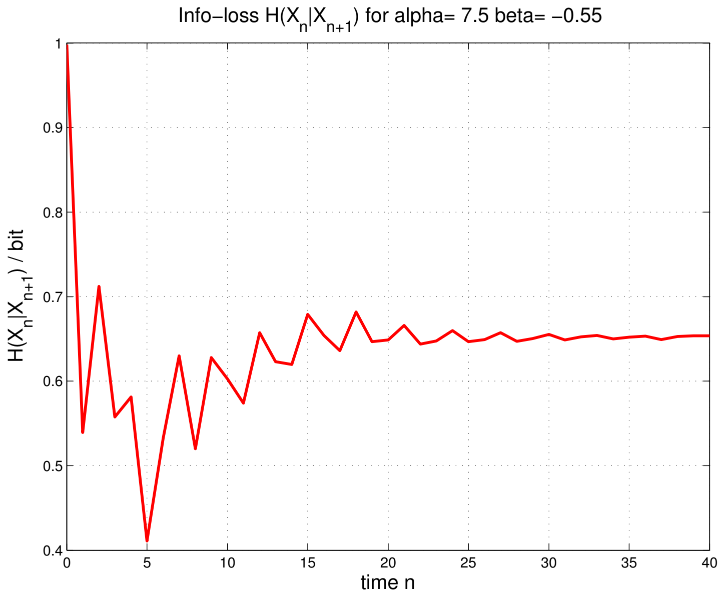


Figure 13: Information Loss $H(X_n|X_{n+1})$ over time $n$ for parameter set $(\alpha, \beta) = (7.5, -0.55)$.

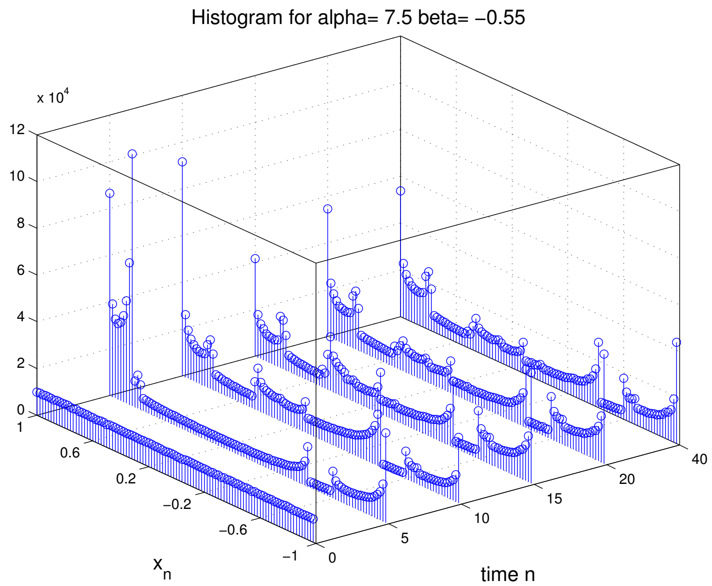


Figure 14: Histogram for the Gaussian map with $(\alpha, \beta) = (7.5, -0.55)$.

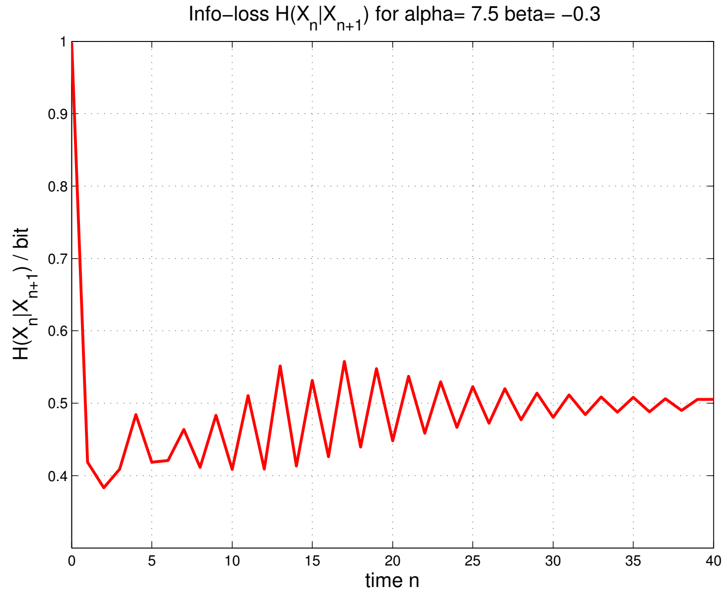


Figure 15: Information Loss $H(X_n|X_{n+1})$ over time $n$ for parameter set $(\alpha, \beta) = (7.5, -0.3)$. The information loss converges slower than in the other experiments.

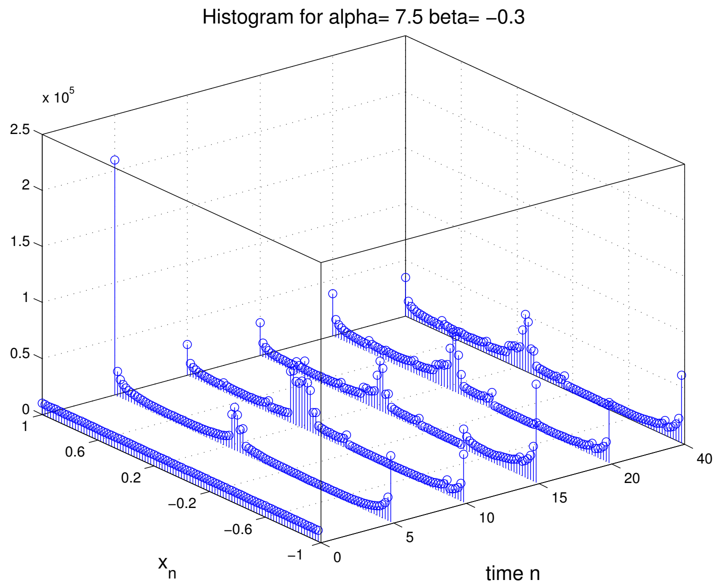


Figure 16: Histogram for the Gaussian map with $(\alpha, \beta) = (7.5, -0.3)$.

## 6.3 Sine-Circle Map

The sine-circle map takes here name from its nonlinearity, which has the specific form of a sine-function. This map has become a standard model to investigate the quasi-periodic route to chaos, since it exhibits frequency-locking for certain choices of the parameter set $(K, \Omega)$.
The sine-circle map is defined by [17, pp. 221],

$$g(x) = x + \Omega - \frac{K}{2\pi} \sin 2\pi x \mod 1. \tag{88}$$

The parameter $K > 0$ is the damping factor of the nonlinearity which gets normalized through the $2\pi$ in the denominator. The frequency-ratio at which the trajectory is moved around the circle is given by parameter $\Omega$. Thus, $\Omega$ is a rational number and is restricted to be $0 < \Omega < 1$.

The behaviour of the sine-circle map changes completely at $K = 1$. As the map is bijective for the range of $0 < K < 1$, it gets non-injective for $K > 1$. This change of behaviour is illustrated at Figure 17 and Figure 18.

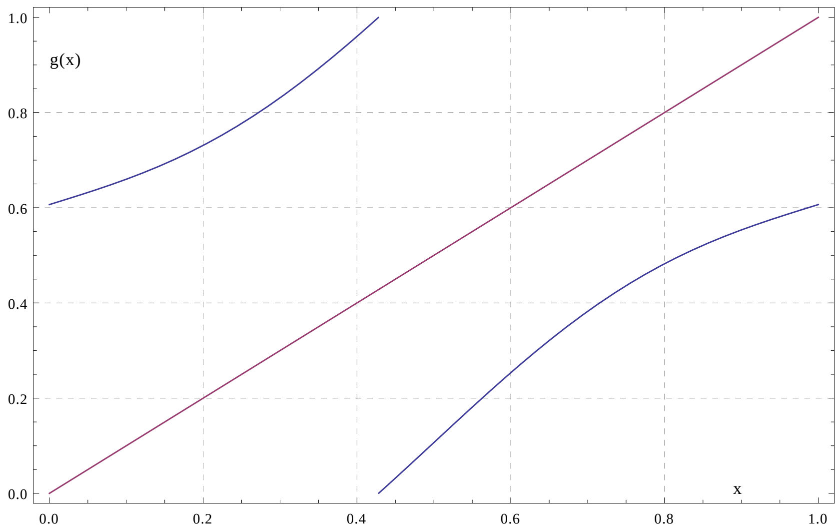


Figure 17: The sine-circle map is bijective for $K = 0.5$.

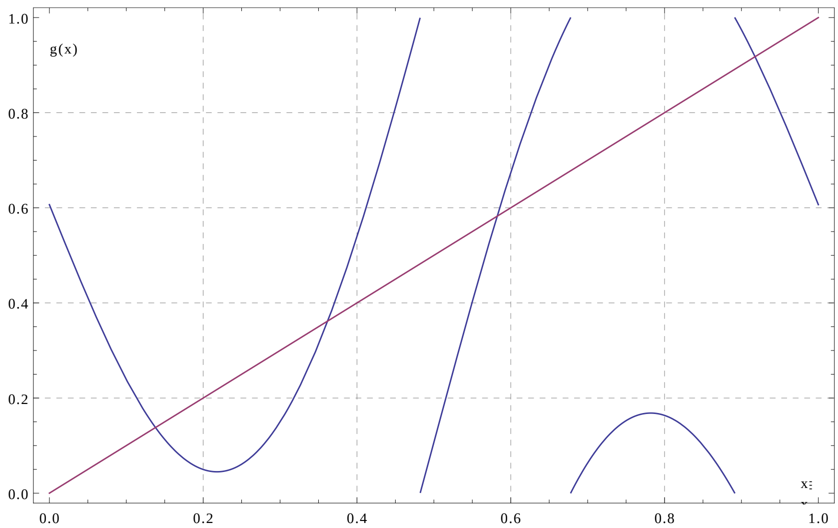


Figure 18: The sine-circle map is non-injective for $K = 5$.

A look at the bifurcation diagram for the sine-circle map [17, pp. 237-238] tells us that chaotic behaviour of the sine-circle map only appears, with parameter $K$ clearly above the critical limit $K = 1$.

Using the method of transformation [8, pp. 130], the marginal PDF of RV $X_{n+1}$ is given as,

$$f_{X_{n+1}}(x_{n+1}) = \sum_{x_i \epsilon g^{-1}(x_{n+1})} \frac{f_{X_n}(x_i)}{|g'(x_i)|}. \tag{89}$$

The derivative of the sine-circle map (88) is

$$g'(x) = \frac{dg(x)}{dx} = 1 - K\cos(2\pi x). \tag{90}$$

Since the sine-circle map hast the constellation $y = x + \sin x$, it is quite a tough job to find the inverse $g^{-1}(x)$ analytically, and can only be numerically estimated for $0 < K < 1$. Unfortunately, not the case we are concerned about.

### 6.3.1 Lyapunov exponent $\lambda$ and K-S entropy $H_{KS}$

Before distinguishing between the case $0 < K < 1$ and $K > 1$, the Lyapunov exponent $\lambda$ should be defined as,

$$\lambda = \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \log |1 - K\cos(2\pi x_n)|. \tag{91}$$

**(i) $\mathbf{0 < K < 1}$**: By the convexity of the logarithm of quantity (91), and by using the Jensen's inequality [5, pp. 25], this quantity can be rewritten as,

$$\lambda \leq \log \left( \underbrace{\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} |1 - K\cos(2\pi x_n)|}_{=1,\ N\to\infty} \right). \tag{92}$$

Since the average calculation inside the logarithm tends to be one, the whole quantity goes to zero. An experiment with random input values $x_n$ between zero and one, and different values of $K$ confirms this assumption.
As a consequence of that, the K-S entropy is to $H_{KS} = 0$ due to Pesin's identity.

**(ii) $\mathbf{K > 1}$** : Again, the Lyapunov exponent can be written as,

$$\lambda \leq \log \left( \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} |1 - K\cos(2\pi x_n)| \right). \tag{93}$$

The Lyapunov exponent is now a function of parameter $K$.

Figure 19 shows a plot of the estimated Lyapunov exponent as a function of $K$ (figure taken from [17, pp. 239]). It can be clearly seen, the Lyapunov exponent is zero for the case (i). As parameter $K$ increases above the limit $K = 1$, the Lyapunov exponent $\lambda$ tends to be greater than zero and the map show chaotic behaviour. As soon as a stable fixed point or a frequency-locked state will be reached, $\lambda$ drops below zero. Note, the Lyapunov exponent plotted at Figure 19 has unit *Nats*. To get the Lyapunov exponent in *Bits*, the value hast to be divided by $\ln 2$.

### 6.3.2 Information Loss (Analytic)

Since the PDF of $X_{n+1}$ denoted by (89) can not be properly calculated and is unknown to us, the information loss can not be analytically determined. An estimation of the upper limit to the information loss can be given as follows,

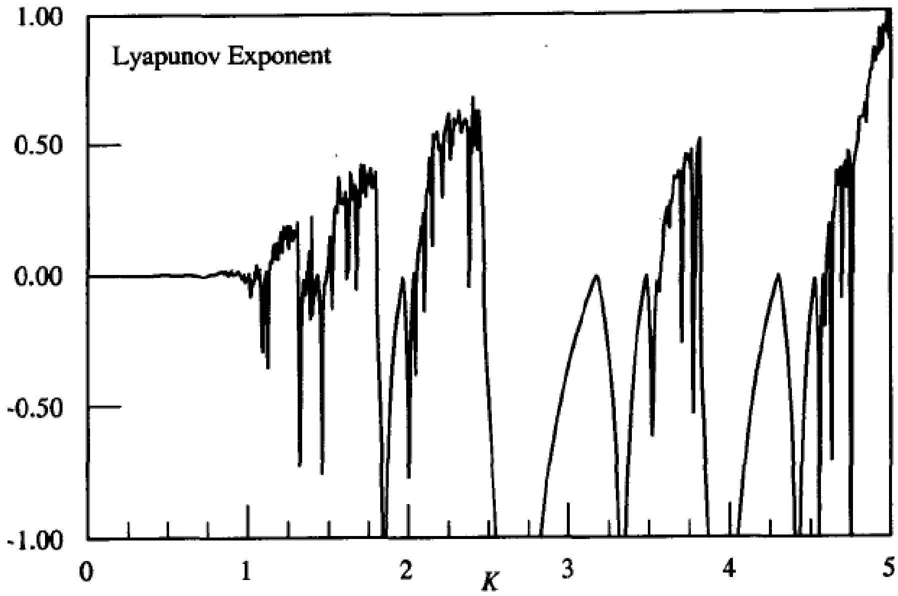


Figure 19: Lyapunov exponent of the sine-circle map for $\Omega = 0.606661$ over parameter $K$. Note, the Lyapunov exponent has unit *Nats*. To get the Lyapunov exponent in *Bits*, the value hast to be divided by $\ln 2$. Figure taken from [17, pp. 239].

$$H(X_n|X_{n+1}) = \int_{\mathcal{X}} H(X_n|X_{n+1} = x) dP_{X_{n+1}}(x) \tag{94}$$

$$\leq \int_{\mathcal{X}} \log card(g^{-1}(x)) dP_{X_{n+1}}(x) \tag{95}$$

$$\leq \max_x \; card(g^{-1}(x)) \tag{96}$$

where $card(g^{-1}(x))$ is the cardinality of the preimage. By Figure 18 can be seen, that the sine-circle map can admit up to five preimages for $K = 5$.

According to the cases mentioned above at subsection 6.3.1, the following can be said:

**(i) $\mathbf{0 < K < 1}$**: The sine-circle map is bijective, therefore the information loss is $H(X_n|X_{n+1}) = 0$. Thus, these results for the K-S entropy and the information loss coincide with the main result (58),

$$\underbrace{H_{KS}}_{=0} = \underbrace{H(X_n|X_{n+1})}_{=0} \tag{97}$$

**(ii) K > 1** : As numerical results (provided later) show, it can be stated

$$\underbrace{H_{KS}}_{\leq 1.44\text{Bit}} \leq \underbrace{H(X_n|X_{n+1})}_{\leq 1.6\text{Bit}}. \tag{98}$$

#### 6.3.3 Information Loss (Numeric Upper Bound)

The Sine-Circle map is defined by (88) on $\mathcal{X} = [0, 1]$, where $K$ is controlling the damping of the non-linearity and $\Omega$ the frequency-ratio. In the following the outcome of the experiments from different parameter $K$ and constant $\Omega = 0.606661$ will be presented. Generally can be stated, that $H_{KS} \leq H(X_n|X_{n+1})$ and the uniform distribution is not invariant PDF.

Let $(K, \Omega) = (5, 0.606661)$, a case where the map is strongly chaotic with $\lambda \approx 1.44$ Bits (see section 6.3). By Figure 20 can be seen, that the information loss tends to be a little bit less than 1.4 Bit. This matches with the theoretical aspect. The sine-circle map has more than two branches of possible preimages $g^{-1}(x_n)$, therefore the information loss can be more than 1 Bit. The histogram taken at different time steps, shows the transformation of the PDF $f_{X_n}$ under the map starting by an uniformly distribution. This is illustrated by Figure 21.

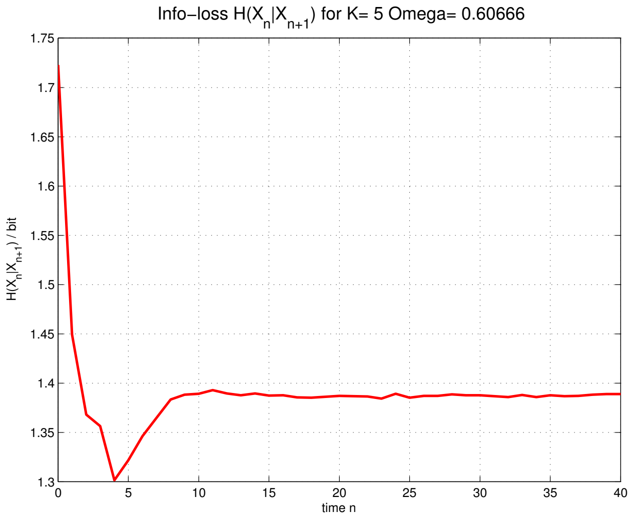


Figure 20: Information Loss of the sine-circle map for $K = 5$ and $\Omega = 0.606661$. The information loss is slightly lower than 1.4 Bit after the transient has gone.

Now, let the set of parameters be $(K, \Omega) = (2.3, 0.606661)$ at which the sine-circle map shows not such strong chaotic behaviour with $\lambda \approx 0.72$ Bits. The only difference to the case above can be seen, that the information loss is $H(X_n|X_{n+1}) \approx 0.85$ Bit. The graph of the information loss is plotted at Figure 22. Figure 23 shows the histogram.

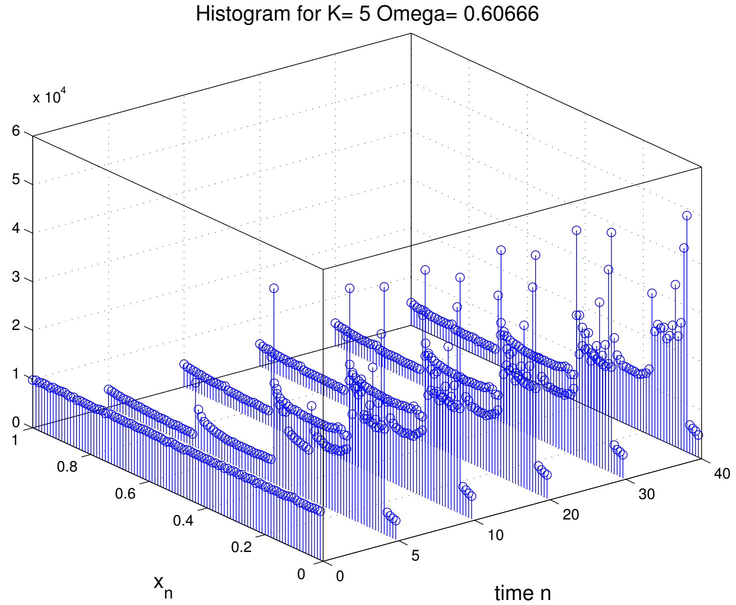


Figure 21: Histogram at certain time steps for the uniformly distributed RV $X_n$, which gets transformed by the sine-circle map ($K = 5$, $\Omega = 0.606661$).

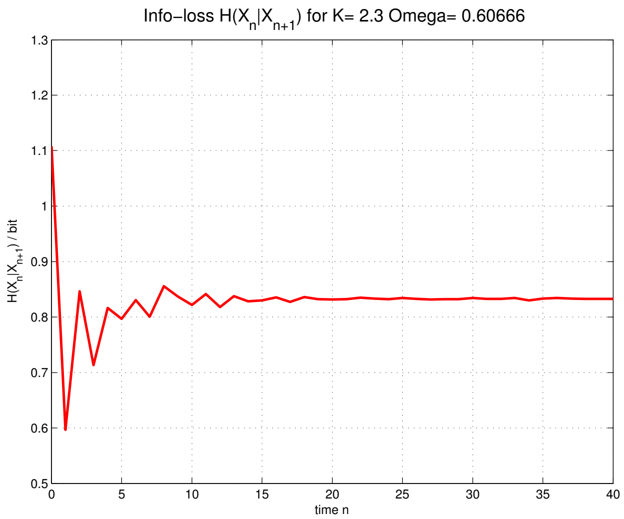


Figure 22: Information Loss of the sine-circle map for $K = 2.3$ and $\Omega = 0.606661$. The information loss is approximately $H(X_n|X_{n+1}) \approx 0.85\,\mathrm{Bit}$.

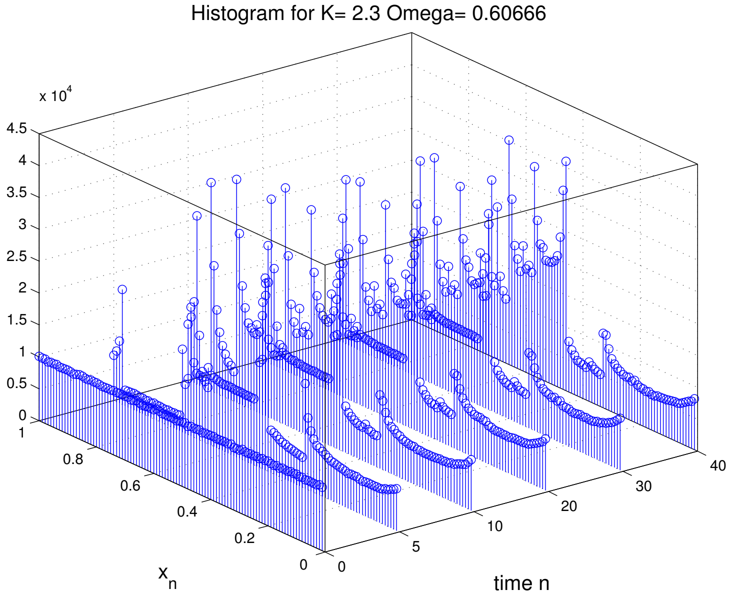


Figure 23: Histogram of uniformly distributed RV $X_n$ at different time steps ($K = 2.3$, $\Omega = 0.606661$).

Figure 24 takes us to the limit case where $K = 1$. At this case the map can be quasi-periodic or chaotic, where $\lambda \approx 0$. As it is supposed, the map changes its behaviour. The map gets bijective and the information loss tends to be zero. This can be clearly seen by Figure 24 with $K = 1$ and by Figure 25 with $K = 0.5$. The information loss is dropping from the vicinity of 0.012 Bit down to number of $10^{-4}$ Bit, which is due to nummerical errors. The histogramm of $K = 0.5$ plotted at Figure 26 shows a quite interesting transformation of the PDF. Starting with a uniform distribution of RV $X_n$, the PDF gets shaped like the normal distribution through the transformation of the sine-circle map.

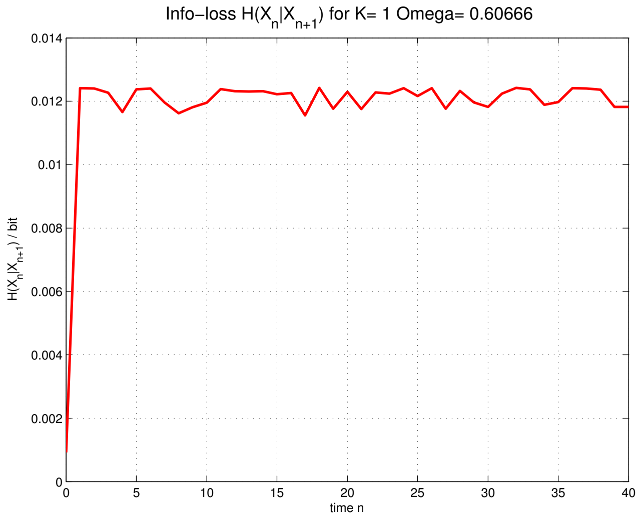


Figure 24: Information Loss of the sine-circle map for $K = 1$ and $\Omega = 0.606661$.

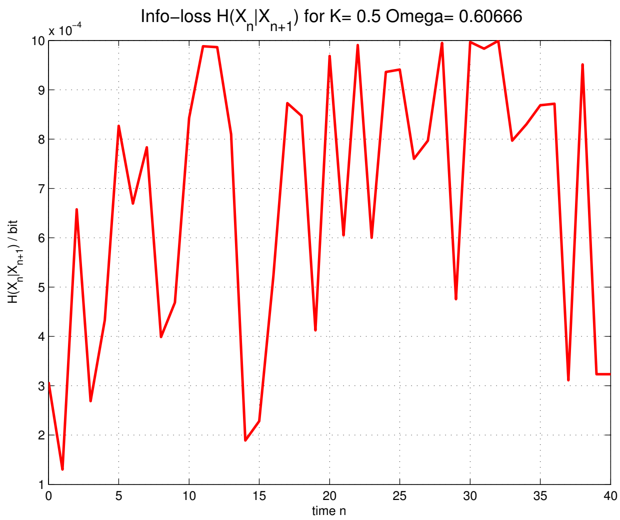


Figure 25: Information Loss of the sine-circle map for $K = 0.5$ and $\Omega = 0.606661$. The information loss can be seen as to be zero, since the map is bijective and therefore no information loss occurs.

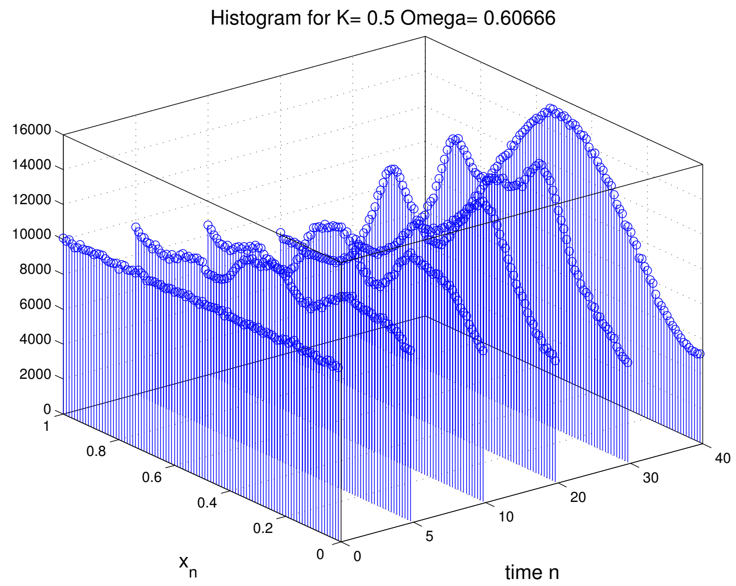


Figure 26: Histogram of uniformly distributed RV $X_n$, that gets transformed under the action of the sine-circle map ($K = 0.5$, $\Omega = 0.606661$).

## 6.4 Generalized $2x$ *mod* 1 map

This map is an extended version of the $2x$ *modulo* 1 map (Bernoulli shift), which can be treated as the tent map. The generalization of this map is made through the additional parameter $\eta$. The Generalized $2x$ *modulo* 1 map is defined as,

$$g(x) = \begin{cases} 2\eta x, & \text{if } -\infty \leq x < 1/2 \\ 2\eta(x-1)+1, & \text{if } 1/2 \leq x \leq +\infty \end{cases} . \tag{99}$$

Figure 27 shows a plot of the Generalized $2x$ *mod* 1 map. It can be seen, that parameter $\eta$ is acting like a valve in between the two branches of the map. If $\eta > 1$, the valve is open and mass will be lost. At each iteration of the map, the interval $\Delta = 1 - 1/\eta$ will be cut out of the remaining intervals [11, pp. 73]. As the map iterates forward in time, the remaining set forever will be a symmetric Cantor set with zero Lebesgue measure. It is said, that this map is an open system and the Pesin's identity can not be applied.

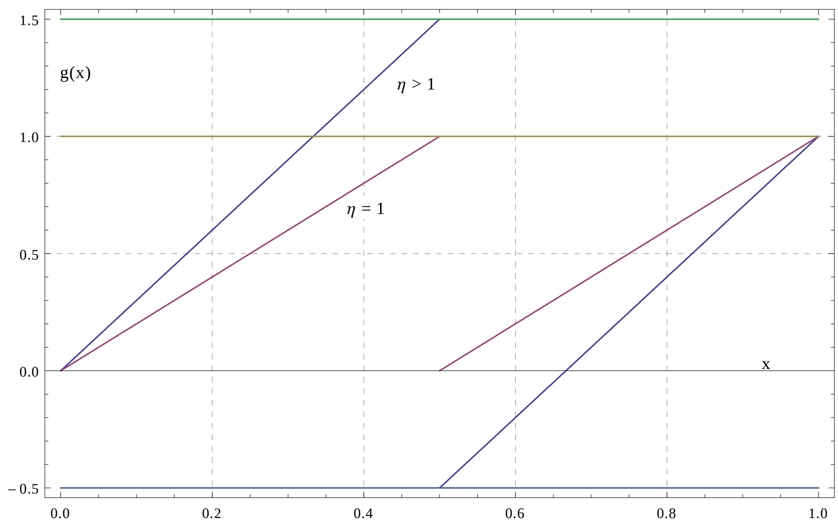


Figure 27: Generalized $2x$ *mod* 1 map. If $\eta = 1$, the map has same behaviour as the symmetric tent map. If $\eta > 1$, probability mass will be lost under the action of the map.

**Construction of a symmetric Cantor Set:** Consider the set $\mathcal{X} = [0,1]$ on which the map (99) is defined. Assume two intervals of length $I_0$ and $I_1$ such that $I_0 + I_1 < 1$. After one iteration of the map, the total length $l_i$, $i = 1, ..., n$ through time $n$ are denoted as,

$$\begin{aligned} l_1 &= I_0 + I_1 \\ l_2 &= (I_0 + I_1)^2 = I_0^2 + 2I_0I_1 + I_1^2 \\ &\vdots \\ l_n &= (I_0 + I_1)^n \end{aligned} \tag{100}$$

As a result of that, there are $Z(n,m) = \frac{n!}{n!(n-m)!}$ intervals of length $I_0^m I_1^{n-m}$ with $m = 0, ..., n$. The overall length of the remaining set at time $n$ decays exponentially,

$$l_n = e^{n \ln{(I_0+I_1)}} \tag{101}$$

If $\eta$ is set to $\eta = 2/3$, then map $g(x)$ exhibit the middle-third Cantor set. Figure 28 illustrates the development of the cantor set as the map evolves in time.

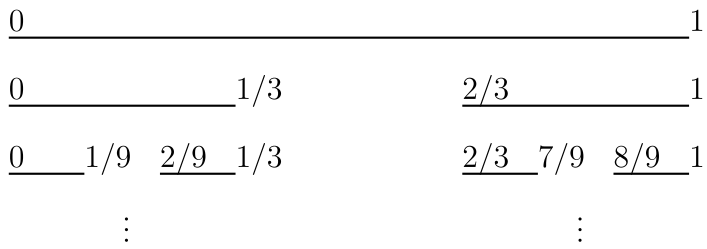


Figure 28: Middle third cantor set, exhibited by $g(x)$ for $\eta = 3/2$.

As we have seen above, we have to distinguish between two cases $\eta = 1$ and $\eta > 1$.

**(i)** $\eta = 1$**:** In this case the map (99) simplifies to the regular modulo 1 map, $g(x) = 2x \ mod \ 1$. This map is similar to the symmetric tent map, which is treated in section 6.1.

**(ii)** $\eta > 1$**:** Let the length of two intervals be $I_0$ and $I_1$ such that $I_0 + I_1 < 1$.
Thus, we rewrite the map (99) as (see [19]),

$$g(x) = \begin{cases} x/I_0, & \text{if } 0 \leq x < 1/2 \\ (x-1)/I_1 + 1, & \text{if } 1/2 \leq x \leq 1 \end{cases} \tag{102}$$

where the derivative is,

$$|g^{'}(x)| = \begin{cases} 1/I_0, & \text{if } 0 \leq x < 1/2 \\ 1/I_1, & \text{if } 1/2 \leq x \leq 1 \end{cases}. \tag{103}$$

### 6.4.1 Lyapunov exponent $\lambda$ and K-S entropy $H_{KS}$

The following derivations are taken from [19, pp. 138].

Let $n_0/N$ and $n_1/N$ be the relative frequency of being in the corresponding intervals of $I_0$ or $I_1$. Assume the map is ergodic, the Lyapunov exponent can be written as,

$$\begin{aligned} \lambda &= \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \log \left| g^{'}(x_n) \right| \\ &= \lim_{N\to\infty} \left( \frac{n_0}{N} \log\left(\frac{1}{I_0}\right) + \frac{n_1}{N} \log\left(\frac{1}{I_1}\right) \right) \end{aligned}. \tag{104}$$

By introducing the natural measure [11, pp. 81], the time dependency will abolish:

$$\lim_{N\to\infty} \frac{n_0}{N} = \frac{I_0}{I_0 + I_1} \tag{105}$$

$$\lim_{N\to\infty} \frac{n_1}{N} = \frac{I_1}{I_0 + I_1} \tag{106}$$

Thus the Lyapunov exponent computes to

$$\lambda = \frac{1}{I_0 + I_1}\left(I_0 \log\left(\frac{1}{I_0}\right) + I_1 \log\left(\frac{1}{I_1}\right)\right) \tag{107}$$

Again, using the context of the natural measure the entropy of the $2^n$ partitions at time $n$ can be written as,

$$h(n, \mu_i) = -\sum_{i=1}^{2^n} \mu_i \log(\mu_i) \tag{108}$$

where the natural measure is,

$$\mu_i = \frac{I_0^{n_0} I_1^{n-n_0}}{(I_0 + I_1)^n},\ n = n_0 + n_1. \tag{109}$$

After some assumptions made through Newton's binomial expansion and introducing the combinatorial factor $Z(n, m)$ (see above) the K-S entropy is,

$$H_{KS} = \lim_{n\to\infty} \frac{1}{n} \sup_{\mu_i} h(n, \mu_i) \tag{110}$$

$$= \log(I_0 + I_1) + \frac{I_0}{I_0 + I_1} \log\left(\frac{1}{I_0}\right) + \frac{I_1}{I_0 + I_1} \log\left(\frac{1}{I_1}\right). \tag{111}$$

Combining (107) and (111), the K-S entropy is,

$$H_{KS} = \underbrace{\log(I_0 + I_1)}_{\gamma} + \lambda \tag{112}$$

where $\gamma$ is the escape rate, which closes the gap between the K-S entropy and the sum of positive Lyapunov exponents. Note, this map is an open system that looses probability mass. Thus Pesin's identity nor $H_{KS} < \lambda$ necessarily hold. Therefore our main result does not apply here.
The interested reader is referred to [19, pp. 136],[20],[21] about the formalism for the escape rate.

If the intervals have same length $I_0 = I_1 = 1/2\eta$ , the K-S entropy is 1 Bit. Note, the parameter $\eta$ cancels out due to the escape rate. For sake of completeness, the Lyapunov exponent calculates to $\lambda = \log(2\eta)$. Taking the middle third cantor set with $\eta = 3/2$ the Lyapunov exponent calculates to 1.58 Bit and the escape rate $\gamma = \log(1/\eta) = \log(2/3) = -0.58$ Bit.

### 6.4.2 Information Loss (Analytic)

An estimation on the information loss can be made due to symmetry around $x = 1/2$. Let $P_{X_n}$ be the probability distribution $X_n$ and let $\mathcal{X}_i$ be the set of the remaining values. Thus the overall information loss can be estimated as,

$$H(X_n|X_{n+1}) = \int_{\mathcal{X}} H(X_n|X_{n+1} = x) dP_{X_{n+1}}(x) \tag{113}$$

$$= \int_{\mathcal{X}} H(X_n|X_{n+1} = x) dP_{X_{n+1}}(x) \tag{114}$$

$$= \int_{\mathcal{X}_i} \log 2 dP_{X_{n+1}} + \int_{\mathcal{X}\setminus\mathcal{X}_i} 0 \ dP_{X_{n+1}}(x) \tag{115}$$

$$\tag{116}$$

Let the time tend to infinity $n \to \infty$, the probability $P_{X_n}(\mathcal{X}_i) \to 0$ tends to zero. Thus, the information loss tends to zero $H(X_n|X_{n+1}) \to 0$. This is a result of the fact, that the probability mass gets bijectively mapped outside the defined region $\mathcal{X}$; see the two branches for $\eta > 1$ at Figure 27.

### 6.4.3 Information Loss (Numeric Upper Bound)

This open system is defined by (99) at section 6.4. Since the map looses probability mass over the time for $\eta > 1$ the information loss decreases down to zero in exponential manner. This can be clearly seen in the graphs of the following experiments.

For $\eta = 1.1$ the graph in Figure 29 is decreasing towards zero, starting from around $H(X_n|X_{n+1}) \approx 0.9\,\mathrm{Bit}$. This higher number compared to other known experiments is because the map is loosing not that much probability mass at the beginning. The histogram in Figure 30 illustrates such behaviour. Note, in general all the probability mass is merged to the margins of the domain, given by $\pm\eta$.
By Figure 31 and Figure 33 can be seen that for $\eta = 1.5$ and $\eta = 2.5$ the information loss is decreasing faster down to zero. Note, the starting value of the information decreases as well, since a lot more mass is put out of the system at the first iterations. As the interval $\Delta = 1 - 1/\eta$, is going to be greater as $\eta$ increases, probability mass is is lost very fast. The graph in Figure 32 and Figure 34 give a good picture of loosing the probability mass.

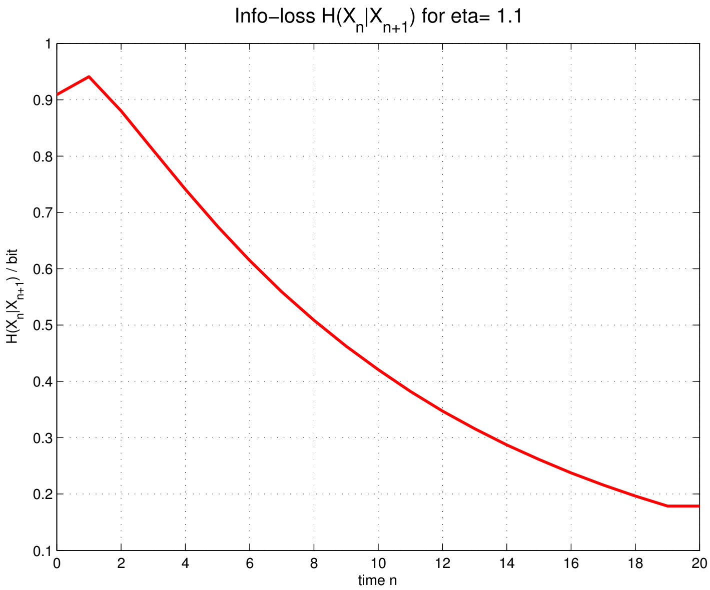


Figure 29: Information Loss $H(X_n|X_{n+1})$ is slowly decreasing towards zero with time $n$.

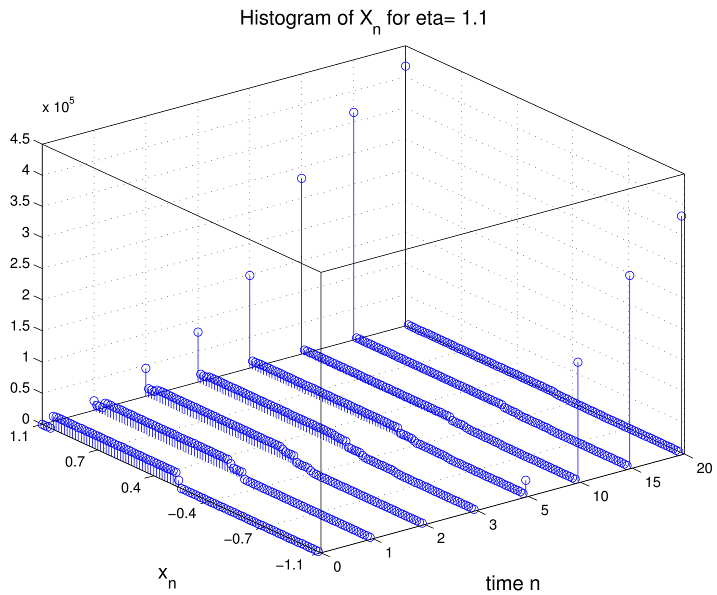


Figure 30: Histogram at certain time steps for the Generalized $2x$ *mod* 1 map. Obtain that probability mass gets lost during the action of the map.

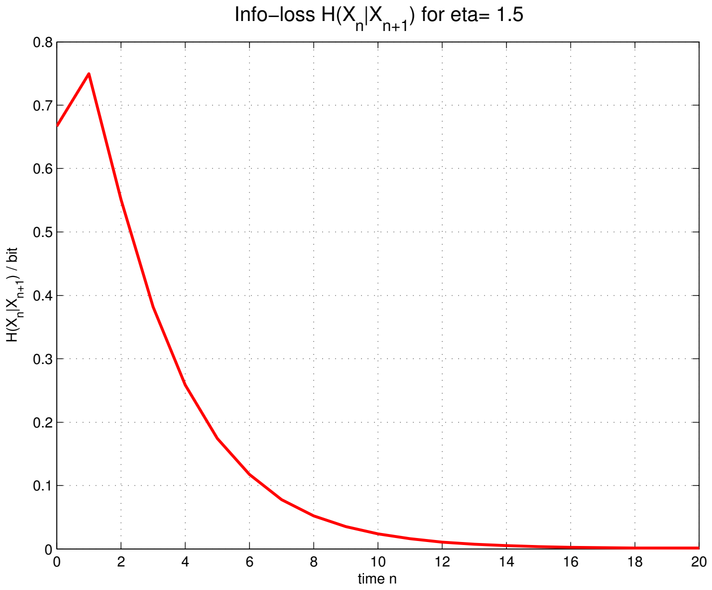


Figure 31: Information Loss $H(X_n|X_{n+1})$ is faster decreasing towards zero with time $n$, since parameter $\eta$ is increased.

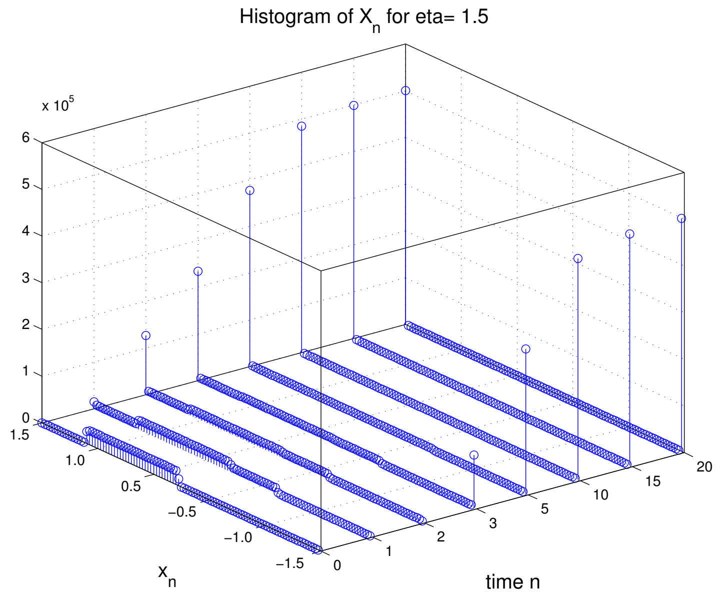


Figure 32: Histogram at certain time steps for the Generalized $2x$ *mod* 1 map. Note that much more probability mass gets lost at the beginning.

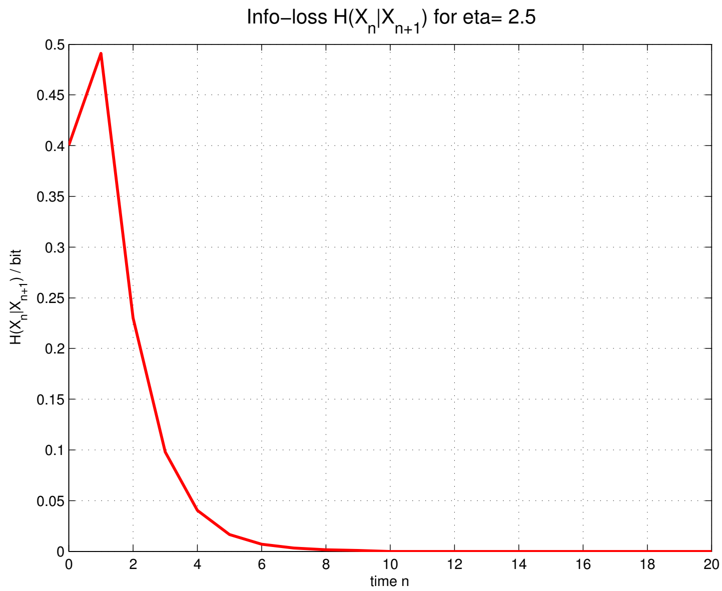


Figure 33: Information Loss $H(X_n|X_{n+1})$ is decreased towards zero within a few time steps.

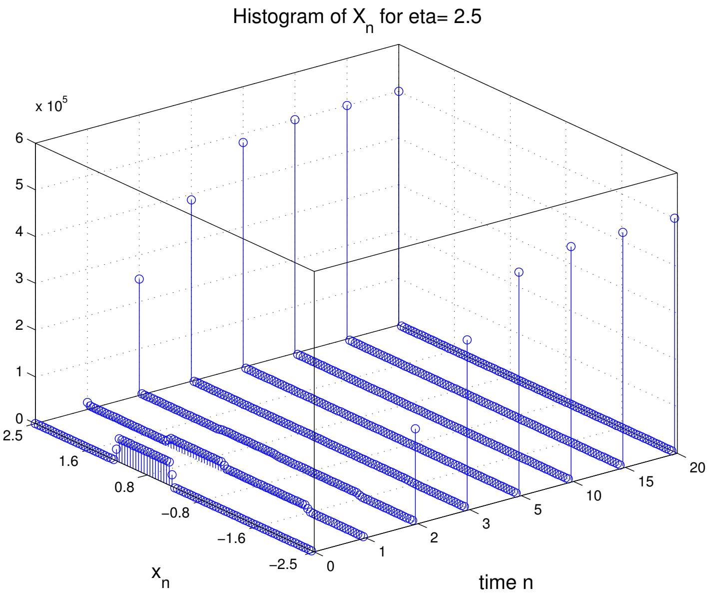


Figure 34: Histogram of the Generalized $2x$ *mod* 1 map for a uniform distribution, as starting distribution. All probability mass gets lost within a few time steps.

# 7 Conclusion and Outlook

In this work, a relationship between the information loss and the information generation (Kolmogorov-Sinai entropy) is proposed from a system-theoretic point of view for the class of measure-preserving maps. In particular, a general inequality (58) is defined which holds for $C^1$-maps and noninvertible maps. Further can be shown, that the information loss is upper semicontinuous and therefore it is an upper bound for the Kolmogorov-Sinai entropy.
For univariate maps, this main result is analyzed and explored for the asymmetric tent map, Gaussian Map (Gauss iterated map), Sine-Circle Map and the Generalized $2x \ mod \ 1$ map which is related to open systems. For open systems the main result does not apply, since these class of systems is not measure-preserving. For univariate maps can be said, that there exists no bijective map which is chaotic. In this case, the information loss is zero and since it is an upper bound for the Kolmogorov-Sinai entropy, which is an interesting point.
As this work is restricted to univariate maps, the multivariate case is not considered. The question arises at this point, whether multivariate maps can be found, which are bijective and where strict inequality between the information loss and the Kolmogorov-Sinai entropy can be obtained? For example, the Baker's map is a two dimensional map, which is measure preserving and bijective as previous results showed. In addition to that, there is the need of evaluating the formalism about the escape rate which can be found in the context of open systems. This escape rate closes the gap between the Kolmogorov-Sinai entropy and the sum of positive Lyapunov exponents. The question is whether the main result can generalized to a larger class of systems by including the escape rate.

# References


[1] J.-P. Eckmann and D. Ruelle, "Ergodic theory of chaos and strange attractors," Rev. Mod. Phys, vol. 57, pp. 617–656, 1985.

[2] D. Ruelle, "Positivity of entropy production in nonequilibrium statistical mechanics," J. Stat.Phys., vol. 85, pp. 1–23, 1996.

[3] ——, "A measure associated with Axiom-A attractors." Amer. J. Math, vol. 98, pp. 619–654, 1976.

[4] L.-S. Young, "What are SRB measures, and which dynamical systems have them?" J. Stat. Phys., vol. 108, pp. 733–754, Sept. 2002.

[5] T. M. Cover and J. A. Thomas, Elements of Information Theory, 2nd edition. Wiley- Interscience Publication, 2005.

[6] B. C. Geiger, C. Feldbauer, and G. Kubin, "Information loss in static nonlinearities," in Proc. IEEE Int. Sym. Wireless Communication Systems (ISWSC), Aachen, Nov. 2011, pp. 799–803, extended version available: `arXiv:1102.4794 [cs.IT]`.

[7] B. C. Geiger and G. Kubin, "On the information loss in memoryless systems: The multivariate case," in Int. Zurich Seminar on Communications, Zurich, Feb. 2012, pp. 32–35.

[8] A. Papoulis and U. S. Pillai, Probability, Random Variables and Stochastic Processes, 4th ed. New York, NY: Mc Graw Hill, 2002.

[9] H. G. Schuster and W. Just, Deterministic Chaos, An Introduction. Fourth, Revised and Enlarged Edition. Wiley VCH Verlag Gmbh&Co. KGaA, Weinheim, 2005.

[10] R. W. Leven, B.-P. Koch, and B. Pompe, Chaos in dissipativen Systemen, 2. berarbeitete und erweiterte Auflage. Akademie Verlag Gmbh, 1994.

[11] E. Ott, Chaos in dynamical Systems. Press Syndicate of the University of Cambridge, 1993.

[12] V. Millionshchikov, "On the theory of characteristic Lyapunov exponents," Mat. Zametki, vol. 7, pp. 503–513, 1970.

[13] D. Ruelle, "An inequality for the entropy of differential maps," Bol. Soc. Bras. Mat., vol. 9, pp. 83–87, 1978.

[14] Y. B. Pesin, “Characteristic Lapunov exponents and smooth ergodic theory,” Rus. Math. Sur., vol. 32:4, pp. 55–144, 1977.

[15] M. Qian and S. Zhu, “SRB measures and Pesin’s entropy formula for endomorphisms,” Transaction of American Mathematical Society, vol. 354:4, pp. 1453–1471, 2001.

[16] F. Ledrappier and L.-S. Young, “The metric entropy of diffeomorphisms Part I: Characterization of measures satisfying Pesin’s entropy formula,” Ann. Math., vol. 122, pp. 509–539, 1985.

[17] R. C. Hilborn, Chaos and Nonlinear Dynamics: An Introduction to Scientists and Engineers 2nd Edition. Oxford University Press.

[18] V. Patidar and K. Sud, “A comparative study on the co-existing attractors in the Gaussian map and its q-deformed version,” Com Non.l.Sc. and Num.Sim., vol. 14, pp. 827–838, 2007.

[19] J. R. Dorfmann, An Introduction to Chaos in Nonequilibrium Statistical Mechanics. Cambridge University Press, 1999.

[20] L.-S. Young, P. Wright, and M. Demers, “Entropy, Lyapunov Exponents and Escape Rates in Open Systems,” Ergodic Theory and Dynamical Systems, vol. 32, pp. 1270 – 1301, July 2012.

[21] L.-S. Young and M. Demers, “Escape Rates and Conditionally Invariant Measures,” Nonlinearity, vol. 19, pp. 377–397, October 2006.